Language Learning & Technology
ISSN 1094-3501 CC BY-NC-ND

2024, Volume 28, Issue 1
pp. 1–27

ARTICLE

Multimodal interactive alignment: Language learners' interaction in CMC tasks through Instagram

Muntaha Muntaha, Curtin University

Julian Chen, Curtin University

Toni Dobinson, Curtin University

Abstract

Technological advancement has enabled language learners to employ verbal and nonverbal cues in computer-mediated communication (CMC). These cues can support language use for learners wishing to communicate more effectively in English. Interactive alignment is one phenomenon that shows how humans tend to collaborate in their language use by adapting, priming, and reusing verbal and nonverbal cues to achieve mutual understanding. Informed by a sociocognitive framework, this study explored and documented English language learners' multimodal interactive alignment during their CMC task engagement through Instagram. We collected data from 30 first-year Indonesian business school learners who participated in seven online CMC tasks using Instagram chat features: text chat, voice chat, and video chat. To examine various interactive alignments (e.g., how interlocutors adapt, prime, and reuse verbal and nonverbal cues to achieve mutual understanding) that occurred during multimodal task communication, we employed multimodal (inter)action analysis. Findings revealed that learners adapted and reused various nonverbal features (e.g., emojis, GIFs, facial expressions, gestures) and verbal cues (e.g., expression, lexical) to convey and comprehend meaning during CMC task completion. Caveats about using various nonverbal alignment patterns for supporting better English online communication were also noted. The study highlights how language learners use the full repertoire of semiotic resources in CMC to maximize their online language learning.

Keywords: *Interactive Alignment, Multimodal (Inter)action Analysis, Instagram, Computer-Mediated Communication (CMC)*

Language(s) Learned in This Study: *English*

APA Citation: Muntaha, M., Chen, J., & Dobinson, T. (2024). Multimodal interactive alignment: Language learners' interaction in CMC tasks through Instagram. *Language Learning & Technology*, *28*(1), 1–27. https://hdl.handle.net/10125/73563

Introduction

Recent trends in computer-mediated communication (CMC) research acknowledge the use of multiple communication modes for online interaction. According to scholars such as Guichon and Cohen (2016), multimodality for meaning-making during online interaction enhances language learning. For example, learners can strategically use multimodality to reinforce the conveyed meaning in text chats by adding emojis (Li & Yang, 2018) or enacting gestures to negotiate meaning during videoconferencing (Lee et al., 2019). The widespread use of CMC in supporting language learning has changed the complexity and dynamics of how humans use their language to exchange ideas and messages in online communication, including the way they align interactively in online conversation. *Interactive alignment* is one phenomenon that shows how humans tend to collaborate in their language use by adapting, priming, and reusing verbal and nonverbal cues to achieve mutual understanding (Nishino & Atkinson, 2015; Pickering & Garrod, 2004). In an additional language (henceforth LX, see Dewaele, 2017) learning context, multimodality and alignment have become central to a sociocognitive approach, which is based on the belief that the human

body, mind, and the environment around the sites of communication operate collaboratively in the process of language learning, not just in human cognition (Atkinson, 2011). Learners naturally adapt to the learning environment by performing interactive alignment (Atkinson, 2014). Therefore, a sociocognitive approach recognizes the involvement of multimodalities, such as gestures, images, sounds, animations, and videos, in language learning.

Using nonverbal cues for interactive alignment during conversation is the natural outcome of interactions in many situations, either in offline, face-to-face discussions, or online conversations (Oben & Brône, 2016; Zhou & Wang, 2021). For the last two decades, studies have mainly explored verbal alignment in a language learning context both in offline, face-to-face, and online settings (e.g., Dao et al., 2018; Kim et al., 2019; Michel & Cappellini, 2019; Michel & Smith, 2018; Uzum, 2010; Zhou & Wang, 2021). However, studies on interactive alignment involving verbal and nonverbal cues in their analysis simultaneously are scarce. Oben and Brône (2016) explored alignment process at lexical and gestural levels during task completion in offline face-to-face conversation. Given that the rise of multimodality in CMC today might have created more complex and diverse alignment due to the emergence of new features in the digital platform, research investigating interactive alignment entailing verbal and nonverbal cues in online interaction is needed.

Hence, this study offers multimodal (inter)action analysis as the analytical tool for better capturing learner interaction dynamics among modes during online interactions. It responds to the call for further research on interactive alignment suggested by Michel and Cappellini (2019). Further, using a sociocognitive framework, this study explores how verbal and nonverbal interactive alignments occurred during CMC tasks in three online communication channels afforded by Instagram: text chat, audio chat, and video chat because Instagram is one of the three biggest communication apps among youth in Indonesia. Additionally, only a limited study has investigated Instagram interaction. The in-depth approach taken in this study complements holistic approaches to language learning. It illuminates how verbal and nonverbal alignment can support LX learning in a CMC environment. Thus, the current study was guided by the following research questions (RQs):

1. In what ways did learners display multimodal interactive alignment in CMC tasks through Instagram?
2. What modes, other than verbal cues, contributed to the interactive alignment in CMC tasks through Instagram?

Literature Review

Interactive Alignment in LX Learning from a Sociocognitive Approach

Historically, the term interactive alignment, in a language learning context, refers to the phenomenon where speakers reuse, adapt, and prime their language to each other at the level of expressions, structures, and sounds (Costa et al., 2008). This helps speakers simplify their production and comprehension during interaction by supporting explicit inference mechanisms and enables them to develop and reuse routine expressions in dialogue (Pickering & Garrod, 2004; Zhou & Wang, 2021). Informed by sociocognitivism, Atkinson (2014) expanded the scope of alignment beyond the linguistic level by including how learners adapt to their environment and coordinate their mind and body actions. In other words, learners align with all aspects of the learning process, including verbal, nonverbal, or mediated learning tools (e.g., laptop, whiteboard, screen) in any environment and social practice. The sociocognitive approach also considers alignment as part of the learning process, whereby learners build moment-to-moment social relations and cooperative social action in an LX environment. Through alignment, learners can engage in any social activities that support target language use and development in any social situation (Atkinson, 2014).

CMC studies have explored alignment in various language learning contexts. For example, Uzum (2010) investigated the occurrence of verbal alignment in CMC interaction through the text chat transcription and stimulated recall interview, and found that the alignments were manifested in fluency and speed, negotiation of meaning, and lexical and grammatical choices. Michel and Cappellini (2019) explored linguistic

alignment in synchronous video and text chat and found that learners performed structural alignments by imitating the grammatical patterns used by their counterparts more often than lexical alignments such as applying similar word choices. Evidence of alignment was also noted by Zhang (2017), who measured alignment quantitatively and found that learners perform higher alignment during the continuation task compared to the summary activity by reusing the same phrase acquired from the input task in the sentences they wrote during the project. Thus, reusing words/phrases affected learners' lexical acquisition and enhanced interactive alignment. These studies provide evidence for verbal alignment as a valuable source of 1) language exposure or 2) input from learners during conversation, or 3) stimulation from a text in input task completion.

Currently, the involvement of nonverbal cues along with verbal cues to foster interaction in LX learning via CMC has attracted growing research attention. Lee et al. (2019) examined the role of gesture in videoconferencing interaction using multimodal analysis and noted that learners extended their verbal negotiation of meaning by showing iconic gestures (representing object/action) and deictic gestures (pointing hand) to enhance mutual understanding. Regarding the role nonverbal cues play in text-based CMC, Maa and Taguchi (2022) investigate emojis as pragmatic resources in text chat between L2 Japanese learners and their native speaker peers. Their findings revealed that learners adaptively noticed and incorporated emojis into their text messages in order to add expressiveness to sentences, adjust tone of conversation, or build interpersonal relationship with the interlocutors. Despite the promising findings above, our study attempted to expand the investigation of alignment from verbal alignment to both verbal and nonverbal behaviors across multiple communication channels (i.e., text, audio, video). The co-occurrence of multidimensional alignment could help us better understand the impact of semiotic resources on facilitating LX learning.

Multimodality and LX Interaction

In communication, people simultaneously use multiple semiotic resources to co-construct meaning (Kress & Van Leeuwen, 2001; O'Halloran, 2004), including verbal/linguistic cues and other non-linguistic elements such as gestures, eye gaze, intonation, or images (Norris, 2004). Jewitt (2014, pp. 6-7) postulated four underpinnings that conceptualize multimodality:

1) communication draws on a diversity of modes, all of which have the potential to contribute equally to meaning;
2) all modes are shaped through their cultural-historical and social uses to realize social function;
3) people orchestrate meaning through their choice and configuration of modes because the sense of each mode was created and interwoven with the meaning of other modes co-present and cooperating in the communicative event; and
4) meanings of signs created by humans are social because they are shaped by norms and rules, and the motivation and interest of sign-makers influence them in a specific social context.

Multimodality is an integral part of LX teaching, learning, and communication. For example, Faraco and Kida's (2008) study revealed that, alongside their verbal cues in managing learning sequences, teachers' nonverbal behaviors (e.g., eye gaze, gestures) offered metalinguistic commentaries to learners' signs or messages that could not be formulated verbally. Similarly, Olsher (2008) indicated that gestures, eye gaze, and posture helped adult learners repair turns in a communication breakdown to achieve the ultimate task goal. Interestingly, language beginners use gestures, as a nonverbal mechanism, to satisfy lexical and meaning-making needs (Rosborough, 2014). This is verified by Negueruela and Lantolf (2008), who suggested that the use of gestures in LX communication is spontaneous and indispensable since communication is the product of social activity.

Given the salient multimodal features (e.g., video, images, emojis, GIFs) undergirding the online communication landscape, the roles of nonverbal cues in CMC to support LX are even more paramount in social networking (Calvo-Ferrer et al., 2016). Studies have found positive evidence of multimodal use as

part of LX instruction in online distance learning. The evidence accentuates the potential of CMC for communication strategies, negotiation of meaning, and fostering communication (Satar, 2016). For example, Hampel and Stickler (2012) reported that utilizing multimodal online tools such as text, voices, images, and live video may increase interactions and that learners may carry out better LX communication by enacting these CMC functionalities. Vandergriff (2013) highlighted that emoticons in LX text chat are often used as politeness markers that help users convey socio-emotional information such as sender stance, relation, and position to co-participants in online communication. Furthermore, Satar and Wigham (2017) revealed that teacher trainees used multimodal resources, such as word stress, gaze, and text, to enhance their teaching instruction when engaging in a role-play task as an online teacher.

Multimodal (Inter)action Analysis in CMC

The concept of multimodal (inter)action analysis as a framework was initially introduced by Norris (2004, 2011, 2019). It refers to "a holistic analytical framework that understands the multiple modes in (inter)action as all together building one system of communication" (Norris & Pirini, 2016, p. 24). The analysis considers that all learners' activities are interactions with other learners, tools, objects, or the environment. Norris (2004) outlined key analytical tools for enacting this analysis: mediated action, communication mode, and engagement site. The mediated action, as a unit of analysis, is the acting of learners with/through mediational means in different settings. This framework classifies mediated action into two levels: lower-level action and higher-level action. Norris (2019, pp. 42-44) defines lower-level action as "the smallest pragmatic meaning unit of a mode" (e.g., verbal cues, pointing gestures, emojis, or images) and higher-level action as "chains of lower-level mediated actions come together to build the higher-level mediated action" (e.g., task opening, negotiation of agreement, or content discussion). A communication mode is a system of mediated actions; for example, an utterance is a lower-level action in the verbal mode, and a gesture unit is a lower-level action in the nonverbal mode. The site of engagement, the final analytical tool in this framework, is the place, media, or moment where social practices and mediational means enable mediated actions to occur (Jones & Norris, 2005).

In operating multimodal (inter)action analysis, we consider the contribution of nonverbal cues in the engagement site. Similar to other social media (e.g., Facebook, WhatsApp), Instagram chat tools afford visual cues such as emojis, GIFs, and images in text chat, or gestures, proxemics, and gaze in video chat to enrich the users' experiences in online communication. Developed by Kurita Shigetaka, a Japanese telecommunication worker, emojis are pictorial characters and pictographs in the digital writing system (Giannoulis & Wilde, 2020). They are used as a replacement for an emoticon to make pictographs and visual representations of emotions and sentiments more visually salient (Danesi, 2017). The graphic interchange format (GIF) is an image format that enables the display of an animated picture with a series of movements in a short time (Veszelszki, 2015). A gesture is a conscious/spontaneous body movement orchestrated by the speaker to manifest expressiveness and facilitate the conversation (McNeill, 2005, 2012). Proxemics express the speaker's physical position toward other interlocutors or relevant objects during the conversation (Satar & Wigham, 2017). Finally, gaze refers to the direction of orientation displayed by the speaker through the positioning of the head, particularly the eyes looking at the interlocutor or environment around (Satar, 2013).

Instagram as an Informal LX Learning Platform

Instagram was chosen as a CMC platform in this study because it has been recognized as a digital application that provides its users with multimodal features. Aghayi and Christison (2021) argued that Instagram provided the users with multimodal features that connected students' formal learning to their real-life situations. Given its high popularity and familiarity for everyday communication practice, Instagram is also found to be an effective LX learning tool for building autonomous and social learning, thus enabling learners to cooperate, collaborate, and share knowledge with each other outside of the classroom (Erarslan, 2019). Furthermore, the affordances of Instagram that allow users to mash up different modalities such as posting images and text simultaneously were also proven to heighten learner engagement in LX writing activities since multimodal components might attract multisensory systems which then

stimulated them to be more actively engaged in the task interaction (Prasetyawati & Ardi, 2020). On top of that Instagram is the most common social media used among Indonesian youth groups (Nurhayati-Wolff, 2021). Therefore, the use of Instagram in this study was considered fit for supporting language learning outside the classroom in an authentic environment for Indonesian LX learners.

Research Methods

Setting and the Participants

The study was conducted in a private university in Central Java Province, Indonesia. Due to COVID-19 pandemic restrictions in the academic year of 2020-2021, all teaching deliveries moved to the massive open online course (MOOC), *OpenLearning* platform (see https://myedu.ums.ac.id/), which allowed teachers to share their materials, create interactive forums, or conduct quizzes and assignments. Based on their individual needs and preferences, teachers could also blend the MOOC with other CMC tools, such as Zoom and Google Meet, to support online teaching and learning.

Thirty first-year college learners (F=22, M=8, average age 18.5 years) enrolled in the *English for Communication* unit at a business school were involved in this study. This unit was compulsory for all the freshmen entering the university. The goal of this unit was to provide students with basic skills for performing everyday English communication in many different real-life scenarios. Given the limited class time, we supplemented the course with CMC tasks to allow learners to have more time to continue practicing English with their peers outside of their regular/formal online classroom. Abiding by the ethics, this project was neither part of their formal learning activities, nor would affect their official scores. Tasks were designed based on the unit goal which aimed to develop student communicative skills whereas the topics were selected based on learners' preferences indicated in their responses to the needs analysis survey conducted before the study. The result of this survey also showed that learners conceded Instagram as a preferred social media platform for learning English communication since it provided multimodal live chats and was already part of their daily communication means compared to other applications (e.g., TikTok, Twitter or Google Hangouts).

The level of participants' English proficiency was assessed at a minimum A2 based on the Common European Reference Framework (CEFR) since it was the minimum entry requirement for university enrolment. Most of them considered their local language, such as Javanese, Sundanese, or Buginese, as their first language, and Bahasa Indonesia was their dominant language for communication. Following ethics codes of the institution and country in which the research was conducted pertaining to human subject research, we ensured that all participants involved in this study voluntarily signed the consent forms. They also provided consent for their task interactions to be audio/video recorded and gave permission for their photographs (including their faces) to be published for academic purposes. No coercion was exercised in the study.

Data collection

The learners completed seven communicative tasks with their peers assigned to them in a dyad or group (three people) on a weekly basis (see Table 1). Assigning learners into dyads or groups was the strategy to examine the quality of engagement created during online interaction since Instagram was open for many users to be involved in the discussion. The tasks were divided into three types: *information gap, reasoning gap,* and *opinion gap* (Ellis, 2018; Prabhu, 1987); they completed all sessions via their preferred Instagram communication channel (e.g., text chat, audio chat, video chat, or free channel) using their smartphones outside their regular meeting on the MOOC. In doing so, an Instagram chat group was first created to manage the flow of the tasks. Then, the facilitator (one of the researchers) led the task session by giving the task instructions and randomly assigning the dyad or group. Finally, the learners created a small chat group and started to perform the task within 20–30 minutes. Prior to task performance, learners had been informed that the facilitator would only deliver and monitor task activities, but they would not interfere in their task interaction.

Table 1

The Task Type, Topic, and Channel

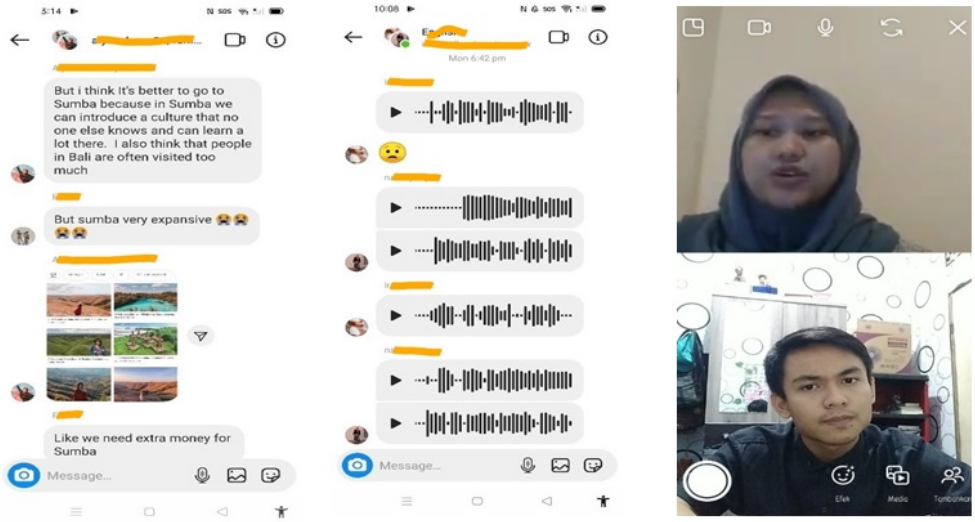
Week	Type	Topic and process	Channel
1	Information gap	*Story and movie:* Rearranging random short videos becomes a full story in a dyad	Video call
2	Information gap	*Story and movie:* Rearranging random pictures becomes a complete story in a group	Free channel
3	Reasoning gap	*Travel:* Sharing information about tourist destinations and deciding the site to go on holiday in a dyad	Text chat
4	Reasoning gap	*Travel:* Selecting only 12 kg survival kits from the provided list to carry during the journey in a group	Free channel
5	Opinion gap	*Family and friends:* Sharing and discussing opinion about 'how to build a strong friendship' in a dyad	Voice chat
6	Opinion gap	*Family and friends:* Commenting, sharing, and discussing two pictures showing contrasting life phenomena (happy and sad family pictures) in a group	Free channel
7	Information gap	*Story and movie:* Describing and guessing six different characters taken from famous novels and movies in a dyad	Video call

As Figure 1 illustrates, three communication channels were available to learners on Instagram: text, audio, and video chat. Through text chat, learners could post text, images, GIFs, short-recorded voice notes, and videos. Those who wished to post could touch the message area at the bottom of the screen and select the kind of messages they wanted to send. If they wanted to send a short video or picture, they could choose the camera icon. For the voice note, they could choose the microphone icon, and for inserting images, they could choose the storage image icon. Meanwhile, the camera video recording icon on the top right corner was used to play synchronous video chats/calls. All task sessions were saved automatically in the Instagram archive, except video calls. However, in this study, the quality of audio chat logs was bad. The sounds were sometimes louder and slower which affected the intonation produced, so we could not analyze it. It might happen due to the variation in smartphone brands or the quality of the microphone used. The learners recorded their video-based task activities through the screen recording application and sent them to the facilitator's email at the end of the task. Despite the fact that learners could choose their preferred channel to carry out the task in free channel sessions, they tended to select text chat over the other two modes. Text

chat could be possibly the most familiar chat channel to them, similar to Facebook or WhatsApp and it is easier to utilize; audio/video chats, on the other hand, are not commonly used by these participants. In addition, it might also be influenced by their current English proficiency as Satar and Ozdener (2008) argued that text chat was commonly chosen by less proficient learners (elementary level) because it provided more time to think.

Figure 1

Screenshot of Instagram Chat from Left to Right (Text, Audio, and Video Chat)



Data Analysis

In operationalizing multimodal (inter)action analysis, we divided the data into categories based on the communication channels. The text and audio chat datasets were collected from the Instagram archive and transcribed manually. All nonverbal elements in text and audio chats (e.g., emojis, pictures, images, and intonation) were included in the transcript in their original form on Instagram. Meanwhile, the learners recorded the video chat data using a screen-capture program from their smartphone and sent it to the researcher (the first author) through email.

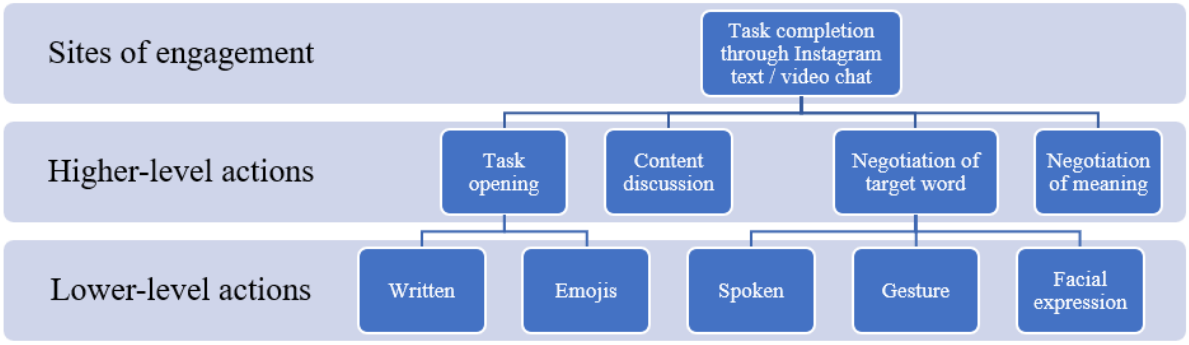
Initially, the verbatim data of the video recordings were transcribed using ELAN https://archive.mpi.nl/tla/elan, a software package for text annotation to audio/video recording developed by Max Planck Institute for psycholinguistics. This software was chosen because its features allowed transcription of verbal elements and a wide variety of nonverbal elements, including gestures, gaze, and proxemics, to be simultaneously displayed in different layers on a timeline. The scripts showed all the elements of the verbal and nonverbal cues deployed during online interaction by transcribing multimodal data. Each verbal turn was followed by a nonverbal description and numbered starting from the beginning of the video.

Multimodal (inter)action analysis was used to analyze the conversation transcription (see. Norris & Pirini, 2016). Norris's (2019, p. 164) suggestion that "a lower-level mediated action does not ever exist by itself" was also considered in the data analysis because generally, humans would produce the utterance in higher-level action (since it always involved many different modes (e.g., spoken, gesture, and facial expression). We categorized data based on available engagement sites (e.g., text chat, voice chat, or video chat). Then, learners' utterances displaying higher-level action were coded by turn-taking to address both research questions (see Appendix A & Appendix C). A microanalysis of lower-level action interplay (e.g., verbal cues, gestures, emojis, or images) within a particular higher-level action turn (e.g., task closing, negotiation

of target words, or content discussion) was conducted to enable an understanding of how multimodal alignment had been achieved and the contribution of each mode to the success of the LX communication, as in Appendix B (e.g., Wigham & Satar, 2021). An example of hierarchy between higher- and lower-level action in task interaction can be seen in Figure 2. Moreover, to categorize higher-level actions, we adapted the discourse functions of synchronous communication employed by Hampel and Stickler (2012) in their study such as social interaction, on-task negotiation of meaning, off-task conversation, and technical discussion. We used these categories as our initial analysis but modified them by specifying social interaction patterns generated from the data. That is, we broke down the categories into *social interaction* (task opening and task closing), *on-task negotiation meaning* (negotiation of meaning, negotiation of words, negotiation of agreement), *off-task conversation*, and *technical discussion*. We also identified additional functions as suggested by Liang (2010) such as *task management*, *error correction,* and *content discussion* (see Appendix A).

Figure 2

The Hierarchy of Higher- to Lower-Level Action in Task Interaction



To illustrate, we used the interactive alignment coding scheme derived from Dao et al. (2018) that classified the alignment based on the utterance produced by the speakers (prime), which is then reused in the next following turns (target) as shown in the following example:

A: . . . uh ***the guy who wants to steal the money*** (Prime)

B: Ok I think the first is ***the man who …wear…wear glasses*** (Target)

In this example, speaker B adopted a similar structure to that produced by speaker A in the previous utterance (relative clause), illustrating the interactive alignment pattern of primes → target sequence in terms of structure. The repetition occurred between speakers (alignment to interlocutor) or within the same speakers' utterances (self-alignment). Michel and Cappellini (2019) suggested that multimodal alignment might arise if the utterances produced by learners contained verbal and nonverbal cues in either prime or target utterances. However, due to the space constraint, we will only discuss and present the multimodal interactive alignment at two engagement sites, text, and video chat. All names of learners displayed below are pseudonymous.

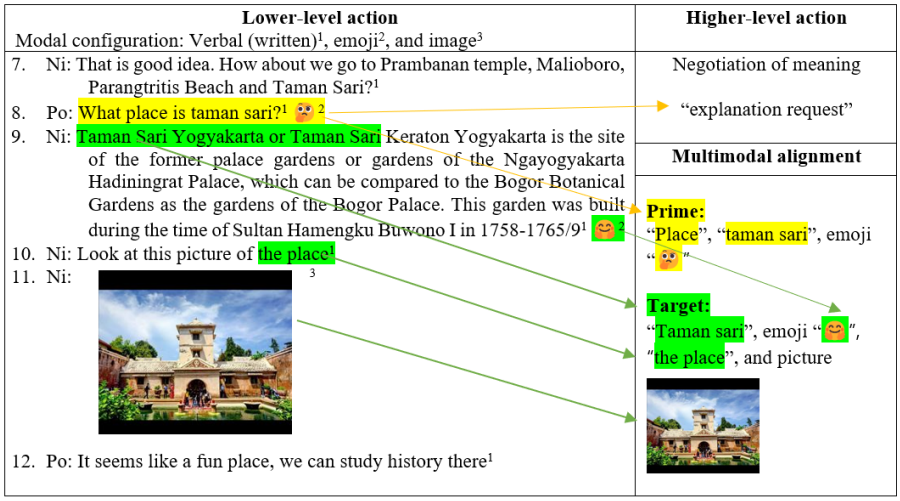
Figure 3 further demonstrates the multimodal interactive alignment coding scheme of higher-level mediated action of "negotiation of meaning". In this conversation, learners utilized three communication modalities (verbal, emojis, and pictures). One of the learners (Poppy [PO]) displayed a prime pattern by incorporating lower-level actions of verbal written and thinking face emoji (verbal + emoji) in turn 8. Noraini (NI) attempted to make her lower-level actions of verbal utterances and the choice of nonverbal cues aligned with those of her interlocutor. In this example, learners discussed and decided upon tourism destination sites for their holiday (task 3). NI recycled the words "Taman Sari" and "place" as her verbal lower-level actions in turns 9, 10, and 11 to align her language with her partner's. Besides this, NI also included a

picture of "Taman Sari" as her nonverbal lower-level action in turn 11, along with a detailed description in turn 9, to respond to PO's question indicated by a 🤔 (thinking face) emoji.

NI's lower-level actions, in turns 9, 10, and 11, both verbal and nonverbal, are examples of multimodal alignment to the question proposed by her partner PO. From this example, it seems that the verbal mode (written) has high modal intensity since it plays a great role in this alignment while nonverbal (emoji and images) modes serve to bolster visual representation of the message delivered.

Figure 3

Example of Multimodal Interactive Alignment Coding Scheme in Text Chat



Findings

Based on the total 64 task performances (40 text chats and 24 video chats), constituting 2743 turns, the multimodal analysis of chat transcriptions indicated that learners exhibited multimodal interactive alignment while producing higher-level actions in various ways. Some learners did it by *replicating emojis, reproducing GIFs, mimicking facial expressions*, and others performed by *imitating proxemics, facial expressions, and gestures*. We summarized the frequencies of interactive alignment that occurred across higher-and lower-level actions in Table 2.

Table 2

The Frequency of Multimodal Interactive Alignment Occurred Across Higher- and Lower-Level Actions

Higher-level action	Frequencies			
	Replicating emoji	**Reproducing GIFs**	**Mimicking facial expressions**	**Imitating proxemics, facial expressions, and gestures**
Task opening	9	0	2	3
Negotiation of meaning	0	0	0	1

Negotiation of agreement	1	0	0	0
Negotiation of target words	0	0	0	1
Content discussion	19	2	6	8
Error correction	1	0	0	0
Task closing	8	2	5	8
Off task conversation	0	0	1	0

For more specific details, in the following sections, we present examples of how learners employed multimodal cues to form interactive alignment in particular higher-level actions.

Alignment Through Replicating Emojis

Learners seemed to collaboratively use emoji features along with verbal cues to express their feelings and moods during task interaction. They utilized the appropriate emoji to amplify their verbal messages. The combination of both verbal and nonverbal modes indicated that interactive alignment occurring in the task opening is crucial for learners to show a phatic expression and a friendly face to their partners before starting an Instagram chat, as shown in Figure 4.

Figure 4

Excerpt of Task Opening (Task 2)

Lower-level action Modal configuration: Verbal (written), emoji, and GIF	**Higher-level action**
1. RY: hello guys 2. RY: 😁 3. EV: Hii 😊 4. ST: hai 5. RY: have you got a picture of each? 6. EV: Yes, I have 7. ST: Yes 8. RY: Okay, who wants to describe the picture first? 9. RY:  10. EV: I will tell about my picture	Task opening

In Figure 4, learners demonstrated nonverbal cues contextually to align with the context of the interaction in the task opening. The exchange took place in a group of three. Learners discussed the correct order of six random pictures to make a complete story. Rudi (RY) started the conversation by greeting all learners (turns 1-2) after receiving the task procedures from the facilitator. He displayed the combination of two modes of verbal cues: "hello guys" and 😁 (beaming face with smiling eyes emoji) to show his greeting by emphasizing a happy feeling (see Appendix D, for identified nonverbal functions). This prime pattern of verbal cue + nonverbal cue was reused by Elva (EV) in turn 3, as she posted "hi" + 😊 (smiling face with smiling eyes emoji). EV captured the positive, phatic signal and aligned her utterance by reusing the same communication pattern. Although Sinta (ST) attempted to align with EV by reusing "hai", it did not align with the feeling of happiness and friendliness sent through emojis because she did not respond to the signal

in the same way. Further, RY showed self-alignment with his previous utterances by repeating the same pattern of verbal + nonverbal cues in turn 8-9, when he initiated a topic and invited others to describe their assigned pictures by posting "okay, who wants to describe the picture first?" and continued with a GIF showing a man asking "Who?" with open hands.

Another example occurred when learners discussed the reasoning gap task in a group in Figure 5. They had to choose only 12 kilograms (kg) of the essential survival kit from the list given for their journey into the middle of the rainforest. Fit (FI) answered the question in turn 23, but she was unsure. She thought that their baggage was maybe 10 kg and used a 🤔 (thinking face) emoji to tell the others that she was not sure about the weight of the baggage. By adding this emoji, she wanted the others to recheck the importance of their added baggage to reduce weight. Meanwhile, in turn 24, Denisha (DE) asked the others to add something to their baggage because they still had 2 kg of space left. DE also added a 🤔 (thinking face), indicating that she did not know the item that should be added; she wanted the others to suggest it. Duta (DU), in turn 25, suggested bringing a half packet of biscuits, and DE, in turn 26, agreed to the suggestion with a 😆 (smiling face with open mouth and tightly closed eyes) emoji to indicate that she was happy with DU's recommendation. It was apparent that learners aligned their emoji use in turns 23 and 24. The learners' verbal messages with nonverbal "thinking face" emojis strengthened the illocutionary force of their messages and assisted the negotiation of the agreement (Li & Yang, 2018).

Figure 5

Excerpt of Negotiation of Agreement (Task 4)

Lower-level action Modal configuration: Verbal (written) and emoji	**Higher-level action**
22. DU: how much was our total baggage? 23. FI: I think maybe 10 kg? 🤔 24. DE: then want to bring what else? 🤔 25. DU: how about ½ packets of biscuits? 26. DE: yes of course I agree 😍	Negotiation of Agreement

Alignment Through Reproducing GIFs

Learners were observed to strategically select an appropriate GIF within a specific context during their online conversations. Since Instagram text chat features enable learners to use GIFs, learners had to carefully choose the GIFs cues to build coherence within the context. Figure 6 below illustrates how learners used GIFs to close the reasoning gap task through text chat. This task required them to decide on a tourism destination for their New Year's Eve holiday. After agreeing to go to Raja Ampat Island, Irina (IR) proposed the time for the trip in turn 20. She showed her joy and enthusiasm with a GIF displaying a moving car with a "HAPPY HOLIDAY" phrase in turn 21. Nita (NT), in turn 22, agreed with the time offered by her partner by saying, "Okey good idea". NT then posted a GIF showing a flying jet plane to align with Irina's visual cue. From this extract, it was observed that NT carefully selected a flying jet plane to achieve interactive alignment with IR's utterance, which showed enthusiasm for starting the trip. The use of a flying jet plane GIF functions as a visual co-speech demonstration to strengthen her own talk stating an interest in starting the travel as soon as possible (see Tolins & Samermit, 2016).

Figure 6

Excerpt of Task Closing (Task 3)

Lower-level action Modal configuration: Verbal (written), emoji and GIF	**Higher-level action**
19. NT: So when we would be leaving? 20. IR: 2 days before the new year we otw, so that our new year's eve will be there 21. IR:  22. NT: Okey good idea, I will prepare our holiday to the maximum 😊 23. NT:  24. IR: See you tomorrow 😘	Task closing

Alignment Through Mimicking Facial Expressions

Another alignment shown in the online video chat was through mimicking facial expressions. As shown in Figure 7 below, Feline (FE) and Ayla (AY) were paired up to describe and guess the names of the fictional characters in the pictures they were given. In turn 75, 76, and 77, they negotiated the character's name in one picture. In this phase, AY needed to think about the clue to guess the character's name correctly. In turn 76, AY assumed the character to be "Pinocchio", which FE confirmed and emphasized the clue for this character as the long nose. In turn 78, AY responded by saying 'long nose' while moving her index finger from her nose to the screen. In turn 79, seeing her partner's action, FE laughed, and AY also laughed with her. In this conversation, the learner changed the tone from being serious to light-hearted through laughing together. During an exchange, Uzum (2010) noted that learners sometimes develop their conversation style to align with others. In this case, AY aligned her facial expression to change the tone of her conversation with that of her partner by laughing together to not only lighten up the mood but also maintain the flow of the conversation.

Figure 7

Excerpt of Content Discussion (Task 7)

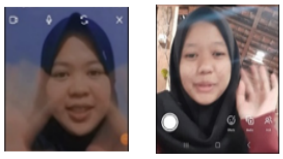
Lower-level action Modal configuration: Verbal (spoken), facial expression, and gesture	**Higher-level action**
75. FE: Okay the last is . . when I am lying my nose getting long, so even you lying your nose getting long. who I am? you can guess 76. AY: Uhm Pinocchio (FE vs. AY)  FE: Looks at the screen while describing the specific clues of the fictional character and she looks at to the right side AY: Looks at the screen and pays attention seriously to her partner's description and moves her index finger in front of her face forward to the screen when saying 'long nose'	Content discussion
77. FE: Yes, that's true because the clue is with nose and 78. AY: Nose long 79. FE: Yes, that's true  FE: Looks at the screen while laughing AY: Looks at the screen and laughs aligning with her partner's laugh	

Alignment Through Imitating Proxemics, Facial Expressions, and Gestures

The analysis also showed that learners used the nonverbal cues of proxemics, facial expressions, and gestures to align with their interlocutor to foster communication. In the example of Figure 8, learners closed their discussion by thanking each other for being cooperative during task completion in turns 12-16. They expressed thanks in turns 13-14 and leave-taking turns 15-16. Noraini (NI), in turn 13, closed the conversation by giving positive remarks on the task that they had just finished. Alevi (AV) agreed and praised her partner before leaving the conversation by saying, "you did really good job, thank you" and showed a thumbs up. Aligning to AV's gesture, NI raised her thumbs when saying, "yeah, thank you". AV also used a waving hand gesture when she left the conversation, saying "bye bye". NI aligned with these gestures by waving her right hand and saying, "bye bye". This excerpt provided a good example of the pattern of multimodality achieved by the learners. They initiated the prime verbally then finally completed it nonverbally using facial expressions and gestures. This showed that learners collaboratively aligned with each other by using both facial expression and gestures as well as their utterances. The common sequence pattern observed was the prime (verbal + nonverbal) → target (verbal + nonverbal), which was also common in face-to-face conversation (Dings, 2014).

Figure 8

Excerpt of Task Closing (Task 1)

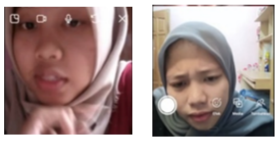
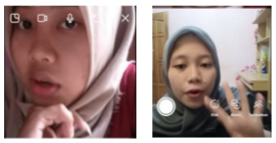
<table>
<tr><th>Lower-level action
Modal configuration: Verbal (spoken), facial expression and gesture</th><th>Higher-level action</th></tr>
<tr><td>12. NI: Yes, that's right. It is a meaningful story actually
13. AV: Oh... yes, I also think the same ... way. Okay, I think that's all for our task today. You did really good job. Thank you
14. NI: Yeah, thank you
(AV vs. NI)
 
AV: Looks at the screen with a smile. shows a thumbs up while saying 'you did really good job, thank you'
NI: Looks at the screen with a smile and shows a thumbs up aligning her partner's compliment</td><td rowspan="2">Task Closing</td></tr>
<tr><td>15. AV: Bye bye
16. NI: Bye bye
 
AV: Looks at the screen and waved both her hands to the screen while saying 'bye bye'
NI: Looks at the screen and waves her right hand aligning her partner's leave-taking</td></tr>
</table>

Moreover, learners also demonstrated interactive alignment by employing the proxemics of head and body movements with questioning faces for requesting clarification from their partner(s) in the negotiation of meaning. In Figure 9, learners were required to describe to their partner the jumbled short videos assigned to them and discuss the correct order of the videos to create a complete story. Denisha (DE) showed a questioning face and queried the information in turn 63. She again strategically moved her head closer to the camera and showed a questioning face (prime) to request more clarification when she could not get adequate responses to her queries in turns 65 and 67. Hansa (HS) aligned her proxemics to the prime

displayed by DE by moving her head close to the screen with a questioning face (target) and requesting clarification to DE's questions in turn 66. She also used iconic gestures to represent the word 'fourth', putting her four fingers up close to the camera in turn 68 (see McNeill, 2012). In this excerpt, HS aligned her proxemics, facial expressions, and gestures to ask and respond to her partner's clarification which is in line with the study by Oben and Brône (2016) which revealed that during an interaction, the speaker tended to adjust and match their verbal and nonverbal cues to the interlocutors.

Figure 9

Excerpt of Negotiation of Meaning (Task 1)

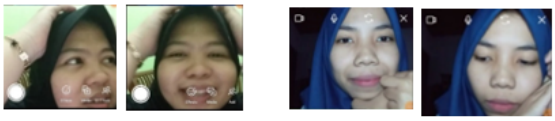
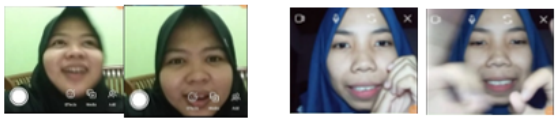
Lower-level action Modal configuration: Verbal (spoken), facial expression and gesture	**Higher-level action**
63. DE: Okay okay I remember. You say is the third or the fourth video? 64. HS: Yeah 65. De: The third or the fourth sequence 66. HS: Pardon (DE vs. HS) 	Negotiation of Meaning
67. DE: Third or fourth? you said that the third video can be the fourth 68. HS: Fourth yes yeah 69. DE: The last or . . . 70. HS: The last I think 71. DE: Okay I agree with you  De: Looks at the screen and places her hand under the chin Hs: Looks at the screen, listens to her partners' questions, moves her head a bit to the camera while saying 'pardon' with a questioning face, and raises her four fingers aligning her partner's clarification request	

Finally, multimodal alignment also occurred when learners attempted to search for target words by priming the interlocutor's facial expression and employing gestures to corroborate another speaker's confirmation check. In Figure 10, learners were required to do an information gap task through video chat, where they described the pictures of fictional characters for their counterparts to guess the characters' names on the pictures. The first alignment flourished from turn 45 when Alevi (AV) felt perplexed by the characters: "oh my God, I am not sure what is that?", and she placed her hands on her head, showing a confused face and looking upward. AV continued demonstrating her iconic gestures and facial expression of confusion in turns 47, 49, and 51. This prime influenced Amal (AM) to align with her interlocutor. AM changed her gestures and facial expression to seek other clues by looking downward with a thinking face. She put her hand across her right cheek and chin between turns 50 and 52. In this excerpt, we observed nonverbal alignment as AM adjusted her gestures and facial expression to accommodate her partner in negotiating the target words. The prime displayed by AV was received as a signal by AM to make more effort in searching for clues to describe the fictional character so that AV could retrieve the target words and finally achieve the task goal. The second alignment was discovered when AM performed the iconic hand gesture of "OK" in turn 56 to corroborate Alevi's confirmation check "Is that Batman?" in turn 55, and it aligned with Alevi's previous utterance in turn 53 "Okay, the last name is bla bla man." In this case, the verbal cue of "okay" was a prime for the iconic hand gesture "OK". The finding above verified that learners often used nonverbal cues of proxemics, gestures, facial expression, and emojis to align with their interlocutor in negotiation during online dialogue. During communication breakdown, gestures were crucial to assist with giving more information and resolve misunderstandings for negotiating meaning in an online conversation

(see Lee et al., 2019).

Figure 10

Excerpt of Negotiation of Target Words (Task 7)

Lower-level action Modal configuration: Verbal (spoken), facial expression and gesture	Higher-level action
45. AV: Oh my God, I am not sure what is that 46. AM: I think 47. AV: Black body? 48. AM: Yeah 49. AV: Black body, Uhm. what is that? 50. AM: What is the clue. Uhm. Black body Vi. 51. AV: I don't know what is that? (AV vs. AM)  AV: Puts both her hands on the heads and looks up somewhere while saying 'Oh my God'. She moves her head to the right and left and finally faces the camera with a smiling face while saying 'I don't know what is that?' AM: Looks at the screen smiling and touches her chin with her hand. she looks down and places her hand on her cheek. It seems she aligns with her partner confusion by attempting to find the words as a clue.	Negotiation of Target Words
52. AM: He is a hero like iron man. With the last name is man too. 53. AV: Okay, the last name is bla bla bla man 54. AM: Yeah yeah yeah 55. AV: Is that Batman? 56. AM: Yeah, that's it, Oh my god  AV: Looks up with a smiling face and she looks at the screen with a questioning face screen while saying 'Is that Batman?' AM: Looks at the screen and places her hand on her cheek. She moves both her hand to create an OK shape while saying 'Yeah, that's it' aligning to her partner confirmation check	

Discussion and Implications

The findings revealed that learners displayed multimodal interactive alignment in the CMC tasks they attempted on Instagram (RQ1). They used priming mechanisms in an online conversation where they imitated each other's verbal and nonverbal cues (Zhou & Wang, 2021). Naturally, speakers tended to align their language during interaction because of "the automatic tendency of interactants to reuse each other's morphosyntactic structures and lexical choices" (Michel & Cappellini, 2019, p. 189). Multimodal alignment using nonverbal cues such as emojis, gestures, facial expressions, gaze, and GIFs can also be naturally aligned in form and function. Learners managed to reveal quite similar forms of nonverbal such as waving hands, smiling facial expressions, or thinking faces. However, they interacted at different sites of engagement in text and video chats. This shows that learners can strategically use nonverbal elements during interaction based on their function in the discourse.

Alignment performed by the learners served as a small part of the big picture of verbal and nonverbal modes used to create a particular discourse in the interpersonal CMC setting. It was displayed through learners' gestures, gaze, proxemics, searches for target words, and means of reaching an agreement. The study confirms previous findings that gestures and gaze, as additional visual support, are conducive to input

enhancement by making conveyed meaning more comprehensible, thus helping the interlocutor understand messages correctly (Lee et al., 2019; Satar, 2013). In addition, emojis can build rapport by assisting learners in expressing their feelings in the text chat (Vandergriff, 2014). Further, GIFs in text chat helped learners attain communicative fluidity. They could organically choose the best means to convey meaning and emotion in real-time despite not seeing the interlocutors' faces (Lim, 2015). In addition, this evidence of multimodal alignment also provides information for filling the gap left by the previous research, which analyzed alignment mainly from a linguistics point of view, including the alignment that occurred at the lexical and structural level (see Zhou & Wang, 2021).

The findings also indicated that learners used various modes, aside from verbal cues, in their interactive alignment in CMC tasks via Instagram (RQ2). The modes were diverse and based on the availability of each engagement site such as emojis, GIFs, and images during text chat interaction, and gestures, facial expressions, and proxemics for video chat. These modes collectively built a meaningful conversation in context. Although the salient findings revealed that the verbal mode had high intensity of usage in task interaction compared to the nonverbal mode, nonverbal cues afforded learners to enact more positive emotion and task engagement in the conversation and alignment. For example, 😊 (smiling face with smiling eyes), and hand waving gestures conveyed positive and friendly signals at the beginning and end of the conversations, building a positive atmosphere and increasing understanding between speakers (Li & Yang, 2018; McNeill, 2005). In other cases, gestures, proxemics, and facial expressions complemented attention and assisted interlocutors in meaning negotiation (Lee et al., 2019). This provides evidence that in the natural setting outside the classroom (e.g., social media), learners might adapt and adjust their communicative behavior to reach their communicative goal in the environment by utilizing any semiotic resources available to them at the engagement site. This finding echoed Atkinson (2014) that language learning is a holistic process of humans fulfilling their social action by utilizing language and other semiotic resources as communicative tools within the environment.

The implications of multimodal alignment for language learning are twofold. First, evidence of multimodal alignment proves that learners continuously adapt their cognition and behavior to their environment. They can use the possible semiotic resources offered by the engagement site (Instagram) to reach their communicative goal through adaptation and adjustment to the environment (including technological tools and interlocutors' utterances) for the LX communication purpose. Language learning is not limited to a specific setting (e.g., classroom, school); however, it can take place in any social moment as long as learners can engage in the environment that promotes target language use. Hence, learners can learn the target language outside the classroom informally by interacting with people around the globe through CMC as part of their everyday activities.

Secondly, multimodal alignment strengthens the vital role that nonverbal cues play in LX online pedagogy. The findings encourage language educators to acknowledge the use of nonverbal cues along with verbal cues in learners' interaction in formal and informal language learning contexts. Indeed, to create meaningful communication, humans need to chain their utterances to one another both verbally and nonverbally (Oben & Brône, 2016). Since the CMC environment has limited contextual cues compared to face-to-face settings, the use of nonverbal elements is crucial to preserve the flow of dialogue, maintain smooth communication, and help interlocutors communicate effectively (Lim, 2015; Satar, 2013; Uzum, 2010). Such aspects encourage learners to be more engaged with learning and lead them to achieve better learning outcomes. However, despite the insights into the role of multimodality in achieving alignment, this study is not without limitations. Firstly, it only looks at Instagram as a communication platform within the Indonesian context, and secondly it includes only a small number of homogenous learners. Looking ahead, it needs to be further explored how language learners from various cultural backgrounds, and/or with a higher English language proficiency level, utilize the nonverbal cues available to them on Instagram to accomplish CMC tasks.

Conclusion

This study aimed to enrich and broaden our understanding of LX interactive alignment in the CMC tasks

interaction context, particularly in the case of Instagram as one of the three biggest communication apps for the young generation in Indonesia. The findings shed light on many aspects. First, they reveal and support the notion of interactive and multimodal alignments as central to LX interaction online, with interaction being key to learning (Allwright & Hanks, 2009; Long, 2015). Second, they demonstrate how learners manage and adapt to the new technological features of their virtual learning contexts to overcome the problems of not being face-to-face and add value to their means of communication. Learners can strategically use their language, embodied actions, and the affordances of the available technological tools to achieve the communicative goals of the task successfully. They can use various nonverbal and verbal cues for their communication in different channels such as text chat (e.g., emojis, images, and GIFs) and video calls (e.g., gestures, proxemics, and facial expression). The use of semiotic resources such as emojis, GIFs, and images compensate for the absence of visual cues in text chat. In addition, nonverbal cues in video chat help them to convey their emotion and this aspect is particularly integral for remote learning and teaching amid the pandemic.

These semiotic resources enable learners to tap into multimodality, thus minimizing the psychological (virtual) distance that is usually felt in distance learning. They also provide additional learning support for language learners besides 2D textual chat interaction. Moreover, employing multimodal (inter)action analysis as a research tool might contribute to the development of current CMC research within the SLA context (Wigham & Satar, 2021). This study lends empirical support for and explanation about how language teachers can maximize the affordances of new communication technology features and encourage students to tap into multimodality (e.g. emojis) in order to support their comprehension and interaction through (a)synchronous online exchanges such as videoconferencing (Gutiérrez et al., 2021). Lastly, the findings expand on the existing phenomenon of interactive alignment and document evidence of learners' multimodal alignment, which may have remained unnoticed without this study.

Acknowledgements

We would like to thank the anonymous reviewers for their constructive feedback on the earlier version of this article. This project was funded by *Lembaga Pengelola Dana Pendidikan* (LPDP) in cooperation with the Ministry of Religious Affairs, Republic of Indonesia under *"Beasiswa Indonesia Bangkit"* and Curtin University, Australia.

References

Aghayi, A. A., & Christison, M. (2021). Instagram as a tool for professional learning. In K. Kelch, P. Byun, S. Safavi, & S. Cervantes (Eds.), *CALL theory applications for online TESOL education* (pp. 82–99). IGI Global. https://doi.org/10.4018/978-1-7998-6609-1.ch004

Allwright, D., & Hanks, J. (2009). *The developing language learner: An introduction to exploratory practice*. Palgrave Macmillan.

Atkinson, D. (2011). A sociocognitive approach to second language acquisition: How mind, body, and world work together in learning additional languages. In D. Atkinson (Ed.), *Alternative approaches to second language acquisition* (pp. 143–166). Routledge.

Atkinson, D. (2014). Language learning in mindbodyworld: A sociocognitive approach to second language acquisition. *Language Teaching, 47*(4), 467–483. https://doi.org/10.1017/s0261444813000153

Calvo-Ferrer, J. R., Melchor-Couto, S., & Jauregi, K. (2016). ReCall special issue: Multimodal environments in CALL editorial multimodality in CALL. *ReCALL, 28*(3), 247–252. https://doi.org/10.1017/s0958344016000136

Costa, A., Pickering, M. J., & Sorace, A. (2008). Alignment in second language dialogue. *Language and*

Cognitive Processes, 23(4), 528–556. https://doi.org/10.1080/01690960801920545

Danesi, M. (2017). *The semiotics of emoji: The rise of visual language in the age of the Internet*. Bloomsbury

Dao, P., Trofimovich, P., & Kennedy, S. (2018). Structural alignment in L2 task-based interaction. *ITL - International Journal of Applied Linguistics, 169*(2), 293–320. https://doi.org/10.1075/itl.17021.dao

Dewaele, J.-M. (2017). Why the dichotomy 'L1 versus LX user' is better than 'native versus non-native speaker. *Applied Linguistics*. https://doi.org/10.1093/applin/amw055

Dings, A. (2014). Interactional competence and the development of alignment activity. *The Modern Language Journal, 98*(3), 742–756. https://doi.org/10.1111/j.1540-4781.2014.12120.x

Ellis, R. (2018). *Introducing task-based language teaching*. Curtin University.

Erarslan, A. (2019). Instagram as an education platform for EFL learners. *TOJET: The Turkish Online Journal of Educational Technology, 18*(3), 54–69. http://www.tojet.net/articles/v18i3/1835.pdf

Faraco, M., & Kida, T. (2008). Gesture and the negotiation of meaning in a second language classroom. In S. G. McCafferty & G. Stam (Eds.), *Gesture: Second language acquisition and classroom research* (pp. 280–297). Routledge.

Giannoulis, E., & Wilde, L. R. A. (2020). Emoticons, kaomoji, and emoji: The transformation of communication in the digital age. In E. Giannoulis & L. R. A. Wilde (Eds.), *Emoticons, kaomoji, and emoji: The transformation of communication in the digital age* (pp. 1–22). Routledge.

Guichon, N., & Cohen, C. (2016). Multimodality and CALL. In F. Farr & L. Murray (Eds.), *The Routledge Handbook of Language Learning and Technology* (pp. 509–521). https://doi.org/10.4324/9781315657899

Gutiérrez, B. F., Glimäng, M. R., O'Dowd, R., & Sauro, S. (2021). *Mentoring handbook for virtual exchange teachers: Strategies to help students achieve successful synchronous and asynchronous online intercultural communication*. Stevens Initiative. https://www.stevensinitiative.org/resource/mentoring-handbook-for-virtual-exchange-teachers/

Hampel, R., & Stickler, U. (2012). The use of videoconferencing to support multimodal interaction in an online language classroom. *ReCALL, 24*(2), 116–137. https://doi.org/10.1017/s095834401200002x

Jewitt, C. (2014). An introduction to multimodality. In C. Jewitt (Ed.), *The Routledge handbook of multimodal analysis* (2nd ed., pp. 15–30). Routledge.

Jones, R. H., & Norris, S. (2005). *Discourse in action: Introducing mediated discourse analysis*. Routledge. https://doi.org/10.4324/9780203018767

Kim, Y., Jung, Y., & Skalicky, S. (2019). Linguistic alignment, learner characteristics, and the production of stranded prepositions in relative clauses. *Studies in Second Language Acquisition, 41*(5), 937–969. https://doi.org/10.1017/s0272263119000093

Kress, G., & Van Leeuwen, T. (2001). *Multimodal discourse: The modes and media of contemporary communication*. Arnold

Lee, H., Hampel, R., & Kukulska-Hulme, A. (2019). Gesture in speaking tasks beyond the classroom: An exploration of the multimodal negotiation of meaning via Skype videoconferencing on mobile devices. *System, 81*, 26–38. https://doi.org/10.1016/j.system.2018.12.013

Li, L., & Yang, Y. (2018). Pragmatic functions of emoji in internet-based communication--A corpus-based study. *Asian-Pacific Journal of Second and Foreign Language Education, 3*(1), 1–12. https://doi.org/10.1186/s40862-018-0057-z

Liang, M.-Y. (2010). Using synchronous online peer response groups in EFL writing: Revision-related

discourse. *Language Learning & Technology, 14*(1), 45–64. http://doi.org/10125/44202

Lim, S. S. (2015). On stickers and communicative fluidity in social media. *Social Media + Society, 1*(1). https://doi.org/10.1177/2056305115578137

Long, M. H. (2015). *Second language acquisition and task-based language teaching*. Wiley-Blackwell.

Maa, J., & Taguchi, N. (2022). Using L2 interactional-pragmatic resources in CMC: A case of Japanese orthography and emoji. *Language Teaching Research, 26*(2), 190–212. https://doi.org/10.1177/13621688211064934

McNeill, D. (2005). *Gesture and thought*. University of Chicago Press.

McNeill, D. (2012). *How language began: Gesture and speech in human evolution*. Cambridge University Press.

Michel, M., & Cappellini, M. (2019). Alignment during synchronous video versus written chat L2 interactions: A methodological exploration. *Annual Review of Applied Linguistics, 39*, 189–216. https://doi.org/10.1017/s0267190519000072

Michel, M., & Smith, B. (2018). Measuring lexical alignment during L2 chat interaction: An eye-tracking study. In S. M. Gass, P. Spinner, & J. Behney (Eds.), *Salience in second language acquisition* (pp. 244–268). Routledge.

Negueruela, E., & Lantolf, J. P. (2008). The dialectics of gesture in the construction of meaning in second language oral narratives. In S. G. McCafferty & G. Stam (Eds.), *Gesture: Second language acquistion and classroom research* (pp. 88–106). Routledge.

Nishino, T., & Atkinson, D. (2015). Second language writing as sociocognitive alignment. *Journal of Second Language Writing, 27*, 37–54. https://doi.org/10.1016/j.jslw.2014.11.002

Norris, S. (2004). *Analyzing multimodal interaction: A methodological framework*. Routledge.

Norris, S. (2011). *Identity in (inter)action: Introducing multimodal (inter)action analysis*. De Gruyter Mouton.

Norris, S. (2019). *Systematically working with multimodal data: Research methods in multimodal discourse analysis*. John Wiley & Sons.

Norris, S., & Pirini, J. (2016). Communicating knowledge, getting attention, and negotiating disagreement via video conferencing technology: A multimodal analysis. *Journal of Organizational Knowledge Communication, 3*(1), 23–48. https://tidsskrift.dk/jokc/issue/view/3562

Nurhayati-Wolff, H. (2021). *Share of Instagram users in Indonesia as of April 2021, by age group*. Statista. https://www.statista.com/statistics/1078350

O'Halloran, K. L. (2004). Introduction. In K. L. O'Halloran (Ed.), *Multimodal discourse analysis: Systemic functional perspectives* (pp. 1–10). Continuum.

Oben, B., & Brône, G. (2016). Explaining interactive alignment: A multimodal and multifactorial account. *Journal of Pragmatics, 104*, 32–51. https://doi.org/10.1016/j.pragma.2016.07.002

Olsher, D. (2008). Gesturally-enhanced repeats in the repair turn: Communication strategy or cognitive language learning tool? In S. G. McCafferty & G. Stam (Eds.), *Gesture: second language acquisition and classroom research* (pp. 109–130). Routledge.

Pickering, M. J., & Garrod, S. (2004). Toward a mechanistic psychology of dialogue. *Behavioral and Brain Sciences, 27*(2), 169–226. https://doi.org/10.1017/s0140525x04000056

Prabhu, N. S. (1987). *Second language pedagogy*. Oxford University Press.

Prasetyawati, O. A., & Ardi, P. (2020). Integrating Instagram into EFL writing to foster student

engagement. *Teaching English with Technology, 20*(3), 40–62. https://tewtjournal.org/volume-2020/issue-3/

Rosborough, A. (2014). Gesture, meaning-making, and embodiment: Second language learning in an elementary classroom. *Journal of Pedagogy, 5*(2), 227–250. https://doi.org/10.2478/jped-2014-0011

Satar, H. M. (2013). Multimodal language learner interactions via desktop videoconferencing within a framework of social presence: Gaze. *ReCALL, 25*(1), 122–142. https://doi.org/10.1017/s0958344012000286

Satar, H. M. (2016). Meaning-making in online language learner interactions via desktop videoconferencing. *ReCALL, 28*(3), 305–325. https://doi.org/10.1017/s0958344016000100

Satar, H. M., & Ozdener, N. (2008). The effects of synchronous CMC on speaking proficiency and anxiety: Text versus voice chat. *The Modern Language Journal, 92*(iv), 595–613. https://doi.org/10.1111/j.1540-4781.2008.00789.x

Satar, H. M., & Wigham, C. R. (2017). Multimodal instruction-giving practices in webconferencing-supported language teaching. *System, 70*, 63–80. https://doi.org/10.1016/j.system.2017.09.002

Tolins, J., & Samermit, P. (2016). GIFs as embodied enactments in text-mediated conversation. *Research on Language and Social Interaction, 49*(2), 75–91. https://doi.org/10.1080/08351813.2016.1164391

Uzum, B. (2010). An investigation of alignment in CMC from a sociocognitive perspective. *CALICO Journal, 28*(1), 135–155. https://doi.org/10.11139/cj.28.1.135-155

Vandergriff, I. (2013). Emotive communication online: A contextual analysis of computer-mediated communication (CMC) cues. *Journal of Pragmatics, 51*, 1–12. https://doi.org/10.1016/j.pragma.2013.02.008

Vandergriff, I. (2014). A pragmatic investigation of emoticon use in nonnative/native speaker text chat. *Language@Internet, 11*(article 4), 1–17. https://www.languageatinternet.org/articles/2014/vandergriff

Veszelszki, A. (2015). Emoticons vs. reaction-Gifs non-verbal communication on the internet from the aspects of visuality, verbality and time. In A. Benedek & K. Nyíri (Eds.), *Beyond words: Pictures, parables, paradoxes* (pp. 131−145). Peter Lang.

Wigham, C. R., & Satar, H. M. (2021). Multimodal (inter)action analysis of task instructions in language teaching via videoconferencing: A case study. *ReCALL, 33*(3), 195–213. https://doi.org/10.1017/s0958344021000070

Zhang, X. (2017). Reading–writing integrated tasks, comprehensive corrective feedback, and EFL writing development. *Language Teaching Research, 21*(2), 217–240. https://doi.org/10.1177/1362168815623291

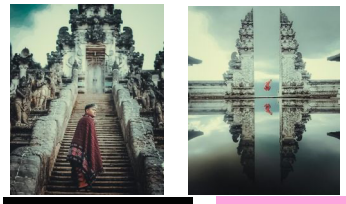
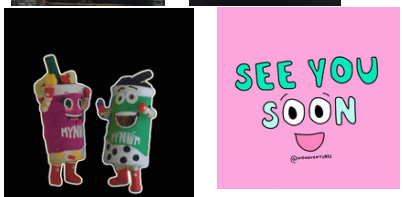
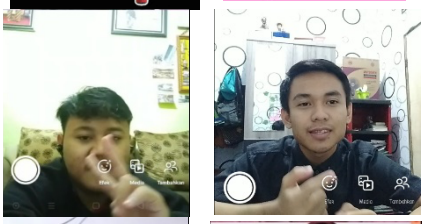
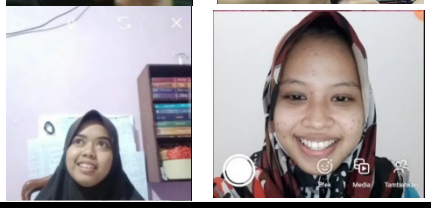
Zhou, X., & Wang, C. (2021). Effects of interactive alignment on L2 vocabulary learning by Chinese EFL learners. *Language Teaching Research*. https://doi.org/10.1177/13621688211004629

Appendix A. Definition and Example of Higher-Level Actions

Higher-level action	Definition	Example
Social interaction (1&2)		
1. Task opening	Opening moves in task discussion	AV: Hallo AM: Hallo AV, how are you today? AV: How are you today? Eh, I am fine, I am fine. How about you? AM: I am doing great
2. Task closing	Closing moves in task discussion	AV: Okay, good job, thank you AM AM: Thank you AV, see you next time AV: Bye, goodbye AM: Bye bye
On-task negotiation meaning (3, 4, & 5)		
3. Negotiation of meaning	Moves where learners check understanding or ask for clarification/explanation of the meaning	DU: And the second is uhm the . . . it is come from Disney again, it is love story other uhm . . . two person AF: Two persons DU: Boy and girl it's very famous writing by William Shakespeare, what…? AF: love story . . . Romeo and Juliet, that's right? DU: That's right
4. Negotiation of agreement	Moves where learners make requests for agreement	DE: How about the second day we go to Malioboro? RY: It sounds good. And then on the 3rd day what if we go to Borobudur temple? DE: Yes, I agree
5. Negotiation of target words	Moves where learners make requests for a clue to find/retrieve specific words	AW: That's right, that's right, good good. And then we next to the third character is about the one of the family of avengers. He have a . . he have a hammer, hammer, you know? AW: In Indonesia, hammer is *Palu Palu*. You know? ST: Thor
6. Content discussion	Moves where learners propose opinion, thought, comment, or response to the negotiations.	DE: From the first video, in the first video I saw there was a grandmother and a man sitting on the chair beside of the road and I saw a package of cookies in the middle of them and but there was still one cookie left HS: Yeah DE: And then the man took the cookie HS: Uh-huh

7. Task management	Moves where learners talk about task requirements and procedure	NI: Okay for the task seven we will describe the picture from MT and we will guess the name of the image and you will go first and I go PO: Yes
8. Technical action	Moves where learners talk about technical issues	RY: So guys, what channel do we want to use? HA: eemm maybe text chat FR: text chat
9. Error correction	Moves where learners correct others or themselves	AM: Yess! And I think my las pictures too AM: *last
10. Off-Task conversation	Moves where learners talk about an issue outside the required task	FE: My favorite superhero FE: Is FE: Wonder woman FE: Wkwkwk AY: Yes like a wonder woman 🤣

Appendix B. Definition and Example Lower-Level Actions

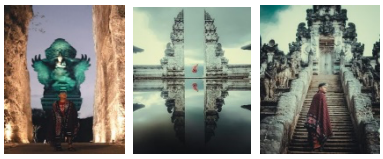
Lower-level action	Definition	Example
1. Written verbal	Written language posted in the text chat	FI: I think I will also bring 5 kg of white rice
2. Spoken verbal	Spoken language posted in the audio chat or uttered in video chat	IR: "Uhm, love problem. I think Romeo and Juliet"
3. Emoji	Pictogram or ideogram posted in the text chat	😆 👏 😁 🤣
4. Image	Digital image of a thing posted in the text chat	
5. GIF	Animated pictures representing feelings or actions posted in the text chat	
6. Gesture	Hand movement to express an idea or meaning usually accompanying speech in the video chat	
7. Facial expression	Expression of one's face to convey meaning in the video chat	

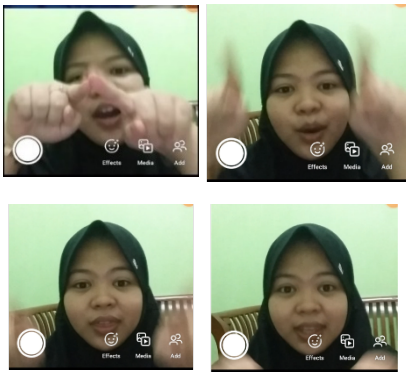
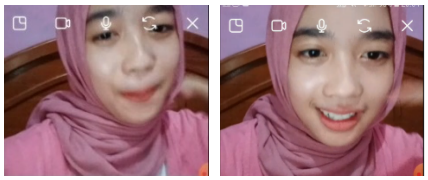
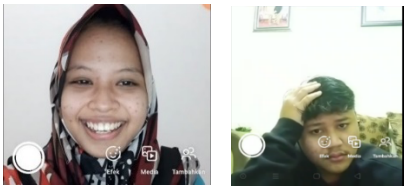
Appendix C. Frequencies of Lower- and Higher-Level Actions

Lower-level action	Frequencies
Written	1672
Spoken	1315
Emoji	428
GIFs	36
Images	16
Gestures	258
Facial expressions	550

Higher-level action	Frequencies
Task opening	211
Task closing	157
Task management	178
Negotiation of meaning	29
Negotiation of agreement	17
Negotiation of target words	21
Content discussion	2291
Technical action	52
Error correction	14
Off task conversation	17

Appendix D. Function of Nonverbal Cues in Synchronous Communication*

Nonverbal cue	Function	Example
1. Emoji	Emotion signal (to show speaker's attitude or emotion)	TI: I totally agree, we can stock photos to post on our Instagram feed 😄
	Emotion intensity enhancer (to emphasize the speaker's emotion or attitude)	NT: If I were a child, I would be very sad. The condition of the loss of family attention or lack of parental affection is very painful 😭
	Illocutionary force modifier (to lessen the illocutionary force of the speaker's messages)	HQ: Stove? To cook the rice FA: Yeah of course, how can we eat the rice without cooking them 😂
	Backchannel device (to shorten the response or as a conversation closure)	IR: Okay see you too guys 🙌 EK: 👋👋👋 IR: 🤷👋
2. GIF	Co-speech demonstration (to provide visual elaboration of speaker's own talk)	NT: Very sad when I discuss picture B NT: 
	Affected response (to visually represent a response to the interlocutor's prior talk)	EV:  SH: Okay Ev 🌈👋
3. Image	Adding information with visual	DU: its good, we can go to kuta beach, melasti, GWK and many place in Bali DU: 

4. Gesture	Iconic gesture (to present images of concrete entities and/or actions)	AV: He always brings like uhm. I don't know how to explain it, a round thing. 
	Deictic gesture (pointing hand or any extensible body/held object to locate entities or actions)	IR: Ok . . uhm . . next . . uhm, the black mask . . character has a . . . black cloth or sayap eh apa in on the back . . . black mask 
5. Facial expression	Expressing emotion	

Note. * Adopted from Li and Yang (2018), Tolins and Samermit (2016), and McNeill (2005)

About the Authors

Muntaha Muntaha is a doctoral student in Applied Linguistics Program, School of Education at Curtin University, Australia. His research interests include technology-supported language teaching/learning, multimodality in language teaching/learning, and task-based instruction for language teaching/learning. Muntaha Muntaha is the corresponding author.

E-mail: muntaha.abdulqodir@postgrad.curtin.edu.au

ORCiD: https://orcid.org/0000-0001-7795-7167

Julian Chen is an Associate Professor of Applied Linguistics/TESOL and Course Coordinator of Asian Languages at the School of Education, Curtin University. Julian's research involves technology-mediated task-based language teaching, 3D virtual learning, netnography, teacher identity and action research.

E-mail: julian.chen@curtin.edu.au

ORCiD: https://orcid.org/0000-0001-7788-0462

Toni Dobinson is an Associate Professor at Curtin University where she coordinates and teaches the Post Graduate Programmes in Applied Linguistics. She is also Discipline Lead for Applied Linguistics/TESOL and Languages. Her research interests include language teacher education, language and identity, language and social justice and classroom research.

E-mail: T.Dobinson@curtin.edu.au

ORCiD: https://orcid.org/0000-0003-1790-0016