Discourse Processes

ISSN: (Print) (Online) Journal homepage: www.tandfonline.com/journals/hdsp20

The role of pointing gestures and eye gaze in second language vocabulary learning

Paula Janjić, Gozdem Arikan, Harmen B. Gudde, Joseph J.C. Murphy, Laksha Sivaram & Kenny R. Coventry

To cite this article: Paula Janjić, Gozdem Arikan, Harmen B. Gudde, Joseph J.C. Murphy, Laksha Sivaram & Kenny R. Coventry (08 May 2024): The role of pointing gestures and eye gaze in second language vocabulary learning, Discourse Processes, DOI: 10.1080/0163853X.2024.2343625

To link to this article: https://doi.org/10.1080/0163853X.2024.2343625

Published online: 08 May 2024.

DISCOURSE PROCESSES
https://doi.org/10.1080/0163853X.2024.2343625

Routledge
Taylor & Francis Group

The role of pointing gestures and eye gaze in second language vocabulary learning

Paula Janjić[a], Gozdem Arikan[a], Harmen B. Gudde[a,b], Joseph J.C. Murphy[a], Laksha Sivaram[a], and Kenny R. Coventry[a]

[a]School of Psychology, University of East Anglia, Norwich, UK; [b]Helmholtz Institute, Department of Experimental Psychology, Utrecht University, Utrecht, The Netherlands

ABSTRACT

Learning a second language is recognized as a necessity for social, political, and economic development. However, the processes contributing to initial vocabulary learning have not been explicated. In a series of experiments, this study examines the role of deictic gestures and gaze in second language vocabulary learning. Such cues have been shown to be fundamental in first language learning, but their efficacy in second language learning has not been established. In three experiments 435 participants learned pseudowords by watching images of a teacher naming objects placed on a table while systematically manipulating pointing and gaze. Moreover, manipulating the position of the object relative to the teacher (within or out of reach) served to establish the possible importance of these cues as social versus attentional constructs in second language vocabulary learning. Results show that gaze and gesture did not affect vocabulary learning, but object position did. We discuss implications of these results for theories of first language and second language vocabulary learning.

Introduction

One of the first challenges learners come across in second language (L2) learning is acquiring vocabulary. The dominant linguistic approach to vocabulary learning treats language as an informationally-encapsulated system in which learners need to acquire distributional information within a language (Mitchell et al., 2019). However, language theories recognize language as social, embodied, and multimodal (Barsalou, 2008; Coventry & Garrod, 2004; Pulvermüller et al., 2005), emphasizing how language is deeply situated in a physical and communicative context (Murgiano et al., 2021). A prerequisite to developing a functional communicative context is establishing joint attention between interlocutors. This is typically achieved using pointing gestures and eye gaze, cues that are regarded as fundamental to language processing and learning (Goldin-Meadow, 2007; Tomasello et al., 2005).

It has been established that shared attention and presence of multimodal cues such as eye gaze affect language processing (Grzyb & Vigliocco, 2020; Kreysa et al., 2018). In development, understanding pointing gestures and eye gaze is tightly coupled to the emergence of joint attention, which sets the stage for early language acquisition (Brooks & Meltzoff, 2008; Goldin-Meadow, 2007). Numerous studies have shown that gaze and pointing in infancy are correlated with later language and vocabulary growth (e.g., Brooks & Meltzoff, 2008; Çetinçelik et al., 2021; Law et al., 2012; McGillion et al., 2017).

The role of attention and the accompanying subprocesses of alertness, orientation, and detection have also been identified as crucial in the context of L2 learning (Tomlin & Villa, 1994). Likewise,

CONTACT Paula Janjić p.janjic@uea.ac.uk; Kenny R. Coventry k.coventry@uea.ac.uk School of Psychology, University of East Anglia, Norwich Research Park, Norwich NR4 7TJ, UK

theories of L2 learning have become more situated, multimodal, and embodied (Atkinson, 2010; Douglas Fir Group, 2016; Mitchell et al., 2019). However, the topic of eye gaze and pointing gestures in L2 vocabulary learning has not been much considered. To date, research in L2 word learning relating to gestures has mostly been focused on iconic gestures (García-Gámez & Macizo, 2020; Macedonia et al., 2011; Morett, 2014).

The main goal of the present study is to examine the role of eye gaze and pointing gestures in L2 vocabulary learning. Specifically, we systematically vary gaze and pointing gestures of a "teacher" naming objects placed on a table in front of him or her in a vocabulary learning task, where participants learn pseudowords they hear in combination with presented objects. To tease apart the mechanisms involved in possible effects of gaze and gesture on word learning, this series of studies varied where the named object was positioned on the table with respect to the teacher (within peripersonal [reachable] vs. extrapersonal [nonreachable] space) to establish whether the social/interactional component of deictic communication facilitates word learning or if gesture/eye gaze merely draw attention to an object (Caldano & Coventry, 2019; Coventry et al., 2008; Rocca et al., 2019). We also varied the number of objects presented on the table to examine whether nonlinguistic cues are especially helpful in word learning when the referent is potentially ambiguous. Next, we briefly review evidence for the role of nonverbal cues in first language (L1) and L2 learning before detailing the proposed experiments.

Role of eye gaze and pointing gestures in L1 processing and learning

The ability to establish joint attention in infancy by using and following eye gaze and pointing gestures is a crucial step in children's language acquisition (Goldin-Meadow, 2007; Tomasello et al., 2005).

Such "deictic cues" have been hypothesized to promote learning by enhancing the child's attention to the referent and/or by enhancing the child's understanding of the speaker's intentions (Booth et al., 2008). Çetinçelik et al. (2021) support the notion that eye gaze plays a greater role than simply acting as a mere deictic pointer, such as arrows. In a systematic review, they argue that it facilitates infants' learning by enhancing their alertness, memory, and attentional capacities, which potentially benefit the encoding process during word learning in a unique way.

The importance of multimodal cues varies as a function of stage of development (Golinkoff & Hirsh-Pasek, 2006). For example, Paulus and Fikkert (2014) found that 14- to 24-month-old infants paid more attention to an agent's eye gaze when learning to map a novel word to a referent, whereas older children and adults relied more on pointing cues, suggesting there is a developmental change in the use of different cues during word learning. In 28- to 31-month-olds, however, Booth et al. (2008) showed that gaze and gesture together led children to learn more new words compared to the single cue condition.

The number of studies relating to gaze and pointing gestures in vocabulary learning beyond infancy is sparse. Bang and Nadig (2020) tested children aged 6 to 11 years by comparing gaze to a nonsocial directional cue (e.g., a pointing arrow) in word learning. Children learned from both cues equally well but with qualitative differences in the way they attended to the gaze as opposed to the arrow cue. In the gaze condition, children spent more time looking at the cued area and engaged in more back-and-forth looking between the referent and the cue compared to the arrow cue condition.

In language processing in adults, indexical cues such as eye gaze and pointing gestures appear together with speech, with evidence that the combination of nonverbal and verbal cues facilitates language comprehension by reducing potential ambiguity and cognitive effort (Holler & Levinson, 2019; Holler et al., 2014; Murgiano et al., 2021). Knoeferle et al. (2018) propose three possible mechanisms by which gaze may enhance language comprehension. First, it could serve the role of a simple pointer to a relevant location in the same way that an arrow directs attention to a specific object or location ("deictic account"). Alternatively, the presence of social attention in the form of speakers' gaze could help listeners more successfully link spoken words to their referents by making object representations more grounded and salient ("referential account"). In that way

gaze does not only direct attention but also contributes to encoding processes through enriching the listener's mental representations. Finally, the "syntactic-thematic account" suggests that gaze informs not only deictic and/or referential processes but also other comprehension processes such as syntactic structure building and thematic role assignment by highlighting certain parts of the utterance.

While eye gaze and pointing are both clearly fundamental in early L1 vocabulary learning, their role in L2 learning has received less attention (Cooperrider & Mesh, 2022). If L2 learning is based on similar processes as L1 (the "fundamental continuity hypothesis"; Ellis & Wulff, 2015; Krashen, 1988; MacWhinney, 2012), it would be expected that eye gaze and gesture are instrumental in L2 learning as well.

Eye gaze and pointing gesture in L2 learning

Newly emerging sociocognitive and sociocultural approaches to L2 learning place stronger emphasis on multimodality, embodiment, and the situated nature of L2 learning (Atkinson, 2010; Douglas Fir Group, 2016; Ortega, 2013), leading to a greater focus on nonlinguistic cues during L2 learning. Multiple L2 learning studies (Comesaña et al., 2012; García-Gámez & Macizo, 2020; Poarch et al., 2015) have explored aspects of multimodality in L2 vocabulary learning within the framework of the Revised Hierarchical Model (Kroll & Stewart, 1994). In this model, learning of new words in L2 is mediated by words in L1, which are used as a bridge to reach the underlying concepts. As learners increase their proficiency, they access concepts directly from L2, without the need for mediation from L1 words, thus enhancing retrieval and understanding.

Accordingly, a focus on strengthening the links between new words and concepts should enhance initial vocabulary learning. This can be achieved through presenting multimodal materials, such as pictures (Morett, 2014) and gestures (García-Gámez & Macizo, 2020; Macedonia et al., 2011) that enrich the linguistic stimuli, consistent with Dual-Coding Theory (Paivio, 1971), or through mental manipulations of words in activities like categorizing and describing, in line with the Levels of Processing framework (Craik & Lockhart, 1972). For example, Paivio proposed that learners create mental representations through both linguistic and nonlinguistic channels. Therefore combining linguistic elements with additional cues enriches the learning experience and facilitates better retention of vocabulary.

However, eye gaze and pointing gestures—key cues in L1 vocabulary learning and processing that guide attention—have thus far not been considered, although attention itself has been shown to be a crucial part of L2 learning (Tomlin & Villa, 1994). In the classification of L2 teachers' nonverbal behavior, Allen (1999) pointed out that 82% of all teachers' communication is nonverbal. Consistent with these findings, native English speakers spontaneously produce more deictic and representational gestures when speaking to L2 English learners as compared to English L1 speakers (Adams, 1998 as cited in Morett, 2014). Iconic gestures have been found to facilitate instructed L2 vocabulary learning across different age groups and settings (Allen, 1995; Macedonia et al., 2011; Morett, 2014).

Deictic gestures in L2 learning have been studied thus far only in classroom L2 grammar learning (Matsumoto & Dobs, 2017; Stam & Tellier, 2022) and in the context of L2 vocabulary learning where different nonhuman agents (e.g., virtual agents, robots, avatars) act as tutors (Belpaeme et al., 2018; Bergmann & Macedonia, 2013; Demir-Lira et al., 2020; Tsuji et al., 2021; Vogt et al., 2019). However, findings have been inconsistent, with the final emphasis put on the need to further investigate the role of specific cues and their interactions.

Present studies

In this series of experiments the goal was to examine the role of pointing gestures and gaze in adult foreign language vocabulary learning. Participants learned new vocabulary (nouns in a pseudolanguage) through watching an instructor in varying conditions of gaze and gesture.

Specifically, the experiments included the impact of eye gaze, pointing, and the combined effects of pointing plus gaze on noun learning compared to a baseline condition.

Based on the relevance of the social cues during processing and learning observed in past (L1) studies (Booth et al., 2008; Caruana et al., 2021), the hypothesis was that word learning performance will be optimal in the gaze and pointing condition. Both in adults (Caruana et al., 2021) and in children (Booth et al., 2008), the appearance of these two cues simultaneously was shown to enhance language processing and/or learning compared to when they would appear individually. Differences were hypothesized to arise between each individual cues as well. It was predicted that pointing gestures might lead to better outcomes than gaze only (Paulus & Fikkert, 2014). Since gaze can be used to both send and receive social information, it is more ambiguous than a pointing gesture, which indicates both the communicative intent and the referent more explicitly (Caruana et al., 2021). Finally, the condition with no cue was expected to lead to the poorest outcomes.

The series of experiments was planned in advance, with later experiments contingent on the results of the previous one. The decision tree with the main experiment and associated follow-up experiments as well as the data and stimuli are available on OSF (https://osf.io/q5uzw/).

Experiment 1

The first experiment was set up to investigate the role of gaze and pointing gestures in the basic learning condition where each item to be learned was individually placed in front of an instructor (Figure 1). The procedure was based on different vocabulary learning studies that use similar methodology (García-Gámez & Macizo, 2020; Morett, 2014).

Methods

Participants

Participants were monolingual English-speaking adults between 18 and 30 years old, with normal or corrected-to-normal vision and no record of learning disabilities or hearing impairments. Participants were recruited through online recruitment platforms. Students were recruited via the institutional recruitment system (Sona Systems, https://www.sona-systems.com/) and were rewarded in credits in accordance with the university's regulations. Participants from the Prolific platform were compensated financially, following the Prolific guidelines. The target sample size of 145 paticipants was

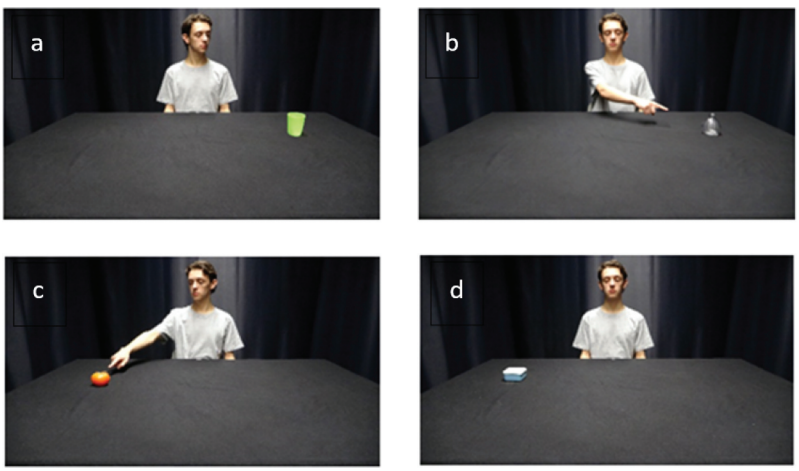

Figure 1. Examples of stimuli illustrating the four main conditions: gaze, pointing only (eyes closed), gaze + pointing, and no cue (eyes closed).

determined based on Brysbaert and Stevens' (2018) sample size recommendations for a power of 90%, alpha of 0.05, and effect size of d = .4 (r = .2). This effect size has consistently been shown to be the average effect size in psychological science and is therefore considered to be a reliable guide when setting up new research (Brysbaert & Stevens, 2018). We note that this target sample size is noticeably larger than similar studies in the field.

Design and procedure

Participants learned pseudowords in an online experimental setting (using the Gorilla Experiment Builder; www.gorilla.sc). Participants saw pictures of target items with a simultaneous auditory presentation of the pseudoword followed by its translation to English (as in previous studies; Macedonia et al., 2011; Morett, 2014). Both the pseudoword and translation were presented in the auditory modality, whereas images were presented with no text. Across four within-participant conditions, the instructor's cue(s) toward the target object in the image was varied to show either gaze (Figure 1a), a pointing gesture (eyes closed; Figure 1b), both gaze and a pointing gesture (Figure 1c), or no cue at all (closed eyes and straight head direction from the teacher; Figure 1d). In the gaze-only condition, the head was positioned to be aligned with the direction of gaze, thereby adding to the saliency, visibility, and ecological validity of the cue, since the perception of gaze depends also on the orientation of the head (Böckler et al., 2015). Participants learned 32 words, distributed across four conditions of interest, each condition consisting of 8 words. To avoid effects of confounding variables and potential carryover effects, the order of conditions was randomized, and words were not tied to a certain condition but appeared in all four conditions across participants. Also, there were two instructors and different pointing hands. Throughout the whole experiment each participant would see a randomly assigned instructor that would point with both hands.

Experimental stimuli

Picture stimuli. Thirty-two stimuli were selected based on two existing datasets of objects (Brodeur et al., 2014; Geusebroek et al., 2005), initially resulting in 250 objects across 4 categories and 10 subcategories. The list of objects was initially narrowed down by having two independent native English-speaking raters remove the items that stood out in terms of size (objects that would be too large to fit on a table or objects that would potentially be too small to identify, e.g., blueberry), asymmetry (asymmetrical objects or objects with a handle and tools were removed due to their effect on visual processing), and familiarity (less familiar objects or objects that are named by using two words, e.g., hair dryer, toilet paper), leaving a total of 108 items. The final selection was made based on psycholinguistic properties of the 108 remaining words, matching them based on frequency, age of acquisition, familiarity, imageability, and concreteness taken from available norms (SUBTLEX-UK; van Heuven et al., 2014) for frequency and The Glasgow Norms (Scott et al., 2019) (Table 1).

Pictures were taken of the 32 resulting items, half of which belonged to the "food" category and the other half to "household items" (Appendix A). The instructors (one man, one woman) and item images were created separately and merged in Photopea (www.photopea.com) editing software. Items were positioned to the left and right of the instructor, equally distanced from the middle point of the table, where the instructor was seated. The items were positioned to always be within the reach of the pointing hand. The pointing hand therefore always corresponded to the side on which the item was positioned on the table. The dimensions of the table were 122 cm × 166 cm, and the distance of the items from the middle point of

Table 1. Psycholinguistic properties of the selected stimuli.

	Mean	*SD*	Min	Max
Frequency	4.334	0.470	3.318	5.210
Age of acquisition	2.553	0.542	1.529	3.382
Familiarity	6.199	0.412	5.286	6.783
Imageability	6.553	0.237	5.971	6.933
Concreteness	6.542	0.231	5.912	6.903

the table edge depended on the height and arm length of the instructors (60 cm × 67 cm). The camera was centered 30 cm from the other side of the table, at a height of 97 cm. For the purpose of counterbalancing, all different condition combinations we created resulted in 512 different pictures that showed the 32 items across two positions (left and right) with four different cue conditions (gaze, gesture, gaze + gesture, and no cue) and two instructors (man and left-handed, women and right-handed). The images that included gaze as a condition also included a congruent head turn (such that gaze and head direction were identical). This was done to ensure visibility/salience of the gaze cue, thus minimizing ambiguity related to the cue direction. Moreover, having the head follow the gaze is more natural and ecologically valid than having only the eyes moving from left to right while having the head face forward. This decision was supported by the literature showing that head turn is an essential part of eye gaze (Atkinson et al., 2018; Böckler et al., 2015., Langton et al., 2000).

Pseudowords stimuli. A list of 32 pseudowords was created in the Wuggy software (Keuleers & Brysbaert, 2010). The words were four to seven letters long and contained two to three syllables. Three native English raters ensured the words did not bear an obvious resemblance to any existing English words. The final 32 pseudowords were randomly assigned to the target items (see Appendix B for a final list). The audio files were created in Google's text-to-speech software.

Procedure

Participants were informed that they were taking part in an experiment targeting learning of words in a new language. They were briefed on the set up of the experiment, its duration, and structure that consisted both of a learning and a testing part. Before starting, participants were presented with practice trials. At the beginning of the learning phase, participants were presented with a list of the words to be learned and asked to read through and retype them. The goal of this familiarization task was to facilitate learning outcomes during the following short learning session. The learning phase of the experiment took approximately 40 minutes in total, with each item presented for 6 seconds. The procedure was repeated twice within each of the four blocks corresponding to four cue conditions. The whole cycle across the blocks was presented three times with a 90-second break in between (see trial procedure in Figure 2). As an additional task to control for the attention, the participants were asked to press the space bar every time they spotted a mismatch between the word and the image. For every dozen (real) trials, a mismatch trial would appear once at a random point. The items or pseudowords that would appear in the attention checks were different from the target stimuli to avoid interference in learning.

The testing phase took place immediately after the learning phase and consisted of two tasks, the first one targeting production of newly learned words and the second one testing their recognition. In the production task, participants were shown the target items and were asked to name the stimuli by typing their answer into the bracket below the picture. In the recognition task, instead of writing down the response, participants chose between four possible options. The incorrect options were the other pseudowords that were included in the learning session. The experiment ended with a brief questionnaire asking about the learning strategies used during the experiment.

Figure 2. Trial procedure that introduced an instructor with a straight gaze (0.5 s), followed by a target appearing in one of two positions on the lateral plane along with the instructor's cue (depending on condition) and audio naming the item and its translation to L1.

The outcome measure was accuracy score (binary: correct or incorrect). To account for potential spelling mistakes in the production tasks, answers that contained a spelling error of one Levenshtein's distance from the target answer (addition, substitution, or omission of one letter) were still scored as correct.

Results and discussion

Participants' scores were included in the analysis only if they scored above chance level in the recognition task (over 25% correct answers), passed the attention checks during the experiment, did not self-report use of any additional strategies in the final questionnaire (e.g., writing words down), and were monolingual. After applying the exclusion criteria to 195 recruited participants, the final sample consisted of 145 monolingual English-speaking participants between 18 and 30 years old (M = 24.07; SD = 5.11) who were exposed to 32 pseudowords during the learning session. Table 2 shows the learning outcomes on the production and recognition tasks.

To analyze the data and investigate the effect of different cues on vocabulary learning, generalized linear mixed models (Baayen, 2008) with binomial error structure and logit link function were used. Pointing, gaze, and their interaction were included in the model as fixed effects with two levels (absent and present), and the intercepts of items and participants were included as random effects factors to account for the repeated observations of the same individuals and items. The outcome measure was the accuracy score (correct or incorrect). The contrast coding method used was sum coding.

The p values were estimated using the significance criterion of $\alpha < .05$. In analysis of each experiments' data, two separate models were run for two testing tasks (production task and recognition task).

The analysis showed no effects of pointing (production: $F(1, 4464) = .701$, $p = .403$; recognition: $F(1, 4465) = .191$, $p = .662$), gaze (production: $F(1, 4463) = .860$, $p = .354$; recognition: $F(1, 4464) = .417$, $p = .519$), and their interaction (production: $F(1, 4463) = 2.849$, $p = .092$; recognition: $F(1, 4464) = .520$, $p = .471$) on L2 word learning outcomes. A table with the summary statistics for both models can be found in Appendix C.

This set of results did not confirm the hypotheses, which expected to find effects of cue. One potential explanation for the absence of effects of cues could be that a single object was used in the first experiment in each trial. Some studies on L1 processing and learning found that the effect of gaze differs depending on whether or not it is used to disambiguate a message (Kreysa et al., 2018; Macdonald & Tatler, 2013; Jachmann et al., 2019). Having only one item present in the scene potentially made the cues redundant for the adult L2 vocabulary learners, since attention was already directed toward the target object that stood out in the scene even without the cue. For that reason, the second experiment presented two objects on each trial.

Table 2. Mean percentage correct scores, SDs, and SEs on each of the four conditions for both production and recognition tasks in the first experiment.

	Production task			Recognition task		
Condition	Mean	*SD*	*SE*	Mean	*SD*	*SE*
No cue	36.9%	0.483	0.015	75.1%	0.433	0.013
Gaze only	33.2%	0.471	0.015	74%	0.439	0.013
Pointing only	35.7%	0.479	0.015	74.3%	0.437	0.013
Gaze + pointing	36.4%	0.481	0.015	74.5%	0.436	0.013

Experiment 2

Following the null results from Experiment 1, where one item was presented during learning, the planned modification for the second experiment was to include an additional item in the learning setup. This would create ambiguity, where receiving a cue would potentially bring a more tangible advantage to the learner. The hypotheses remained the same.

Methods

Participants

One hundred seventy-two participants were recruited through the Sona Systems and Prolific recruitment platforms. After excluding the participants based on the same exclusion criteria as in the previous experiment, the final sample consisted of 146 monolingual English-speaking participants between 18 and 30 years old (M = 24.6; SD = 4.98).

Experimental stimuli

The 32 target items remained the same as in the previous experiment. To create ambiguity, distractor items were added into the learning setup. The distractor item was always one of the objects appearing in the same condition as the target object. They were counterbalanced to always appear for the same number of trials. The distractor items always appeared simultaneously with the target items and at the same distance from the instructor (Figure 3). The cues were always congruent with the position of the target item.

Procedure

The procedure was the same as described in the first experiment, only this time participants were told that two items would be on the table accompanied by audio indicating the correct word. They were also informed about the testing phase afterward as well about the task of identifying mismatches that would serve as an attention check in the analysis. Before starting, participants were presented with practice trials. The structure and the duration of the procedure remained the same as in the first experiment.

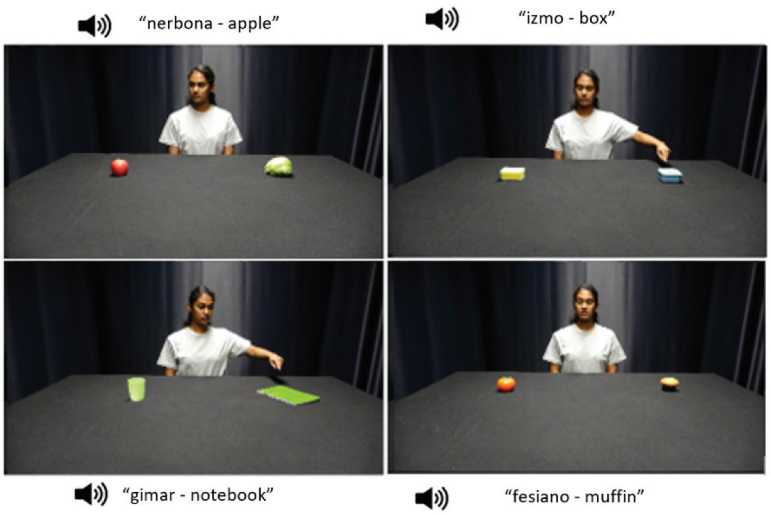

Figure 3. Images in four modified conditions for Experiment 2: gaze only, pointing only, gaze + pointing, and no cue.

Table 3. Mean percentage scores, SDs, and SEs on each of the four conditions for both production and recognition tasks in the second experiment.

	Production task			Recognition task		
Condition	Mean	*SD*	*SE*	Mean	*SD*	*SE*
No cue	30.8%	0.013	0.014	73.4%	0.442	0.013
Gaze only	34%	0.474	0.014	72%	0.449	0.013
Pointing only	34.4%	0.475	0.014	75.5%	0.431	0.013
Gaze + pointing	35.9%	0.48	0.014	71.4%	0.452	0.013

Results and discussion

The model was built in the same way as in the previous experiment. It included gaze and pointing and their interaction as fixed factors and their random intercepts. Table 3 shows the learning outcomes on the production and the recognition tasks.

The production data showed that none of the predictors had an impact on the dependent variable measuring the learning outcome (pointing, $F(1, 4326) = 1.086$, $p = .297$; gaze, $F(1, 2886) = 1.568$, $p = .211$; interaction pointing × gaze, $F(1, 4326) = .515$, $p = .473$). Please refer to Appendix D for the detailed results.

However, the analysis of the data from the recognition test showed a main effect of gaze, $F(1, 2283) = 11.12$, $p < .001$, $\eta^2 = 0.005$. The presence of gaze had a negative effect on the learning outcome, meaning that the presence of the gaze cue resulted in lower accuracy compared to the baseline (Table 4). There was no observed effect of pointing, $F(1, 4125) = .846$, $p = .358$, or the interaction, $F(1, 4122) = .018$, $p = .893$.

Although an effect of deictic cue on L2 vocabulary learning was expected, it was surprising to find only an effect of gaze, because among the three conditions involving cues (pointing, gaze, and pointing + gaze), gaze was hypothesized to have the weakest effect. Although small (d = −0.04), the direction of the effect was also unexpected, since we assumed that all cues would have a positive effect on learning.

One possible explanation for the hindering effect of gaze is that gaze may capture the attention of the learner, perhaps therefore distracting rather than facilitating attention toward the target object. Support for this possible interpretation comes from an eye-tracking study of word learning in school-aged children (6–11 years) in which the gaze condition produced a different pattern of eye movements compared to an arrow-cue condition (although it did it not produce a different learning outcome; Bang & Nadig, 2020). In the gaze condition,

Table 4. Summary of the generalized linear mixed-effects model for word learning outcomes measured by the recognition test (Experiment 2).

	Estimate	*SE*	Z value	*p* value	Lower CI (2.5%)	Upper CI (97.5%)
Fixed effects						
Intercept	1.570	0.190	8.268	<.001	1.199	1.957
Gaze	**− 0.330**	**0.106**	**− 3.106**	**.002**	− 0.542	− 0.119
Pointing	− 0.081	0.090	− 0.886	.376	− 0.259	0.099
Gaze × pointing	0.01	0.181	0.055	.956	− 0.351	0.372
	Variance	*SD*				
Random effects						
ID	2.633	1.623				
Item	0.483	0.695				
	R^2 marginal	R^2 conditional				
Model fit						
	0.005	0.489				

Model equation: model_r <- glmer(score ~ gaze × pointing + (1|id) + (1|item), data = total_recognition2, family = binomial()).

children engaged in more back-and-forth looking between the referent and the cue. The different pattern observed requiring additional time and drawing attention away from the target might explain the negative effect that it had on learning in the current study. When comparing to the other three conditions, in the gaze condition the cue was at the farthest point from the target item.

To further explore the mechanisms behind the unexpected negative effect of the gaze, a third experiment was conducted manipulating the distance of an object from instructors and the congruence or incongruence of cues. By varying the distance, it is possible to test if the cost of gaze indeed comes from the position of the target object compared to the cue, as speculated. Manipulating congruence would help to unpack the hindering effect of gaze and to test if it can be replicated in a similar experimental set up.

Experiment 3

The third experiment introduced incongruency between the deictic cue and the position of the target item to further elucidate the results from the previous experiments. Based on the previous results, we expected to find no difference between learning conditions in the number of words learned with congruent or absent cues, which would confirm that having gaze and gestures as cues during L2 word learning does not bring a facilitative effect compared to no cue at all. However, while our previous results indicate that gaze and pointing cues do not facilitate word learning, we wanted to establish if there might be a cost to word learning if cues are incongruent with the target object. If gaze is distracting when aligned with the target object (congruent), then we expected that such distracting effects might be magnified when the cues are pointing to the distractor object (i.e., incongruent with the target object). Thus, in line with the results of the previous experiment, it would be expected that the effect of congruency would show that the learning outcomes would be best in the condition with no cues where there is no distracting effect of gaze and the worst when the gaze is present and pointing to the distractor item (incongruent).

Following the results of Experiment 2 where the gaze cue had a negative effect on word learning outcomes, one explanation could be that the distance of the target objects from the cues was affecting word learning. The effect of pointing did not come across because the hand cue is much closer to the target object than the gaze cue, thereby avoiding the need to look back-and-forth. With the addition of the distance manipulation (within-reach vs. outside-reach), it could be established if the negative effect of gaze can be attributed to the close proximity of the target object that is drawing the learner's attention. Finding an interaction between distance and condition in this study could confirm whether the negative effect of gaze is due to distance between the face and target object. Moreover, whether an object is within reach of a speaker (in their peripersonal space) or not has been shown to be integral to spatial communication (Caldano & Coventry, 2019; Coventry et al., 2008). When an object is out of reach, we argue that it is more detached from the speaker, and hence it is more distal both physically and mentally (Piwek et al., 2008).

In summary, we expected to find a difference between the number of words learned depending on different cue-related conditions (congruent, incongruent, or absent) and different item positions (within-reach or outside-reach).

Methods

Participants

Participants were recruited and compensated in the same way as in the previous two experiments. After applying the same exclusion criteria on 188 recruited participants as in the previous experiments, the final sample consisted of 145 monolingual English-speaking participants between 18 and 30 years old (M = 22.9; SD = 6.69).

Experimental stimuli

The third experiment used a within-subjects design with two factors (3 × 2): cue type (with three levels: congruent, incongruent, and absent) and distance (with two levels: within-reach and outside-of-reach). Gaze and pointing appeared together, but when present, they were either congruent or incongruent with the target object. In the first two experiments, the objects always appeared within reach, whereas in Experiment 3 they appeared both within and outside of reach. Other than in the new (distal) position of the objects (Figure 4), the images were similar to those used in previous experiments.

The number of items was increased to fit the new 3 × 2 design with an equal number of items per cell. The new number of items was 36. The additional target items, as well as 20 new distractor items, were chosen from the same selection of words that was matched on different psycholinguistic properties for the purposes of the first two experiments.

Procedure

The instructions were modified to clarify that trials could have a mismatch occurring between the cue and the target item. Participants were told that ultimately the audio cue always correctly indicated the target item. For the attention check task, instead of spotting a mismatch as in previous experiments, participants pressed the space bar whenever they saw a blue dot over one of the items. To avoid the interference with learning of the target items, separate trials were introduced with additional words/items. The rest of the procedure remained the same as described before.

Results and discussion

One hundred forty-five participants were exposed to 36 pseudowords during the learning session. The model included cue type and distance as well as their interaction as fixed factors and participants and items as random factors. Table 5 shows the learning outcomes on the production and the recognition tasks.

In the third experiment the model was built with the fixed factors "cue type" and "position" as well as their interaction. The random effects remained the same as in the previous model, accounting for the variation in the intercept of items and participants. Sum coding was used to contrast the categorical variables, allowing us to interpret the model coefficients as deviations from the overall mean of the dependent variable. As can be seen in Table 5, the incongruent-outside condition produced poorer learning outcomes in both testing tasks. However, the analysis of the production task showed no significant effects (cue type: $F(2, 3248) = .360$, $p = .698$; position: $F(1, 1410) = 1.050$, $p = .306$) (Appendix E), while the analysis of the recognition task data showed a main effect of position, $F(1, 2114) = 6.547$, $p = .011$, $\eta^2 = 0.003$ (Table 6). Cue type (congruence) as a predictor was not found to be significant, $F(1, 3959) = 1.413$, $p = .244$.

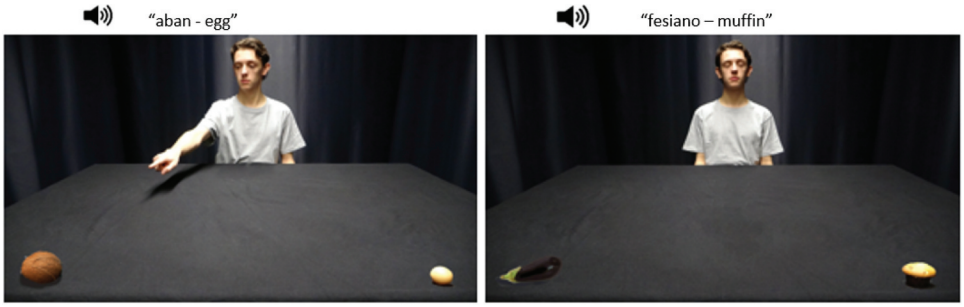

Figure 4. New (distal) position of the items combined with the incongruent and absent cues.

Table 5. Mean percentage scores, SDs, and SEs on each of the six conditions for both production and recognition tasks in the third experiment.

Condition		Production task			Recognition task		
Cue type	Position	Mean	*SD*	*SE*	Mean	*SD*	*SE*
Absent	Within	32.8%	0.470	0.016	74.4%	0.437	0.015
Congruent	Outside	30.2%	0.460	0.016	75.6%	0.430	0.015
Incongruent	Within	33.4%	0.472	0.016	74.9%	0.434	0.015
Absent	Outside	30.1%	0.459	0.016	71.5%	0.452	0.015
Congruent	Within	34.6%	0.476	0.016	75.9%	0.428	0.015
Incongruent	Outside	28.9%	0.453	0.015	62.1%	0.485	0.017

The results of Experiment 3 help to partially elucidate findings from the previous experiments. Although the expected interaction between the type of cue and position did not come out, a main effect of position showed that the words were learned better when they were in the instructor's peripersonal space. This could be related to the negative effect of gaze from the previous experiment. Since the gaze cue was most distant from the object (the pointing hand was in the previous experiment always close to the object), the distance of the cue from the target item might play a role in explaining the learning outcomes. Having an on-screen cue that is placed away from the object might require additional time to switch and direct attention from the cue to the target. An alternative possibility is that this was a result of the effect of peripersonal versus extrapersonal space, which was already shown to affect language processing (Caldano & Coventry, 2019) but might also affect language learning. Further experiments are needed to explore this avenue and disentangle the effects from distance from the effects of peripersonal versus extrapersonal space. It would be interesting to see whether the effect of distance would translate into nonsocial cues, such as arrows or pointers, which are often compared to social cues in word learning studies (Barry et al., 2015; Michel et al., 2019). Unexpectedly, the effect of congruence was not found to be significant. This could be due to the incongruent condition leading the participants to dismiss the cues altogether after realizing they are unreliable. Thus, one possible future modification could be to only use the congruent and the absent condition without using the incongruent condition. This would still serve to obtain the information if the negative effect of gaze can be replicated and lead to poorer learning outcomes compared to the no cue condition.

Table 6. Summary of the generalized linear mixed-effects model for word learning outcomes measured by the recognition test (Experiment 3).

	Estimate	*SE*	Z value	*p* value	Lower CI (2.5%)	Upper CI (97.5%)
Fixed effects						
Intercept	1.441	0.234	6.165	<.001	0.983	1.899
Cue Type (congruent)	0.184	0.115	1.610	.107	− 0.040	0.409
Cue Type (incongruent)	0.143	0.122	1.171	.242	− 0.096	0.382
Position (outside)	**0.330**	**0.127**	**2.602**	**.009**	**0.081**	**0.579**
Cue_type×Position (Congruent×Outside)	0.209	0.269	0.777	.437	− 0.318	0.735
Cue_type×Position (Incongruent×Outside)	0.156	0.237	0.667	.511	− 0.308	0.619
	Variance	SD				
Random effects						
ID	2.44	1.56				
Item	1.27	1.13				
	R^2 marginal	R^2 conditional				
Model fit						
	0.005	0.532				

Model equation: model_r <- glmer(score ~ cue_type × position + (1|id) + (1|item), data = total_recognition3, family = binomial()).

General discussion

In the context of overwhelming evidence showing the importance of gaze and pointing gestures in L1 learning (Booth et al., 2008; Golinkoff & Hirsh-Pasek, 2006), we set out to systematically investigate if the same attentional cues would also have an impact on adult L2 vocabulary learning. Across three experiments, participants had the task of learning (pseudo)words in a new language by viewing an instructor in an image naming and pointing and gazing (not at all, congruently, incongruently) at a target object. Despite the arguments behind the fundamental continuity hypothesis that similar processes are at play both in L1 and L2 learning (Ellis & Wulff, 2015; MacWhinney, 2012), the overall results of the experiments failed to find support for a facilitative effect of pointing gesture and gaze in L2 word learning.

The starting point for the series of studies was an experiment set up with only one target item present. In this experiment with null results, we suggested that the absence of an effect might be due to cues having a different effect when there is ambiguity in the learning situation (i.e., where the referent is unambiguous), in line with findings from studies on language processing (Kreysa et al., 2018). Accordingly, Experiment 2 introduced ambiguity with the presentation of two objects on each trial. However, this experiment once again confirmed the previous finding that gaze and pointing gesture do not provide an advantage for L2 word learning. On the contrary, the results showed an unexpected negative impact of gaze on word learning (compared to the no cue condition).

Experiment 3 manipulated target object distance from the instructor and congruence of cues (congruent, incongruent, absent) to further examine the (negative) effect of gaze found in Experiment 2 and to test if incongruence of cues might lead to a drop in word learning performance. No effects of cue type were found, failing to replicate the (negative) effect of congruent gaze in the previous experiment. However, the experiment did find a main effect of object distance, with word learning rates lower overall for objects distal to the instructor in the images. Below we explore explanations for the effects we found across the studies (distance, negative effect of gaze) before considering why we failed to find facilitative effects of gaze and pointing overall on L2 vocabulary learning.

The effect of distance in Experiment 3, taken together with the negative effect of gaze in Experiment 2, speak in favor of divided attention between the instructor in the image and the target object in the present experiments. When the object was further away from the instructor, the object was bigger on the screen and hence more visually salient, yet word learning rates were nevertheless lower than when the object was smaller and positioned within the reach of the instructor. It is well known that faces generally grab a lot of attention in images (e.g., Gliga et al., 2010; Ro et al., 2001; Theeuwes & Van der Stigchel, 2006). Therefore, it is possible that it was easier for participants to process objects placed closer to the instructor compared to dividing attention between the face of the instructor and the target object. The results of Experiment 2 are consistent with the view that the eyes are one of the most attention-grabbing facial elements, with open eyes diverting attention away from the target object. Indeed, there is much evidence for selective attention to eyes from early infancy, forming one of the bases of intention understanding (see, e.g., Langton et al., 2000). So open eyes attract attention more than closed eyes, taking away attention from processing the target object and hence reducing the likelihood of binding the new object name to the object presented. Of course, one way of testing this hypothesis would be to rerun the studies using eye tracking to see if distance and eye gaze indeed mediate looking time to the target object.

We now turn to the failure to find facilitative effects of instructor cues. One possibility is that gaze and gesture are important cues to L2 vocabulary learning, but only when learning takes place in more ecologically realistic situations. The use of static images in the present experiments meant that the dynamics of eye gaze and pointing were not captured. Still images omit important information about the dynamics of gaze and pointing that may influence action perception and therefore looking behavior (Haxby et al., 2020; Welke & Vessel, 2022). For example, the trajectory (prestroke, stroke,

poststroke hold, and retreat) of a referential gesture is both rich and elaborate, making it a reliable and strong cue to a referent (Shattuck-Hufnagel et al., 2016; Kendon, 1980; McNeill, 1992; Norris, 2011). So it might be predicted that using dynamic gesture and gaze might divert attention away from the face of the instructor to focus on the target object more decisively, thus potentially enhancing word learning. All that being said, many online L2 learning applications use static images, and therefore the present findings indicate that pointing and eye gaze may offer little aid to word learning in that context.

The alternative possibility is that the present results may be the first evidence that gesture and eye gaze, so central to L1 acquisition, may not facilitate L2 acquisition. There are many possible reasons why this might be the case. First, one of the key differences between learning vocabulary in L1 versus L2 is that the object name is known in the L1 when learning an L2, but that is of course not the case in L1 acquisition. A child acquiring language does not know a priori what a new word refers to. In that case a caregiver can use pointing and eye gaze to define for the child what the word does refer to. For instance, "car," "wheel," and "bonnet" can all be used to refer to a car, and pointing and eye gaze can be essential to disambiguate what is being referred to. In contrast, in L2 acquisition the word "car" can be used to make it clear what is being referred to, obviating the need for gaze and gesture to do that work. Indeed, in the present experiments, the object name was always given in English at the start of each trial, making it clear what the new word to be learned referred to. While participants processed the images (performing above chance on the catch trials and affected by distance in Experiment 3), the heard word may have reduced the need for attentional visual processing during the task.

Second, it may be the case that cues are only effective for some types of processing. For example, in an eye-tracking study that looked at the effects of speaker's gaze on L1 comprehension and processing (Kreysa et al., 2018), gaze was found to help direct attention to the target object more rapidly but did not lead to improved memory performance. This might suggest that a speaker's gaze might be of limited importance for slower processes such as consolidating word-object relations in memory or might even come with a cost of dividing attention while learning.

Overall, the results across experiments did not find clear support for the effect of attentional cues in adult L2 vocabulary learning. The results of this study contribute to the body of research discovering the role of attention processes in L2 learning (Tomlin & Villa, 1994) by adding a comprehensive and novel investigation of gaze and pointing gesture in this context. The results could be used to further motivate the discussion on the continuity of the learning processes across L1 and L2 learning. Finally, the unexpected adverse effect of features such as gaze and position that were discovered could be applied to the industry setting to serve as a reminder for creators of digital language learning materials on the importance of basing their design-related decisions for L2 learning applications on research findings.

Acknowledgments

This work was supported by the European Union's Horizon 2020 research and innovation programme under the Marie Sklodowska-Curie Actions grant agreement No 857897.

Disclosure statement

No potential conflict of interest was reported by the author(s).

ORCID

Paula Janjić http://orcid.org/0000-0003-2795-0105

Data availability

All data files and stimuli for the experiments reported in this study will be made publicly available at Open Science Framework (https://osf.io/q5uzw/).

References

Allen, L. Q. (1995). The effects of emblematic gestures on the development and access of mental representations of French expressions. *The Modern Language Journal*, *79*(4), 521–529. https://doi.org/10.1111/j.1540-4781.1995.tb05454.x

Allen, L. Q. (1999). Functions of nonverbal communication in teaching and learning a Foreign language. *The French Review*, *72*, 469–480.

Atkinson, D. (2010). Extended, embodied cognition and second language acquisition. *Applied Linguistics*, *31*(5), 599–622. https://doi.org/10.1093/applin/amq009

Atkinson, M. A., Simpson, A. A., & Cole, G. G. (2018). Visual attention and action: How cueing, direct mapping, and social interactions drive orienting. *Psychonomic Bulletin & Review*, *25*(5), 1585–1605. https://doi.org/10.3758/s13423-017-1354-0

Baayen, R. H. (2008). *Analyzing Linguistic data: A practical introduction to statistics using R*. Cambridge University Press. https://doi.org/10.1017/CBO9780511801686

Bang, J. Y., & Nadig, A. (2020). An investigation of word learning in the presence of gaze: Evidence from school-age children with typical development or autism spectrum disorder. *Cognitive Development*, *54*, 100847. https://doi.org/10.1016/j.cogdev.2020.100847

Barry, R. A., Graf Estes, K., & Rivera, S. M. (2015). Domain general learning: Infants use social and non-social cues when learning object statistics. *Frontiers in Psychology*, *6*, 551. https://doi.org/10.3389/fpsyg.2015.00551

Barsalou, L. W. (2008). Grounded cognition. *Annual Review of Psychology*, *59*, 617–645.

Belpaeme, T., Vogt, P., van den Berghe, R., Bergmann, K., Göksun, T., de Haas, M., Kanero, J., Kennedy, J., Küntay, A. C., Oudgenoeg-Paz, O., Papadopoulos, F., Schodde, T., Verhagen, J., Wallbridge, C. D., Willemsen, B., de Wit, J., Geçkin, V., Hoffmann, L., Kopp, S., Pandey, A. K. (2018). Guidelines for designing social robots as second language tutors. *International Journal of Social Robotics*, *10*(3), 325–341. https://doi.org/10.1007/s12369-018-0467-6

Bergmann, K., & Macedonia, M. (2013). A virtual agent as vocabulary trainer: Iconic gestures help to improve learners' memory performance. In R. Aylett, B. Krenn, C. Pelachaud, & H. Shimodaira (Eds). *Intelligent virtual agents. IVA 2013. lecture notes in computer science* (Vol. 8108, pp. 139–148). Springer. https://doi.org/10.1007/978-3-642-40415-3_12

Böckler, A., van der Wel, R. P. R. D., & Welsh, T. N. (2015). Eyes only? Perceiving eye contact is neither sufficient nor necessary for attentional capture by face direction. *Acta Psychologica*, *160*, 134–140. https://doi.org/10.1016/j.actpsy.2015.07.009

Booth, A., McGregor, K., & Rohlfing, K. (2008). Socio-Pragmatics and attention: Contributions to gesturally guided word learning in toddlers. *Language Learning and Development*, *4*, 179–202.

Brodeur, M. B., Guérard, K., & Bouras, M. (2014). Bank of standardized stimuli (boss) phase II: 930 new normative photos. *PLOS ONE*, *9*(9), e106953.

Brooks, R., & Meltzoff, A. N. (2008). Infant gaze following and pointing predict accelerated vocabulary growth through two years of age: A longitudinal, growth curve modeling study. *Journal of Child Language*, *35*(1), 207–220. https://doi.org/10.1017/s030500090700829x

Brysbaert, M., & Stevens, M. (2018). Power analysis and effect size in mixed effects models: a tutorial. *Journal of Cognition*, *1*(1), 9. http://doi.org/10.5334/joc.10

Caldano, M., & Coventry, K. R. (2019). Spatial demonstratives and perceptual space: To reach or not to reach? *Cognition*, *191*, 103989. https://doi.org/10.1016/j.cognition.2019.06.001

Caruana, N., Inkley, C., & Nalepka, P. (2021). Gaze facilitates responsivity during hand coordinated joint attention. *Scientific Reports*, *11*, 21037. https://doi.org/10.1038/s41598-021-00476-3

Çetinçelik, M., Rowland, C. F., & Snijders, T. M. (2021). Do the eyes have it? A systematic review on the role of eye gaze in infant language development. *Frontiers in Psychology*, *11*, 589096. https://doi.org/10.3389/fpsyg.2020.589096

Comesaña, M., Soares, A. P., Sánchez-Casas, R., & Lima, C. (2012). Lexical and semantic representations of L2 cognate and noncognate words acquisition in children: Evidence from two learning methods. *British Journal of Psychology*, *103*, 378–392.

Cooperrider, K., & Mesh, K. (2022). Pointing in gesture and sign. In A. Morgenstern & S. Goldin-Meadow (Eds.), *Gesture in language: Development across the lifespan* (pp. 21–46). American Psychological Association. https://doi.org/10.1037/0000269-002

Coventry, K. R., & Garrod, S. C. (2004). *Saying, seeing, and acting: The psychological semantics of spatial prepositions*. Psychology Press.

Coventry, K. R., Valdés, B., Castillo, A., & Guijarro Fuentes, P. (2008). Language within your reach: Near-far perceptual space and spatial demonstratives. *Cognition*, *108*(3), 889–895. https://doi.org/10.1016/j.cognition.2008.06.010

Craik, F. I. M., & Lockhart, R. S. (1972). Levels of processing: A framework for memory research. *Journal of Verbal Learning and Verbal Behavior*, *11*(6), 671. https://doi.org/10.1016/S0022-5371(72)80001-X

Demir-Lira, Ö. E., Kanero, J., Oranç, C., Koskulu, S., Franko, I., Göksun, T., & Küntay, A. C. (2020). L2 Vocabulary teaching by social robots: The role of gestures and on-screen cues as scaffolds. *Frontiers in Education*, *5*, 599–636. https://doi.org/10.3389/feduc.2020.599636

Douglas Fir Group. (2016). A transdisciplinary framework for SLA in a multilingual world. *Modern Language Journal*, *100*, 19–47.

Ellis, N. C., & Wulff, S. (2015). Second language acquisition. In E. Dabrowska & D. Divjak (Eds.), *Handbook of cognitive linguistics* (pp. 409–431). De Gruyter Mouton.

García-Gámez, A. B., & Macizo, P. (2020). The way in which foreign words are learned determines their use in isolation and within sentences. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *46*(2), 364–379. https://doi.org/10.1037/xlm0000721

Geusebroek, J. M., Burghouts, G. J., & Smeulders, A. W. M. (2005). The Amsterdam library of object images. *International Journal of Computer Vision*, *61*(1), 103–112. https://doi.org/10.1023/B:VISI.0000042993.50813.60

Gliga, T., Elsabbagh, M., Andravizou, A., & Johnsn, M. (2010). Faces attract infants' attention in complex displays. *Infancy*, *15*(5), 550–562.

Goldin-Meadow, S. (2007). Pointing sets the stage for learning language—and creating language. *Child Development*, *78*, 741–745. https://doi.org/10.1111/j.1467-8624.2007.01029.x

Golinkoff, R. M., & Hirsh-Pasek, K. (2006). The emergentist coalition model of word learning in children has implications for language in aging. In E. Bialystok & F. I. M. Craik (Eds.), *Lifespan cognition: Mechanisms of change* (pp. 207–222). Oxford University Press. https://doi.org/10.1093/acprof:oso/9780195169539.003.0014

Grzyb, B. J., & Vigliocco, G. (2020). Beyond robotic speech: Mutual benefits to cognitive psychology and artificial intelligence from the study of multimodal communication. In S. Muggleton & N. Chater (Eds.), *Human-Like Machine intelligence* (pp. 274–294). Oxford Academic. https://doi.org/10.1093/oso/9780198862536.003.0014

Haxby, J. V., Gobbini, M. I., & Nastase, S. A. (2020). Naturalistic stimuli reveal a dominant role for agentic action in visual representation. *NeuroImage*, *216*, 116561. https://doi.org/10.1016/j.neuroimage.2020.116561

Heuven, W. J. B., Mandera, P., Keuleers, E., & Brysbaert, M. (2014). Subtlex-UK: A new and improved word frequency database for British English. *Quarterly Journal of Experimental Psychology*, *67*(6), 1176–1190. https://doi.org/10.1080/17470218.2013.850521

Holler, J., & Levinson, S. C. (2019). Multimodal language processing in human communication. *Trends in Cognitive Sciences*, *23*(8), 639–652. https://doi.org/10.1016/j.tics.2019.05.006

Holler, J., Schubotz, L., Kelly, S., Hagoort, P., Schuetze, M., & Özyürek, A. (2014). Social eye gaze modulates processing of speech and co-speech gesture. *Cognition*, *133*(3), 692–697. https://doi.org/10.1016/j.cognition.2014.08.008

Jachmann, T. K., Drenhaus, H., Staudte, M., & Crocker, M. W. (2019). Influence of speakers' gaze on situated language comprehension: Evidence from event-related potentials. *Brain and Cognition*, *135*, 103571. https://doi.org/10.1016/j.bandc.2019.05.009

Kendon, A. (1980). Gesticulation and Speech: Two aspects of the process of utterance. In Mary R. Key (Ed.), *The relationship of verbal and nonverbal communication* (pp. 207–228). De Gruyter Mouton.

Keuleers, E., & Brysbaert, M. (2010). Wuggy: A multilingual pseudoword generator. *Behavior Research Methods*, *42*(3), 627–633. https://doi.org/10.3758/BRM.42.3.627

Knoeferle, P., Kreysa, H., &Pickering, M. (2018). Effects of a speaker's gaze on language comprehension and acquisition: Studies on the role of eye gaze in dialogue. In G. Brone & B. Oben (Eds.), *Advances in interaction studies: Eye-tracking in interaction*(pp. 47–66). Benjamins.

Krashen, S. D. (1988). *Second language Acquisition and second language learning*. Prentice Hall.

Kreysa, H., Nunnemann, E. M., & Knoeferle, P. (2018). Distinct effects of different visual cues on sentence comprehension and later recall: The case of speaker gaze versus depicted actions. *Acta Psychologica*, *188*, 220–229. https://doi.org/10.1016/j.actpsy.2018.05.001

Kroll, J. F., & Stewart, E. (1994). Category interference in translation and picture naming: Evidence for asymmetric connections between bilingual memory representations. *Journal of Memory and Language*, *33*, 149–174.

Langton, S. R. H., Watt, R. J., & Bruce, V. (2000). Do the eyes have it? Cues to the direction of social cognition. *Trends in Cognitive Science S*, *4*(2), 50–59. https://doi.org/10.1016/S1364-6613(99)01436-9

Law, B., Houston-Price, C., & Loucas, T. (2012). Using gaze direction to learn words at 18 months: Relationships with later vocabulary. *Language Studies Working Papers*, *4*, 3–14. https://www.reading.ac.uk/elal/our-research/LSWP

Macdonald, R. G., & Tatler, B. W. (2013). Do as eye say: Gaze cueing and language in a real-world social interaction. *Journal of Vision*, *13*(4), 6. https://doi.org/10.1167/13.4.6

Macedonia, M., Müller, K., & Friederici, A. D. (2011). The impact of iconic gestures on foreign language word learning and its neural substrate. *Human Brain Mapping*, *32*(6), 982–998. https://doi.org/10.1002/hbm.21084

MacWhinney, B. (2012). The Logic of the Unified Model. In S. Gass & A. Mackey (Eds.), *Handbook of Second Language Acquisition* (pp. 211–227). Routledge.

Matsumoto, Y., & Dobs, A. M. (2017). Pedagogical gestures as interactional resources for teaching and learning tense and aspect in the esl grammar classroom. *Language Learning*, *67*, 7–42. https://doi.org/10.1111/lang.12181

McGillion, M., Herbert, J. S., Pine, J., Vihman, M., dePaolis, R., Keren-Portnoy, T., & Matthews, D. (2017). What paves the way to conventional language? The predictive value of babble, pointing, and socioeconomic status. *Child Development*, *88*(1), 156–166. https://doi.org/10.1111/cdev.12671

McNeill, D. (1992). *Hand and mind: What gestures reveal about thought*. University of Chicago Press.

Michel, C., Wronski, C., Pauen, S., Daum, M. M., & Hoehl, S. (2019). Infants' object processing is guided specifically by social cues. *Neuropsychologia*, *126*, 54–61. https://doi.org/10.1016/j.neuropsychologia.2017.05.022

Mitchell, R., Myles, F., & Marsden, E. (2019). *Second language learning theories* (4th ed.). Routledge.

Morett, L. M. (2014). When hands speak louder than words: The role of gesture in the communication, encoding, and recall of words in a novel second language. *Modern Language Journal*, *98*(3), 834–853. https://doi.org/10.1111/modl.12125

Murgiano, M., Motamedi, Y. S., & Vigliocco, G. (2021). Situating Language in the Real-World: The Role of multimodal iconicity and indexicality. *Journal of Cognition*, *4*(1), 1–18. https://doi.org/10.5334/joc.113

Norris, S. (2011). Three hierarchical positions of deictic gesture in relation to spoken language: A multimodal interaction analysis. *Visual Communication*, *10*(2), 129–147. https://doi.org/10.1177/1470357211398439

Ortega, L. (2013). SLA for the 21st Century: Disciplinary progress, transdisciplinary relevance, and the Bi/multilingual Turn. *Language Learning*, *63*, 1–24. https://doi.org/10.1111/j.1467-9922.2012.00735.x

Paivio, A. (1971). *Imagery and Verbal Processes*. Holt, Rinehart & Winston.

Paulus, M., & Fikkert, P. (2014). Conflicting social Cues: Fourteen- and 24-month-old infants' reliance on gaze and pointing cues in word learning. *Journal of Cognition and Development*, *15*(1), 43–59. https://doi.org/10.1080/15248372.2012.698435

Piwek, P., Beun, R. J., & Cremers, A. (2008). 'Proximal' and 'distal' in language and cognition: Evidence from deictic demonstratives in Dutch. *Journal of Pragmatics*, *40*(4), 694–718. https://doi.org/10.1016/j.pragma.2007.05.001

Poarch, G. J., Van Hell, J. G., & Kroll, J. F. (2015). Accessing word meaning in beginning second language learners: Lexical or conceptual mediation? *Bilingualism: Language and Cognition*, *18*(3), 357–371. https://doi.org/10.1017/S1366728914000558

Pulvermüller, F., Hauk, O., Nikulin, V. V., & Ilmoniemi, R. J. (2005). Functional links between motor and language systems. *European Journal of Neuroscience*, *21*, 793–797. https://doi.org/10.1111/j.1460-9568.2005.03900.x

Ro, T., Russell, C., & Lavie, N. (2001). Changing faces: A detection advantage in the flicker paradigm. *Psychological Science*, *12*, 94–99. https://doi.org/10.1111/1467-9280.00317

Rocca, R., Wallentin, M., Vesper, C., & Tylén, K. (2019). This is for you: Social modulations of proximal vs. distal space in collaborative interaction. *Scientific Reports*, *9*, 14967. https://doi.org/10.1038/s41598-019-51134-8

Scott, G. G., Keitel, A., Becirspahic, M., Yao, B., & Sereno, S. C. (2019). The Glasgow norms: Ratings of 5,500 words on nine scales. *Behavior Research Methods*, 1258–1270. https://doi.org/10.3758/s13428-018-1099-3

Shattuck-Hufnagel, S., Ren, A., Mathew, M., Yuen, I., & Demuth, K. (2016). Non-referential gestures in adult and child speech: Are they prosodic. In *Proceedings from the 8th international conference on speech prosody* (pp. 836–839). Boston: Boston University.

Stam, G., & Tellier, M. (2022). Gesture helps second and Foreign Language learning and teaching. In A. Morgenstern & S. Goldin-Meadow (Eds.), *Gesture in Language: Development Across the Lifespan* (pp. 335–364). De Gruyter Mouton. https://doi.org/10.1515/9783110567526-014

Theeuwes, J., & Van Der Stigchel, S. (2006). Faces capture attention: Evidence from inhibition of return. *Visual Cognition*, *13*, 657–665. https://doi.org/10.1080/13506280500410949

Tomasello, M., Carpenter, M., Call, J., Behne, T., & Moll, H. (2005). Understanding and sharing intentions: The origins of cultural cognition. *Behavioral and Brain Sciences*, *28*, 675–735. https://doi.org/10.1017/S0140525X05000129

Tomlin, R. S., & Villa, V. M. (1994). Attention in cognitive science and second language acquisition. *Studies in Second Language Acquisition*, *16*, 183–203. https://doi.org/10.1017/S0272263100012870

Tsuji, S., Fiévet, A., & Cristia, A. (2021). Toddler word learning from contingent screens with and without human presence. *Infant Behavior & Development*, *63*, 101553. https://doi.org/10.1016/j.infbeh.2021.101553

Vogt, P., van den Berghe, R., de Haas, M., Hoffman, L., Kanero, J., Mamus, E., Montanier, J., Oranç, C., Oudgenoeg-Paz, O., García, D. H, Papadopoulos, P., Schodde, T., Verhagen, J., Wallbridge, C. D., Willemsen, B., de Wit, J., Belpaeme, T., Göksun, T., Kopp, S., Krahmer, E., . . . Pandey, A. K. Second Language Tutoring Using Social Robots: A Large-Scale Study. In *Proceedings of the 2019 14th ACM/IEEEInternational Conference on Human-Robot Interaction (HRI), Daegu, Korea*, 11–14 March 2019.

Welke, D., & Vessel, E. A. (2022). Naturalistic viewing conditions can increase task engagement and aesthetic preference but have only minimal impact on EEG quality. *NeuroImage*, *256*, 119218. https://doi.org/10.1016/J.NEUROIMAGE.2022.119218

Zhang, Y., Frassinelli, D., Tuomainen, J., Skipper, J. I., & Vigliocco, G. (2021). More than words: The online orchestration of word predictability, prosody, gesture, and mouth movements during natural language comprehension. *Proceedings of the Royal Society B*, *288*(1955), 20210500. https://doi.org/10.1101/2020.01.08.896712

Appendix A. List of pseudowords

Wordlist A (4 letters, 2 syllables)	Wordlist B (5 letters, 2 syllables)	Wordlist C (6 letters, 3 syllables)	Wordlist D (7 letters, 3 syllables
saro	bueda	neveto	bafioro
bava	tromo	dekido	tavilco
abes	firpa	isalto	nerbona
ibre	yarza	conveo	deroxte
laga	conzo	tovasa	fesiano
izmo	milbo	seanos	almieso
poad	sigre	temiga	parensa
bipa	resel	pamatu	acrivar
luvo	gimar	civeno	sovosta
azme	zindo	revato	inzagio

Appendix B. List of items

Word	Category
Egg	Food
Bread	Food
Muffin	Food
Mushroom	Food
Orange	Food
Apple	Food
Lemon	Food
Banana	Food
Pear	Food
Potato	Food
Pepper	Food
Onion	Food
Tomato	Food
Carrot	Food
Corn	Food
Lettuce	Food
Book	Household_items
Tape	Household_items
Ruler	Household_items
Notebook	Household_items
Box	Household_items
Clock	Household_items
Vase	Household_items
Battery	Household_items
Candle	Household_items
Cup	Household_items
Glass	Household_items
Plate	Household_items
Bowl	Household_items
Bottle	Household_items
Tray	Household_items
Envelope	Household_items

Appendix C. Model output: Experiment 1

Summary of the generalized linear mixed-effects model for word learning outcomes measured by the production test (Experiment 1)

	Estimate	*SE*	Z value	*p* value	Lower CI (2.5%)	Upper CI (97.5%)
Fixed effects						
Intercept	− 0.979	0.21	− 4.658	<.001	− 1.392	− 0.567
Gaze	− 0.079	0.077	− 1.022	.307	− 0.233	0.072
Pointing	0.077	0.077	0.999	.318	− 0.074	0.228
Gaze × pointing	0.257	0.154	1.663	.096	− 0.046	0.559
	Variance	*SD*				
Random effects						
ID	3.524	1.877				
Item	0.563	0.750				
Fixed effects						
Intercept	1.532	0.174	8.822	<.001	1.191	1.872
Gaze	− 0.039	0.077	− 0.500	.617	− 0.190	0.113
Pointing	− 0.036	0.077	− 0.470	.638	− 0.188	0.115
Gaze × pointing	0.159	0.154	1.028	.304	− 0.144	0.461
Random effects						
ID	1.835	1.355				
Item	0.48	0.693				

Model equation: model_p <- glmer(score ~ gaze × pointing + (1|id) + (1|item), data = total_production1, family = binomial()).
Model equation: model_r <- glmer(score ~ gaze × pointing + (1|id) + (1|item), data = total_recognition1, family = binomial()).

Appendix D. Model output: Experiment 2

Summary of the generalized linear mixed-effects model for word learning outcomes measured by the production test (Experiment 2)

	Estimate	*SE*	Z value	*p* value	Lower CI (2.5%)	Upper CI (97.5%)
Fixed effects						
Intercept	−1.106	0.200	−5.534	<.001	−1.505	−0.713
Gaze	0.142	0.107	1.325	.185	−0.071	0.355
Pointing	0.086	0.085	1.006	.314	−0.084	0.256
Gaze × pointing	−0.142	0.17	−0.834	.405	−0.482	0.197
	Variance	*SD*				
Random effects						
ID	2.869	1.694				
Item	0.577	0.759				

Model equation: model_p <- glmer(score ~ gaze × pointing + (1|id) + (1|item), data = total_production2, family = binomial()).