Marsden, H., Marsden, E., Hopp, H., Jackson, C., Jochum-Critchley, T., & Rogers, V. (accepted for publication, 16 Feb 2026). The effects of prediction training on the real-time processing, offline comprehension and oral production of case in L2 German. *Language Teaching Research*.

The effects of prediction training on the real-time processing, offline comprehension and oral production of case in L2 German

I Introduction

The learning and teaching of foreign-language morphosyntax presents considerable challenges, in particular for inflectional forms with low salience and low communicative value, like gender and case. In this paper, we explore whether current psycholinguistic research into the processing of morphosyntax can inform teaching, to facilitate grammar learning. Psycholinguistic research has shown that in native language processing, speakers use grammatical information (among other cues) to predict upcoming input. Evidence for this comes from visual-world eye tracking studies, in which a participant listens to linguistic input while a camera tracks their eye movements as they view pictures on a screen. Such research shows—for example—that in languages with grammatical gender (e.g. Spanish), on hearing a sentence containing a feminine article and noun (*la casa* “the.FEM house”) while viewing a screen depicting images of a feminine noun and a masculine noun, the participant’s eye gaze moves to the image representing the feminine noun earlier than when the screen depicts two feminine nouns (Lew-Williams & Fernald, 2007). This is taken as evidence of prediction in real-time comprehension and immediate use of the grammatical gender expressed by the article.

While evidence of predictive processing is found reliably in native speakers for both gender and case morphology, results are mixed for second language (L2) speakers. Most of this research has focused on gender. With advanced or near-native L2 speakers, there is some evidence of predictive processing based on gender in L2 Spanish (Dussias et al., 2013) and German (Hopp, 2013), even when the first language (L1) is English, which lacks grammatical gender; though other studies found no such evidence (Grüter et al., 2012, Spanish). Studies of intermediate-level L2 speakers have generally found no evidence of grammar-based predictive processing (e.g., Lew-Williams & Fernald, 2010), though Hopp's (2016) investigation of the effect of training on predictive processing of gender in German is an exception. Hopp found that, after training to 100% accuracy on the gender of 80 German nouns, upper-intermediate-level L2-German speakers came to use gender predictively in a visual-world eye tracking experiment, though they had not shown evidence of prediction prior to the training. This suggests that increasing the robustness of L2 speakers' grammatical gender knowledge led to prediction, at least immediately after the training.

Turning to case, a review by Schlenter (2023) identifies previous studies of predictive processing in L2, covering German (Henry et al, 2022; Hopp, 2015; Schlenter & Felser, 2021), Japanese (Mitsugi, 2017; Mitsugi & MacWhinney, 2016) and Korean (Frenck-Mestre et al. 2020). Schlenter reports that the results generally show case-based prediction in L1 groups but not in L2. For example, using subject-initial and object-initial transitive sentences in German, Hopp (2015) found that even advanced L1-English L2-German speakers did not make anticipatory looks to the target final argument of the sentence based on the case of the first argument's article, in contrast to native German speakers. However, there is some evidence of case-based prediction by L2 speakers when additional factors, namely prosody or L1 similarity, were considered. Prosody was manipulated in Henry et al.'s (2022) investigation of L1-English L2-German speakers, revealing that, when the stimuli included prosodic cues alongside case-marking cues, some evidence of prediction emerged in higher proficiency speakers. Schlenter & Felser (2021) also found predictive use of case-marking by highly proficient L2-German speakers, but with Russian—another case-marking language—

as their L1. Hence, Schlenter (2023, p. 258) suggests that, while predictive use of case-marking is clearly difficult in L2 processing, it may be achievable under “favourable conditions”, such as facilitative prosody or L1 influence, at least among high proficiency learners.

This paper presents a preliminary investigation of whether classroom-based training on prediction can also facilitate development of predictive processing based on case information in L2 German. We build on Hopp’s (2016) finding that, for gender, lab-based training via massed input on the object of prediction (i.e., the article and noun) facilitated predictive processing. We apply this logic to case. However, in addition to targeting the object of prediction in training (here, case marking), the present study also aimed to train the mechanism of prediction directly, via training materials that could be incorporated within the context of a regular German course.

II Background and motivation

1 German case and predictive processing

Studies of case processing in L2 German typically focus on the distinct morphology of nominative and accusative masculine definite articles (*der* and *den*, respectively), along with the fact that German allows both subject-initial and object-initial word order. In a transitive sentence with two masculine nouns, case marking on the initial noun can contribute to disambiguation, as illustrated in (1–2) from Hopp (2015) (adapted from Kamide et al. 2003). In (1), the sentence-initial article *der* provides a strong cue that the masculine noun *Wolf* is the subject, while in (2) the sentence-initial accusative masculine article *den* indicates that *Wolf* is an object.

1. Subject-verb-object (SVO):

Der	Wolf	tötet	gleich	den	Hirsch.
the.NOM	wolf	kills	soon	the.ACC	deer

“The wolf will soon kill the deer.”

2. Object-verb-subject (OVS):

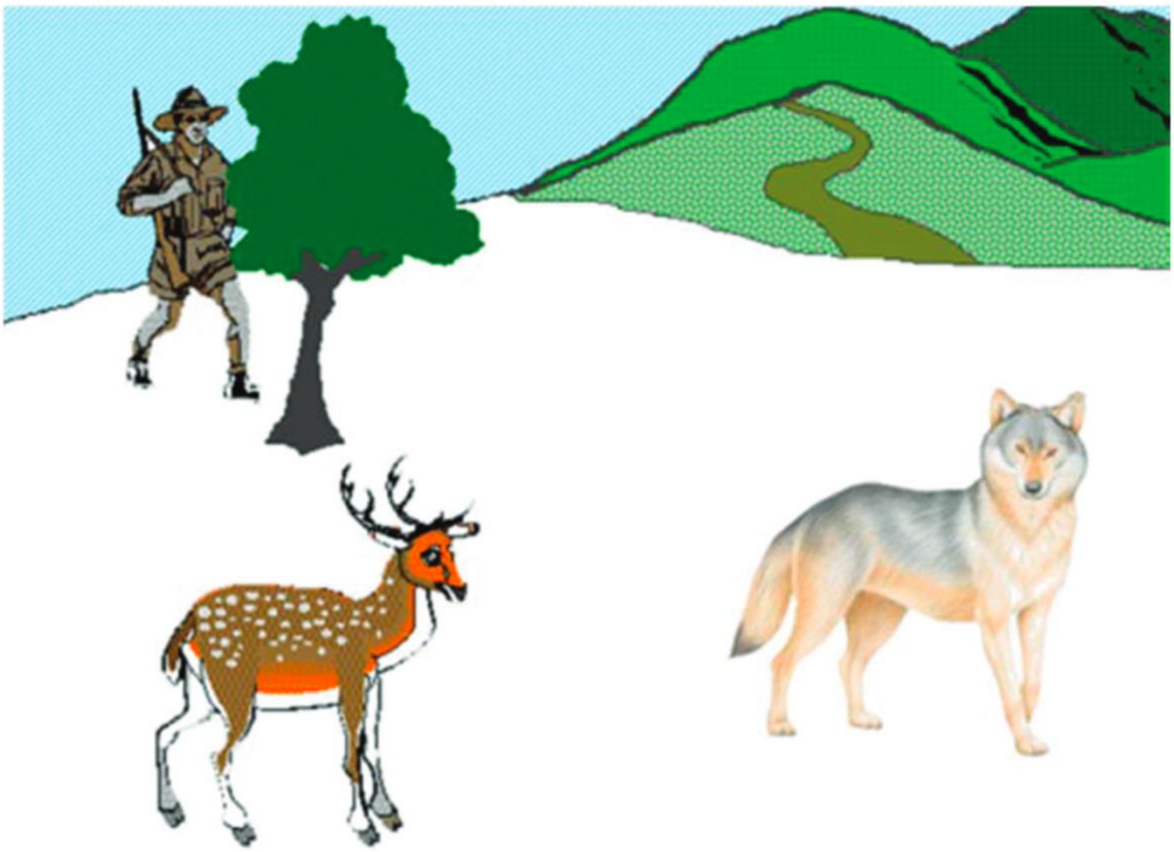
Den	Wolf	tötet	gleich	der	Jäger.
the.ACC	wolf	kills	soon	the.NOM	hunter

“The hunter will soon kill the wolf.”

The object-initial order in (2) is non-canonical: it serves to topicalize the object, along the lines of an English cleft, *It is the wolf that the hunter will soon kill* (Hopp et al., 2020).

All German nouns belong to a grammatical gender class: masculine, feminine or neuter. The surface forms of determiners, adjectives, pronouns and nouns depend on a combination of gender, case and number. There is considerable syncretism within the system. For example, feminine definite articles have the same form, *die*, for both nominative and accusative case, while masculine articles are always unambiguous for case. Consequently, while case-marking is, in principle, a reliable cue to meaning, this cue is often unavailable in practice, due to overlap in surface article forms in many genders (Kempe & MacWhinney, 1998). Nonetheless, as noted above, L1-German speakers use case-marking predictively. In Hopp’s (2015) study, L1-German participants heard sentences such as (1) or (2) while viewing a depiction of the three nouns from those sentences (Figure 1). The results showed that participants looked towards the second noun (i.e., the hunter, in the OVS sentence in (2)) before the onset of the second article. This shows that the participants rapidly integrated case information on the first article with the semantics of the verb to identify that *Wolf* in (2) was the object of the verb rather than the subject, and hence looked towards the hunter as the likely subject of “kills” (rather than interpreting the first noun *Wolf* as the subject, which would have incurred predictive looks to *deer* as object).

Figure 1: Image accompanying (1–2) in Hopp (2015)



Hopp (2015) additionally tested low-intermediate to advanced L2-German speakers but found no evidence of case-driven predictive looks to target nouns before the second noun onset. Rather, for both SVO and OVS sentences, participants looked towards a potential patient of the verb before they heard the second noun. Henry et al. (2022) used the same pictures and sentences, but incorporated a prosodic manipulation into their experiment. They created two variations for each word order in (1–2): one with no prosodic emphasis on any word; and one with intonation as if in response to an object-wh-question, which yielded a pitch accent (i.e., prosodic focus) on the object noun in the OVS condition. The results showed that on OVS sentences with prosodic focus, prediction accuracy increased in higher proficiency L2-German participants—though the level of erroneous prediction in OVS remained high.

Taken together, the existing research suggests that, while L1-German speakers naturally use case morphology to make predictions about upcoming sentence structure in real time, intermediate–advanced L2-German speakers tend not to, unless additional cues are available. In this way, L2 development appears similar to children's acquisitional trajectory in L1 German, where, initially, support is needed from other cues (e.g., prosody) before children use case alone (Grünloh et al., 2011).

2 Textbook approaches to teaching case in German

L2 textbooks typically teach case in a scaffolded way, introducing individual cases one by one, rather than presenting the whole case system at once. This is because, in addition to nominative and accusative case, Instruction on case necessarily entails some prior instruction on gender, and most textbooks (e.g., *DaF kompakt neu* [Braun et al., 2016], *Willkommen!* [Coggle & Schenke, 2012]) introduce the nominative forms of articles early on, as part of their exposition of the three German noun gender classes. Explanations of case itself—as opposed to incidental exposure to case-marked forms—typically begin with the accusative. Definite and indefinite article forms for nominative and accusative case are often presented in a table, accompanied by explanations of their grammatical functions (i.e., nominative represents the verb subject; accusative represents the direct object).

However, the focus is usually on the forms of the articles, with no information about the contribution to sentence interpretation that the accusative case brings via its grammatical function of marking the object. In terms of skills, case teaching typically focuses on production of the correct article form, be it practised through pattern drills or more open-ended communicative or task-oriented activities. For example, in *DaF kompakt neu*, after introducing the nominative and accusative forms of definite and indefinite articles for all genders, students practise talking in pairs about things they have. An example dialogue includes—but does not highlight or make obligatory for comprehension—two different forms of the indefinite article: *Hast du einen Laptop?* / *Nein, ich habe ein Tablet* ("Have you a.MASC.ACC laptop" / "No I have a.NEUT.ACC tablet") (Braun et al., 2016, p. 25).

The non-canonical OVS word order is rarely presented in textbooks. Thus, the potential for case-marking to contribute to the learning of this word order—or for word order to contribute to the learning of case marking on articles—is not addressed. In relation to research that investigates L2 processing of case-marking in OVS word order, there are two key learning issues. Learners must come to know that case-marking can be used to identify grammatical roles, and that the order of subjects and objects can be flexible. The former is highlighted in typical German language teaching while the latter is not.

3 Teaching intervention studies on case

Several previous studies have investigated the effects of different teaching interventions on the L2 acquisition of case. Most of these employ Processing Instruction (PI; VanPatten & Cadierno, 1993), which differs from typical textbook instruction precisely in that it highlights the relationship between the case-marking form and its meaning. PI builds on VanPatten's Input Processing model of L2 acquisition, which proposes that learners tend to use salient contributions to meaning, such as lexical cues and word order, and filter out less salient cues, such as case (VanPatten 2004). This can result in non-target processing strategies, such as tending to process the first noun in a sentence as the subject (VanPatten, 2020, p. 112). PI was designed to establish and develop effective processing of morphosyntactic cues that may otherwise get filtered out. It was originally conceived as having three elements: (i) a short explicit information component (EI) that includes an explanation of the relevant form-meaning (or form-function) mapping and information about potential unreliable processing strategies; (ii) referential activities, comprising reading and listening exercises designed so that learners must attend to the form-meaning mapping at hand (rather than relying on an unreliable strategy) to complete the exercise successfully; and (iii) affective activities, which provide a flood of exemplars of the target form, but ask learners to express an affective reaction to the content or demonstrate more global-level understanding, instead of requiring them to attend to form-meaning mapping. PI has been shown to be effective for a range of grammatical properties, using a range of offline measures (reviewed in Shintani, 2015; DeKeyser & Prieto Botana, 2015). Furthermore, referential activities alone—without the provision of EI—have been found to be effective on offline measures (Marsden & Chen, 2011). EI about L1–L2 differences along with referential activities has also been shown to have a beneficial effect on *online* processing and oral production (McManus & Marsden, 2017; 2018; 2019).

For German case, specifically, PI research has investigated questions relating to the role of (i) attending to meaning and form (relative to form alone); (ii) explicit instruction; and (iii) prosodic cues on learning outcomes; and it has compared PI with traditional instruction. Concerning the first of these, Hanan (2015; also Kasprowicz & Marsden, 2017) examined the value of connecting a form (i.e., the German definite article) with its meaning (agent/patient) versus “spotting the form”. In training, both types of activities incorporated EI before practice. Training for the form-meaning participants involved selecting the correct picture from two that differed only in their depiction of SVO and OVS—as in referential activities. In the “spot-the-form” training, participants simply clicked on the word for “the”. The study found that both approaches led to gains in comprehension and production by child L2-German learners, whereas a control group that received no intervention related to case made no gains. In short, EI combined with practice attending to the form (or form and meaning) was beneficial.

Turning to the role of EI, Henry et al. (2009) examined whether provision of EI with referential activities facilitates earlier processing of accusative case in OVS sentences than referential activities alone. The referential activities consisted of participants hearing either an SVO or an OVS sentence, then selecting from two scenes, of which only one matched what they had heard. Learning was measured during the training, operationalised as the number of sentences participants needed to hear before managing to choose the correct scene four times in a row. The +EI group reached this criterion considerably sooner than the -EI group. These findings were broadly replicated by VanPatten & Borst (2012).

Focusing on the roles of prosody and EI, Henry et al. (2017) used a four-group design to tease out the effects of these two variables alone and combined on reading comprehension and written production of SVO and OVS sentences. The referential activity again involved identifying which picture matched an SVO or OVS sentence, but sentences were presented aurally, and the cues for picture identification were either case-marking alone or case-marking with prosody. Findings suggested that reading comprehension improved, including at a delayed post-test four weeks after the treatment, for learners who had received EI (whether with or without prosodic cues) and for those who had received prosodic cues alone (without EI). However, immediate benefits for written production were not sustained four weeks later. A further study by Henry (2023) investigated whether inclusion of prosodic cues in a PI intervention would benefit real-time processing of case-marking in OVS sentences. Processing was measured using a self-paced reading task with comprehension questions after each sentence. Analysis of reading times for the correctly-comprehended sentences showed differences between OVS and SVO conditions, whereby for masculine nouns in OVS, reading times were longer on the first noun phrase (marked with *den* "the.ACC"), and—in the PI-with-prosody group—following the nominative-marked second noun phrase. These longer reading times imply a surprisal effect due to processing of accusative case at the first noun phrase and nominative case at the second noun phrase. Though there were no interactions of Word Order (OVS vs. SVO) with Time (Pretest vs. Posttest), Henry argues, based on descriptive differences in estimated marginal means, that the OVS–SVO differences were driven by longer reading times in OVS at the posttest. Consequently, the findings provide indicative evidence of PI-training having an effect on processing.

Another study by Henry (2022) investigated the effectiveness of PI compared with traditional instruction (TI) based on textbook materials on German accusative case. Results from the same self-paced reading task as in Henry (2023) showed a pretest-to-posttest increase in reading times for accusative-marked first nouns in OVS sentences in the PI group but not in the TI group. The PI group also made gains in an interpretation task; and both groups improved in a written production task. These data provide further evidence of PI having an effect on L2 case processing. See also Henry (2025) for further evidence of structured input leading to gains in comprehension and production of case in German OVS sentences.

Prior research thus suggests that practice in connecting form to meaning can facilitate acquisition of case in L2 German, and learning gains are likely enhanced by provision of EI. While most previous studies used offline tasks as outcome measures, Henry's (2022, 2023) studies used an online task. There, the longer reading times in OVS sentences may reflect that, after training, participants perceived the atypical case-marking (i.e., accusative on the first noun) in real time. However, the findings do not necessarily show whether the function of that case-marking was processed and incorporated into the comprehension of the sentence. Andringa & Curcic (2015) aimed to test this by investigating whether EI about case-marking in an artificial language would lead to case-driven prediction in beginner learners. The target form was direct-object marking (DOM) on animate—but not inanimate—object noun phrases, by means of a pre-object particle *al* (similar to DOM with animate nouns in Spanish). For both the EI group and an implicit instruction group that received no EI, the training involved viewing and hearing descriptions of 104 scenes depicting everyday activities. Half of the transitive training items contained an animate object with DOM; the other half had inanimate objects and no DOM. After 15 training items, the EI group—but not the implicit group—received an explanation of DOM. After training, participants completed an eyetracking task. In each trial, they heard a sentence and saw pictures corresponding to two nouns. One of these was always the final noun heard in the sentence, and participants were asked to choose the image that matched the final word. In the critical trials, one image was animate and one inanimate, such that the DOM particle could potentially facilitate earlier fixations on the target image. However, the results showed no evidence of earlier prediction of the final noun in the different-animacy trials, in either group, though the EI group made gains in an offline grammaticality judgment test.

The study by Andringa & Curcic is the only prior investigation of the effects of training on L2 predictive processing using case, to our knowledge. The study targeted beginner learners, but, as noted above, previous research has found little evidence of predictive processing using case in even intermediate–advanced L2 speakers, unless additional cues were available. The low proficiency of the learners in Andringa & Curcic could have been a limiting factor in development of predictive processing. Additionally, the learners were not given *practice* in the predictive mechanism itself; rather, they received EI followed by exposure to multiple exemplars. These factors, along with the positive results for PI on case acquisition in offline comprehension and online—but not predictive—processing, motivate the present study's research questions and design.

III Research questions

Building on the research outlined above, our goal was to investigate whether prediction training, incorporating EI and PI-inspired referential activities, can lead to gains in the real-time processing of German case by intermediate-level L2-German learners. Our primary research question was:

- To what extent does training on prediction using case lead to increased prediction during real-time processing? (RQ 1)

Two further research questions asked:

- To what extent does training on prediction using case lead to more accurate:
    - comprehension of case? (RQ 2)
    - production of case? (RQ 3)

IV Methods

1 Participants

The participants were recruited from comparable intermediate-level (B1–B2) German language classes at two UK universities. The classes ran through the whole academic year, and the research began half-way through that period. All participants attended German classes regularly throughout the participation duration. Two separate participant groups were recruited in each of two consecutive years. In the first year, a total of 28 learners across both sites participated in the prediction training intervention (henceforth, the training group). In the second year, a new group of 13 learners, across both sites, participated in a comparison study that did not include any experimental intervention other than the tests (henceforth, the comparison group).[1]

To ascertain equivalence between the experiment and comparison groups, we examined the results of a 30-item written German proficiency test (Goethe Institut, 2004) and a language background questionnaire, which participants completed at the start of the research. Results are summarised in Table 1. An independent t-test on the proficiency scores showed no difference between the groups ($t(39) = –1.612$, $p = .115$, 95% CI: -4.615, 0.522).[2] Most participants (31/41) were L1-English speakers, while others had a variety of L1s. We included participants with any L1 for two reasons. First, our goal was to find out

[1] We use "comparison" rather than "control" group, in acknowledgement that the number of participants is rather small for this to be a formal, experimental "control" group. The small group size reflects the combined constraints of the small population of German learners we could recruit from and the time-limited project funding. Our research questions are not expressed in terms of a control group, and only one of our four predictions refers to the comparison group.

[2] We also checked whether location (i.e., university) modulated proficiency. Results of a two-way ANOVA (Location × Group) yielded no significant interaction or effects (Location: $F(1, 37) = 0.402$, $p = .530$; Group: $F(1, 37) = 2.891$, $p = .098$; Location × Group: $F(1, 37) = 1.037$, $p = .315$).

whether intermediate-level learners can benefit from instruction on prediction with case among groups that reflect ecologically valid classes where mixes of L1s are normal. Thus, exclusion of participants on the basis of L1 would reduce the potential usefulness of the research. Similarly, we included participants with a range of language learning backgrounds.[3] Second, although there is some evidence that L1–L2 similarity may facilitate L2 processing of case in advanced L2 speakers (Schlenter & Felser, 2021), at intermediate level, there is little evidence of learners using grammatical cues predictively at all. Our study thus represents a first step of investigating whether training at intermediate level could lead to predictive use of case cues irrespective of L1*,* with potential for follow-up research to tease out L1 influence.

Table 1. Participant information

	Training group (*n* = 28)	**Comparison group** (*n* = 13)
Mean age in years (range, SD)	19.7 (18–28, 2.17)	19.8 (18–25, 1.99)
L1	English (*n* = 22); Dutch; French; Hungarian; Korean; Rumanian; Russian	English (*n* = 9); Polish (*n* = 2); English + Gujurati; Spanish + Catalan
Mean proficiency score /30 (range, SD)	15.11 (7–24, 4.0)	17.15 (11–22, 3.24)

[3] Other foreign languages studied by at least one participant were Dutch, English, French, Greek, Hindi, Italian, Japanese, Latin, Russian, Spanish, Turkish and Welsh. Two participants in the training group had lived in a German-speaking country for one year, and one in the comparison group for four years.

2 Training materials

The prediction training intervention used EI along with PI-inspired referential structured input activities, on the grounds that previous intervention studies targeting case in German found learning gains with such an approach. The training group had two training sessions lasting 40–50 minutes each. The first began with EI, which reviewed nominative and accusative case, then used SVO and OVS sentences to show how the upcoming nouns can be predicted on the basis of case-marking on their articles.[4] This instruction was delivered by a member of the research team, using a uniform protocol and set of instructional slides. After the instruction, the learners individually worked through six training activities, using tablets or PCs, and headphones. The activities were built using the OpenSesame experiment builder (Mathôt et al. 2012). Each training activity comprised a set of around 30 items. Some activities were adapted from Cadierno (1992), Henry et al. (2017), and Kasprowicz & Marsden (2017), while others were created using original artwork or clipart licensed for use in education.[5] In line with PI referential activities, getting the right answer in the training activities required processing of case-marking on articles during reading or listening. However, the activities differed from those reviewed above, in that they all required participants to *predict* the correct answer, based on case-marking, prior to hearing (or reading) the end of the sentence. Figure 2 illustrates one of the training activities. The item starts with the written and audio presentation of the verb for that item: *ziehen* "pull". The second screen shows two masculine referents doing the action of the previously-introduced verb. The learner then hears either *der* ("the.NOM") or *den* ("the.ACC") as the first word of

[4] The instruction slides and further examples of training items, can be viewed at https://osf.io/pa4q3/?view_only=768f25fc456642ee826350d17964edab.
[5] The experimental task pictures (see next section) were created similarly.

a sentence that will describe the picture. Learners were asked to select which noun should come next in the sentence by clicking one of the circled options. In Figure 1, since the sentence begins with the accusative *den* (appearing in a speech bubble here, to indicate the auditory presentation in the actual exercise), and the picture shows the thief pulling the dog, participants should predict "dog" as the next word, because "dog" is the object of the verb "pull".

Figure 2: Example training item, targeting the OSV sentence *Den Hund zieht der Dieb*. "The.ACC dog pulls the.NOM thief"

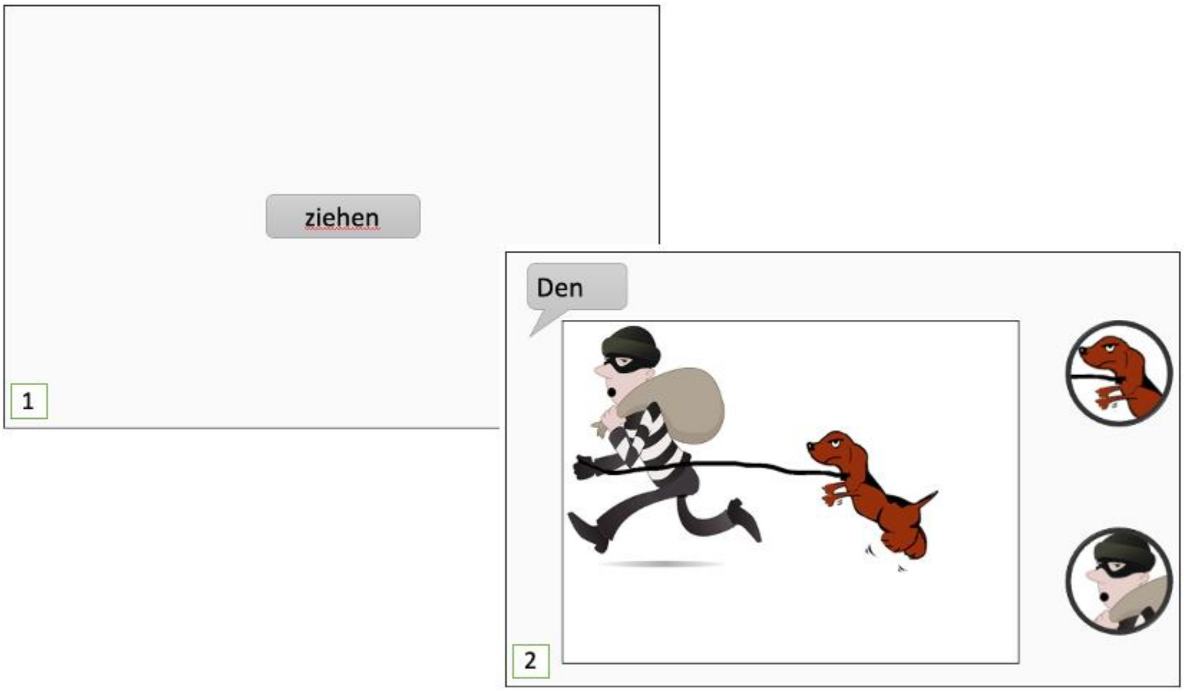


Image credit (thief): Jorgenmac

Feedback was presented on screen immediately after selecting the answer. The feedback for selection of the wrong noun in Figure 2 was: "The sentence began with **den**, so the character that is the object of **ziehen** must come next. In the picture, the dog (not the thief) is the one that another character pulls, so **Hund (dog)** comes after **den**”. A second training activity had the same format but presented the article in written form above the picture, instead of in audio form.

Another pair of training activities included hearing or reading an article-noun-verb sentence fragment and deciding whether a transitive action depicted on the screen would match the completed sentence. For example, if the picture showed a dog searching for a bird (both masculine nouns) and the sentence fragment was *Den Hund sucht*… “The.ACC dog searches for…”, then comprehension of the accusative article would lead to understanding that the dog should be the object and responding “No” (the picture cannot match the sentence). A third pair of activities involved selecting the correct picture from a pair on the basis of hearing (or reading) *Den* or *Der*. The pictures depicted the same transitive action performed by two animate referents: one masculine, and one feminine or neuter. In one picture the masculine noun is the agent of the action and in the other, the patient. In all the training activities, feedback was provided after every item. Participants were encouraged to answer each item as quickly as they could. At the second training session, the participants worked through a second set of the six training activities that used different items.

No specific materials were created for the comparison group. They followed their regular curriculum, broadly based on a communicative language teaching approach, with their regular teachers (as did the training group, except for the intervention outlined above). The regular curriculum focused on developing skills and meaning-making through engagement with authentic materials. While grammar topics were addressed, there was no intensive practice of a particular grammar property comparable to the case training in the training group.

3 Test measures

Participants' performance was assessed using visual-world eye tracking incorporating a comprehension task, and an oral production task.[6]

a. Eye tracking task

Twenty-four sentence pairs were created, with an SVO and OVS variant of the same meaning within each pair, as illustrated in (3).

3.	a.	Der	Archäologe	begrüsst	gerade	den	Segler. (SVO)
		The.NOM	archaeologist	greets	now	the.ACC	sailor
	b.	Den	Segler	begrüsst	gerade	der	Archäologe. (OVS)
		The.ACC	sailor	greets	now	the.NOM	archaeologist

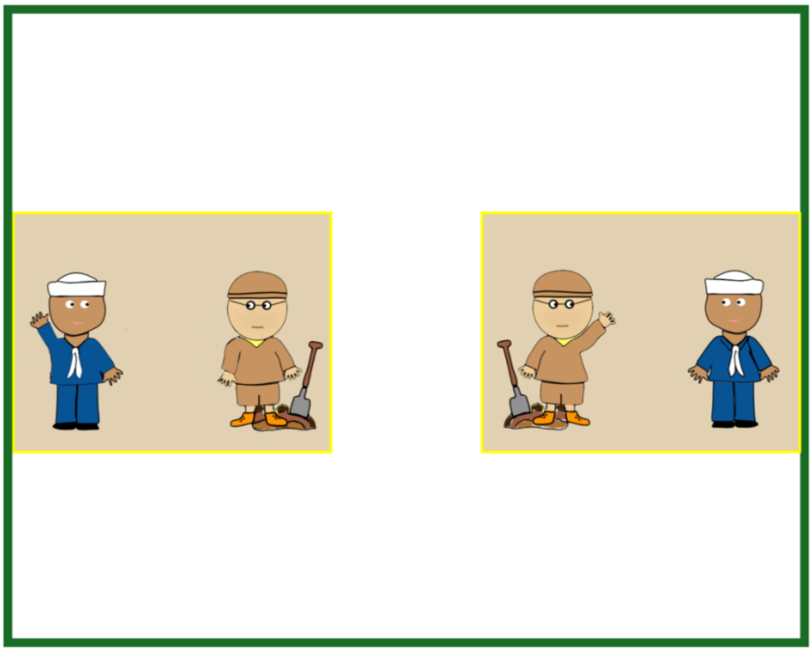
Meaning of both (a) and (b): "The archaeologist is greeting the sailor just now."

The critical sentences were reversible in meaning, so that either noun could plausibly be the subject or object. The two nouns were always masculine in gender. Accordingly, comprehension of the article case leads to the correct interpretation of the sentence. Participants' comprehension was measured using a picture-selection task. In each critical trial, two pictures appeared on the screen, each depicting the same transitive event but with the subject and object from one picture reversed in the other, as in Figure 3. After 560 ms, an audio recording of one target sentence was played (either 3a or 3b for Figure 3), and

[6] The test materials are shared at [ADD URL ON ACCEPTANCE].

participants had to select the picture that matched that sentence. In Figure 3, the correct picture is the right-hand image for both the SVO or the OVS variants in (3), since, in both cases, the sailor is the object of the verb "greet".

Figure 3: Picture array for Sentences (3a–b)



The critical sentences included the adverb *gerade* "just now" between the verb and second noun in order to increase the time available for participants to make predictive looks to the target picture between the end of the verb and the beginning of the second noun.

This design, which presents two pictures of an action and measures looks to the target scene, differs from the design in most of the predictive processing studies reviewed in the Background section (e.g., Hopp 2015), which presented one scene and measured looks to a target noun within the scene. We adopted the two-scene design because it offers efficient measurement of both comprehension (via the picture-sentence matching task) and processing (via eye gaze) within the same task. (See Cristante & Schimke, 2020, and Wong & Ito, 2018, for further examples using two-scene designs.)

In addition to the critical sentences, 24 pairs of filler sentences were created, in which the verb (not the nouns) differentiated the two sentences, as in (4). All three grammatical genders were used across the fillers. No filler included two masculine nouns.

4. Die Oma	(a) wäscht /	(b) trocknet	gerade das Kind.
The grandmother	(a) washes /	(b) dries	now the child

The pictures for each filler sentence depicted the two verbs in the pair: a grandmother washing a child, and a grandmother drying a child, for (4a–b). Thus, for (4a), the picture that illustrated washing was the correct one.

The 96 sentences (48 critical, 48 filler) were recorded by a male, native speaker of German using a natural, neutral prosody that did not emphasize any of the words and was intended to provide no prosodic cues to interpretation. The critical and filler sentences were evenly distributed across three 32-item lists. No list contained both items from a pair. The correct picture was displayed with equal frequency on the left and the right. The individual pictures were also balanced in terms of the agent position (left/right). Trials were presented randomly within each iteration of the experiment. Participants experienced a different list at each of the three test times. (See Procedure for details of test timings.)

The eye-tracking set-ups at the two research locations were slightly different: at one, eye movements were recorded using an SR Research Eyelink 1000 Plus eye-tracker, sampling at 1000 Hz; at the other, an Eyelink Portable Duo eye-tracker was used, sampling at 500 or 250 Hz. All data was resampled at 250 Hz prior to analysis. At both locations, participants were seated about 800 mm from the monitor that displayed the picture arrays. Viewing was binocular, but eye movements were recorded from just the dominant eye, determined using the standard simple eye dominance test. To begin each trial, participants looked at a fixation point in the centre of the monitor screen. Subsequently, the picture array appeared, then after 560ms, a sentence was played. Participants were instructed to select the picture that matched the sentence by clicking a “left” or “right” button on a keyboard. The aural presentation of each sentence lasted 2.1–3.4 seconds. The pictures remained visible until the participant responded. A fixation cross appeared then before the next trial. Two practice items preceded the experiment trials. Participation took around 15 minutes.

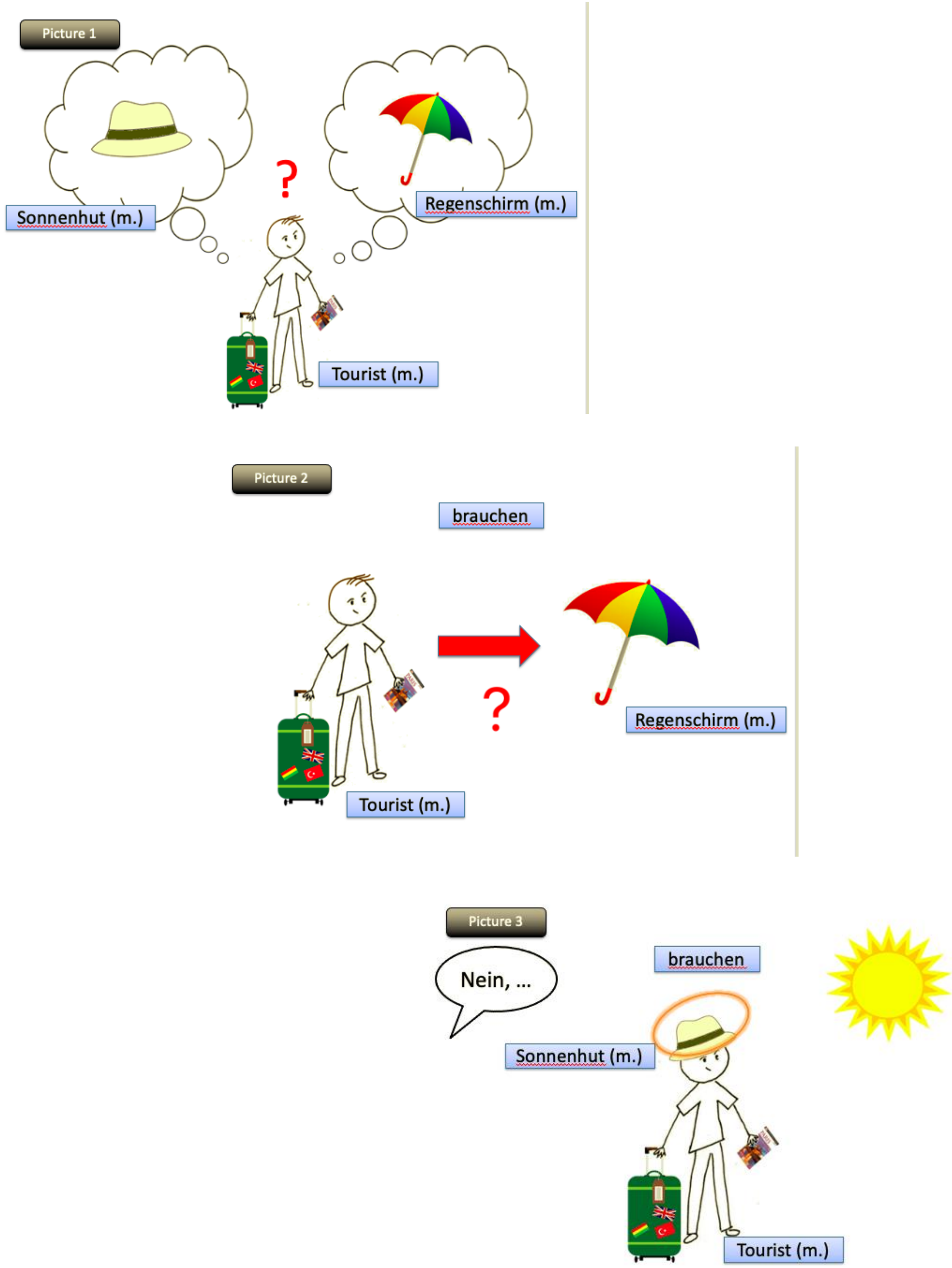
b. Oral production task

The production task was designed to elicit OVS and SVO sentences in response to picture contexts. Each trial involved a series of three pictures, with the sequence designed to motivate either an OVS or SVO sentence in a spoken response on the third picture. A set of pictures for a critical OVS trial is given in Figure 3. Picture 1 in critical OVS trials always depicted one person with two thought bubbles, each containing a masculine-gender object. The German words for all the nouns were included in the picture, along with the noun gender (i.e., “m.” = masculine in Figure 4). Participants were told to take their time looking at these nouns, which would be relevant to the subsequent pictures, and then click to advance to Picture 2. Here, they were instructed to think of a question using the transitive verb given at the top of the picture and the nouns depicted. The verb in Figure 3 is *brauchen* “need”, yielding the question in (5a). Once the participant felt they had the question in mind, they clicked to reveal Picture 3, which provides the answer to the question.

Participants were asked to say the answer out loud, beginning with the circled noun. A speech bubble with the word *Ja* (“Yes”) or *Nein* (“No”) provided a prompt as to whether the answer should affirm or negate the question. In Figure 4, the sunhat is circled, with the intention of eliciting the OVS sentence in (5b).

5. a.	Braucht	der	Tourist	den	Regenschirm?	
	Needs	the.NOM	tourist	the.ACC	umbrella?	
	“Does the tourist need the umbrella?					
b.	(Nein.)	Den	Sonnenhut	braucht	er.	
	No	the.ACC	sunhat	needs	he	
	“No. He needs the sunhat.” (“It’s the sunhat that he needs.”)					

Figure 4: Example of production task trial designed to elicit the OVS sentence given in (5b)



This design using a series of pictures was selected rather than simply presenting the elicitation picture (Picture 3) in isolation, in order to provide a more authentic context for use of OVS word order. The sequence of presenting a choice between two objects (Picture 1), asking a question about one of them (Picture 2), then prompting an answer that selects the *other* object (Picture 3) aims to establish contrastive focus on that object, thereby making OVS order more natural.

Eighteen OVS trials were created and counterbalanced by 18 SVO trials. Within the OVS trials, 12 were designed as critical items to elicit a No response (as in (5b)), and 6 as fillers eliciting a Yes response. The SVO trials were all fillers. Twelve had a Yes response and 6 a No response. Among the 24 fillers, 12 had masculine nouns as the object and 10 as the subject. Elsewhere, the filler nouns were feminine or neuter. The 36 trials were divided across two counterbalanced lists, with 6 critical OVS-No trials in each list. Within each list, the trials were presented in a quasi-randomised order that was fixed for all participants. Participants were presented with List 1 in the pre-test, List 2 at the post-test, and List 1 again at the delayed post-test.

The task was created in Powerpoint. Participants sat at a laptop and clicked through the slides at their own pace. Instructions were pre-recorded and played automatically as participants clicked through. Four practice items were presented before the experiment. Participants' responses were audio-recorded from the start of the experiment trials. Participation took around 10 minutes.

4 Procedure

The research was approved by the ethics committee in the first author's department. All participants gave informed consent before participating. Participation took place across 15 weeks. In Week 1, both the training group and the comparison group completed the background questionnaire, proficiency test, eye-tracking task, and oral production task. In Weeks 2 and 3, the training group received the prediction training. In Week 4 then Week 15, both groups completed immediate-posttest and delayed-posttest versions of the eye-tracking and oral production tasks.

All testing and training took place in person. Testing was conducted in a lab in the presence of a researcher. The training took place during regular class time at one of the universities, but in a lab outside class time at the other. Participants received a thank-you payment after each test session.

All participants completed the pretests and posttest, but some attrition was experienced at the delayed post-test, with 18 out of 28 participants remaining in the training group and 10 out of 13 in the comparison group.

5 Predictions

Predictions for each test measure are formulated for the training group as follows:

- Prediction 1: Eye tracking task, picture selection

  On hearing an OVS sentence, the training group will select the target OVS picture more frequently at the posttests than at the pretest.

- Prediction 2: Eye tracking task, eye gaze behaviour

  On hearing an OVS sentence, the training group will fixate earlier on the target OVS picture at the posttests than at the pretest.

- Prediction 3: Oral production task

The training group will produce OVS sentences more frequently at the posttests than at the pretest.

Since the comparison group participants were also enrolled on German courses at the time of participation and they were exposed to sentence-initial accusative case-marking via the experiment, it was expected that they would also make gains in German, potentially including gains relating to case. As such, the comparison group's performance serves as a baseline measure which may reflect incidental learning and learning through completing the tasks. However, we predicted a between-group difference:

- Prediction 4: Between-group comparison

  Across all measures, gains made by the training group will be larger than gains made by the comparison group.

V Results

1 Eye Tracking task

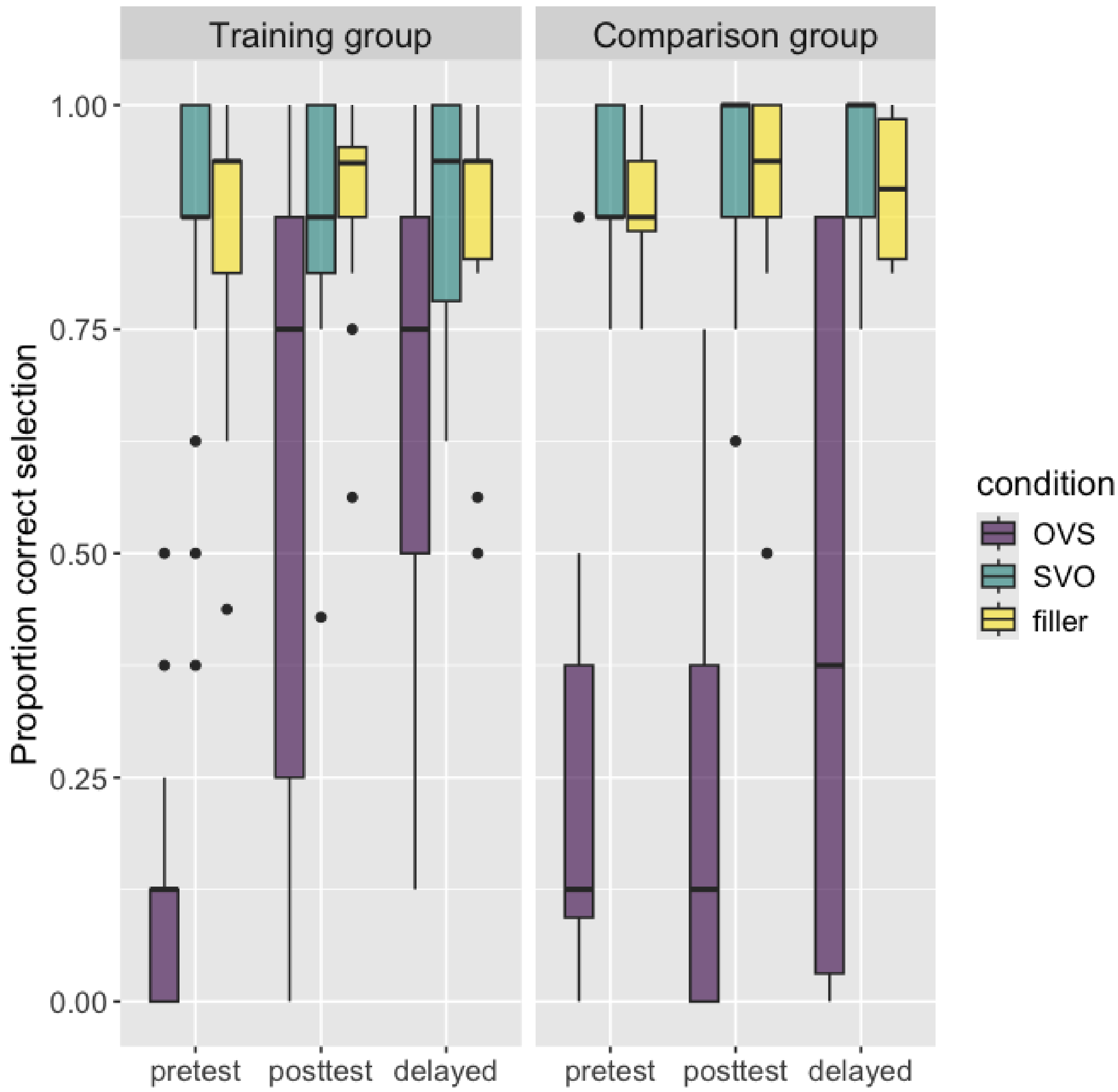
Two measures were produced by the eye tracking task: comprehension accuracy, and eye gaze fixations. Analysis details are outlined for each task separately, below. Across both measures, data from two participants (one experiment, one comparison) were not recorded at the pretest, due to error. The participant numbers included for analysis were thus 27, 28, and 18 in the training group at the pretest, posttest and delayed posttest, respectively, and 12, 13 and 10 in the comparison group. Our data analysis includes mixed-effects modelling, which is argued to be effective even for unbalanced data sets (Baayen et al., 2008).

a. Comprehension data

For comprehension accuracy, each correct picture selection scored 1, and incorrect scored 0. Proportions of correct picture selection were then calculated for the SVO and OVS conditions, and the fillers, at each test time (Figure 5). For further analysis of the critical data (excluding the fillers), we used the lme4 package (Bates et al., 2015) to fit logistic mixed-effects models to the binary accuracy scores, with fixed effects and interaction of group (experiment, comparison; sum-coded as 1, –1), condition (SVO, OVS; 1, –1), and time (pretest, posttest, delayed posttest). For the time variable, Helmert coding was used, so that the pretest accuracy was first compared with both posttests together ("pre-v-posts"), then the immediate posttest was compared with the delayed posttest ("post-v-delayed"). The maximal random-effects model failed to converge, so we excluded correlations among random effects (following Singmann et al., 2021) then iteratively excluded the random effects associated with the least amount of variance until convergence was achieved.[7]

[7] All models reported in this paper code the group and time variables as reported here. Models that failed to converge followed the same procedure to reach a model that converged. All statistical analyses were conducted in the R statistical environment (R Core Team, 2024).

Figure 5: Proportion selection of the correct picture by condition, time, and group



It is clear from Figure 5 that, for the SVO condition and fillers, there was high accuracy in target picture selection by both groups across test times. However, for the OVS condition a different pattern emerged. At the pretest, accuracy was low for both groups. Then at the posttest, accuracy increased notably in the training group, with further increase at the delayed posttest. In the comparison group, by contrast, there was little change from pretest to posttest, though the delayed posttest accuracy showed some increase.

Table 2. Logistic mixed-effects model for accuracy by group, condition and time[8]

Fixed effects	Estimate	SE	z	p
Group	0.080	0.154	0.520	.603
Condition	1.776	0.189	9.398	<.001
Time[pre-v-posts]	-0.637	0.147	-4.337	<.001
Time[post-v-delayed]	-0.211	0.108	-1.954	.051
Group × Condition	-0.284	0.160	-1.774	.076
Group × Time[pre-v-posts]	-0.368	0.122	-3.015	.003
Group × Time[post-v-delayed]	0.127	0.108	1.170	.242
Condition × Time[pre-v-posts]	0.651	0.177	3.673	<.001
Condition × Time[post-v-delayed]	0.089	0.108	0.823	.410
Group × Condition × Time[post-v-posts]	0.418	0.122	3.439	.001
Group × Condition × Time[post-v-delayed]	-0.050	0.108	-0.463	.643

Turning to the model output (Table 2), of particular note is the three-way interaction of group, condition and time[pre-v-posts]. Considered in light of Figure 5, this interaction is likely driven by the considerably bigger change in accuracy on the OVS condition between pretest and posttests in the training group than in the comparison group. Further exploration of the interaction by means of a model that nested condition within group within time confirms that in the OVS condition, there was a significant change in accuracy between the pretest and posttests in the training group (b = –1.760, SE = 0.163, $p < .001$, odds ratio [henceforth, *OR*] = 0.172, CI = 0.125–0.237).[9] In the comparison group, the

[8] The random effects included by-participant and by-item intercepts, by-participant random slopes for condition and time[pre-v-posts], and by-item random slopes for condition, time[pre-v-posts] and their interaction.
[9] The closer *OR* is to 0, the lower the likelihood of incorrect pictures being selected at Time *y* relative to Time *x*. If *OR* = 1, the likelihood of incorrect picture selection is the same at Time *y* as Time *x*.

change was statistically much smaller (b = –0.382, SE = 0.211, p = .07, OR = 0.683, CI = 0.451–1.032).

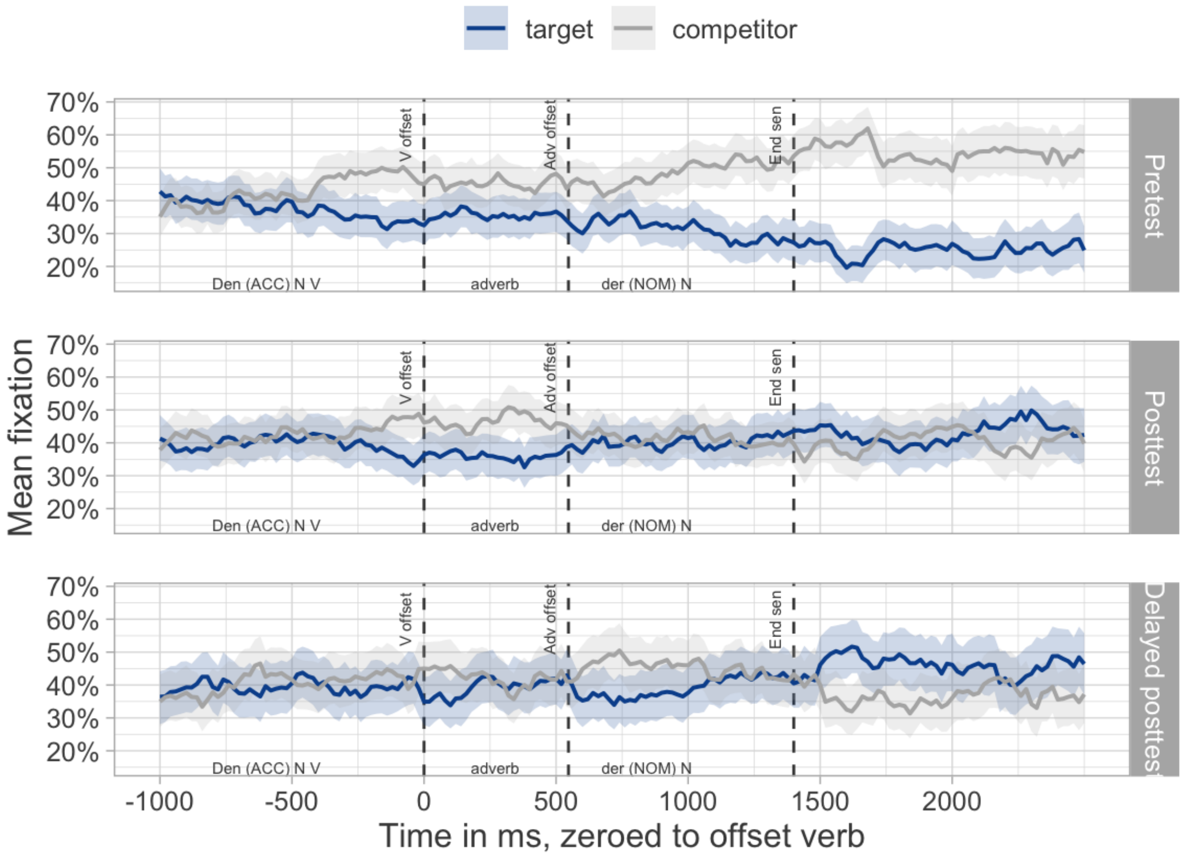
In sum, in the OVS condition (i.e., on hearing an OVS sentence), (i) the training group selected the target OVS picture more frequently at both of the posttests than at the pretest; and, (ii) gains were greater in the training group than in the comparison group. In addition, the comparison group made gains between the immediate and delayed posttests.

b. Eye gaze data

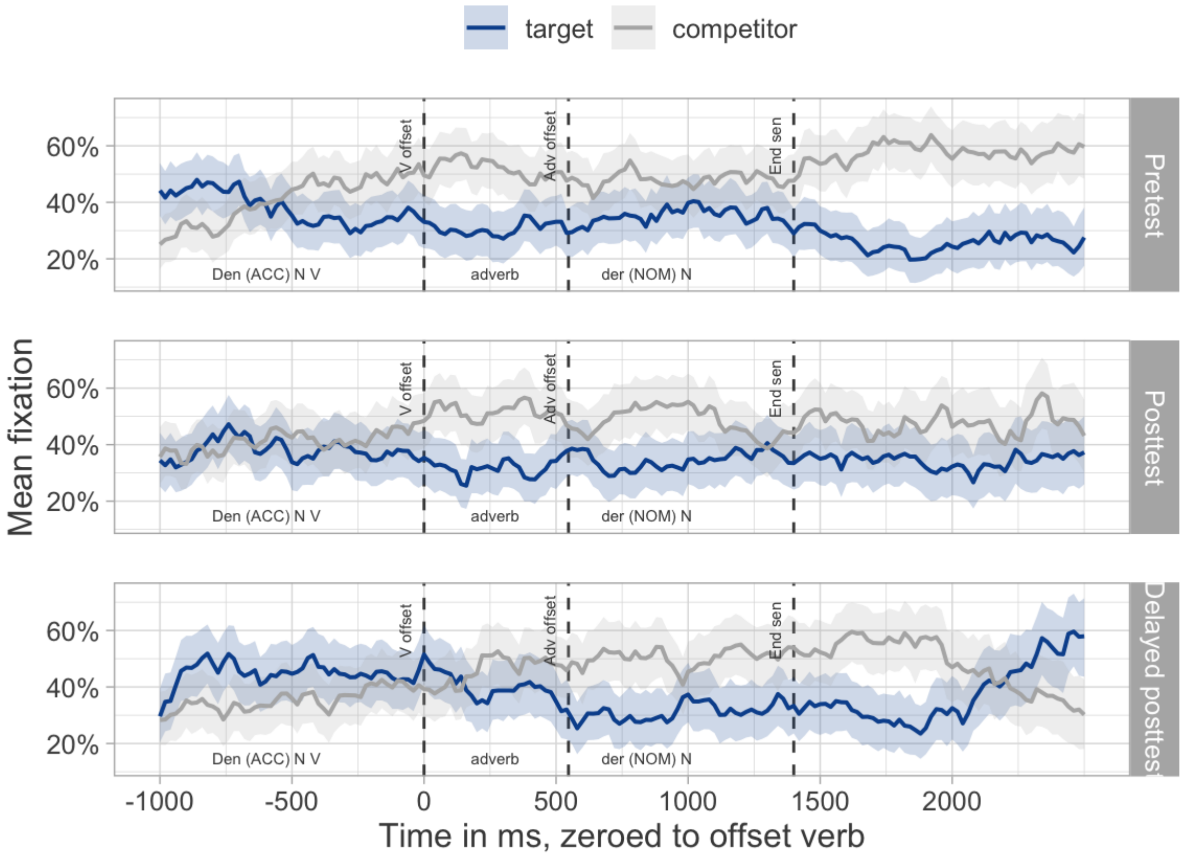
Figure 6 illustrates the proportions of fixations in the target and competitor picture areas within a 3250 ms window as participants listened to the OVS sentences. The zero point in each plot represents the offset of the audio-presentation of the verb. The prediction window (i.e., from offset verb to offset adverb) lasted, on average, 547 ms. The end of the sentence was, on average, at 1400 ms after the verb offset. These plots were created using code adapted from Stone et al. (2020).

Figure 6: Average fixation proportions in the OVS condition, to the target (OVS) and competitor (SVO) pictures over time, at each test time, by the training group (A) and the comparison group (B). (Lines show mean fixation proportions; shading shows 95% bootstrapped confidence intervals.)

A. Training Group



B. Comparison Group



From Figures 6A–B, we see that the two groups showed similar patterns at the pretest, with fixations to the SVO competitor scene increasing over the time period, and diverging from fixations to the OVS target scene before the end of the sentence. This suggests that participants tended to (wrongly) interpret the OVS sentences as SVO. At the posttest and delayed posttest, both groups' proportions of fixations to the target were generally below the proportions to the competitor within the prediction window (verb offset–adverb offset), suggesting that neither group predicted the target picture. Nonetheless, interesting between-group differences can be observed. In the training group, proportions of fixations to the target and competitor were approximately equal before the end of the sentence, whereas in the comparison group, there were more fixations to the competitor at this point.

For further analysis, we tested whether fixation proportions to the target picture differed between group and between test time at two regions of interest defined in relation to the beginning of the prediction window and the end of the sentence. One region of interest (the "prediction window") was defined as running from the offset of the verb to 2 SD beyond the average offset of the adverb: 0–644 ms. The second region (the "end-of-sentence window") was defined as 2 SD before to 2 SD after the average end of sentence (970 –1831 ms). Following the method outlined in Ito and Knoeferle (2023), we calculated log-ratios of fixation proportions to the target versus competitor, using the formula log((Target fixation + .5)/(Competitor fixation + .5)). This quantifies any fixation bias towards the target or competitor, with positive values indicating a bias to the target and negative values a bias to the competitor (Ito & Knoeferle, 2023). Table 3 shows the mean log ratios by group and test time for each region of interest.

Table 3. Mean log ratios of fixation proportions to target versus competitor, in two regions of interest

Time	Prediction window (adverb)		End of sentence	
	M	*SD*	*M*	*SD*
	Training group			
Pretest	-0.132	1.05	-0.321	1.01
Posttest	-0.110	1.06	0.008	1.06
Delayed	-0.043	1.05	0.035	1.07
	Comparison group			
Pretest	-0.255	1.04	-0.266	1.04
Posttest	-0.220	1.04	-0.138	1.05
Delayed	-0.102	1.06	-0.250	1.04

The mean log ratio values in Table 3 are predominantly negative, indicating a bias towards fixations to the competitor throughout the sentence presentation. However, in the training group, the end-of-sentence mean log ratios are positive at the two posttests. To test for effects of group and test time on the log ratios of fixation proportions, we ran linear mixed-effects models for each time window (Table 4), with Time and Group as fixed effects.

Table 4. Linear mixed-effects model results for effects of test time and group on the log-ratios of fixation proportions to target versus competitor, by window[10]

Fixed effects	Estimate	*SE*	*df*	*t*	*p*
			Prediction window		
(Intercept)	-0.140	0.053	42.639	-2.663	0.011
Time[pre-v-posts]	-0.056	0.050	47.056	-1.116	0.270
Time[post-v-delayed]	-0.059	0.053	51.787	-1.103	0.275
Group	0.052	0.044	49.208	1.190	0.240
Time[pre-v-posts] × Group	0.010	0.046	49.210	0.223	0.825
Time[post-v-delayed] × Group	0.018	0.063	42.081	0.278	0.783
			End-of-sentence window		
(Intercept)	-0.153	0.049	49.086	-3.137	0.003
Time[pre-v-posts]	-0.150	0.068	57.233	-2.191	0.033
Time[post-v-delayed]	0.032	0.048	32.881	0.672	0.506
Group	0.055	0.032	35.763	1.731	0.092
Time[pre-v-posts] × Group	-0.083	0.051	37.804	-1.629	0.112
Time[post-v-delayed] × Group	-0.047	0.048	32.972	-0.980	0.334

Unsurprisingly, Table 4 shows no significant effects or interactions in the prediction window. However, in the end of sentence window, there was an effect of Time [pretest vs posttests], and a marginally significant effect of Group. These results indicate that eye gaze behaviour

[10] Maximal models did not converge (with by-participant and by-item random intercepts and slopes for Time and Group, and the by-item interaction for Time × Group). After excluding correlations among random effects, the prediction window model converged. For the end-of-sentence model, the by-item random effects were reduced to (1 + time[pre-v-posts] | item).

was different at the two posttests compared with the pretest, and that there was some degree of between-group difference. However, there was no interaction of Time by Group. This means that our prediction of larger gains in the training group than the comparison group cannot be confirmed. Nonetheless, it seems likely that the main effects of Time and Group in the end-of-sentence window were driven by the increasing looks to the target by the training group in the posttests, as seen in the fixation data in Figure 6 and Table 3. To probe this further, we ran separate mixed-effects linear models for the end-of-sentence window within each group. These confirmed an effect of Time from the pretest to the posttests in the training group (b = -0.232, SE = 0.079, t = -2.941, p = .005) but none in the comparison group (b = -0.061, SE = 0.099, t = -0.618, p = .542). This separate exploratory analysis supports the prediction that the training group would fixate earlier on the target OVS picture at the posttests than at the pretest.

2 Production task

The production task data were transcribed, then coded for type of structure produced (SVO, OVS, passive, or other), and for the determiner used on the object noun. The analysis focused on the 12 critical OVS-No trials (i.e., where the context motivated OVS). Due to a technical problem in the pretest data collection with the comparison group, the responses of eight participants were not recorded correctly and could not be analysed, leaving only five participants in this group for the pretest.

For each test time, we calculated the overall proportion of each structure produced, and the proportion of OVS structures with the target accusative article on the first noun (Figure 7). For further analysis, we coded all instances of OVS structures with an accusative article (henceforth, “target production”) as 1, and all other structures as 0, as our aim was to examine whether training led to increased production of accurate OVS. We then fitted a logistic mixed-effects model to this measure of target production, with fixed effects and interactions of group and time (Table 5).

Figure 7. Proportions of each response type produced at each test time, by group

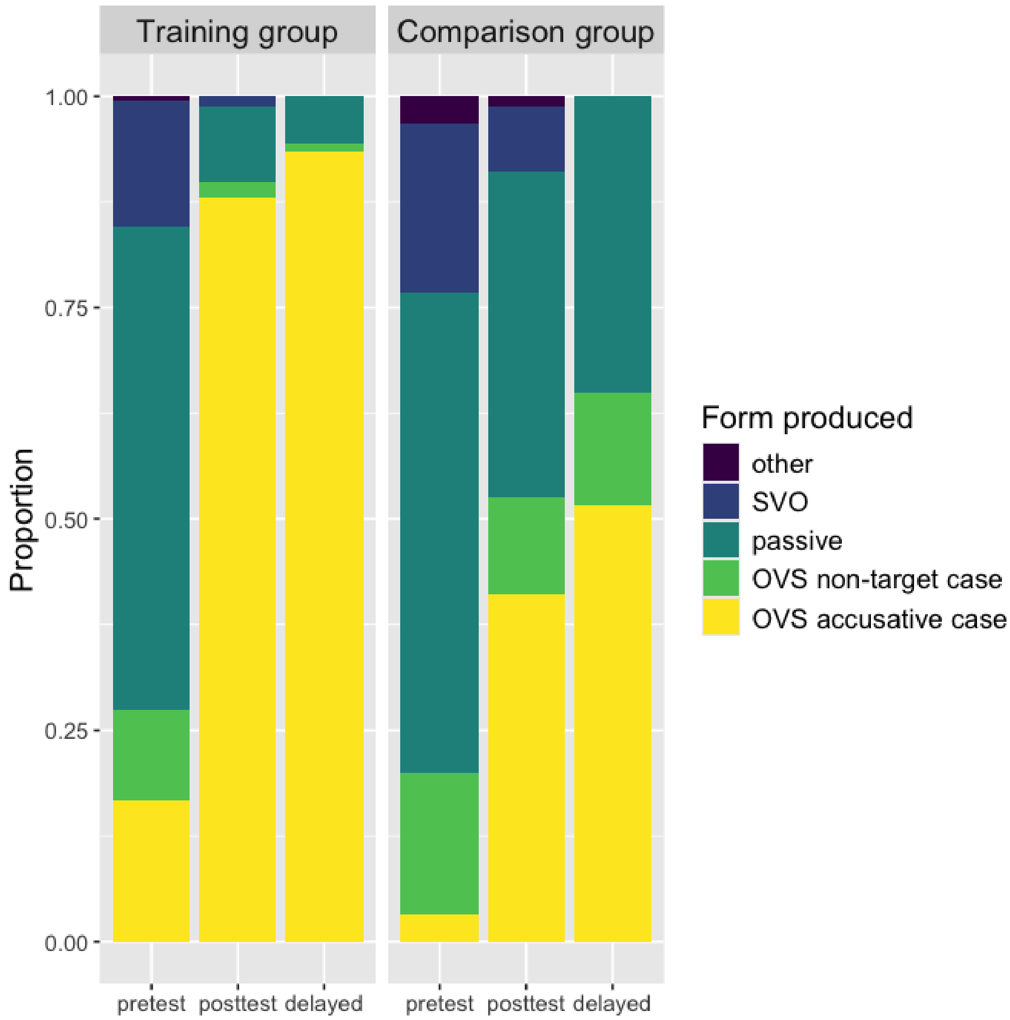


Figure 7 shows that, at the pretest, proportions of OVS were relatively low: <0.17 with the target accusative article, or <0.28 including those with a non-target article (most frequently, the nominative *der*). Instead, both groups tended to produce passive forms as a means of beginning the sentence with the non-agent entity that was circled. For example, for the trial illustrated in Figure 4, some participants produced (9):[11]

6.	Nein,	der	Sonnenhut	wird	von	dem	Tourist	gebraucht.
	No	the.NOM	sunhat	PASS.AUX	by	the.DAT	tourist	needed

“No, the sunhat is needed by the tourist.”

However, rates of accurate OVS production increased dramatically in the training group at the posttest (0.88) and delayed posttest (0.93). Rates of accurate OVS also increased in the comparison group, but to no higher than 0.52 at the delayed posttest.

Table 5: Logistic mixed-effects model for target OVS production by group and time[12]

Fixed effects	Estimate	*SE*	*z*	*p*
(Intercept)	-2.416	1.684	-1.435	0.151
Group	8.643	1.980	4.366	<.001
Time[pre-v-posts]	-10.099	2.958	-3.414	0.001
Time[post-v-delayed]	-1.660	0.809	-2.052	0.040
Group × Time[pre-v-posts]	-8.055	2.969	-2.713	0.007
Group × Time[post-v-delayed]	-1.903	0.877	-2.169	0.030

[11] The non-nominative form for *Tourist* should be *Touristen*, but leaving such minor errors aside, use of the passive is not ungrammatical, though it is unnatural in the scenarios given.

[12] The maximal model did not converge. The model that converged included by-participant and by-item random intercepts, a by-participant random slope for Time(pre-v-posts) and a by-item random slope for Group.

The logistic mixed-effects regression model (Table 5) shows main effects of Group and Time, and interactions of Group with each Time contrast. Considering these results in conjunction with the proportions of accurate OVS responses in Figure 7, it is clear that the first interaction reflects the much greater increase in target production from pretest to posttests by the training group relative to the comparison group. The source of the second interaction is not obvious since both groups increased slightly in target production from the immediate to delayed posttest. A follow-up model that nested time within group indicates that the posttest-to-delayed-postest increase is significant within the training group (b = –2.729, SE = 0.653, p < .001, OR < .001), but not in the comparison group (b = 0.156, SE = 0.438, p = .357, OR = 1.169). This could suggest continued gains in target production within the training group, but less robust growth in the comparison group. However, since participant numbers decreased at the delayed posttest, exploration of findings that focus on that test should be cautious. Moreover, the missing pretest data from eight comparison group participants could affect the results. To investigate this, we ran a supplementary analysis that replaced the missing data with simulated data based on the pretest results of the 33 participants (across both groups) with intact pretest data. The mixed-effects model (Table §1) based on this data also yielded a significant interaction of Group x Time[pre-v-posts] but (unlike in Table 5) no Group x Time[post-v-delayed] interaction. Taken together, the original results (Figure 7, Table 5) and supplementary model confirm the predictions that the training group would produce OVS sentences more frequently at the posttests than the pretest, and that gains in the training group would be greater than in the comparison group.

VI Discussion

The primary goal in this study was to investigate whether training on prediction using case could lead to gains in predictive processing among intermediate-level L2-German learners. The visual-world eye tracking results provide suggestive evidence that the training brought about a change in processing of OVS sentences. Descriptively, the training group's looks to the target picture—indicating an OVS interpretation—increased at the end of the sentence, relative to the pretest where looks were predominantly towards the competitor picture (the SVO interpretation). This contrasted with the comparison group, whose fixations were predominantly towards the competitor picture throughout presentation of OVS sentences across all three test sessions. A within-group model confirmed the training group's increase in target fixations after training. This processing pattern suggests emergent integration of the case-marking information that motivates an OVS interpretation. However, because, at posttest, the looks to the target picture did not begin before the onset of the second noun (i.e., within the *prediction* window), they do not represent case-driven prediction.[13]

One explanation of the training group's change in eye gaze behaviour could be that, instead of beginning to use case during processing, the participants' tendency to process the first noun as the subject simply decreased overall. To try to tease apart whether the present data suggest attenuation of a first-noun bias or the beginnings of use of case, we plotted the difference in looks to the picture corresponding to the SVO interpretation between the SVO

[13] Given slower processing by L2 speakers generally, one might argue that our prediction window may not allow enough time to capture evidence of prediction. However, in our two-picture design, prediction could begin even on hearing just the first noun, before presentation of the verb. Since our prediction window begins from the verb offset, we suggest that it is conservative enough to potentially capture prediction. Future studies may be able to design a longer prediction window that could further increase the chance of capturing any L2 prediction.

and the OVS sentences. The plot (Figure §1) showed that, in the posttests, the training group's relative difference in looks to the SVO picture between SVO and OVS sentences got bigger. This suggests that real-time interpretation of case led them to predict the SVO picture less in OVS sentences than in SVO. If the change in behaviour were due simply to attenuation of their first-noun bias alone, then the relative difference between SVO and OVS sentences should remain constant, while only the absolute proportion of looks to SVO should go down from pretest to posttests.

Our subsequent research questions asked about the extent to which prediction training would lead to more accurate comprehension and production of case. In both the picture selection task and the oral production task, the prediction training was followed by large increases in accurate comprehension and production of case in OSV sentences. These gains were retained and even increased at the delayed posttest. The comparison group also made gains from pretest to posttests, though they were smaller than in the training group, and, in the comprehension task, were only evident at the delayed posttest. These results show that the prediction training effectively led to gains in both comprehension and production of case in OVS sentences. It might be argued that task effects could explain the training group's comprehension gains, because the training activities included a task in which participants had to select the correct image from a choice of two, similar to the eyetracking comprehension task. The training activity differed in that participants heard (or read) just the first determiner of the sentence rather than whole sentences, and the pictures included one masculine noun and one feminine or neuter, whereas, in the eyetracking experiment, all nouns in the critical items were masculine. Nonetheless, it is possible that the training group's familiarity with this kind of task may have contributed to their gains. Importantly, though, since this listen-then-predict activity is integral to prediction training, task-effects do not qualify the effectiveness of the method.

Nonetheless, since the comparison group also made gains in both measures, the training group's gains cannot be ascribed solely to the training. Contributing factors could be the participants' ongoing regular German classes (possibly interacting with the experimental training) and, perhaps more plausibly, undertaking the tests themselves. Each test session essentially provided a flood of input on accusative case-marking and OVS word order. Input flooding has generally been shown to have a positive effect on grammar acquisition (though not always: e.g., Marsden [2006]. See Hernández [2018] for a review). The method aims to trigger implicit learning mechanisms or draw learners' attention to a target grammar form through providing multiple examples of that form within meaningful contexts but without giving explicit information. Though not designed as input flood materials, the present study's test instruments partially meet these aims. Furthermore, the tests may have raised participants' conscious awareness or understanding of the target forms. For example, noticing the presence or absence of OVS was needed to determine picture selection. Moreover, in the oral production task specifically, where the comparison group showed the greatest within-group gains, the design intentionally used contexts in which OVS sentences were felicitous, and it required the participants to begin their utterances with the non-agent noun. This design may have prompted participants to think explicitly about word order and case, or to "notice a gap" (Swain, 1995) in their ability to fulfil the task requirements if they tried to begin with a subject. A study by Koch et al. (2021) investigating predictive processing using German verb morphology provides evidence that L2 participants may become aware during testing that they can use a form predictively.

Such contributions to learning gains via the testing itself notwithstanding, it is clear that the present study's prediction training promoted at least some gains in the processing, comprehension and production of case in the training group, given that group's greater gains relative to those of the group that only did the tests. The evidence of processing gains (i.e., earlier integration of case-marking information after the training) adds weight to Henry's (2023) argument that PI training on case can instigate development towards more nativelike case processing in intermediate-level learners. In Henry's study, this argument hinged on indicative evidence of an effect of training on processing, whereas in the present study, the training group's pretest-to-posttests change in processing was statistically robust and it was maintained at the delayed posttest, even though it did not represent development of the fully predictive processing that the training targeted.

Three key limitations of the present study should be noted. First, the comparison group was small. This reflects the real-world constraints of conducting teaching intervention research with participants from the very limited population of university-level learners of German in the UK. The small sample size means that confidence in the results of between-group statistical models is necessarily reduced, and these results should be interpreted cautiously. Second, the study did not include an additional intervention group, so we cannot assess the effect of the prediction training relative to a different teaching method that also provided a comparable amount of training on case. Finally, the study did not control for possible cross-linguistic influence. Development of case-based predictive processing could be facilitated if participants' L1 (or other previously acquired language) also uses case (see Castle et al., 2025, who found evidence of prior-language influence in offline case comprehension in an artificial language learning study). Additional studies could address these limitations. Further research could also probe the relationship between the modality of the training (both reading and listening in this study) and the modality of outcome measures, neither of which incorporated written language.

VII Conclusion

The key finding from this study is that targeted prediction training using case can affect early implicit processing of case-marking. Though the training did not result in *predictive* processing, it led to real-time processing of (accusative) case morphology that had been absent prior to the training. In other words, borrowing Schlenter's (2023) term, the prediction training created a "favourable condition" at least for the development of earlier processing of case-marking, if not for prediction. L2 prediction based on case thus remains elusive among predominantly L1-English learners, but the indicative gains in processing, and the robust and durable gains in comprehension and oral production following the prediction training, testify to the potential utility of incorporating this approach to teaching certain elements of case and word order in German.

References

Andringa, S., & Curcic, M. (2015). How explicit knowledge affects online L2 processing: Evidence from differential object marking acquisition. *Studies in Second Language Acquisition*, 37(2), 237–268. https://doi.org/10.1017/S0272263115000017

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, 59(4), 390–412. https://doi.org/10.1016/j.jml.2007.12.005

Bates, D., Mächler, M., Bolker, B. M., & Walker, S. C. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67. https://doi.org/10.18637/jss.v067.i01

Braun, B., Doubek, M., Fügert, N., Kotas, O., Marquardt-Langermann, M., Nied Curcio, M., Sander, I., Schäfer, N., Schweiger, K., Trebesius-Bensch, U., Vitale, R., Walter, M. (2016). *DaF kompakt neu A1-B1. Deutsch als Fremdsprache für Erwachsene*. Ernst Klett Sprachen.

Cadierno, T. (1992). *Explicit Instruction in Grammar: A Comparison of Input Based and Output Based Instruction in Second Language Acquisition* Available from Linguistics and Language Behavior Abstracts (LLBA). (58273515; 9402253). https://www.proquest.com/dissertations-theses/explicit-instruction-grammar-comparison-input/docview/58273515/se-2

Castle, C. M., Jensen, I. N., Mitrofanova, N., & Westergaard, M. (2025). Investigating crosslinguistic influence (CLI) in L3 morphosyntax through artificial languages. *Second Language Research*. https://doi.org/10.1177/02676583251332128

Coggle, P., Schenke, H. (2012). *Willkommen! German Beginner's course (2nd edition)*. Hodder Education.

Cristante, V., & Schimke, S. (2020). The processing of passive sentences in German: Evidence from an eye-tracking study with seven- and ten-year-olds and adults. *Language, Interaction and Acquisition*, 11(2), 163–195. https://doi.org/10.1075/lia.19013.cri

DeKeyser, R., & Prieto Botana, G. (2015). The effectiveness of processing instruction in L2 grammar acquisition: A narrative review. *Applied Linguistics*, *36*(3), 290–305. https://doi.org/10.1093/applin/amu071

Dussias, P. E., Valdés Kroff, J. R., Guzzardo Tamargo, R. E., & Gerfen, C. (2013). When gender and looking go hand in hand. *Studies in Second Language Acquisition*, *35*, 356–387. https://doi.org/10.1017/S0272263112000915

Frenck-Mestre, C., Kim, S. K., Choo, H., Ghio, A., Herschensohn, J., & Koh, S. (2019). Look and listen! The online processing of Korean case by native and non-native speakers. *Language, Cognition and Neuroscience*, *34*(3), 385–404. https://doi.org/10.1080/23273798.2018.1549332

Goethe Institut (2004) *Einstufungstest* [Placement test]. Goethe Institut. https://www.goethe.de/en/spr/kup/tsd.html (accessed September 2020).

Grünloh, T., Lieven, E., & Tomasello, M. (2011). German children use prosody to identify participant roles in transitive sentences. *Cognitive Linguistics, 22*(2), 393–419. https://doi.org/10.1515/cogl.2011.015

Grüter, T., Lew-Williams, C., & Fernald, A. (2012). Grammatical gender in L2: A production or a real-time processing problem? *Second Language Research*, *28*(2), 191–215. https://doi.org/10.1177/0267658312437990

Hanan, R. E. (2015). *The effectiveness of explicit grammar instruction for the young foreign language learner: A classroom-based experimental study*. [Doctoral dissertation, University of York]. White Rose eTheses Online. https://etheses.whiterose.ac.uk/9059/

Henry, N. (2022). The offline and online effects of processing instruction. *Applied Psycholinguistics*, 43(4), 945–971. https://doi.org/doi:10.1017/S0142716422000200

Henry, N. (2023). The additive use of prosody and morphosyntax in L2 German. *Studies in Second Language Acquisition*, *45*(2), 348–369. https://doi.org/10.1017/S0272263122000092

Henry, N. (2025). The effects of structured input and output on the acquisition of the accusative case in L2 German. Instructed Second Language Acquisition, 9(1). https://doi.org/10.3138/isla-2024-0003

Henry, N., Culman, H., & VanPatten, B. (2009). More on the effects of explicit information in instructed SLA: A Partial Replication and a Response to Fernández (2008). *Studies in Second Language Acquisition*, *31*(4), 559–575. https://doi.org/10.1017/S0272263109990027

Henry, N., Jackson, C. N., & Dimidio, J. (2017). The role of prosody and explicit instruction in processing instruction. *The Modern Language Journal*, *101*(2), 294–314. https://doi.org/10.1111/modl.12397

Henry, N., Jackson, C. N., & Hopp, H. (2022). Cue coalitions and additivity in predictive processing: The interaction between case and prosody in L2 German. *Second Language Research*, *38*(3), 397–422. https://doi.org/10.1177/0267658320963151

Hernández, T. A. (2018). Input Flooding. In J. I. Liontas & M. Dellicarpini (Eds.), *The TESOL Encyclopedia of English Language Teaching* (pp. 1–7). John Wiley & Sons, Ltd. https://doi.org/10.1002/9781118784235.eelt0073

Hopp, H. (2013). Grammatical gender in adult L2 acquisition: Relations between lexical and syntactic variability. *Second Language Research*, *29*(1), 33–56. https://doi.org/10.1177/0267658312461803

Hopp, H. (2015). Semantics and morphosyntax in predictive L2 sentence processing. *IRAL - International Review of Applied Linguistics in Language Teaching*, *53*(3), 277–306. https://doi.org/10.1515/iral-2015-0014

Hopp, H. (2016). Learning (not) to predict: Grammatical gender processing in second language acquisition. *Second Language Research*, *32*(2), 277–307. https://doi.org/10.1177/0267658315624960

Hopp, H., Bail, J., & Jackson, C. N. (2020). Frequency at the syntax–discourse interface: A bidirectional study on fronting options in L1/L2 German and L1/L2 English. *Second Language Research*, *36*(1), 65-96. https://doi.org/10.1177/0267658318802985

Ito, A., & Knoeferle, P. (2023). Analysing data from the psycholinguistic visual-world paradigm: Comparison of different analysis methods. *Behavior Research Methods*, *55*(7), 3461–3493. https://doi.org/10.3758/s13428-022-01969-3

Kamide, Y., Scheepers, C., & Altmann, G. T. M. (2003). Integration of syntactic and semantic information in predictive processing: Cross-linguistic evidence from German and English. *Journal of Psycholinguistic Research*, *32*(1), 37–55. https://doi.org/10.1023/A:1021933015362

Kasprowicz, R., & Marsden, E. (2017). Towards ecological validity in research into input-based practice: Form spotting can be as beneficial as form-meaning practice. *Applied Linguistics*, 1–27. https://doi.org/10.1093/applin/amw051

Kempe, V., & MacWhinney, B. (1998). The acquisition of case marking by adult learners of Russian and German. *Studies in Second Language Acquisition*, 20(4), 543–587.

Koch, E. M., Bulté, B., Housen, A., & Godfroid, A. (2021). Using verb morphology to predict subject number in L1 and L2 sentence processing: A visual-world eye-tracking experiment. *Journal of the European Second Language Association*, 5(1), 115–132. DOI: https://doi.org/10.22599/ jesla.79

Lew-Williams, C., & Fernald, A. (2007). Young children learning Spanish make rapid use of grammatical gender in spoken word recognition. *Psychological Science*, *18*(3), 193–198. https://journals.sagepub.com/doi/full/10.1111/j.1467-9280.2007.01871.x

Lew-Williams, C., & Fernald, A. (2010). Real-time processing of gender-marked articles by native and non-native Spanish speakers. *Journal of Memory and Language*, *63*(4), 447–464. https://doi.org/10.1016/j.jml.2010.07.003

Marsden, E. (2006). Exploring input processing in the classroom: An experimental comparison of processing instruction and enriched input. *Language Learning*, 56(3), 507-566. https://doi.org/10.1111/j.1467-9922.2006.00375.x

Marsden, E. J., & Chen, H.-Y. (2011). The roles of structured input activities in processing instruction and the kinds of knowledge they promote. *Language Learning*, *61*(4), 1058–1098. https://doi.org/10.1111/j.1467-9922.2011.00661.x

Mathôt, S., Schreij, D., & Theeuwes, J. (2012). OpenSesame: An open-source, graphical experiment builder for the social sciences. *Behavior Research Methods*, *44*(2), 314-324. doi:10.3758/s13428-011-0168-7

McManus, K., & Marsden, E. (2017). L1 explicit instruction can improve L2 online and offline performance. *Studies in Second Language Acquisition*, *39*(3), 459–492. https://doi.org/10.1017/S027226311600022X

McManus, K., & Marsden, E. (2018). Online and offline effects of L1 practice in L2 grammar learning: A partial replication. *Studies in Second Language Acquisition*, *40*(2), 459–475. https://doi.org/10.1017/S0272263117000171

McManus, K., & Marsden, E. (2019). Using explicit instruction about L1 to reduce crosslinguistic effects in L2 grammar learning: Evidence From oral production in L2

French. *The Modern Language Journal*, *103*(2), 459–480. https://doi.org/10.1111/modl.12567

Minor, S., Mitrofanova, N., & Westergaard, M. (2025). The interaction of linguistic and visual cues for the processing of case in Russian by Russian-German bilinguals: An eye tracking study. *Topics in Cognitive Science*, 17, 855–867. https://doi.org/10.1111/tops.12724

Mitsugi, S. (2017). Incremental comprehension of Japanese passives: Evidence from the visual-world paradigm. *Applied Psycholinguistics*, *38*(4), 953–983. https://doi.org/10.1017/S0142716416000515

Mitsugi, S., & Macwhinney, B. (2016). The use of case marking for predictive processing in second language Japanese. *Bilingualism: Language and Cognition*, *19*(1), 19–35. https://doi.org/10.1017/S1366728914000881

R Core Team. (2024). A language and environment for statistical Computing. R Foundation for Statistical Computing. https://www.R-project.org/.

Schlenter, J. (2023). Prediction in bilingual sentence processing: How prediction differs in a later learned language from a first language. *Bilingualism: Language and Cognition*, *26*(2), 253–267. https://doi.org/10.1017/S1366728922000736

Schlenter, J., & Felser, C. (2021). Second language prediction ability across linguistic domains: Evidence from German. In E. Kaan & T. Grüter (Eds.), *Bilingual Processing and Acquisition: Vol. 12. Prediction in second language processing and learning.* (pp. 48–68). John Benjamins.

Shintani, N. (2015). The effectiveness of processing instruction and production-based instruction on L2 grammar acquisition: A meta-analysis. *Applied Linguistics*, *36*(3), 306–325. https://doi.org/10.1093/applin/amu067

Stone, K., Lago, S., & Schad, D. J. (2020). Divergence point analyses of visual world data: Applications to bilingual research. *Bilingualism: Language and Cognition*, 1–9. https://doi.org/10.1017/s1366728920000607

VanPatten, B. (2004). Input processing in second language acquisition. In B. VanPatten (Ed.), *Processing instruction: Theory, research, and commentary* (pp. 5–31). Lawrence Erlbaum.

VanPatten, B. (2020). Input processing in adult SLA. In B. VanPatten. G. D. Keating & Stefanie Wulff (Eds.), *Theories in second language acquisition. 3rd edition* (pp. 105–127). Routledge.

VanPatten, B., & Borst, S. (2012). The roles of explicit information and grammatical sensitivity in processing instruction: Nominative-accusative case marking and word order in German L2. *Foreign Language Annals*, *45*(1), 92–109. https://doi.org/10.1111/j.1944-9720.2012.01169.x

VanPatten, B., & Cadierno, T. (1993). Explicit instruction and input processing. *Studies in Second Language Acquisition*, *15*(2), 225–243. https://doi.org/10.1017/S0272263100011979

Wong, W., & Ito, K. (2018). The effects of processing instruction and traditional instruction on L2 online processing of the causative construction in French: An eye-tracking study. *Studies in Second Language Acquisition*, 40(2), 241–268. doi:10.1017/S0272263117000274