Learning morphology from cross-situational statistics: Exploring cross-linguistic "bottlenecks"

Liuqi Zhu[1], Padraic Monaghan[1,2], Jessie S. Nixon[3], and Patrick Rebuschat [1, 3]

[1] Lancaster University, UK

[3] Max Planck Institute for Psycholinguistics, The Netherlands

[3] University of Tübingen, Germany

Correspondence concerning this article should be addressed to Liuqi Zhu, Department of Linguistics and English Language, Lancaster University, Lancaster LA1 4YL, United Kingdom. Email address: l.zhu8@lancaster.ac.uk.

Author Note

Our power analysis, materials, anonymized data and data analysis scripts are available on our project site (https://osf.io/dvpgq/?view_only=d66af363ddd04ae6bfd2d9a766d21b30) on the Open Science Framework (OSF) platform. We gratefully acknowledge the financial support provided by Lancaster University's Camões Institute Cátedra for Multilingualism and Diversity. Padraic Monaghan and Patrick Rebuschat jointly supervised this project and thus share senior authorship.

Liuqi Zhu, https://orcid.org/0000-0001-5277-9093

Padraic Monaghan, https://orcid.org/0000-0003-3965-2682

Jessie Nixon, https://orcid.org/0000-0001-9698-1445

Patrick Rebuschat, https://orcid.org/0000-0002-5903-3478

Abstract

Vocabulary can be learned by tracking cross-situational statistics between words and aspects of the environment. Nouns and verbs – with observable referents in the environment – can be acquired rapidly, but whether more abstract morphological features can also be acquired cross-situationally, and whether such learning is affected by prior language experience, remain open questions. One hundred and seventeen adults with native languages varying in morphological richness (Chinese, English, German) were exposed to an artificial language comprising intransitive sentences with morphology indicating number, tense, and subject-verb agreement. For each sentence, two scenes appeared on a screen, and participants selected which matched the sentence. Participants learned both vocabulary and morphology from cross-situational statistics, but we found no evidence for an effect of language background. Cross-situational learning is a sufficiently powerful mechanism to drive acquisition of morphological features as well as vocabulary, and is robust to differences in morphological expression from speakers' native languages.

***Keywords*:** cross-situational learning, statistical learning, morphology, **bottleneck hypothesis**, cross-linguistic influence

Introduction

Learning to map sounds onto co-occurring referents in the environment poses an impressive challenge. It is often not possible to figure out the meaning of a novel word based on one scene due to the potential ambiguity of possible mappings that can be made (Quine, 1960). However, recent research has suggested a way to solve the problem of referential ambiguity. After being exposed to multiple scenes, learners are able to determine the mapping between the sound and its referent, by keeping track of cross-situational statistics between words and referents that regularly appear together (Smith & Yu, 2007; Smith & Yu, 2008). To date, cross-situational learning (CSL) has been shown to be effective for vocabulary learning of words with explicit, observable referents in the environment, such as nouns (Dal Ben et al., 2023; Ge et al., 2024; Suanda & Namy, 2012; Smith & Yu, 2008; Vlach & DeBrock, 2019). Learning verbs has also been observed, but in far fewer studies (e.g., Monaghan et al., 2015; Scott & Fisher, 2012; see Roembke et al., 2013, for review). While the learning of the word-world mapping is widely observed in CSL studies, these studies have largely focused on learning referents that can be directly observed in the environment.

However, there have been a small number of studies of CSL that have investigated mappings of morphological features that express relations to more abstract, and harder to observe, properties of the environment to which the language relates. In a CSL study, Rebuschat et al. (2021) investigated whether morphology that marked the agent and patient of the sentences could be acquired from complex sentence-scene correspondences. In their CSL paradigm, adult participants were presented with a scene comprising two aliens interacting with one another in each learning trial, which co-occurred with hearing a transitive sentence in an artificial language that described the scene. After exposure, learning of nouns, verbs, adjectives, marker words, as well as sensitivity to word order, were then tested. They found that verbs and word order were first acquired, followed by nouns, then adjectives, and finally the morphological case markers. However, in a follow-up study, they found that the learning of the case markers was only at a chance level when referential ambiguity was increased by presenting two different scenes (one matching and one a foil) in each learning trial. The learning of case markers was also found challenging in Walker et al. (2020) and Monaghan et al. (2024) using similar CSL paradigms.

Finley (2023) tested the effect to which morphological cues relating to semantic categories of nouns could be acquired through CSL. In each experiment trial, participants heard a novel word with CVCV stem and CV suffix. The inflectional suffix cues were consistently mapped to semantic categories (e.g. [-ke] for fruits) in the experimental group while in the control group the inflectional cue did not carry any meanings. Along with the word, participants saw three items, one of which related to the word they heard. Results showed that the experimental group significantly outperformed the control group, indicating that the morphological cue could be rapidly acquired from CSL, and furthermore, the cue could then be used to improve learning of the stems (see also Monaghan & Mattock, 2012). The beneficial effect of the morphological cues for learning, however, was only realised when referential ambiguity was initially low, enabling the role of the stem and the morpheme to be discerned. Finley's (2023) experiment tested morphological learning in single words, and the complexity of natural language learning involves determining the role of vocabulary and morphology within longer and more complicated sentences with richer syntactic or morphological rules, and also potentially without the benefit of initial bootstrapping from lower referential ambiguity in situations to facilitate initial CSL of morphology.

However, taken together, there is evidence that, under certain circumstances, morphological markers can be acquired from cross-situational statistics. However, the question remains why these markers are more challenging to acquire than other aspects of the language, under conditions where both vocabulary and morphology are learned in parallel (e.g., Mirkovic & Gaskell, 2016; Walker et al., 2020). There are numerous potential explanations for this, which could operate individually or cumulatively to explain learning (Dekeyser, 2005; Ellis, 2022; Slabakova, 2019), and that could limit the effectiveness of CSL to support acquisition of morphology.

First, the immediacy, or transparency, of a cue's referent may influence natural language learning. The function of the case markers was not immediately available within a single scene in Rebuschat et al.'s (2021) study, but instead had to be interpreted from the interoperation among words within the sentence, and between potential agent and patient actors in the environment. This meant that the case markers were potentially more opaque as a referent to the markers than were the nouns, verbs, or adjectives, both because of their lower visual salience (van Zoest & Donk, 2005) and because they required relations among words to be first established (Gleitman et al., 2005) before the mappings could be constructed. Relatedly, the difficulty of learning functional morphology, according to the Bottleneck Hypothesis (BH) (Slabakova, 2019), is proposed to be due to the requirement that it requires integration with a variety of linguistic knowledge (e.g., syntax-morphology, or morphology-semantics) that affect the acceptability when processing the whole sentence. For example, the placement of the functional marker switches between the subject and the verb in English (e.g., Panda cook*s*, Panda*s* cook), which links morphological and syntactic variation. Relatedly, Ellis (2022) identified low contingency (low reliability), of morphology as a contributor to difficulty of learning, another aspect of transparency. For example, in English, the form -s can be interpreted as plural or third-person singular as in the Panda examples above, while the meaning of a word stem is much more stable. Thus, transparency is affected both within the language and in terms of mapping to the environment because of variation in the immediacy of the target referent (Dekeyser, 2005).

If opacity of referents and of mappings are contributors to CSL of morphology, then this is unlikely to be to the same degree across all targets for morphological features. For instance, tense morphology may be difficult to acquire because of difficulties in isolating temporal order in events (Tünnermann & Scharlau, 2018), whereas number (e.g., singular or plural) may be easier to acquire because of its more immediate appearance in environmental stimuli – if there is one or more than one panda in the environment then that is easily observed, but whether the panda walked yesterday or will walk tomorrow cannot be determined from observation of the (current) environment. In contrast, subject-verb agreement may be harder to acquire both because it is not apparent in the environment, and also because of the integration of syntactic and morphological constraints that are required, potentially explaining why subject-verb agreement is widely observed to be difficult to learn in second language acquisition studies (e.g., Jensen et al., 2020; Lecouvet et al., 2021; Wu et al., 2022).

A second possible contributor to lower levels of learning of morphology from CSL was that, in previous studies of case marking learning, the learning target (case marking) was largely absent from participants' native languages. The participants in Rebuschat et al. (2021), for instance, were native English speakers, and English marks agent and patient roles only for pronouns. Learning an additional language can be influenced by previously learnt languages, the phenomenon of which is called cross-linguistic influence (CLI; McManus, 2015; Pajak et al., 2016; Rothman et al., 2019; Suethanapornkul, 2020). A multitude of empirical evidence shows that L1 plays a dominant role in influencing additional language learning, both in classroom (Ahn, 2015; Chen, 2016; Choi & Ionin, 2021; Finn & Hudson Kam, 2015; Hermas, 2010; Hwang & Lardiere, 2013; McManus, 2015) and immersion (Diaubalick & Guijarro-Fuente, 2019) settings.

The extent to which transfer occurs from previously learned languages to affect learning of an additional language has been proposed to depend on structural linguistic similarity on a language feature by language feature basis (e.g., Linguistic Proximity Model: Westergaard, 2019; Westergaard et al., 2017; Scalpel Model: Slabakova, 2017). In particular, similarities of morphological features have been widely documented as affecting additional language learning (Gardner et al., 2021; Hawkins & Hattori, 2006; Jiang, 2007; Van der Slik et al., 2019). For instance, in an implicit language learning task, Ellis (2007) found that adults with L1 that has little morphological inflection pay more attention to lexical cues rather than morphological cues. Similarly, a large dataset of L2 Dutch learners in Van der Slik et al. (2019)'s study revealed that those adults with morphologically less complex L1s performed worse in acquiring Dutch than those with morphologically richer L1s. Jiang (2007) found that intermediate and advanced adult learners were better at detecting errors for morphemes that were not congruent with the learners' L1. Hence, whether the speaker's native language may influence the degree to which different morphological features are detected by the speaker. For example, Mandarin indicates simple present and future tense with prepositions (e.g.,□□□□□□□ tā yībān zhōuyī yóuyǒng (He usually on Monday swim), □□□□□ wǒ míngtiān xuéxí (I tomorrow learn)), whereas English and German tend to indicate tense using morphology or short auxiliary verbs (e.g., I learned/Ich lernte, I will learn/Ich werde lernen), sometimes in addition to prepositions.

A third potential explanation for lower levels of learning of morphology compared to vocabulary learning from CSL is that morphological features require more cognitive effort to learn. Firstly, morphological features tend to be lower in saliency than content words (Dekeyser, 2005; Ellis, 2022). In Finley (2023) and Rebuschat et al. (2021), for instance, morphological markers were shorter than word stems (monosyllabic versus bisyllabic), which could reduce the ability of participants to detect them. The structure of these languages was such so as to respect the natural language property of inflectional and derivational morphology tending to be shorter than word stems (Bybee, 2010). Secondly, morphological features tend to be less contingent in terms of form-meaning mappings. For example, the form of “-s” could indicate both plural and third person singular. Thirdly, the information that morphological features carry is often redundant and thus results in the learning of the morphological cues being blocked as it is cognitively more difficult to detect and to interpret. Due to the universality of this property of natural language, we do not investigate this potential contributor further.

In this study, we grasp this issue of the extent to which adults can learn the meaning of morphological cues where the referent is not immediately available from the environment, determining whether this varies for morphological cues with different functional targets and whether the ability to acquire the morphological cues from CSL depends on the speakers' affinity with morphological cues present within their native language. In this respect, we test two of the theoretical contributors to difficulties in learning morphology. The first is transparency, both within the language in terms of the extent to which the morphology is dependent or independent of syntactic constraints – thus, contrasting subject-verb agreement with tense and number features. But also within the environment, in terms of whether the referent is immediately detectable or not – thus contrasting number with subject-verb agreement and tense).

Previous studies of morphological learning have been shown to be challenging for learners (Jiang, 2007), even when learning is explicitly trained (Ellis & Sagarra, 2010). Learning of morphology implicitly, from cross-situational statistics has also been shown to be difficult, and poorer than learning of nouns, verbs, or adjectives (e.g., Rebuschat et al., 2021). Our first research question was therefore to determine which morphological features can be learned from cross-situational statistics, whether there is an overall difference in learning of morphology compared to other vocabulary, and if so, whether there are differences among different features in terms of their learnability. We thus tested the limits of cross-situational learning in supporting participants' acquisition of tense, number, and subject-verb agreement, in addition to learning of nouns and verbs. If learning is successful then this would indicate that morphological learning is feasible from implicit language exposure and does not require prior learning of word stems. If learning is not successful for morphology, but is effective for vocabulary, then this indicates that morphological learning may be sequential – after vocabulary is acquired – or that more salient cues might need to be made available to ensure learning. We predicted that, overall, nouns and verbs would be learned more accurately than tense, number, and subject-verb agreement. If transparency affects learning of morphology, then we also predicted that number would be learned more easily than tense and subject-verb agreement. However, in our study design, singular and plural morphemes each occur in half. Because tense was cued three-way, each of the past, present, and future inflectional cues appeared in only one-third of the trials. The discrepancy in frequency might lead to different learning outcomes as the role of frequency is considered one of the major factors in second language learning (Larsen-Freeman, 1976; Larsen-Freeman, 1978; Ellis, 1996; Ellis,1998; Ellis, 2002; Ellis, 2008).

The second contributor to morphological learning difficulty that we test is language background – the extent to which the participants' first language expresses few or many morphological features. For this second research question, we tested whether differences in morphological expressiveness in learners' L1 affected acquisition of different morphological features from cross-situational statistics. The results of the current study enable us to detect "bottlenecks" (the difficult linguistic features) for language learners from different native language backgrounds (Slabakova, 2019), with implications for the teaching focus of L2 grammar in a classroom.

For the language groups we selected speakers with L1 that varied in terms of morphological inflectional richness. We therefore tested participants with few inflectional morphological features in their L1 (Chinese), and speakers of L1 with several morphological features in the language but which are relatively poor (English) compared to speakers of languages with relatively rich morphological systems (German). The artificial language in our study included inflections for number and tense, which exist in both English and German languages but not in Chinese. However, the German language contains richer morphological markers that are more complex than English (e.g., grammatical gender; case system). If there is a holistic transfer effect from L1 to additional language learning of morphology, then L1 morphological richness would affect learners' sensitivity towards (all of) the morphological features in the novel language, such that the German group should outperform both the English and Chinese L1 groups. However, if transfer is feature-by-feature according to structural similarity (e.g., Westergaard et al., 2017) then the English and German groups should be similar to one another, and outperform the Chinese L1 group.

Methods

Participants

Sample size was estimated by means of Monte Carlo simulations of data, which predicted that 35 participants per language group would be sufficient for power of 0.8. The detailed description of our power analysis can be found in our preregistration, https://osf.io/dvpgq/. One hundred and seventeen native speakers of Chinese, English and German volunteered for this study. However, ten participants had to be excluded either because they took written notes during the experiment or because their language background did not meet the inclusion criteria of being a native speaker of either Chinese, English or German. Due to unexpected technical issues, two participants had to be excluded due to their missing data in the CSL task and debriefing questionnaires. Our final sample thus consisted of 105 participants (74 female, 31 male), which were distributed into three groups, based on their native language(s).

Thirty-five participants each spoke Chinese, English, and German as their native language. None of the participants reported having learned Portuguese, on which the phonetics of the artificial language were based.

However, all participants in L1 Chinese and L1 German groups reported to have learnt a second language that marked number, tense, and subject-verb agreement, mostly English, whereas 66% (23/35) of participants in the English-native-speakers group reported being monolingual. The mean age in our sample was 23.27 (SD = 4.72, range 18 to 36 years), and there were no significant differences between the groups in terms of age.

The study was approved by the ethics review panel of the Faculty of Arts and Social Sciences at Lancaster University and conducted in accordance with the provisions of the World Medical Association Declaration of Helsinki. None of the participants were remunerated in this study. However, the English native speakers received course credits at their home institution for taking part.

Materials

Artificial language

Vocabulary. There were 21 pseudowords in the language (see Table 1). Sixteen bisyllabic pseudowords served as either nouns or verbs: eight of these referred to distinct animal cartoon characters (panda, pig, lion, mouse, sheep, rabbit, dog, cow) and the remaining eight to different actions (cook, work, swim, run, sleep, walk, sing, paint). Five monosyllabic pseudowords were morphological inflections for either singular or plural number, or past, present, or future tense. We created four different versions of the mappings between words and referents to avoid any biases for particular preferences between words and meanings, shown in Appendix 2.

In order to control for phonotactic preferences across language groups we constructed words from syllables that exist in all three languages with those syllables read by a female native speaker of Portuguese in a monotone. Every pseudo morpheme was read and recorded individually by a female native speaker of Portuguese in a monotone, aiming to avoid accents sound familiar to any of the language groups. The sound files of morphemes were formed into two-word sentences abiding grammar rules based on natural languages on Gorilla Experiment Builder (https://gorilla.sc), with a 250ms pause between different words.

Table 1. The artificial language vocabulary for one of the four randomised mappings. There were 8 nouns, 8 verbs, and 5 morphological inflections (tense and number marking).

Category	Pronunciation	Meaning
Nouns	/faʊlu/	panda
	/fima/	pig
	/fuki/	lion
	/jitu/	mouse
	/kitə/	sheep
	/lipə/	rabbit
	/luʧi/	dog
	/ʃaji/	cow
Verbs	/naɪpə/	cook
	/paʧu/	work
	/paʃə/	swim
	/siʃə/	run

		/pulə/	sleep
		/suli/	walk
		/masə/	sing
		/tusi/	paint
Morphemes	Number	/saɪ/	singular
		/ti/	plural
	Tense	/nɑ/	past
		/kə/	present
		/paʊ/	future

Grammar. Sentences followed subject-verb order conforming with the structure for intransitive sentences in main clauses in all the three languages. We followed the morphological system of a morphologically rich natural language – Portuguese – to design our artificial language. Both singular and plural forms in our artificial language are marked based on Portuguese, whose subject is formed by [stem]+ [suffix for number] and verb is formed by [stem] + [suffix for tense] + [suffix for number]. Similarly, the grammar of our artificial language sentence followed Subject [noun (stem) + inflectional cue for number] + Predicate [verb(stem)+ inflectional cue for tense+ inflectional cue for number].

For example, for the sentence “Fauluti pachunati” meaning “Pandas worked” was constructed as follows:

Faulu ti pachu na ti

Noun plural verb past plural

Visual stimuli. Visual referents for nouns and verbs were presented in static images (See examples in Figure 1). Eight animal cartoon characters (panda, pig, lion, mouse, sheep, rabbit, dog, cow) were used as visual referents for the eight nouns in the artificial language. Each image had either one or two of the same animals performing one of the eight tasks (working, walking, sleeping, singing, swimming, cooking, running, painting).

To reflect time of the event, a visual cue that indicated the time was presented using a written word (past/now/future) in participants` native language appearing within an icon (a left arrow for past, a circle for present, and a right arrow for future). Note that in previous studies teaching tense in artificial languages, researchers have trained participants on the meaning of time adverbs (e.g., Ellis, 2007) prior to exposure to the grammar of the language, whereas we wanted to know whether participants could acquire the tense without this prior training.

Figure 1 illustrates an example of a training trial for an English speaker, where the cues for the meaning of inflectional cues for tenses were written in English (Future, Now, Past), while in native speakers of German group the time visual cues were gestern (yesterday), heute (today), and morgen (tomorrow), and for the Chinese group they were □□(before), □□ (now), □□(after).

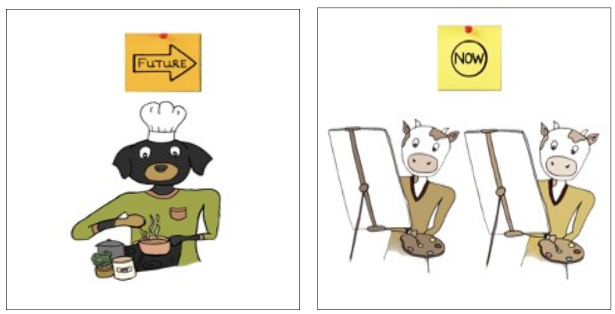

Figure 1 An example of a training trial. The information in each picture contains animal(s) that map with the nouns in artificial language, actions that map with the verbs, and time reference that maps with the suffixes for tense, and number of the animals that maps with the suffixes for number. In the meantime, participants hear a sentence describing one of the pictures: “/luʧisaɪ naɪpəpaʊsaɪ/”, which means “The dog will cook”.

Debriefing questionnaire (retrospective verbal reports)

We also asked questions to investigate whether explicit knowledge was developed during the implicit learning process and whether it predicted the learning outcome by adapting the awareness questionnaire from Rebuschat et al. (2015). The questions ask about what participants were aware of regarding the language structure, gradually providing more explicit information of the language to determine the participants' level of awareness (See appendix for details of the debriefing questionnaire).

Procedure

After providing informed consent and providing background language information, participants were then exposed to the cross-situational learning task, involving training and test trials. Finally, they completed the awareness questionnaire. The entire procedure took around 60 minutes.

Cross-situational learning task

In this task, participants were then instructed to choose one of the two pictures that the sentence was referring to in each trial.

Training trials. During each training trial, participants saw a fixation cross for 500 milliseconds, followed by seeing two static images co-occurring on the screen (See Figure 2). One of the images was the target for the sentence, and the other was a foil, which differed in between two and four of the features (e.g., animal, action, plurality, tense). A thousand milliseconds later, participants heard a sentence of the artificial language describing one of the two images. Right after the sentence finished, participants had to decide as soon as possible which image the sentence was referring to by pressing Q or P on the keyboard. No time limits were set for each trial but there were instructions before and between training trials that remind participants to respond quickly and accurately. By repeatedly experiencing the sentence-scene mappings, we tested whether participants could become sensitive to the co-occurrences of sentences in the language with the referents in the target scene.

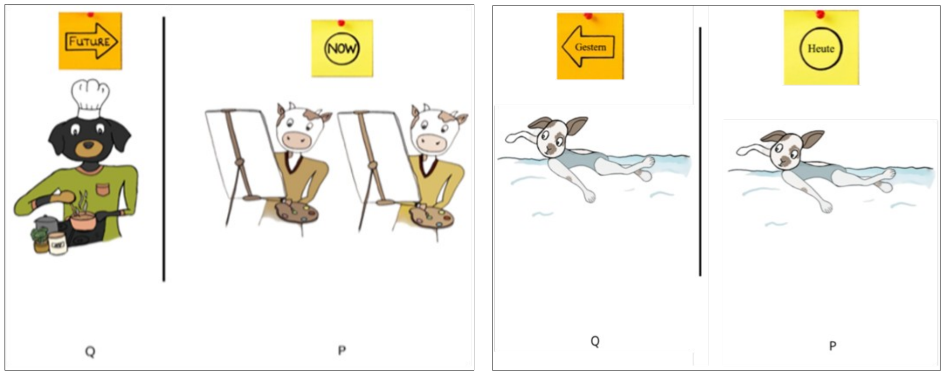

Figure 2 Left: example of a training trial in the native speakers of English group. Two picture choices were presented in each training trial. After 1000 ms, participants also heard a sentence (e.g., “/lutʃisaɪ naɪpəpaʊsaɪ/”, meaning: A dog will cook) that described one of the two pictures. Right: example of a test trial in the native speakers of German group. The pictures differed in only one linguistic feature. For example, participants were presented with the same animals doing the same task at different time. After 1000ms, participants heard a sentence describing one of the pictures (e.g., “/lipəsaɪ/ /paʃənɑsaɪ/”, meaning: A rabbit swam.)

Test trials. Test trials for nouns, verbs, number, and tense were designed in the same way as the training trials, except that one of the four features (referents for noun, verb, tense, or number) differed between the two pictures (see Figure 2). Hence, the correct picture could be selected only if the learner was sensitive to the feature that differed between the two. Thus, the learner's acquisition of the nouns, verbs, inflection cues for tense and number can be isolated through the accuracy of the test trials.

Test trials for the subject-verb agreement test did not include any visual stimuli, and participants heard a sentence in the artificial language. Half the sentences were grammatical, and half presented incompatible number morphemes on the noun and verb (e.g., noun singular, verb plural, or noun plural and verb singular). When the sentence finished playing, participants saw a question mark. They were instructed to press Q if they thought the sentence sounded good or P if it sounded bad to them.

The experiment contained 8 blocks, including 5 training blocks (blocks 1, 2, 3, 5, and 7) and 3 mixed training and testing blocks (blocks 4, 6, and 8). There were 48 training trials in each training block. In each mixed block, there were 48 training trials and 32 test trials comprising 8 trials each for the target features of nouns, verbs, number, and tense, which were randomly inter-mixed. At the end of the mixed block, 8 subject-verb agreement test trials were presented (four grammatical, four ungrammatical).

Nouns, verbs, and morphemes occurred an equal number of times within each block. Animal, action, tense, and number features in the pictures occurred an equal number of times in both target and foil scenes.

Coding and statistical analysis

Cross-situational learning task

We used logistic mixed-effects models (Jaeger, 2008) to analyse the cross-situational learning over time.

The first mixed-effects model was for training trials. Participants' accuracy for training trials was set as a binary dependent variable (correct = 1, incorrect = 0). In the model, Block (1 to 8) was entered as a fixed effect, which was tested for both linear and quadratic effects. To investigate the effect of L1 in the training trials, L1 (Chinese, English and German) was also entered as a fixed effect, with Chinese as the reference level (note we also relevelled with English as reference level to be able to compare English and German L1 groups). We also tested the interaction of L1 and Block as a fixed effect, exploring whether the participants with different L1 learn differently over time. For random effects (Baayen et al., 2008), we included intercepts for subjects and items as well as by-subject and by-item random slopes for Block, L1 and their interaction.

In the next three models, accuracy in the test trials was coded as a binary dependent variable. The second mixed-effects model investigated whether morphology was learnt differently than vocabulary in the test trials. In this model, we included as fixed effect whether the test was for vocabulary or for morphology (noun = 1, verb = 1; tense = -0.67, number = -0.67, subject-object agreement = -0.67). This variable (vocabulary or morphology) was included alongside fixed effects of L1, Block, and the interaction between L1 and vocabulary or morphology. All fixed effects were also included as by-item and by-subject random slopes.

The third mixed-effects model determined whether a specific type of morphological features was more difficult to learn than the others. We included Block, L1, morphology test type (tense, number, subject-object agreement), and the interaction between L1 and morphology test type as fixed effects, with English and tense tests as the reference level (note we also relevelled with Chinese and number tests as reference level to be able to compare Chinese and German L1 groups, number tests and subject-verb agreement tests, and the interactions between the L1 language groups and the morphology testtypes). In terms of subject and item random effects, we started by including intercepts and all fixed effects and their interactions as slopes, but the model did not converge and so we simplified the random effects structure by deleting the interaction slopes and the model then converged.

The fourth mixed-effects model is an exploratory analysis, exploring whether nouns and verbs were learnt differently in cross-situational learning paradigm. We added Block, L1, Nouns or Verbs and interaction between Nouns or Verbs and L1. All fixed effects were also included as by-item and by-subject random slopes.

Debriefing questionnaire

Participants' answers to the debriefing questions were coded following the guidance of Rebuschat et al. (2015)'s coding scheme of awareness, ranking from full awareness to complete unawareness (see appendixes for all debriefing questions). Participants who reported using morphological rules to distinguish words strategically were considered to have "full awareness" (Q1~2), those who mentioned past,present,future (Q3) or singular, plural (Q4) when asking about the patterns of the language or the morphology system were considered "partial awareness", and those who only mention that tense or number in general after the question explicitly asks so were coded as "minimal awareness". Participants who did not report that the morphemes represent tense or number were coded as "unaware". **All participants who reported minimal, partial, or full awareness were coded as "aware" and others "unaware".**

Results

We first determined whether there was overall evidence of cross-situational learning using this paradigm and whether learning was affected by language background, before focusing on whether vocabulary and morphology were learned to different degrees.

Cross-situational learning: Overall learning in training trials

Performance in the training trials is shown in Figure 3. The mixed effects model revealed that participants' accuracy improved over time, with the main effect of Block significant as both a linear (logit estimate = .584, SE = .079, $p < .001$) and a quadratic (logit estimate = - .773, SE = .066, $p < .001$) effect, which reflected the initial steep increase in learning followed by asymptote. Therefore, there was evidence that participants learned as a consequence of exposure to the cross-situational statistics between the language and the scenes.

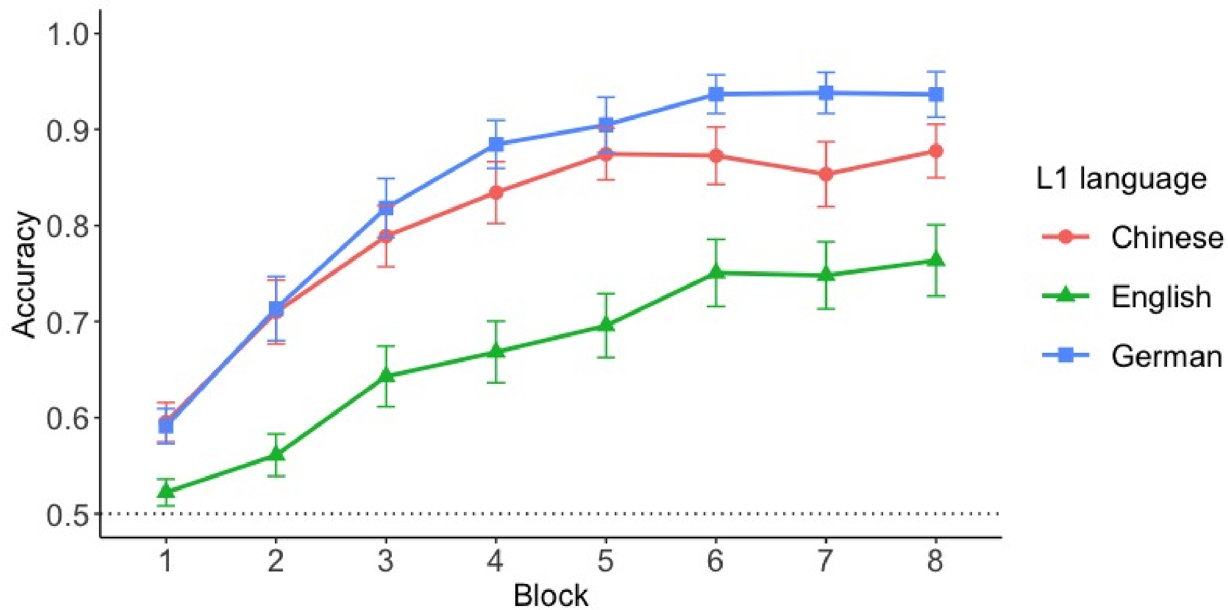

Figure 3 Learning trajectory for each L1 language group over 8 blocks of training. Error bars show standard error of the mean. Dotted line at 0.5 shows chance performance.

The mixed effects model did not indicate a significant main effect of L1. However, the interactions with Block were significantly different between L1 groups. Compared to L1 English group, L1 Chinese (logit estimate = .265, SE = .097, p = .006) and L1 German (logit estimate = .410, SE = .083, p < .001) was interacting positively significantly with Block. This indicates that the groups began at similar levels of accuracy, however, Chinese, and German native speakers learned more rapidly than English native speakers over time. L1 German compared to L1 Chinese group was not a significant interaction with Block (estimate = .147, SE = .099, p = .140). Therefore, no evidence of German native speakers performing significantly better than Chinese native speakers was found in cross-situational learning. The final model is shown in Table 2. In conclusion, while there was not an overall significant difference in performance among the L1 groups, the rate of learning over time differed in the L1 English group from the L1 German or Chinese group.

Table 2. Best fitting model for accuracy in the training trials of CSL tasks, testing the main effect of L1 and Block.

	training_trial_accuracy				
Predictors	*Odds Ratios*	*std. Error*	*CI*	*Statistic*	*p*
(Intercept)	0.94	0.13	0.71 – 1.24	-0.44	0.660
block	1.79	0.14	1.54 – 2.09	7.43	**<0.001**
L1 [English]	0.80	0.13	0.59 – 1.09	-1.39	0.165
L1 [German]	0.79	0.15	0.55 – 1.13	-1.28	0.200
block × L1 [English]	0.77	0.07	0.63 – 0.93	-2.75	**0.006**
block × L1 [German]	1.16	0.11	0.95 – 1.41	1.48	0.140

Number of observations: 42183, Participants: 105, Item: 1036, AIC = 34499.7, BIC = 34914.9, log-likelihood = -17201.8.

R syntax: glmer (training_trial_accuracy ~ block + L1 + block:L1 + (1 + block + L1 + block:L1 | item) + (1 + block + L1 + block:L1 | ppt), data=overall_training_dataset_renamed, family = binomial, control=glmerControl (optCtrl = list(maxfun = 100000), optimizer = "nloptwrap", calc.derivs = FALSE))

Cross-situational learning of vocabulary and morphology

In the training trials, the two scenes differed by two or more properties – so it was not possible to determine which aspects of language learning from cross-situational statistics were driving performance. However, the test trials enable us to distinguish the different aspects of the language. We first determined whether there was a difference in learning the vocabulary and the morphology in the language, and if this was affected by language group. Performance on tests of combined vocabulary (nouns and verbs) and morphology (number, tense, subject-verb agreement) is shown in Figure 4.

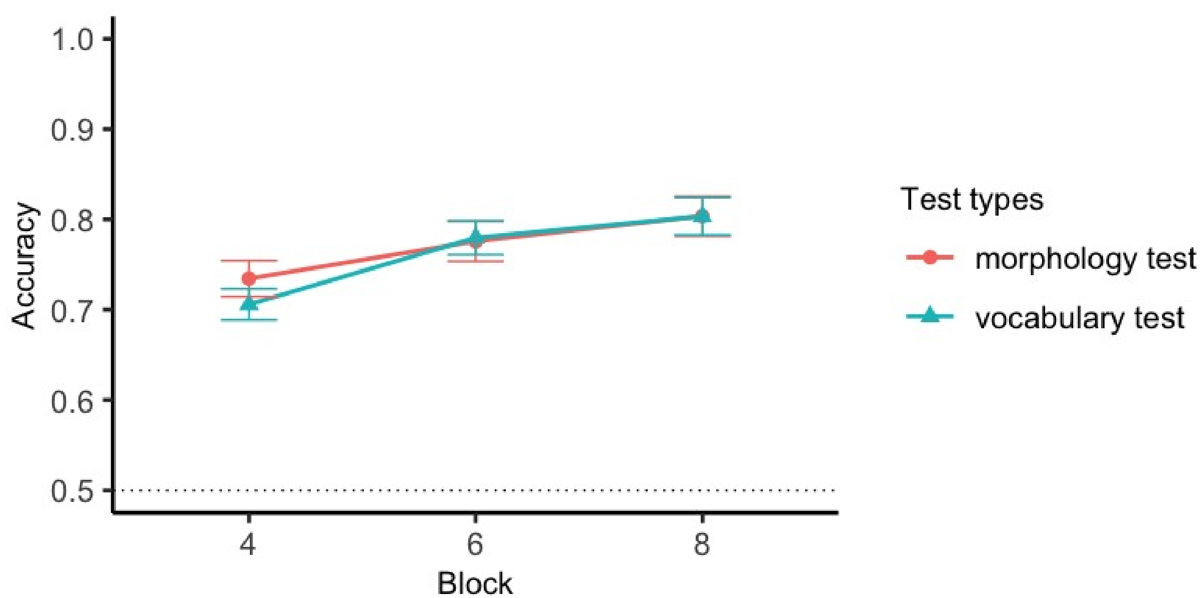

Figure 4 Performance on vocabulary and morphology test trials in the mix-blocks. Error bars show standard error of the mean. Dotted line at 0.5 shows chance performance.

The second mixed-effects model revealed a significant effect of Block (logit estimate = .325, SE =.039, p <.001) and a significant effect of L1 group, with L1 English being significantly less accurate than L1 Chinese (logit estimate = 1.287, SE =.352, p <.001) and L1 German group (logit estimate = 1.115, SE =.330, p <.001). L1 Chinese performed non-significantly better than L1 German group (logit estimate = -.172, SE = .426, p = .686). No significant difference between performance on morphology and vocabulary tests was found in this model (logit estimate = –.055, SE = .114, p = .630). These results demonstrate that overall learning of vocabulary and morphology appeared to be similarly effective from cross-situational statistics as input.

For the interaction between L1 group and vocabulary or morphology test, compared to L1 English group, L1 Chinese group showed a significantly smaller discrepancy in morphology and vocabulary tests (logit estimate = –.554, SE = .244, ***p* = .024**), while the discrepancy in the L1 German group was non-significantly larger than the L1 Chinese group (*p* = .546) and non-significantly smaller than L1 English group (*p* = .141). The final model is shown in Table 3.

Table 3. Best fitting model for accuracy in the test trials of CSL tasks, testing the main effect of test type (vocabulary and morphology tests).

	test_trial_accuracy				
Predictors	*Odds Ratios*	*std. Error*	*CI*	*Statistic*	*p*
(Intercept)	0.60	0.11	0.41 – 0.87	-2.68	**0.007**
L1 [Chinese]	3.62	1.27	1.82 – 7.22	3.66	**<0.001**
L1 [German]	3.05	1.01	1.60 – 5.82	3.38	**0.001**
vocab or morph [vocabulary test]	0.95	0.11	0.76 – 1.18	-0.48	0.630
block	1.38	0.05	1.28 – 1.50	8.23	**<0.001**
L1 [Chinese] × vocab or morph [vocabulary test]	0.57	0.14	0.36 – 0.93	-2.26	**0.024**
L1 [German] × vocab or morph [vocabulary test]	0.71	0.17	0.45 – 1.14	-1.42	0.156

Number of observations: 10227, Participants: 105, Item: 346, AIC = 8996.3, BIC = 9452.0, log-likelihood = -4435.2.

R syntax: glmer(test_trial_accuracy ~ L1 + vocab_or_morph + block + L1:vocab_or_morph + (1 + block + L1 + vocab_or_morph + L1:vocab_or_morph |ppt) + (1 + block + L1 + vocab_or_morph + L1:vocab_or_morph| item), data=overall_testing_dataset_renamed, family="binomial", control=glmerControl (optCtrl = list(maxfun = 100000), optimizer = "nloptwrap", calc.derivs = FALSE))

Cross-situational learning of different language features

In order to investigate whether and when adult learners had acquired each specific feature (nouns, verbs, number suffixes, tense suffixes, subject-verb agreement (SV agreement)) via CSL training and test trials, we ran one-sample t-tests for the test trials in the three test blocks to determine the time during training when the performance was greater than chance. Results for L1 Chinese, L1 English and L1 German groups are shown in Tables 4, 5 and 6. Performance was significantly above chance ($p < .001$) for all test types in L1 Chinese and L1 German groups. English native speakers showed slower learning in tense tests, where the accuracy in block 4 was not significantly above chance yet ($t(35) = 1.57$, **$p = .1256$**, $d = 0.27$) but it increased to above chance level in block 6 (t(35) = 2.50, **p = .01738,** d = 0.42) and block 8 (t(35) = 2.76, **p = .009144**, d = 0.47). This indicates that by block 4, participants in all groups were likely to have acquired all linguistic features in the artificial language, except for the acquisition of tense markers in the L1 English group, the learning of which shows later in block 6 and block 8.

Table 4. One-sample t-tests and Cohen's d for performance against chance level (0.5) for each test type, at each test block in native speakers of Chinese group.

Block	Test Type	t(35)	p-value	p-value
Block 4	Nouns	6.58	1.11	< .001
	Verbs	6.61	1.12	< .001
	Number Suffixes	6.93	1.17	< .001
	Tense Suffixes	6.97	1.18	< .001
	SV Agreement	11.65	1.97	< .001
Block 6	Nouns	7.32	1.24	< .001
	Verbs	7.32	1.24	< .001
	Number Suffixes	9.11	1.54	< .001
	Tense Suffixes	6.18	1.05	< .001
	SV Agreement	8.15	1.38	< .001
Block 8	Nouns	7.40	1.25	< .001

	Verbs	7.38	1.25	< .001
	Number Suffixes	9.75	1.65	< .001
	Tense Suffixes	9.70	1.64	< .001
	SV Agreement	10.88	1.84	< .001

Table 5. One-sample t-tests and Cohen's d for performance against chance level (0.5) for each test type, at each test block in L1 English group.

Block	Test Type	t(35)	Cohen's d	p-value
Block 4	Nouns	2.71	0.46	**.01048**
	Verbs	2.08	0.35	**.04559**
	Number Suffixes	4.26	0.72	**< .001**
	Tense Suffixes	1.57	0.27	**.1256**
	SV Agreement	7.78	1.32	**< .001**
Block 6	Nouns	4.66	0.79	**< .001**
	Verbs	5.36	0.91	**< .001**
	Number Suffixes	3.91	0.66	**< .001**
	Tense Suffixes	2.50	0.42	**< .01738**
	SV Agreement	7.48	1.27	**< .001**
Block 8	Nouns	5.42	0.92	**< .001**
	Verbs	4.83	0.82	**< .001**
	Number Suffixes	3.78	0.64	**< .001**
	Tense Suffixes	2.76	0.47	**< .009144**
	SV	8.34	1.41	**< .001**

	Agreement			

Table 6. One-sample t-tests and Cohen's d for performance against chance level (0.5) for each test type, at each test block in L1 German group.

Block	Test Type	t(35)	Cohen's d	p-value
Block 4	Nouns	9.25	1.56	< .001
	Verbs	4.02	0.68	< .001
	Number Suffixes	9.75	1.65	< .001
	Tense Suffixes	5.96	1.01	< .001
	SV Agreement	6.33	1.07	< .001
Block 6	Nouns	12.03	2.03	< .001
	Verbs	9.43	1.59	< .001
	Number Suffixes	14.19	2.40	< .001
	Tense Suffixes	6.51	1.10	< .001
	SV Agreement	8.70	1.47	< .001
Block 8	Nouns	18.18	3.07	< .001
	Verbs	8.57	1.45	< .001
	Number Suffixes	11.24	1.90	< .001
	Tense Suffixes	12.91	2.18	< .001
	SV Agreement	9.92	1.68	< .001

Performance on tests of nouns, verbs, number suffixes, tense suffixes, and SV agreement is shown in Figure 5.

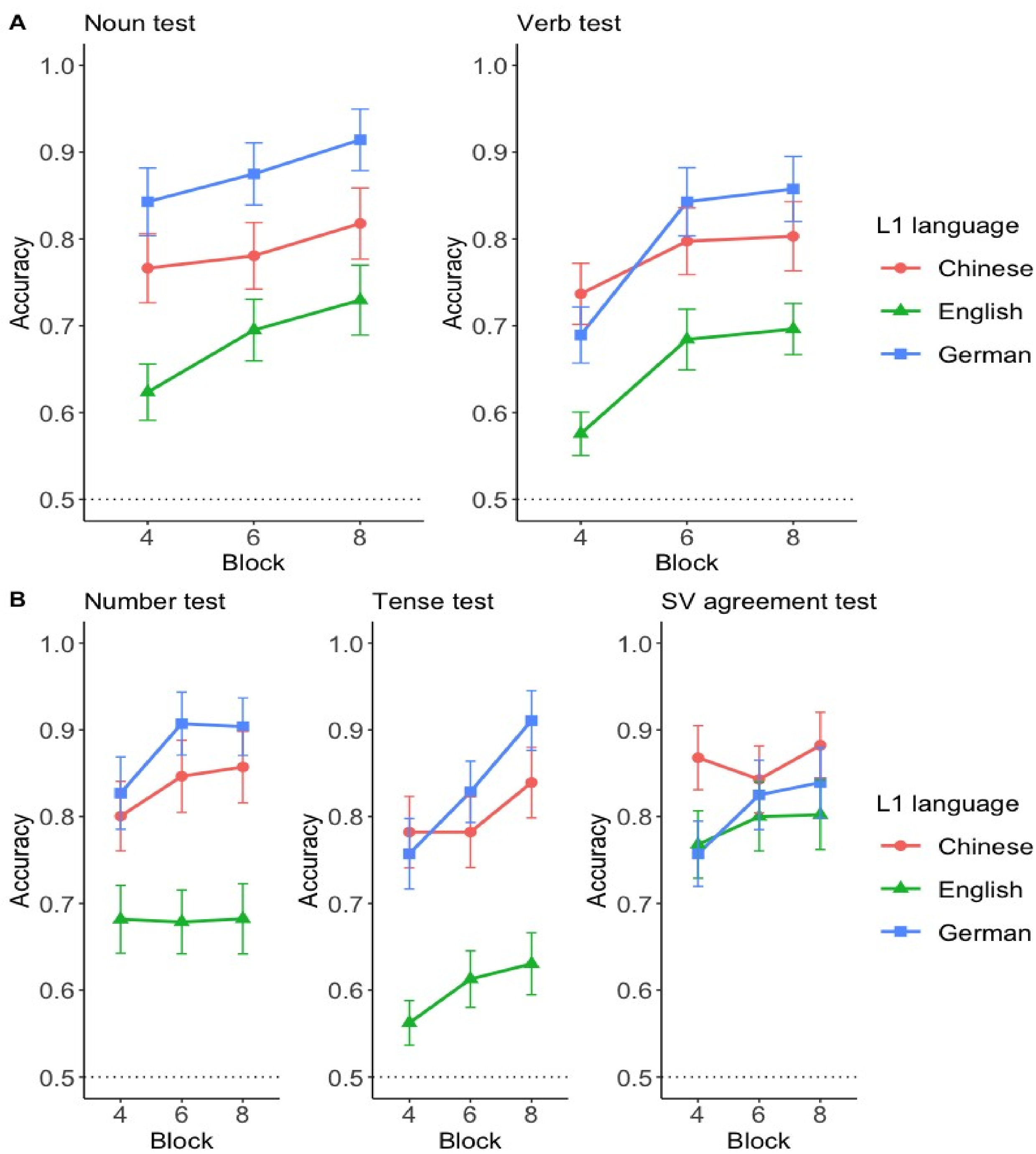

Figure 5 Performance on test blocks for lexical features (nouns, verbs) (A) and morphology features (number suffixes, tense suffixes, SV agreement) (B). Error bars show standard error of the mean. Dotted line at 0.5 shows chance performance.

We next tested whether the native language background had a distinct effect on different aspects of the language, we conducted a mixed effects model just on the morphology tests. Our predictions concerned whether different aspects of the morphological features were affected by L1 and this analysis was pre-registered (the third mixed effect model), however, we also conducted a mixed effects model just on the vocabulary tests as an exploratory analysis (the fourth mixed effect model).

In terms of the performance on the morphology tests, the third mixed effect model revealed a significant main effect of Block (logit estimate = .291, SE = .042, $p < .001$), indicating that the adults are learning morphological features based on cross-situational statistics. The effect of morphology tests is divided into three levels (number test, tense test, subject-verb agreement test). We found that accuracy in the tense tests was significantly lower than in the number (logit estimate = .981, SE = .297, p = .001) and SV agreement tests (logit estimate = 1.362, SE = .353, p <.001). While accuracy in the SV agreement tests was higher than number tests, the difference was not significant (logit estimate = .286, SE = .469, p = .542). For the L1 groups, we found that L1 English is significantly lower than L1 Chinese group (logit estimate = 1.377, SE = .364, p <.001) and L1 German group (logit estimate = 1.356, SE = .373, p <.001), indicating L1 Chinese speakers and L1 German speakers performed significantly better on test trials than L1 English speakers. L1 Chinese group was significantly better than L1 German group (logit estimate = -.902, SE = .433, p = .037**). The accuracy discrepancy between tense tests and subject-verb agreement tests was significantly different between L1 English and L1 German group (logit estimate = -1.586, SE =.527, p = .003) and between L1 German and L1 Chinese group (logit estimate = -.938, SE =.544, p = .085) but not between L1 English and L1 Chinese group (logit estimate = -.657, SE =.519, p = .206). The accuracy discrepancy between number tests and subject-verb agreement tests was significant between L1 English and L1 German group (logit**

estimate = -1.306, SE =.622, p = .036) but not between L1 German and L1 Chinese group (logit estimate = -.698, SE =.648, p = .282) and between L1 English and L1 Chinese group (logit estimate = -.649, SE =.620, p = .295). The accuracy discrepancy between tense tests and number tests was not significant between the L1 groups. The final model is shown in Table 7.

Table 7. Best fitting model for accuracy in the test trials of CSL tasks, testing the main effect of morphology test types (tense, number, SV agreement tests).

	test_trial_accuracy				
Predictors	*Odds Ratios*	*std. Error*	*CI*	*Statistic*	*p*
(Intercept)	0.48	0.12	0.29 – 0.80	-2.85	**0.004**
L1 [Chinese]	3.96	1.44	1.94 – 8.08	3.78	**<0.001**
L1 [German]	3.88	1.45	1.87 – 8.06	3.63	**<0.001**
morphotesttype [number tests]	2.67	0.79	1.49 – 4.77	3.31	**0.001**
morphotesttype [S-V agreement tests]	3.90	1.38	1.95 – 7.80	3.85	**<0.001**
block	1.34	0.06	1.23 – 1.45	6.86	**<0.001**
L1 [Chinese] × morphotesttype [number tests]	0.94	0.42	0.40 – 2.23	-0.14	0.888
L1 [German] × morphotesttype [number tests]	0.74	0.34	0.30 – 1.81	-0.66	0.510
L1 [Chinese] × morphotesttype [S-V agreement tests]	0.52	0.27	0.19 – 1.43	-1.27	0.206
L1 [German] × morphotesttype [S-V agreement tests]	0.20	0.11	0.07 – 0.58	-3.01	**0.003**

Number of observations: 7754, Participants: 105, Item: 302, AIC = 6254.1, BIC = 6615.9, log-likelihood = -3075.1.

R syntax: glmer(test_trial_accuracy ~ L1 + morphotesttype + block + morphotesttype:L1 + (1 + block + L1 + morphotesttype | ppt) + (1 + block

+ L1 + morphotesttype | item), data=overall_testing_add3, family="binomial", control=glmerControl (optCtrl = list(maxfun = 100000), optimizer = "nloptwrap", calc.derivs = FALSE))

For the vocabulary tests, we found a significant main effect of Block (logit estimate = .313, SE = .043, $p < .001$), so learning proceeded effectively on vocabulary items with exposure (see Figure 5). We found that the accuracy of noun tests was significantly higher than verb tests (logit estimate = -.418, SE = .195, p =.032). For language group, we found that accuracy of the L1 Chinese (logit estimate = .698, SE = .317, p = .027) and L1 German (logit estimate = 1.462, SE = .362, $p < .001$) groups were significantly higher than the L1 English group. Accuracy in the L1 German group was significantly higher than L1 Chinese group (logit estimate = .760, SE = .372, p = .041). **The interaction between the L1 German and vocabulary tests (nouns or verbs) was significantly lower than the interaction between L1 Chinese and vocabulary tests (logit estimate = -.757, SE = .344, p = .028), indicating that the difference in accuracy between noun and verb tests are significantly larger in the**

L1 Chinese group, compared to L1 German group. However, the discrepancy was not significant between L1 English and L1 Chinese groups (p= .176) or between L1 English and L1 German group (p = .226). The final model is shown in Table 8.

Table 8. Best fitting model for accuracy in the test trials of CSL tasks, testing the main effect of vocabulary test types (noun and verb tests).

	test_trial_accuracy				
Predictors	*Odds Ratios*	*std. Error*	*CI*	*Statistic*	*p*
(Intercept)	0.62	0.17	0.36 – 1.06	-1.76	0.079
noun or verb [verb test]	0.66	0.13	0.45 – 0.97	-2.14	**0.032**
block	1.37	0.06	1.26 – 1.49	7.21	**<0.001**
L1 [Chinese]	2.01	0.64	1.08 – 3.74	2.21	**0.027**
L1 [German]	4.32	1.56	2.12 – 8.77	4.04	**<0.001**
noun or verb [verb test] × L1 [Chinese]	1.40	0.35	0.86 – 2.27	1.35	0.176
noun or verb [verb test] × L1 [German]	0.65	0.23	0.33 – 1.30	-1.21	0.226

Number of observations: 5111, Participants: 105, Item: 188, AIC = 4734.8, BIC = 5146.7, log-likelihood = -2304.4.

R syntax: glmer(test_trial_accuracy ~ noun_or_verb + block + L1 + noun_or_verb: L1+ (1 + block + noun_or_verb + noun_or_verb: L1| ppt) + (1 + block + noun_or_verb+noun_or_verb: L1 | item), data=overall_testing_add3, family="binomial", control=glmerControl (optCtrl = list(maxfun = 100000), optimizer = "nloptwrap", calc.derivs = FALSE))

Retrospective verbal reports

Participants' answers to the debriefing questions were coded following the guidance of Rebuschat et al. (2015)'s coding scheme of awareness, ranking from full awareness to complete unawareness (see appendixes for all debriefing questions). Participants who reported using morphological rules to distinguish words strategically were considered to have "full awareness" (Q1~2), those who mentioned past, present, future (Q3), singular, plural (Q4) or specified the subject-verb agreement when asking about the patterns of the language or the morphology system were considered "partial awareness", and those who only mentioned tense, number or pattern of sounds were coded as having "minimal awareness" .

Participants who did not report tense, number or subject-verb agreement were coded as “unaware”. **All participants who reported minimal, partial, or full awareness were coded as “aware” and others “unaware”.**

Following the criteria outlined above, we found that sixty out of a hundred and fifty participants were fully aware of the morphological rules (L1 Chinese: 21; L1 English: 12; L1 German: 28). All participants reported guessing at the beginning, but some later used strategies of calculating the number of categories in the pictures (animals, actions, time, number of animals) to figure out their meaning by comparing similar pictures. And some others reported that they learned from the errors when testing and renewing different assumptions. Seventeen participants at some level noticed the morphological cues. When asked about the meaning of inflectional cues, they either gave a generic answer like number or tense, SV agreement or were more specific about the meaning of each sound (e.g., “Yes, I believe 'sai' is highlighting a single character, while 'ti' stands for many.” ; “The structure of the sentense is (subject+number suffixes+verb+tense+number suffixes)”). According to the criteria, nine participants were categorised as partially aware of the morphological rule (L1 Chinese: 4; L1 English: 2; L1 German: 3) and eight were minimally aware (L1 Chinese: 5; L1 English: 2; L1 German: 1). The rest of the twenty-seven participants reported no awareness of the morphological cues (L1 Chinese: 5; L1 English: 19; L1 German: 3). Accuracies of CSL tasks between participants who showed different level of awareness in their debriefing questionnaires are shown in Figure 6.

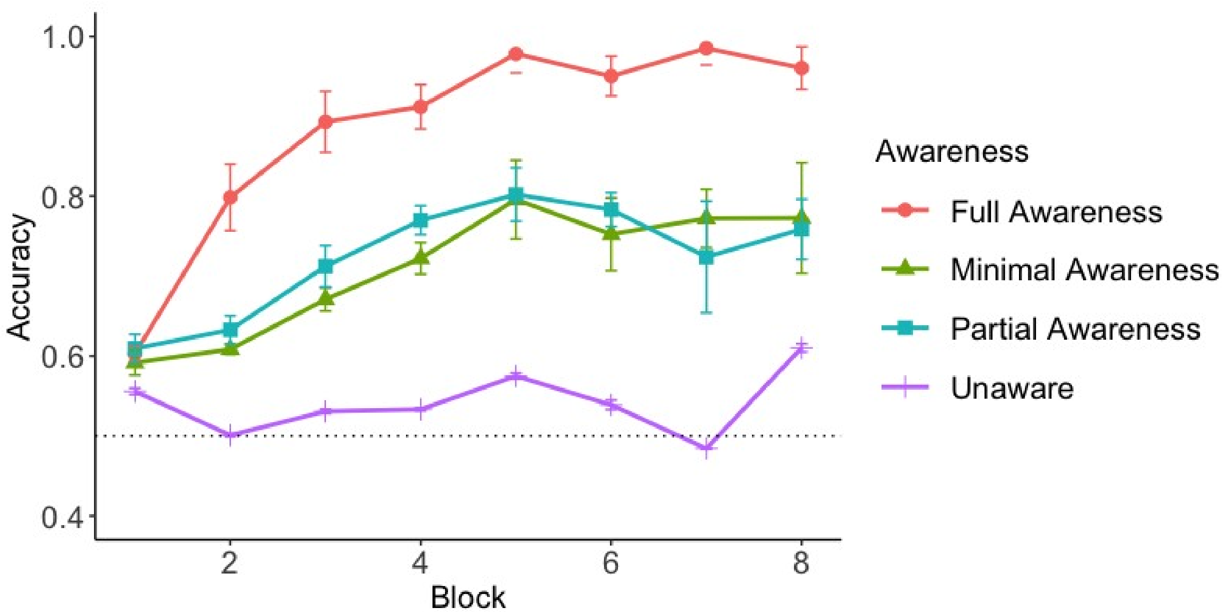

Figure 6. Participants' accuracy on CSL tasks: comparisons between different levels of awareness groups (Full Awareness, Partial Awareness, Minimal Awareness, Unaware).

Note. Error bars represent 95% Confidence Intervals.

Participants who showed full awareness performed significantly better than partial awareness ($t(12490)=20.691$, $p < .001$) and minimal awareness groups ($t(12992)=24.283$, $p < .001$). We did not find significant differences between partial-aware and minimal-aware groups ($t(4526)=1.166$, $p =.244$). The partial-aware and minimal-aware groups showed significantly higher accuracy in CSL tasks than the unaware group ($t(7010)=15.532$, $p < .001$). As there’s a clear discrepancy in CSL performance between people who are aware and unaware of the morphological features, we further explored the performance differences between adults who generate awareness during the immersive learning environment and those who do not. For the following analysis, we included full awareness, partial awareness and minimal awareness groups in the awareness group, in comparison with the unawareness group.

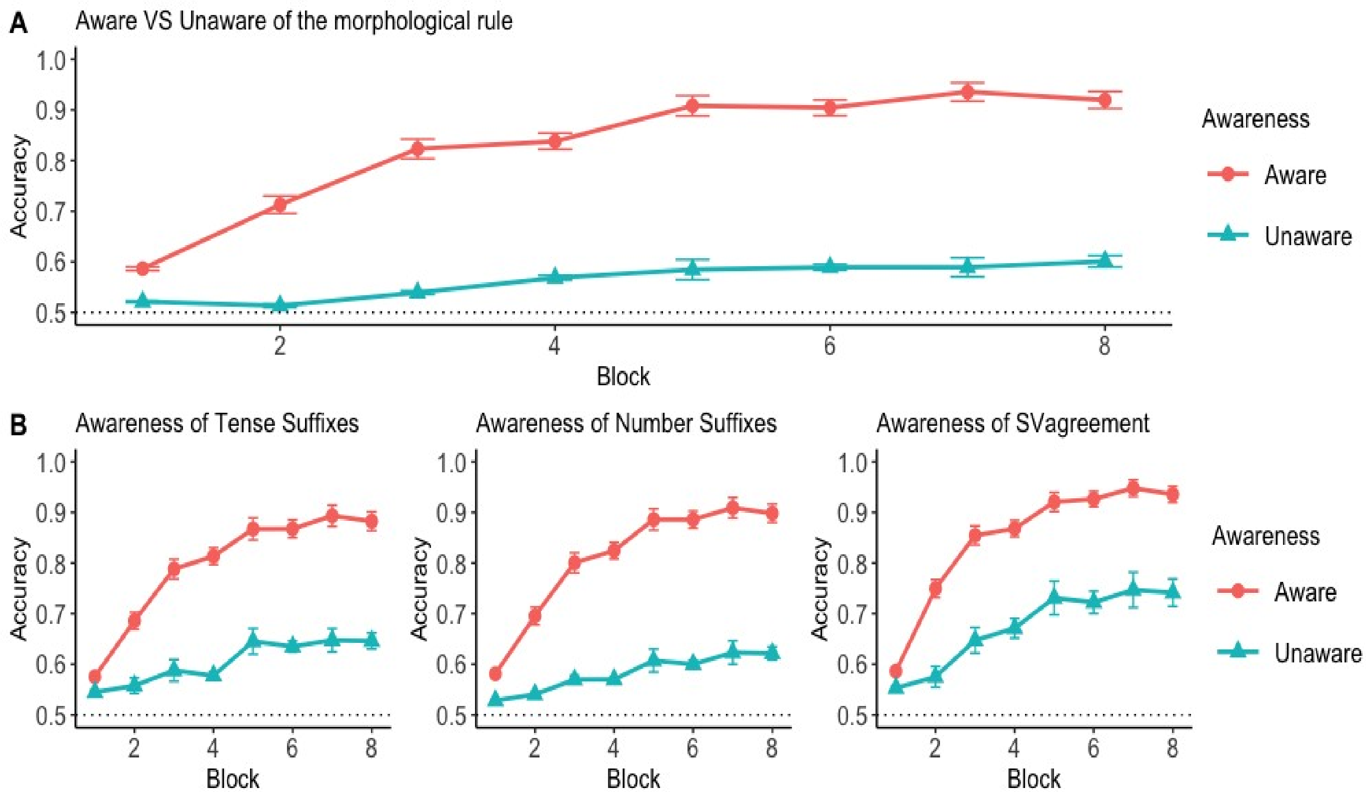

Figure 7. Participants' accuracy on CSL tasks: comparisons between participants who reported to be aware and unaware of the overall morphological rules (A) and specific morphological feature (B).

Note. Error bars represent 95% Confidence Intervals.

As shown in Figure 7(A), the learning trajectories of aware and unaware participants are significantly different from each other. At the end of the CSL task (Block 8), participants who showed awareness (M = .920, SD = .147) of the morphological rules were significantly higher than those who were unaware (M = .601, SD = .058) according to the debriefing questionnaires; $t(14) = 5.929$, $p < .001$. Figure 7 (B) shows that participants being aware of the specific morphological linguistic features (tense suffixes, number suffixes, SV agreement) performed significantly better than participants being unaware of one/more of the morphological features. The learning trajectory of the test trials among participants who showed awareness of tense, number and SV agreement are similar. All of them performed significantly differently to participants who showed unawareness of the related morphological features.

To investigate the influence of awareness on adults' performance in CSL tasks, we employed mixed-effects models with fixed effects for block, L1 language, awareness status (aware vs. unaware), and their three-way interaction sequentially. We found that adding the fixed effects of Awareness significantly improved the model fit ($\chi^2(6) = 3149.19$, $p < .001$), compared to the model with just Block, L1. This suggests that the generalisation of awareness predicts the accuracy of the CSL tasks. However, the inclusion of their interactions did not significantly enhance the model fit ($\chi^2(1) = 1.18$, $p = .277$). The best-fitting model, which included awareness, block, language, and their interactions, is detailed in Table 9. The analysis of the debriefing questionnaire demonstrates the importance of awareness of the morphological features in the early acquisition of second language learning, specifically in an immersive learning setting.

	Correct				
Predictors	*Odds Ratios*	*std. Error*	*CI*	*Statistic*	*p*
(Intercept)	1.23	0.18	0.93 – 1.63	1.47	0.142
general awareness [Unaware]	0.81	0.29	0.40 – 1.62	-0.61	0.545
block	1.68	0.05	1.58 – 1.78	17.75	**<0.001**
L1 [English]	0.57	0.09	0.41 – 0.78	-3.44	**0.001**
L1 [German]	0.93	0.18	0.64 – 1.35	-0.39	0.697
general awareness [Unaware] × block	0.64	0.03	0.57 – 0.71	-8.27	**<0.001**
general awareness [Unaware] × L1 [English]	1.67	0.62	0.80 – 3.47	1.37	0.170
general awareness [Unaware] × L1 [German]	1.45	0.76	0.51 – 4.07	0.70	0.484

Table 9. Best fitting model for accuracy for the overall CSL tasks, testing awareness effect

Number of observations: 55048, Participants: 105, Item: 1152, AIC = 46319.4, BIC = 46569.1, log-likelihood = -23131.7.

R syntax: glmer(Correct ~ general.awareness + block + L1 + general.awareness:block + general.awareness:L1+ (1 +block + L1 | ppt) + (1 +block +L1 | item),

data=Overall_add_awareness, family="binomial", control=glmerControl (optCtrl = list(maxfun = 100000), optimizer = "nloptwrap", calc.derivs = FALSE))

Discussion

Overall, the results show that adults have the cognitive ability to solve the form-meaning mappings of morphological features, along with acquiring the vocabulary by processing sentence-based cross-situational statistics. Participants' performance on both training and test trials significantly improved over blocks. Regarding the effect of L1 background, we found that while participants' performance in training trials was not affected by L1 morphological complexity, Chinese, and German native speakers learned more rapidly than English native speakers over time. In test trials, German and Chinese native speakers have a distinctive higher accuracy than English native speakers. We also investigated whether vocabulary and morphology were learned to different degrees. The results suggest that learners perceive the acquisition of vocabulary and morphology in our artificial language, given their distinctive levels of salience, to involve a comparable degree of cognitive processing difficulty. Results also showed that learning the meaning of cues in CVCV and CV forms are processed at a similar level of cognitive effort in L1 Chinese group but not in L1 English group, which indicates that L1 English groups are more sensitive towards the effect of cue salience. In terms of morphology learning, Subject-verb agreement was learned better than inflectional cues for numbers, followed by tense markers. The form-meaning mappings of tense and number are processed at a similar level of cognitive effort among the three L1 groups. Lastly, as an exploratory analysis, we found an overall noun learning bias in our results. Next, we will examine each of these findings in depth.

For research question 1, we found that adults can rapidly map the form-meaning of both lexical and morphological cues. Our findings suggest that it is likely adults possess the cognitive ability to rapidly map sounds of inflectional cues to their meanings/functions in an immersive learning environment, without any feedback or explicit instruction of vocabulary or grammar. As we have controlled the cooccurrence of nouns and verbs, and their combination with morphemes, it is likely that participants learnt by keeping track of the cross-situational statistics rather than memorising chunking (one part of the sentences). In contrast to findings in previous CSL studies (e.g., Monaghan et al., 2021; Rebuschat et al., 2021; Walker et al., 2020), the success of learning the morphological cues (including the more abstract subject-verb agreement) in our study might be due to that all the targeted linguistic features (noun, verb, inflectional cues for number and tense, subject-verb agreement, SV order) were possessed in participants' native languages or previously learnt languages. The discrepancy of findings between studies tentatively supports that the L1-absent morphological markers limit the short-term and perhaps long-term attainment of L2, aligned with findings in Jiang (2007).

The robust learning might not reflect natural language learning due to the high salience level of morphological cues in our study. Although we manipulated inflectional cues to be lower salience than lexical cues, the inflectional cues in our artificial language were still higher than some inflectional cues in naturalistic languages. For example, the cues for past tense “-ed” in the example of an English sentence, “Yesterday I walked.”, were only pronounced as a single consonant /t/, which was less pronounced than /CV/ form in our artificial inflectional cues. In our study, participants performed very similarly in both vocabulary and morphology tests, which means the differences in perceptual salience between /CV/ and /CVCV/were not enough to reflect the discrepancy of salience between morphological cues and lexical cues in natural language. Although the effect of cue salience can be mitigated by instruction (Cintrón-Valentín and Ellis, 2016), cue salience is one of the determiners of morphology learning difficulty in natural language learning (Ellis, 2022). However, how much salience of a cue, over the spectrum of perceptual salience, distinguishes the lexical and morphological form-meaning mapping, is yet to be further researched.

The high contingency of the morphological cues in our study could also explain the similar accuracy observed in morphology and vocabulary tests. Note that inflectional cues in natural languages are often not only low salience compared to vocabulary but also low contingency (Ellis, 2022). For example, the form “s” in English could mean plural or present singular. The form “-ing” could represent a noun but it could also mean the action is ongoing. The inflectional cues in our study were less complicated. Both inflectional cues and lexical cues were bound with one meaning individually. This finding indicates that low salience of morphological cues alone might not be responsible for morphology learning difficulty, at least not to a large extent. The low contingency of the form-meaning mapping in morphology might.

Our study shows much higher learning of functional markers than previous studies where case marker was found to be difficult to learn (Katharina et al., in prep; Monaghan et al., 2021; Rebuschat et al., 2021; Walker et al., 2020). Compared to the mapping of case markers in Rebuschat et al. (2021) and its replicate studies, the mapping of inflectional cues (number and time reference) in our study is less ambiguous and is visually more available. For example, inflectional cues for the quantity of the subjects were visually easier to spot cross-situationally, while the interpretations of the two case markers were more hidden behind the two objects interacting with each other in the visual world. According to the bottleneck hypothesis (Slabakova, 2019), the functional morphology, such as subject-verb agreement was claimed to be the most difficult part to learn in second language acquisition, due to the property of morphology that is intertwined with syntactical and semantical knowledge. The results in our study show otherwise. Robust learning was also found in the subject-verb agreement tests, which indicates that the difficulty of learning might not come from the "agreement" between the subject and the verb. In other words, the intertwining of morphology and syntax might not cause the learning difficulty. What seemed like a syntactical-semantic problem was a mapping problem. In our study, although the singular/plural markers agree between subject and verb, they always map one meaning in the visual world. In contrast, in the case of third person *-s* in English, the same form "s", the number marker that switches between subject and verb, could indicate either one or two objects in the visual world.

For research question 2, we did not find significant differences between groups, which means we did not find evidence that having a morphologically rich L1 makes a difference in learning L2 morphology. The influence of L1 on performance in our study is nuanced and context-dependent, varying across different levels of the block variable. This finding is not in line with the findings in Van der Slik et all. (2019), where he found L1 morphologically poor speakers tend to have lower proficiency in L2. However, they also found that commanding a morphologically rich L2 facilitates adult L2A. This is consistent with our result which indicates prior language experience might have a larger effect on adult L2A. The much higher learning of the functional markers in our study might be due to that most Chinese native speakers and German native speakers report having a high proficiency in at least another language, which might tune their learning attention because of this explicit learning background of numbers and tense markers (Bono, 2011). Although Chinese native speakers do not have number and tense markers in their L1, the experience of explicitly learning another language tunes their attention to those markers during the CSL.

Our results indicate that learned attention is not only L1-tuned but might be tuned by properties in all previously learned language(s). The role of L1 is less dominant when it comes to second language learning. All previously learnt languages act together as a whole impacting our learned attention in second language learning.

Conclusion

Our study provides evidence that adults possess the cognitive ability to learn vocabulary and morphology under cross-situational learning conditions, implicating that the limited attainment from adult L2A might not be due to biological reasons. The current study also provides more understanding on what Ellies (2022) proposed that cue salience as one of the factors that makes morphology difficult to learn. Further research exploring the role of cue salience is necessary for better understanding of the adult morphology learning.

We also found that L1 does not act as a dominant language to influence L2 morphology learning in cross-situational learning contexts. The result implicates that cross-linguistic might transfer happened property by property and it can be from all previous-learnt language. Further research can focus on whether being a bilingual or receiving explicit instruction of a second language have a larger impact on adult L2A.

Reference

Ahn. (2015). *Second Language Acquisition of Korean Case by Learners with Different First Languages.* ProQuest Dissertations Publishing.

Bornstein, Hahn, C.-S., & Haynes, O. M. (2004). Specific and general language performance across early childhood: Stability and gender considerations. *First Language, 24*(3), 267–304. https://doi.org/10.1177/0142723704045681

Bybee, J. (2010). *Language, usage and cognition*. Cambridge: Cambridge University Press.

Chen. (2016). *Second language perception and production of English regular past tense: L1 influence in phonology and morphosyntax.* ProQuest Dissertations Publishing.

Childers, & Paik, J. H. (2009). Korean- and English-speaking children use cross-situational information to learn novel predicate terms. *Journal of Child Language, 36*(1), 201–224. https://doi.org/10.1017/S0305000908008891

Childers, & Tomasello, M. (2002). Two-Year-Olds Learn Novel Nouns, Verbs, and Conventional Actions From Massed or Distributed Exposures. *Developmental Psychology, 38*(6), 967–978. https://doi.org/10.1037/0012-1649.38.6.967

Choi, & Ionin, T. (2021). Plural marking in the second language: Atomicity, definiteness, and transfer. *Applied Psycholinguistics, 42*(3), 549–578. https://doi.org/10.1017/S0142716420000569

DeKeyser. (2005). What Makes Learning Second-Language Grammar Difficult? A Review of Issues. *Language Learning*, *55*(S1), 1–25. https://doi.org/10.1111/j.0023-8333.2005.00294.x

Ellis, & Sagarra, N. (2010). THE BOUNDS OF ADULT LANGUAGE ACQUISITION: Blocking and Learned Attention. *Studies in Second Language Acquisition*, *32*(4), 553–580. https://doi.org/10.1017/S0272263110000264

Ellis. (2022). Second language learning of morphology. *Journal of the European Second Language Association, 6*(1), 34–59. https://doi.org/10.22599/jesla.85

Endress, & Mehler, J. (2009). Primitive computations in speech processing. *Quarterly Journal of Experimental Psychology (2006), 62*(11), 2187–2209. https://doi.org/10.1080/17470210902783646

Finley. (2023). Morphological cues as an aid to word learning: a cross-situational word learning study. *Journal of Cognitive Psychology, 35*(1), 1–21. https://doi.org/10.1080/20445911.2022.2113087

Finn, & Hudson Kam, C. L. (2015). Why Segmentation Matters: Experience-Driven Segmentation Errors Impair”"Morphem”" Learning. *Journal of Experimental Psychology. Learning, Memory, and Cognition, 41*(5), 1560–1569. https://doi.org/10.1037/xlm0000114

Falk, Lindqvist, C., & Bardel, C. (2015). The role of L1 explicit metalinguistic knowledge in L3 oral production at the initial state. *Bilingualism (Cambridge, England)*, *18*(2), 227–235. https://doi.org/10.1017/S1366728913000552

Gardner, Branigan, H. P., & Chondrogianni, V. (2021). Spoken and written production of inflectional morphology among L1 Mandarin speakers of English. *Journal of Memory and Language, 120,* 104250. https://doi.org/10.1016/j.jml.2021.104250

Girard, & Koenig-Robert, R. (2011). Ultra-rapid categorization of fourier-spectrum equalized natural images: macaques and humans perform similarly. *PloS One*, *6*(2), e16453–e16453. https://doi.org/10.1371/journal.pone.0016453

Gleitman, L. R., Cassidy, K., Nappa, R., Papafragou, A., & Trueswell, J. C. (2005). Hard words. *Language Learning and Development, 1*(1), 23–64.

Hawkins, & Hattori, H. (2006). Interpretation of English multiple wh-questions by Japanese speakers: a missing uninterpretable feature account. *Second Language Research, 22*(3), 269–301. https://doi.org/10.1191/0267658306sr269oa

Hwang, & Lardiere, D. (2013). Plural-marking in L2 Korean: A feature-based approach. *Second Language Research, 29*(1), 57–86. https://doi.org/10.1177/0267658312461496

Jensen, Slabakova, R., Westergaard, M., & Lundquist, B. (2020). The Bottleneck Hypothesis in L2 acquisition: L1 Norwegian learners' knowledge of syntax and morphology in L2 English. *Second Language Research*, *36*(1), 3–29. https://doi.org/10.1177/0267658318825067

Jiang. (2004). Morphological insensitivity in second language processing. *Applied Psycholinguistics, 25*(4), 603–634. https://doi.org/10.1017/S0142716404001298

Jiang. (2007). Selective Integration of Linguistic Knowledge in Adult Second Language Learning. *Language Learning, 57*(1), 1–33. https://doi.org/10.1111/j.1467-9922.2007.00397.x

Laakso, & Calvo, P. (2011). How Many Mechanisms Are Needed to Analyze Speech? A Connectionist Simulation of Structural Rule Learning in Artificial Language Acquisition. *Cognitive Science, 35*(7), 1243–1281. https://doi.org/10.1111/j.1551-6709.2011.01191.x

Lecouvet, Degand, L., & Suner, F. (2021). Unclogging the Bottleneck: The role of case morphology in L2 acquisition at the syntax-discourse interface. *Language Acquisition*, *28*(3), 241–271. https://doi.org/10.1080/10489223.2020.1860056

Marian, Blumenfeld, H. K., & Kaushanskaya, M. (2007). The Language Experience and Proficiency Questionnaire (LEAP-Q): Assessing Language Profiles in Bilinguals and Multilinguals. *Journal of Speech, Language, and*

Hearing Research, 50(4), 940–967. https://doi.org/10.1044/1092-4388(2007/067)

McManus. (2015). L1-L2 differences in the acquisition of form-meaning pairings in a second language. *Canadian Modern Language Review, 71*(2), 51–77. https://doi.org/10.3138/cmlr.2070.51

Mirković, J., & Gaskell, M. G. (2016). Does sleep improve your grammar? Preferential consolidation of arbitrary components of new linguistic knowledge. *PloS one*, *11*(4), e0152489.

Monaghan, & Mattock, K. (2012). Integrating constraints for learning word–referent mappings. *Cognition, 123*(1), 133–143. https://doi.org/10.1016/j.cognition.2011.12.010

Monaghan, Mattock, K., Davies, R. A. I., & Smith, A. C. (2015). Gavagai is as gavagai does: learning nouns and verbs from cross-situational statistics. *Cognitive Science, 39*(5), 1099–1112. https://doi.org/10.1111/cogs.12186

Monaghan, P., Murray, H., & Holz, H. (2024). Redundancy and complementarity in language and the environment: How intermodal information is combined to constrain learning. *Language Learning, in press*. https://doi.org/10.1111/lang.12628

Monaghan, Schoetensack, C., & Rebuschat, P. (2019). A Single Paradigm for Implicit and Statistical Learning. *Topics in Cognitive Science, 11*(3), 536–554. https://doi.org/10.1111/tops.12439

Munoz. (2008). Age-related differences in foreign language learning. Revisiting the empirical evidence. *International Review of Applied Linguistics in Language Teaching, IRAL, 46*(3), 197–220. https://doi.org/10.1515/IRAL.2008.009

Pelucchi, Hay, J. F., & Saffran, J. R. (2009). Learning in reverse: Eight-month-old infants track backward transitional probabilities. *Cognition, 113*(2), 244–247. https://doi.org/10.1016/j.cognition.2009.07.011

Rebuschat, & Williams, J. N. (2012). How implicit is statistical learning? *In Statistical Learning and Language Acquisition* (Vol. 1, pp. 365–382). De Gruyter, Inc. https://doi.org/10.1515/9781934078242.365

Rothman, González Alonso, Jorge, & Puig-Mayenco, Eloi. (2019). *Third Language Acquisition and Linguistic Transfer.* Cambridge University Press.

Peña, Bonatti, L. L., Nespor, M., & Mehler, J. (2002). Signal-Driven Computations in Speech Processing. *Science (American Association for the*

Advancement of Science), *298*(5593), 604–607. https://doi.org/10.1126/science.1072901

Perruchet, & Desaulty, S. (2008). A role for backward transitional probabilities in word segmentation? *Memory & Cognition, 36*(7), 1299–1305. https://doi.org/10.3758/MC.36.7.1299

Rebuschat, Monaghan, P., & Schoetensack, C. (2021). Learning vocabulary and grammar from cross-situational statistics. *Cognition, 206*, 104475–104475. https://doi.org/10.1016/j.cognition.2020.104475

Saffran. (2020). Statistical Language Learning in Infancy. *Child Development Perspectives*, *14*(1), 49–54. https://doi.org/10.1111/cdep.12355

Saffran, Aslin, R. N., & Newport, E. L. (1996). Statistical Learning by 8-Month-Old Infants. *Science (American Association for the Advancement of Science)*, *274*(5294), 1926–1928. https://doi.org/10.1126/science.274.5294.1926

Saffran, J. R., & Kirkham, N. Z. (2018). Infant statistical learning. *Annual review of psychology*, *69*, 181.

Saffran, Newport, E. L., & Aslin, R. N. (1996). Word Segmentation: The Role of Distributional Cues. *Journal of Memory and Language*, *35*(4), 606–621. https://doi.org/10.1006/jmla.1996.0032

Saffran, & Wilson, D. P. (2003). From Syllables to Syntax: Multilevel Statistical Learning by 12-Month-Old Infants. *Infancy,* 4(2), 273–284. https://doi.org/10.1207/S15327078IN0402_07

Scott, & Fisher, C. (2012). 2.5-Year-olds use cross-situational consistency to learn verbs under referential uncertainty. *Cognition*, *122*(2), 163–180. https://doi.org/10.1016/j.cognition.2011.10.010

Slabakova, R. (2019, April). The bottleneck hypothesis updated. In Three streams of generative language acquisition research: Selected papers from the 7th meeting of generative approaches to language acquisition–North America, University of Illinois at Urbana-Champaign (pp. 319-345). Amsterdam: John Benjamins.

Smith, & Yu, C. (2008). Infants rapidly learn word-referent mappings via cross-situational statistics. *Cognition, 106*(3), 1558–1568. https://doi.org/10.1016/j.cognition.2007.06.010

Suanda, & Namy, L. L. (2012). Detailed Behavioral Analysis as a Window Into Cross-Situational Word Learning. *Cognitive Science*, *36*(3), 545–559. https://doi.org/10.1111/j.1551-6709.2011.01218.x

Suethanapornkul. (2020). *Statistical Learning of Predictive Dependencies in the Tense-aspect System of a Miniature Language by English and Thai First Language Adults*. ProQuest Dissertations Publishing.

Tomasello, & Akhtar, N. (1995). Two-year-olds use pragmatic cues to differentiate reference to objects and actions. *Cognitive Development*, *10*(2), 201–224. https://doi.org/10.1016/0885-2014(95)90009-8

Tünnermann, J., & Scharlau, I. (2018). Stuck on a Plateau? A Model-Based Approach to Fundamental Issues in Visual Temporal-Order Judgments. *Vision (Basel)*, *2*(3), 29-. https://doi.org/10.3390/vision2030029

Van der Slik, van Hout, R., & Schepens, J. (2019). The role of morphological complexity in predicting the learnability of an additional language. *Second Language Research*, *35*(1), 47–70. https://doi.org/10.1177/0267658317691322

van Zoest, W., & Donk, M. (2005). The effects of salience on saccadic target selection. *Visual Cognition*, *12*(2), 353-375. https://doi.org/10.1080/13506280444000229

Vlach, & Johnson, S. P. (2013). Memory constraints on infants' cross-situational statistical learning. *Cognition*, *127*(3), 375–382. https://doi.org/10.1016/j.cognition.2013.02.015

Walker, Monaghan, P., Schoetensack, C., & Rebuschat, P. (2020). Distinctions in the Acquisition of Vocabulary and Grammar: An Individual Differences Approach. *Language Learning*, *70*(S2), 221–254. https://doi.org/10.1111/lang.12395

Yu, & Smith, L. B. (2007). Rapid Word Learning under Uncertainty via Cross-Situational Statistics. *Psychological Science*, *18*(5), 414–420. https://doi.org/10.1111/j.1467-9280.2007.01915.x

Appendix

Debriefing questionnaire

Q1: How did you decide which picture was the correct referent? Did you just guess throughout the experiment, or did you follow any particular strategies? If so, what strategies did you follow?

Q2：Did you notice any particular patterns or rules about the grammatical structure of this new language?

Q3： Did you notice the sounds “na”, “ke” or “pau”? If so, what do you think they mean? (The tense morphemes asked in Q3 variants between different versions of mappings.)

Q4： Did you notice the sounds “sai” or “ti”? If so, what do you think they mean? (The number morphemes asked in Q4 variants between different versions of mappings.)

Q5： Do you think the way you made decisions on the pictures changed throughout the experiment?

Q6：What do you think was the aim of this study? The whole experiment takes about 65 minutes to finish.

Artificial language-four random sound-meaning mappings

Noun	Version 1	Version 2	Version 3	Version 4
panda	/faʊlu/	/pulə/	/kitə/	/kitə/
pig	/fima/	/faʊlu/	/patʃu/	/lipə/
lion	/fuki/	/fuki/	/suli/	/fima/
mouse	/jitu/	/fima/	/pulə/	/tusi/
sheep	/kitə/	/siʃə/	/faʊlu/	/patʃu/
rabbit	/lipə/	/ʃaji/	/fuki/	/faʊlu/
dog	/lutʃi/	/tusi/	/jitu/	/jitu/
cow	/ʃaji/	/masə/	/ʃaji/	/paʃə/
Number cue	Version 1	Version 2	Version 3	Version 4
one	/saɪ/	/ti/	/kə/	/saɪ/
two	/ti/	/saɪ/	/paʊ/	/nɑ/
Verb	Version 1	Version 2	Version 3	Version 4
cook	/naɪpə/	/jitu/	/paʃə/	/ʃaji/
work	/patʃu/	/naɪpə/	/lutʃi/	/fuki/
swim	/paʃə/	/lipə/	/masə/	/masə/
run	/siʃə/	/patʃu/	/fima/	/suli/
sleep	/pulə/	/paʃə/	/tusi/	/siʃə/
walk	/suli/	/lutʃi/	/lipə/	/lutʃi/
sing	/masə/	/suli/	/naɪpə/	/naɪpə/
paint	/tusi/	/kitə/	/siʃə/	/pulə/
Tense cue	Version 1	Version 2	Version 3	Version 4
past	/nɑ/	/kə/	/saɪ/	/ti/
now	/kə/	/paʊ/	/nɑ/	/paʊ/
future	/paʊ/	/nɑ/	/ti/	/kə/

Visual stimuli examples

One-sample t-tests and Cohen's d for performance against chance level (0.5) for each test type at each test block.

	Block 1	Block 2	Block 3	Block 4	Block 5	Block 6	Block 7	Block 8
L1 Chinese	(4.90***, 0.85)	(6.27***, 1.11)	(8.81***, 1.56)	(9.64***, 1.70)	(12.81***, 2.26)	(11.37***, 2.01)	(10.00***, 1.77)	(12.64***, 2.23)
L1 English	(1.65, 0.28)	(2.81**, 0.48)	(4.61***, 0.79)	(5.39***, 0.93)	(5.82***, 0.10)	(7.36***, 1.26)	(7.50***, 1.29)	(7.27***, 1.25)
L1	(6.05*	(7.05*	(10.93	(15.50	(13.90	(21.55	(20.02	(18.07

Ger man	**, 1.04)	**, 1.21)	***, 1.87)	***, 2.66)	***, 2.38)	***, 3.70)	***, 3.43)	***, 3.10)

* P < .05

** P < .0125

*** P < .001

Note. L1 Chinese: t(33, d); L1 English: t(34, d); L1 German: t(36, d)

One-sample t-tests and Cohen's d for performance against chance level (0.5) for vocabulary tests at test trials in the mix-blocks.

	Block 4	Block 6	Block 8
L1 Chinese	(7.27***, 1.28)	(7.67***, 1.36)	(7.56***, 1.34)
L1 English	(3.02**, 0.52)	(6.83***, 1.17)	(6.17***, 1.06)
L1 German	(8.90***, 1.53)	(13.69***, 2.35)	(13.69***, 2.35)

* P < .05

** P < .0125

*** P < .001

Note. L1 Chinese: t(33, d); L1 English: t(34, d); L1 German: t(36, d).

One-sample t-tests and Cohen's d for performance against chance level (0.5) for morphology tests at test trials in the mix-blocks.

	Block 4	Block 6	Block 8
L1 Chinese	(7.27***, 1.29)	(7.61***, 1.35)	(10.43***, 1.84)
L1 English	(3.58**, 0.61)	(3.67***, 0.63)	(4.00***, 0.69)
L1 German	(9.82***, 1.68)	(12.44***, 2.13)	(15.33***, 2.63)

* P < .05

** P < .0125

*** P < .001

Note. L1 Chinese: t(33, d); L1 English: t(34, d); L1 German: t(36, d).