Language Learning and Development

ISSN: 1547-5441 (Print) 1547-3341 (Online) Journal homepage: www.tandfonline.com/journals/hlld20

Enhancing second language learning through orthographical changes

Inbal Kimchi, Noga Reshef, Anna Vasileva & Inbal Arnon

To cite this article: Inbal Kimchi, Noga Reshef, Anna Vasileva & Inbal Arnon (28 Apr 2026): Enhancing second language learning through orthographical changes, Language Learning and Development, DOI: 10.1080/15475441.2026.2662295

To link to this article: https://doi.org/10.1080/15475441.2026.2662295

LANGUAGE LEARNING AND DEVELOPMENT
https://doi.org/10.1080/15475441.2026.2662295

Routledge
Taylor & Francis Group

Enhancing second language learning through orthographical changes

Inbal Kimchi[a], Noga Reshef[a], Anna Vasileva[b], and Inbal Arnon[b]

[a]Department of Cognitive Science, The Hebrew University of Jerusalem, Jerusalem, Israel; [b]Department of Psychology, The Hebrew University of Jerusalem, Jerusalem, Israel

ABSTRACT

Adults face many challenges in acquiring a second language, some of which have been attributed to learning from more segmented linguistic units compared to children. Children draw on both single words and multiword units (crossing lexical boundaries) in language learning, with multiword units used to learn about grammatical relations between words. Adults are less likely to learn from multiword units, because of their existing knowledge of words. Indeed, when adults are encouraged to rely on multiword units, by manipulating the auditory input they hear, they show better learning of grammatical relations. Here, we investigate a novel way to increase adults' reliance on multiword units: we show that orthographic changes can encourage reliance on larger units and facilitate learning of grammatical relations in both adults and children learning Greek. In Study 1, we find that adults learn article-noun gender agreement better when the article-noun sequence is written as a single unsegmented word compared to two separate words. In Study 2, we replicated these findings, again showing that learning article-noun agreement from multiword units enhances performance for article-noun agreement, but does improve generalization of the learned grammatical classes to noun-adjective agreement. In Study 3, we find that children (ages 7–11) also show better learning of article-noun agreement when exposed to unsegmented text. These findings provide further support for the benefit of learning grammatical relations from multiword units in both children and adults and point to new ways of enhancing second language instruction.

Introduction

Acquiring a second language is a challenging task. While adults outperform children in a wide range of cognitive tasks, they rarely achieve the same proficiency as children learning a first language. Adults struggle with certain aspects of a new language more than others. They demonstrate good learning abilities during the early stages of learning (Krashen et al., 1979) and can reach high proficiency in learning vocabulary and certain aspects of grammar (like word order, DeKeyser, 2005). However, they have difficulty in learning morphological and syntactic regularities that come easily to children (Clahsen & Felser, 2006; DeKeyser, 2005; Hartshorne et al., 2018; Johnson & Newport, 1989). The discrepancy in learning outcomes between children and adults raises a puzzle: What advantages do children acquiring their first language possess in comparison to adults learning a second language? Addressing this question would shed light on the process of first language acquisition and generate novel insights on how to help adults learn a second language, a task of great importance in an increasingly multi-lingual world (Gorter, 2013; Stein-Smith, 2021).

CONTACT Inbal Kimchi Inbal.Kimchi@gmail.com Department of Cognitive Science, The Hebrew University of Jerusalem, Mount Scopus Campus, Jerusalem 9190501, Israel,

Supplemental data for this article can be accessed online at https://doi.org/10.1080/15475441.2026.2662295.

A large body of research investigates the various factors that impact second language learning. These include learners' predictions based on prior experience (Pickering & Garrod, 2013); transfer from their first language (Jarvis & Pavlenko, 2008; McManus, 2021); their motivation to learn (Ai et al., 2021); and the way adults' cognitive abilities differ in important ways from those of children (Elman, 1993; Felser & Clahsen, 2009; Newport, 1990; Ramscar & Gitcho, 2007), among others. Here, we focus on one of these many factors, which involves the different linguistic building blocks that children and adults learn from. The Starting Big Approach (Arnon, 2010, 2021) suggests that some of adults' difficulty is related to their lesser reliance on multiword units during learning. Multiword units are defined as sequences that contain more than a single lexical word, whose representation is influenced by factors such as frequency and semantics (Arnon, 2021; Arnon & Christiansen, 2017). Under this approach, infants learning their first language extract linguistic units of varying sizes from their input, including both individual words and multiword units. These multiword units can aid learning certain relations between words. Adults learning a second language are less likely to use multiword units as building blocks for multiple reasons: (1) they already know what words are, (2) they know the specific words their own language uses, (3) they are often literature, and (4) they often learn in a classroom setting and from written input. As a result, when approaching a new language, adults may prioritize identifying and encoding individual lexical items, potentially relying less on larger distributional units at the initial stages of learning. The predictions of this theory are that native speakers use both word and multiword building blocks; that children are more likely to rely on multiword units in the process of learning compared to adults; and that learning from such units can help learn certain grammatical relations between words (Arnon, 2010, 2015, 2021).

There is growing evidence supporting these predictions, and showing that both children and adults are sensitive to the distributional properties not only of single words, but also of multiword units, and utilize this information in production, comprehension, and language learning. For example, adults recognize frequent phrases faster than less frequent ones and have better memory of them, even when controlling for the frequencies of the individual words and substrings (Arnon & Snider, 2010; Tremblay et al., 2011). Furthermore, adults process phrases acquired earlier during childhood faster than later acquired ones, mirroring the Age-of-Acquisition effects for individual words (Arnon et al., 2017). Moreover, adults produce words with a shorter phonetic duration when they are part of a more frequent multiword unit, even when controlling for part frequency and predictability (Arnon & Cohen Priva, 2013, 2014; Bybee & Scheibman, 1999).

Similar effects have been observed during development: Infants exhibit quicker recognition of nouns when embedded within familiar phrases compared to when presented in isolation (Fernald & Hurtado, 2006). Pre-verbal infants are sensitive to multiword frequency from early on, looking longer at phrases they hear frequently (e.g., *clap your hands*), compared to phrases they hear rarely (e.g. *take your hands*, Skarabela et al., 2021). This indicates that they extract larger patterns alongside individual words. Toddlers are also sensitive to multiword frequency, showing better repetition and recognition of higher frequency phrases than lower frequency ones (e.g., Bannard & Matthews, 2008; Jones et al., 2020). Children's production errors are also related to the frequency of multiword units. For instance, children are better at producing irregular plurals when they are part of a more frequent sequence (e.g., teeth in "brush your teeth," Arnon & Clark, 2011) and tend to make more errors when the incorrect strings are frequently encountered in other constructions (e.g., more errors like *me do it* when children often hear correct preverbal uses like *let me do it*; Kirjavainen et al., 2009; McCauley et al., 2021). Furthermore, children's early productions include "frozen" or "formulaic" multiword expressions (e.g., Lieven et al., 1992, 2009; Peters, 1983) and are better accommodated using a computational model that extracts both words and multiword units (e.g., McCauley & Christiansen, 2014, 2019). Interestingly, there is also evidence for the use of larger building blocks in children creating language: homesigners make use of holistic gestures which are only then composed into productive parts, alongside the use of productive parts to make larger units, suggesting that whole-to-part processes are a learning bias children bring to language (Goldin-Meadow & Arnon, 2025).

Additional work provides evidence that learning from multiword units can facilitate the mastery of arbitrary grammatical relationships, such as grammatical gender. Grammatical gender is a linguistic phenomenon found in many languages, where nouns are assigned to specific classes and neighboring words are usually marked for agreement (Corbett, 1991). While first language learners master such systems early and with relative ease, leveraging that knowledge for more efficient processing (Grüter et al., 2012; Hopp, 2012, 2016; Hopp & Lemmerth, 2016; Lew-Williams & Fernald, 2007a, 2007b, 2010), second language learners struggle with them, even after long exposure (Scherag et al., 2004). Adult second language learners' ability to use articles proficiently and to rely on them to predict upcoming nouns is inconsistent and dependent on a range of variables, like the presence of grammatical gender in their L1 (Bordag & Pechmann, 2007; Sabourin et al., 2006), their L2 proficiency (Dussias et al., 2013), whether the noun is assigned to the same class across languages (Hopp & Lemmerth, 2016) and more. Under the Starting Big approach, these patterns are related to the more segmented linguistic building blocks adults learn from: while children learn gender agreement from larger units that are then segmented into words, adults learn the words separately and then try to combine them. The prediction is that learning from larger units, as children do, can enhance the acquisition of arbitrary relations between words, like gender agreement (Arnon, 2021).

Support for this prediction comes from a series of artificial language learning studies looking at the acquisition of gender agreement by adults. Adults showed better learning of article-noun agreement in an artificial language when they were exposed first to full sentences and only then to noun labels on their own (sentence-first condition) compared to hearing noun labels first and only then full sentences (label-first condition; Arnon & Ramscar, 2012). Even though their initial input was more complex, in that it required learning segmentation and meaning at the same time, participants in the sentence-first condition showed better learning outcomes. The advantage of learning from multiword units was further investigated using a similar artificial language, but conducting a more subtle manipulation of the input learners were exposed to (Siegelman & Arnon, 2015). Here, participants either heard unsegmented sentences (with no pauses between words) followed by segmented sentences or the other way around. Participants who heard unsegmented-sentences first showed better learning of article-noun agreement and were more likely to treat the article-noun initially as one unit. Furthermore, individual learners who treated the article and noun as one unit showed better learning outcomes. Overall, these findings support the link between larger early building blocks and better learning of grammatical relations. Importantly, learning from multiword units is not expected to benefit all linguistic domains: specifically, such learning is expected to facilitate learning semantically opaque relations that hold between words (such as article-noun agreement), but may not help when learning individual words or semantically dependent relations (like classifiers marking animacy; Siegelman & Arnon, 2015).

Another relevant difference between children learning a first language and adults learning a second is literacy: Adults learning a second language often know how to read and rely on written input, where word boundaries are salient in many orthographies. Learning to read may reduce learners' reliance on multiword units, with negative implications for learning. Indeed, adults with higher literacy levels in their first language showed better segmentation abilities in a second language (Havron & Arnon, 2016). Furthermore, as participants' literacy in their second language improved, their segmentation also improved (Havron & Arnon, 2016). Similarly, literate children outperformed preliterate children in segmentation tasks (Havron & Arnon, 2017). Literacy can also have consequences for learning: Preliterate children learned article-noun pairings better than they learned individual words, suggesting a more chunked representation of the article-noun sequence, while literate children and adults exhibited the opposite pattern (Havron & Arnon, 2020; Havron et al., 2018). Taken together, these findings underscore the influence of literacy on language learning and segmentation, and support the notion that as learners become literate, their reliance on multiword units may decrease, leading to different learning patterns and outcomes.

The current study

The benefit adults show in learning grammatical gender agreement from unsegmented input together with the impact of literacy on segmentation, raises an intriguing question: Could adult learning of grammatical relations between words be improved by simply changing the way they are written? In other words, will exposure to article-noun sequences written as one word facilitate learning article-noun agreement? In this paper, we address this by investigating the impact of subtle orthographical changes on learning and generalization of grammatical gender agreement in adults and children learning Greek. While artificial languages are valuable for experimental control, they often lack the full complexity and variability found in natural languages (Hay et al., 2011). By using Greek, a natural language with grammatical gender and gender-based article-noun agreement, we aim to test the impact of multiword units in a more realistic learning environment.

In Study 1, we look at adults' learning of article-noun agreement in Greek to see if learning is improved when learners are exposed to unsegmented written input – where the article and noun are written as one word – compared to segmented written input – where they are written as two words (as in actual Greek). An additional auditory-only condition is included as a control, allowing us to benchmark learning against prior work that used auditory input alone. Our primary prediction is that exposure to unsegmented written input will lead to better learning of grammatical relations than exposure to segmented written input. We further predict that both written-input conditions will yield better learning than the auditory-only control condition. In Study 2, we ask if exposure to unsegmented written input can also support generalization of the grammatical relations to another linguistic domain. We do this by exposing learners to article-noun agreement but testing them also on noun-adjective gender agreement. The underlying question is whether the advantages gained from exposure to less segmented input extend beyond the immediate domain being learned (in this case article-noun agreement) to facilitate generalization more broadly. In Study 3, we extend our findings to child learners, and ask whether this subtle orthographic manipulation will also facilitate learning in literate children (ages 7–11). Our prediction is that children will also exhibit improved grammatical learning outcomes in the unsegmented written input condition. If borne out, the studies will provide additional support for the benefit of learning from larger units and show the value of utilizing unsegmented written input as an effective instructional strategy, not only for adult language learners but also for literate children.

Study 1: will learning from unsegmented written input facilitate mastery of article-noun agreement in a second language?

In this study, we investigate the impact of manipulating written input on learning grammatical gender in Greek. We do this by teaching Greek to adult Hebrew speakers in one short session. We focus on how well they learn the agreement between gender-marked articles and nouns. We compare the effectiveness of three kinds of input: 1) audio input accompanied by segmented text where the article and noun are separated by spaces (as in standard Greek orthography), 2) audio input accompanied by unsegmented text where the article and noun are written as one word, and 3) audio-only input without written stimuli, as has been used in prior work on learning an artificial gender system (e.g., Siegelman & Arnon, 2015), and is included here as a control baseline. By examining these different conditions, we ask whether adding a written input improves performance, and more importantly, whether exposure to less segmented written input can facilitate learning grammatical relations.

Method

Participants

A total of 88 adult participants participated in Study 1. All participants were undergraduate students, native Hebrew speakers without learning or language disabilities and with no prior

knowledge in Greek. They were randomly assigned to the three conditions. We removed six participants whose native language was not Hebrew (to maintain control over the gender of the nouns we labeled in Greek when translated into Hebrew), six who reported they had ADHD or other learning disabilities, and one who had previous knowledge of Greek. After these exclusions, we had 75 participants in total (mean age: 24 years, sd: 2.39, age range: 20–33, 55 females, 18 males and 2 others), 25 in each condition. The sample size was determined based on previous studies (Havron et al., 2018; Havron & Arnon, 2020; Siegelman & Arnon). Prior to participating, participants signed a consent form agreeing to take part in the experiment. Participants received either 10 NIS or course credit as compensation for their participation.

Materials

Stimuli

Mirroring previous studies, participants were exposed to 12 Greek nouns, belonging to two grammatical classes: six feminine nouns, and six neuter nouns. All nouns had two syllables (see full list in Supplementary Materials S1). Each noun appeared with a gender-marked article ("*mia*" for feminine nouns and "*ena*" for neuter nouns). Participants saw pictures of objects and heard sentences in Greek describing them. All sentences followed a fixed structure, beginning with *vlepo* (I see), followed by the correct article (which varied based on the noun's gender) and the noun (see Example 1). The sentences were generated with Google Translate and edited with WavePad Sound Editor to ensure consistency in the duration of the auditory stimuli. A native Greek speaker went over all sentences to ensure accuracy. (All materials from all three studies are openly available at our OSF page: https://osf.io/qnmsv)

Example 1 : *Greek sentence*

Vlepo	*mia*	*mýti*
I see	a	nose
carrier phrase	article	noun

To eliminate the potential influence of participants' prior knowledge of Hebrew, a language that also has a grammatical gender system, each of the noun classes included an equal number of feminine and masculine nouns in Hebrew (so that knowledge of Hebrew would not help in learning the noun classes). Half of the Greek neuter nouns were feminine in Hebrew, while the other half were masculine, and the same held for the Greek feminine nouns. The objects associated with these nouns were all labeled using frequent Hebrew words. Furthermore, special attention was given to ensure that none of the Greek words resembled Hebrew ones, as well as ensuring that there is no similarity in roots with known European or English words. All pictures of the objects were matched for size and presented on a white background without other distracting elements.

Experimental conditions

We used three conditions to investigate the influence of orthography on learning grammatical gender agreement: (1) an audio-and-segmented-text condition where participants also saw the sentence written, with the article and noun written as two separate words (e.g., *mia mýti*); (2) an audio-and-unsegmented-text condition where the article and noun were written as a single word (e.g., *miamýti*); and (3) an audio-only condition where participants saw pictures and heard a sentence describing them in Greek.

Procedure

The experiment consisted of two stages: an exposure phase and a testing phase. Prior to the start of the experiment, participants were informed that they would be learning a new language and were asked to provide demographic data. To make sure participants did not have prior knowledge of Greek, and to verify the functionality of the speakers or earphones, participants were requested to translate recordings of sentences in various languages, including Greek, Hebrew, and English. Participants were then informed that they would be learning Greek words, which they would be tested on later. The entire experiment lasted less than 15 minutes.

Exposure Phase

Participants in all three conditions were exposed to the same visual and auditory stimuli, with the only difference being the written input. In all conditions, participants were exposed to each of the 12 nouns four times, leading to 48 sentences in total.

Testing Phase

Learning was assessed using two-alternative forced-choice trials. On each trial, participants saw a picture and heard two sentences, with no written input, and were asked to select the correct sentence. Participants could replay each recording up to two times before selecting their answer (the answer buttons were not available until each recording was played at least once). Half of the trials assessed participants' knowledge of the article-noun pairings (article trials): on these trials, the incorrect sentence had the correct noun but an incorrect article (e.g., *vlepo mia mýti* vs. *vlepo ena mýti*). The other half tested knowledge of noun-object matching (noun trials), and the incorrect recording had the wrong noun label but the correct article (e.g., *vlepo mia mýti* vs. *vlepo mia zóni*). Each of the 12 objects appeared as a target once in an article trial and once in a noun trial, resulting in 24 trials in total. The order of trials was randomized for each participant.

Results + discussion

We excluded participants who exhibited low levels of learning. Following Siegelman and Arnon (2015), we removed individuals who scored less than two standard deviations below the mean (in their respective condition) on the noun trials (accuracy was generally high in this condition, suggesting very low performance indicates inattention to the task). This led to the exclusion of three participants, one participant in the audio-only condition and two participants in the audio-and-unsegmented-text condition. Participants in all three conditions showed learning of the noun labels and the article-noun agreement (were above chance, t-test compared to 50% chance, all p's < .001, see Table 1).

We used mixed-effect logistic regression models to test the effect of input condition on learning. In the first model, accuracy was modeled as a function of input condition (audio-only, audio and segmented-text, and audio and unsegmented-text, treatment-coded with the audio-only condition as the reference level), trial type (noun, article, effect-coded, article set as −1), and trial number (scaled and centered), as well as the interaction between input condition and trial type. The model also had random effects for participant and item. The models included random intercepts for participants and

Table 1. Mean accuracy in all conditions and trial types. The table shows the mean proportion of correct answers for participants in the relevant conditions.

	Article	Noun	Total
Audio-only	.65 (sd: 0.15)	.89 (sd: 0.10)	.77
Unsegmented-text	.88 (sd: 0.09)	.97 (sd; 0.06)	.93
Segmented-text	.74 (sd: 0.17)	.97 (sd: 0.06)	.86
Total	.76	.94	

for stimuli (the maximal random effect structure justified by the design that converged; Barr et al., 2013). Models were fitted using the lme4 package in R (Bates, 2010), and the lmerTest package was used to obtain significance test for fixed effects coefficients (Kuznetsova et al., 2015).

Relative to the audio-only condition, accuracy was significantly higher in both written-input conditions. Performance was higher in the audio and segmented-text condition ($\beta = 1.04$, SE = 0.26, $p < .001$) and highest in the audio and unsegmented-text condition ($\beta = 1.44$, SE = 0.26, $p < .001$). As in previous studies, the effect of trial type was also significant, with nouns learned better than article-noun agreement (94% vs. 76%, $\beta = 0.78$, SE = 0.11, $p < .001$). Trial number was not significant ($\beta = -0.03$, SE = 0.07, $p = .72$) (see Table 2; Figure 1). The fact that nouns were learned better than articles indicates that participants were not simply selecting the more familiar sentence: in noun trials, both sentences had been heard before (both were grammatical) while in article trials, one sentence was unfamiliar, yet they performed better overall on the noun trials, suggesting that correct selection indicates learning. The interaction

Table 2. Effects of input condition, trial type, trial number, and the interaction between input condition and trial type on accuracy.

Predictor	Estimate	Std. error	*p*-value
Input-Condition (segmented-text)	1.04	0.26	<.001
Input-Condition (unsegmented-text)	1.44	0.26	<.001
Trial type (noun)	0.78	0.11	<.001
Trial number	−0.03	0.07	.72
Condition (segmented) ×Trial type (noun)	0.53	0.22	.02
Condition (unsegmented) ×Trial type (noun)	−0.08	0.22	.70

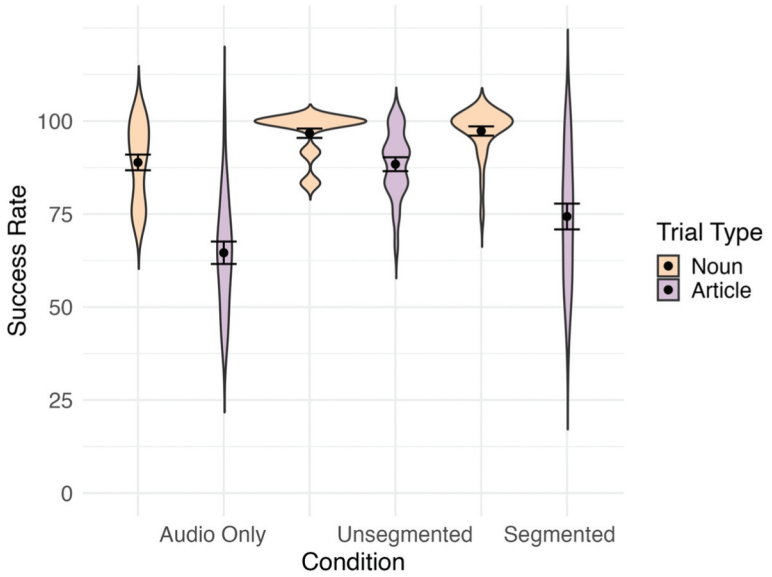

Figure 1. Accuracy by trial type across conditions.

between input condition and trial type differed across written-input conditions. The advantage of segmented text over audio-only was larger for noun trials than for article trials ($\beta = 0.53$, SE = 0.22, $p = .016$), whereas the advantage of unsegmented text did not differ by trial type ($\beta = -0.09$, SE = 0.22, $p = .70$), indicating that unsegmented input facilitated learning more across the two trial types (see Table 2; Figure 1).

Table 3. Effects of input condition, trial type, trial number, and the interaction between input condition and trial type on accuracy.

Predictor	Estimate	Std. error	*p*
Condition (unsegmented)	0.19	0.16	.23
Trial type (noun)	1.00	0.14	<.001
Trial number	0.02	0.10	.85
Condition ×Trial type	−0.31	0.14	.02

To test our main prediction about the facilitative effect of unsegmented text vs. segmented text in learning article-noun agreement, we ran another model, looking only at the two written input conditions. This model had condition (unsegmented-text vs. segmented-text, effect-coded, segmented set as −1), trial type (noun and article, effect-coded, article set as −1), the interaction between them, and trial number (scaled and centered) as fixed effects and random intercept for participant (the maximal random effect structure justified by the design that converged; Barr et al., 2013). As predicted, the effect of condition was not significant (β = .19, SE = 0.16, p = .23), but the interaction between condition and trial type was (β = −.31, SE = 0.14, p = .02): Learning was better in the unsegmented-text condition for articles, but not for nouns, as expected if the relation between them is learned better from multiword units (see Table 3). As in the first model, the effect of trial type was significant with better learning of nouns compared to article-noun agreement in both conditions (β = 1.00, SE = 0.14, p < .001) Trial number was not significant (β = .02, SE = 0.10, p = .85).

To further explore the interaction between input condition and trial type, we ran two additional models, looking at noun learning and article learning separately. Each model had a fixed effect of input condition (unsegmented-text vs. segmented-text, effect-coded, segmented set as −1) and trial number (scaled and centered), and random intercepts for participants and item. Looking only at the article trials, the effect of input condition was significant, with better learning of articles in the unsegmented-text condition (β = .50, SE = 0.14, p < .001). No such difference was found when looking at noun trials (β = −.16, SE = 0.43, p = .70). These findings are in line with previous ones, and with the theoretical prediction: Learning from multiword units facilitates learning of the relation between the article and the noun, but not the noun itself.

In sum, the findings indicate better learning outcomes when participants had both audio and visual input compared to audio alone. More importantly, the subtle change of writing the article and noun as one word led to better learning of the relation between them, as expected if it enhanced treating them initially as one unit.

Study 2: does learning from unsegmented text facilitate generalization of the noun classes?

This study had two goals. The first was to replicate the findings from Study 1, using a larger set of stimuli and another participant sample. The second was to ask if exposure to unsegmented written input will facilitate generalization of the grammatical relations to another domain. That is, we ask whether such learning led to better mastery of the grammatical classes themselves (which noun belongs to which gender class), which will lead to better performance on gender agreement with other elements in the sentence. Specifically, we exposed participants to article-noun agreement during the learning phase (as in the previous study) but tested them also on noun-adjective gender agreement (which they had not been exposed to). We want to see if learning from larger linguistic units fosters better generalization of the noun classes, thereby facilitating agreement patterns with additional elements. To do this, we compared learning from auditory + written input in two conditions: unsegmented-text, where the article

and noun were written as one word, and segmented-text, where they were written as two separate words.

Methods

Participants

A total of 81 participants participated in this study. We excluded one participant whose native language was not Hebrew, resulting in a sample of 80 participants, 40 in each condition (mean age: 24 years, sd: 2.05, age range: 19–31, 59 females and 21 males). All participants were undergraduate students, native Hebrew speakers without learning or language disabilities and with no prior knowledge in Greek. Prior to participating, participants were provided with a consent form and agreed to take part in the experiment. They received either 20 NIS or course credit as compensation for their participation.

Materials

Learning article-noun agreement

As in the previous study, participants learned Greek nouns, belonging to two noun classes (feminine and neuter), and appearing with two different articles (*mia* for feminine nouns, *ena* for neuter nouns). Participants were exposed to 16 nouns (not 12 as in the previous study), evenly split between feminine and neuter genders. As in the previous study, all noun labels had two syllables, were frequent in Hebrew, and the gender of the nouns in Hebrew was balanced (see Supplementary Materials S2 for the full item list). The sentence structure followed the same fixed word order as in Study 1, where all sentences started with the carrier word *vlepo* ("I see") followed by a grammatical article (which varied based on the noun's gender) and finally, a noun. Recordings were obtained from Google Translate and edited with the WavePad Sound Editor application to ensure consistency in the duration of the auditory stimuli. Pictures of objects were matched for size to maintain uniformity. All the pictures were presented on a white background and without other distracting elements. In this Study, unlike in Study 1, the pictures of the nouns in the first exposure stage were entirely white, as shown in Figure 2(A). This was done in preparation for the second exposure learning stage, where participants would learn about colors. As in Study 1, participants in both input conditions were exposed to each of the 16 nouns four times, presented with the same object and audio stimuli, with the only variation being the written stimuli. Then, they were tested on the article-noun pairings (16 article trials, each object appeared once), and the noun-object matching (16 noun trials, each object appeared once).

A: White object used in the first exposure phase

B: Red object used in the second exposure phase

C: Green object used in the second exposure phase

Figure 2. Pictures used in Study 2.

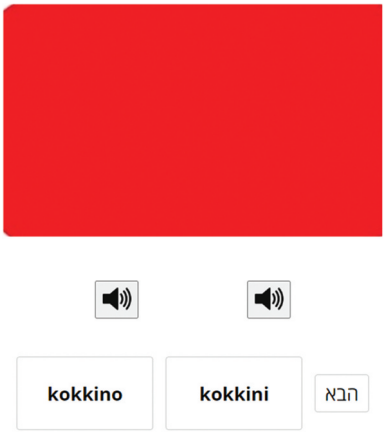

Figure 3. Exposure to the red color labels.

Looking at generalization to noun-adjective agreement

In the second part of the study, we exposed participants to noun-adjective gender agreement, using color adjectives. Greek adjectives vary by gender in a systematic way: the ending "*-o*" is used for neuter gender and the ending "*-i*" for feminine gender (e.g., Green: neuter: *prasino*, feminine: *prasini)*. In this part, participants were explicitly informed that the Greek language offers two distinct ways to express each color, saw the two adjective forms written and listened to them pronounced in Greek (see Figure 3).

After learning the color terms, participants saw a picture of a colored object (green or red) and heard a sentence describing it. The sentence included a carrier phrase, a gender-marked article, a gender-appropriate color adjective and a noun (e.g., *vlepo mia kokkini mýti*, see Example 2). We used the colors green (neuter: *prasino*, feminine: *prasini*), and red (neuter: *kokkino*, feminine: *kokkini*). We modified the white images using the Lunapic color editor application (see Figure 2(B), 2(C)).

After learning the color terms, each participant saw a colored version of four randomly selected objects (two neuter and two feminine) out of the 16 previously learned objects (so that each participant was exposed to different objects) and heard their description in Greek. Each participant saw each object four times, twice in red and twice in green. Unlike in the previous part, participants were only exposed to the auditory stimuli in both conditions. We did not add written input since the adjective appears between the article and the noun in Greek, which would have made the differentiation between the article-noun salient. Each participant heard a total of 16 sentences (four objects heard four times).

Example 2: *Greek Sentence for Second Exposure*

Vlepo	*mia*	*kokkini*	*mýti*
I see	a	red	nose
Carrier phrase	article	color word	noun

We then tested their generalization of the correct gender-marking on the adjectives by testing them on the remaining 12 objects: If they learned the generalization of the noun classes (and the appropriate way to mark that on adjectives) they should be able to select the correct adjective form even for nouns they did not hear with color adjectives before. Participants completed 32 two-alternative forced-choice trials: They saw a picture of a colored object, heard two sentences, and had to select the correct one. Participants could replay each recording up to two times before making their selection (and could not respond before each recording was played at least once). Twenty-four of the trials were adjective-generalization trials: They assessed participants' ability to select the correct gender-marked adjective

for the 12 nouns they had not heard with adjectives. On these trials, the correct sentence had the right gender-marked adjective (e.g., *kokkini mýti*) and the incorrect one had the wrong gender-marked adjective (e.g., *kokkino mýti*). Each of the 12 objects appeared twice, once in each color. Participants also completed eight color-trials which tested their learning of the color terms themselves. On these trials, the incorrect sentence had the wrong color term (e.g., *kokkini mýti vs. prasini mýti*). Eight objects appeared in the color trials, four red objects and four green objects, selected randomly for each participant.

Procedure

Prior to the start of the experiment, participants were informed that they would be learning a new language and were asked to provide demographic data. As in Study 1, participants were requested to translate recordings of sentences in various languages, including Greek (to assess language familiarity), Spanish, and English (as distractors), to assess language familiarity and to verify the functionality of the speakers or earphones. After this, participants were informed that they would be learning Greek, which they would later be tested on. The entire experiment lasted less than 30 minutes. Participants were randomly assigned to one of two input conditions: unsegmented-text or segmented text. The experiment consisted of four parts: the first two were identical to study 1 and involved exposure to the language and testing of article-noun pairings, whereas the third and fourth involved exposure to colored objects and testing of adjective-generalization.

Results + discussion

As in the previous study, we excluded participants who scored less than two standard deviations below the mean noun accuracy within their respective condition. This led to the exclusion of five participants, two in the unsegmented-text condition, and three in the segmented-text condition.

Learning article-noun agreement

Participants showed learning in all conditions and trial types (participants were above 50% chance, see Table 4, t-test comparing to chance, p's < .001). Accuracy in this study was lower when compared to the same conditions in Study 1, which may be driven by the increased number of stimuli (12 nouns in Study 1 vs. 16 here).

We used mixed-effect logistic regression models to test the effect of input condition on learning. The first model had fixed effects of input condition (unsegmented-text, segmented-text, effect-coded, segmented set as −1), trial type (noun, article, effect-coded, article ser as −1), and trial number (scaled and centered), as well as the interaction between condition and trial type. The model also had random effects for participant and item (the maximal random effect structure justified by the design that converged; Barr et al., 2013). As in Study 1, and as predicted, the effect of input condition was not significant (β = −.006, SE = .07, p = .94), but the interaction between input condition and trial type was, with better learning of the article-noun agreement in the unsegmented-text condition (β = −.27, SE = 0.07, p < .001). There was also an effect of trial type, with higher accuracy on noun trials compared to the article trials (β = 1.01, SE = .07, p < .001) (see Table 5; Figure 4). The effect of trial number was also significant (β = −0.12, SE = 0.05, p = .03), with lower accuracy on later trials.

Table 4. Mean accuracy in all conditions and trial types. The table presents the mean proportion of correct answers for participants in the relevant conditions.

	Article	Noun	Total
Unsegmented text	.71 (sd: 0.12)	.91 (sd: 0.09)	.81
Segmented text	.59 (sd: 0.13)	.95 (sd: 0.08)	.77
Total	.65	.93	

Table 5. Effects of input condition, trial type, trial number, and the interaction between condition and trial type on accuracy.

Predictor	Estimate	Std. error	p
Input condition (unsegmented)	−0.006	0.07	.94
Trial type (noun)	1.01	0.07	<.001
Trial number	−0.12	0.05	.03
Input condition ×Trial type	−0.27	0.07	<.001

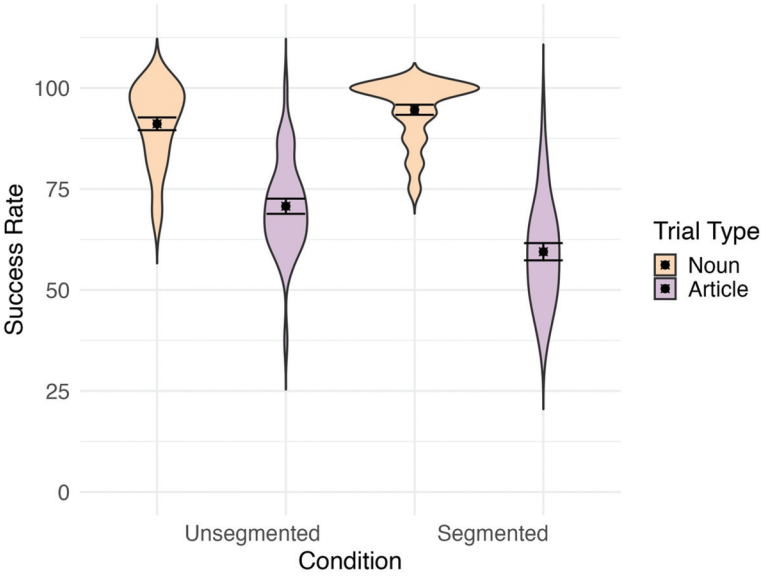

Figure 4. Accuracy by trial type across conditions.

We ran two additional mixed-effect logistic regression models to deepen our understanding of the interaction, looking separately at article and noun trials. The models had input condition (unsegmented-text and segmented-text, effect-coded, segmented set as −1) and trial number (scaled and centered) as fixed effects, with random intercepts for participants and items. Looking at the article trials, the effect of input condition was significant, with better learning in the unsegmented-text condition ($\beta = .26$, SE = 0.07, $p < .001$). No such difference was found for the noun trials, in fact, accuracy was somewhat higher in the segmented-text condition ($\beta = -.30$, SE = 0.17, $p = .07$). In both models, trial number was not significant ($\beta = -.11$, SE = 0.06, $p = .07$; $\beta = -.11$, SE = 0.11, $p = .34$). That is, as in the prior study, and as predicted, learning of article-noun agreement, but not of the nouns themselves, was better when presented as a single word rather than two separate words.

Noun-adjective agreement

Participants showed good learning of the color terms in both conditions (were over 90% accurate at selecting the correct color term in both conditions) but did not show learning of the adjective-noun agreement (were not above chance, $p > .10$, see Table 6).

Again, we ran a mixed-effect logistic regression model to analyze the results, with input condition (unsegmented-text and segmented-text, effect-coded, segmented set as −1), trial type (color and adjective, adjective set as −1), the interaction between them, and trial number (scaled and centered) as fixed effects and random intercepts for participant and item (the maximal random effect structure justified by the design that converged; Barr et al., 2013; see Table 7). The effect of condition was not significant ($\beta = .14$, SE = .09, $p = .10$), and neither was the interaction of condition and trial type (β

Table 6. Mean accuracy in all conditions and trial types. The table presents the mean proportion of correct answers for participants in the relevant conditions.

	Adjective	Color	Total
Unsegmented-text	.54	.94	.74
Segmented-text	.53	.91	.72
Total	.54	.93	

= .12, SE = .08, p = .14). The effect of trial type was significant, with much higher accuracy on color trials than the adjective trials in both conditions (β = 1.24, SE = .08, p < .001) (see Table 7; Figure 5). The effect of trial number was not significant (β = .03, SE = .05, p = .49).

Our findings indicate that learning article-noun agreement from multiword units enhances performance for article-noun agreement but does not extend to an overall improvement in the generalization of gender-based agreement patterns: Accuracy on noun-adjective agreement was not higher in the unsegmented-text condition. One possibility is that the advantage of multiword units does not facilitate noun class generalization. However, this conclusion may be premature given the generally poor performance on adjective-noun agreement. Participants in both conditions were not above chance in selecting the correct adjective form, making it hard to test the effect of the unsegmented vs. segmented written input. The short duration of the learning phrase, along with the only auditory input may have made the task too difficult. It could also be that generalization is possible only if the noun classes were learned during the first phase, that is, only when the article-noun agreement was learned correctly.

Table 7. Effects of input condition, trial type, trial number, and the interaction between input condition and trial type on accuracy.

Predictor	Estimate	Std. error	p
Input Condition (unsegmented)	0.14	0.09	.10
Trial type (color)	1.24	0.08	<.001
Trial number	0.03	0.05	.49
Condition ×Trial type	0.12	0.08	.14

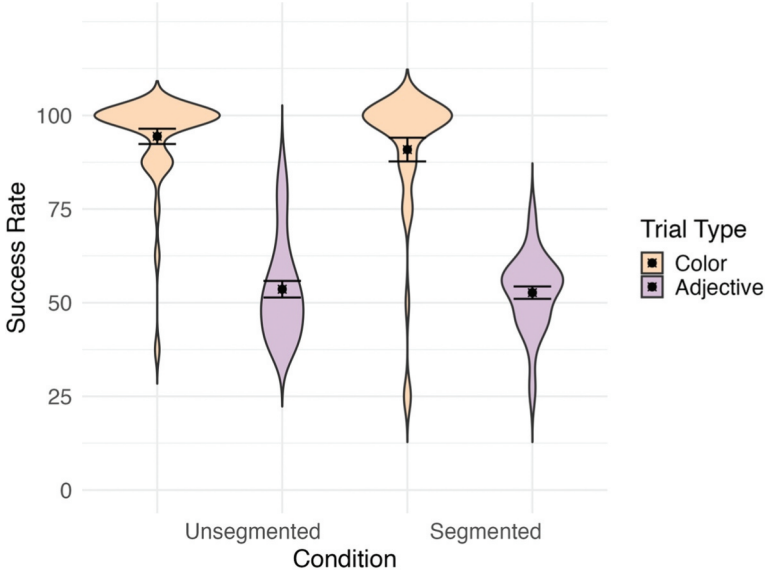

Figure 5. Success rates by trial type across input conditions. Bars represent mean ± standard error.

Table 8. Effects of input condition, noun learnability, trial number, and the interaction between input condition and noun learnability on accuracy in adjective-generalization Trials.

Predictor	Estimate	Std. error	*p*
Input Condition (unsegmented)	−0.01	0.06	.79
Noun Learnability (learned)	0.21	0.05	<.001
Trial number	0.06	0.05	.23
Condition ×Noun Learnability	0.07	0.05	.14

To explore this possibility, we excluded from the adjective-noun analysis, nouns for which participants selected the wrong label in the noun tests or the wrong article in the article tests in the first phase of the experiment. That is, for each participant, we identified the nouns that they learned the noun label and the article-noun agreement for, and only used those in the analysis of the adjective-generalization trials. This left us with 1034 trials for analysis. Looking only at those trials, we saw an interesting pattern. We could now see evidence of learning: mean accuracy on the adjective-generalization trials was 55% ($t(461) = 2.24$, $p = .025$, compared to chance) in the segmented-text condition and 58% ($t(571) = 3.98$, $p < .001$, compared to chance) in the unsegmented-text condition.

We also ran a mixed-effect logistic regression model to analyze the results, with input condition (unsegmented-text and segmented-text, effect-coded, segmented set as −1), noun learnability (trials in which participants correctly identified the noun label and the article-noun agreement were considered trials were the nouns were learned, effect-coded, not learned set as −1), the interaction between them and trial number (scaled and centered) as fixed effects and random intercepts for participant and item (see Table 8). The effect of condition was not significant ($\beta = -.01$, SE = .06, $p = .79$), and neither was the interaction of condition and noun learnability ($\beta = .07$, SE = .05, $p = .14$). However, the effect of noun learning was significant in both input conditions, with higher accuracy on adjective trials when the noun label and noun-article agreement were learned successfully ($\beta = .21$, SE = .05, $p < .001$). The effect of trial number was not significant ($\beta = .06$, SE = .05, $p = .23$). These results suggest that once participants mastered the article-noun agreements, they could generalize this knowledge, to a certain degree, irrespective of the initial learning input. However, this generalization was modest in magnitude and should be interpreted with caution, given that accuracy remained close to chance levels.

Study 3: will learning from unsegmented written input facilitate mastery of article-noun agreement also for children?

In this study, we want to expand our previous findings to children learning a second language, specifically, children at the early stages of literacy acquisition (seven- to eleven-year-olds). These ages were chosen because children at this age in Israel are typically literate and have already started learning a second language in school (usually English). By targeting this age group, we can explore the potential benefits of manipulating written input during the early stages of second language acquisition. We use the same design as in Study 1: A brief 15-minute session of learning Greek, with a focus on article-noun agreement. We hypothesize that, similar to adults, children will show better learning outcomes when exposed to unsegmented text. This study has an additional goal.

Our previous studies did not assess participants' attentiveness to the orthographic manipulation: We do not know if participants noticed the written input during the learning sessions. To address this, the current study incorporates typing trials during exposure as a way of gauging participants' engagement with the presented written stimuli. Every few trials, participants were asked to type in the sentence they just saw/heard. This will tell us whether they noticed the written input: Participants are expected to type the article-noun as one word in the unsegmented-text condition and as two in the segmented-text one. We exposed children to the exact same stimuli used in Study 1, with the following changes. First, each sentence was presented three times, instead of four times as in the previous two studies, to make the study shorter and more child friendly. Second, the written stimuli were presented

in Hebrew letters (and not English), since children at this age are not expected to be fluent readers in English. Finally, we added a reading proficiency test at the end of the study to make sure all children could read (and would thereby be sensitive to the manipulation of the written input).

Method

Participants

Sixty-two children participated in Study 3. We excluded one participant with ADHD, and one who was older than our upper age limit (12 years), leaving us with a sample of 60 participants (mean age: 9.54 years, sd: 1.03, age range: 7.5–11.58, 31 females and 29 males), 32 in the unsegmented condition (mean age: 9.57) and 28 in the segmented condition (mean age: 9.62). The sample size was determined based on previous studies (Havron et al., 2018; Havron & Arnon, 2020; Siegelman & Arnon). All participants were children between the ages of 7 to 11, native Hebrew speakers without learning or language disabilities and with no prior knowledge in Greek. Prior to participating, we obtained signed consent for participation from each child's parent. We also got verbal assent from each child after explaining to about the experiment. Participants received a small token ("reward") as compensation for participation. The study was run at the Israeli Living Lab, a developmental lab based in a children's science museum.

Materials

Stimuli

The stimuli were identical to those of study 1: Children were exposed to the same 12 Greek nouns, 6 neuter and 6 feminine. The written stimuli were presented in Hebrew letters (and not English) and included the diacritic markers of vowels (nikud, see Example 4). Children heard the sentences in one of two input conditions: the unsegmented-text condition, where the audio was accompanied by the article and noun written as one word, and the segmented-text condition where they were written as two words (these are the same two conditions used in the previous two studies).

Example 4: *Greek sentences in Hebrew*

Segmented condition	וְלַפּוֹ מ.יָא מ.יט.י
Unsegmented condition	וְלַפּוֹ מ.יָאמ.יט.י

Procedure

The experiment followed the same structure as Study 1, with the addition of typing trials during the learning phase, and a reading proficiency test at the end. Children were seated in front of a computer with an experimenter next to them. Prior to the start of the experiment, children were informed that they would be learning a new language. To assess language familiarity, they were asked to translate recordings of sentences in various languages, including Greek, Hebrew, and English. This step also served to verify the functionality of the speakers or earphones. Afterward, participants were informed they would be learning Greek, and would then be asked about what they learned. The experiment lasted less than 15 minutes.

The experiment consisted of two parts: an exposure phase where participants saw objects and heard sentences describing them in Greek, and a test phase, where they were tested on their knowledge of the nouns and article-noun agreement. Children heard 36 sentences during the exposure phase (12 nouns heard three times). Every six trials, participants were asked to type the last sentence they heard. If they had difficulty using the keyboard, they wrote the sentence on paper, which was then transcribed by a research assistant. These typing trials were added to assess participants sensitivity to the orthographic manipulation: we wanted to see if their written output matched the written input they saw.

Following exposure, children completed a forced-choice task, identical to that of Study 1 + 2: They saw pictures, heard two sentences describing them, and had to select the correct one. Children completed 24 test trials, half tested noun knowledge and half article-noun agreement. Following the forced-choice task, children completed a reading proficiency assessment. We used a single-word reading test to ensure that the literacy level of all participants was adequate. The test was developed by Shani, Lehman, Shalem, Bahat, and Zeiger in 2006 (as cited in Reiter et al., 2011). It consists of 38 Hebrew words (Supplementary Materials S3) spread across four pages, progressively increasing in difficulty. In our study, each participant was presented with 20 words that were randomly selected from the original test, so that each participant encountered five words from each difficulty level. We recorded participants' responses using the computer's sound system and then had a research assistant code them for accuracy and fluency.

Results + discussion

Greek learning results

We did not have to exclude any participants because of low accuracy (under two standard deviations from the mean noun accuracy in their assigned condition). However, as the orthographic manipulation and the typing trials require sufficient reading abilities, we removed three participants who scored less than two standard deviations below the mean in the reading test, one in the unsegmented-text condition and two in the segmented-text condition. In addition, we excluded five more participants whose reading scores were unavailable due to sound or recording issues or a request not to be recorded (four in the unsegmented condition and one in the segmented condition), leaving us with 52 participants. Participants showed learning in both conditions and trial types (see Table 9, t-test comparing to chance, p's < .001).

We used a mixed-effect logistic regression model to analyze the results, with input condition (unsegmented-text and segmented-text, effect-coded, segmented set as −1), trial type (noun and article, effect-coded, article set as −1), the interaction between them, age in years (scaled and centered) and trial number (scaled and centered) as fixed effects and random intercepts for participant and item (the maximal random effect model that converged; Barr et al., 2013). As found for adults, the effect of condition was not significant ($\beta = -.02$, SE = .07, $p = .82$), but the interaction between condition and trial type was ($\beta = -.24$, SE = 0.06, $p < .001$), indicating that accuracy was higher in the unsegmented-text condition for article trials, but not for noun trials (in fact, accuracy on noun trials was higher in the segmented-condition than in the unsegmented one). The effect of trial type was also significant, with overall higher accuracy in the noun trials compared to the article trials ($\beta = .34$, SE = .06, $p < .001$). The effect of age ($\beta = .12$, SE = .07, $p = .09$) as well as the effect of trial number ($\beta = -.09$, SE = .06, $p = .14$) were not significant (see Table 10; Figure 6).

We ran two additional mixed-effects logistic regression models to deepen our understanding of the interaction, looking separately at article and noun trials. Both models had input condition (unsegmented-text and segmented-text, effect-coded, segmented set as −1), age (scaled and centered), and trial number (scaled and centered) as fixed effects with random intercepts for participants and items. Looking at article trials, the effect of input condition was significant, with higher accuracy in the unsegmented-text condition ($\beta = .21$, SE = 0.09, $p = .02$). The effect of input condition was also significant in noun trials, but in the opposite direction, with better accuracy in the segmented-text

Table 9. Mean accuracy in all conditions and trial types. The table shows the mean proportion of correct answers for participants in the relevant conditions.

	Article	Noun	Total
Unsegmented text	.68 (sd: 0.14)	.72 (sd: 0.16)	.70
Segmented text	.58 (sd: 0.16)	.81 (sd: 0.16)	.70
Total	.63	.77	

Table 10. Effects of input condition, trial type, trial number, and the interaction between condition and trial type on accuracy.

Predictor	Estimate	Std. error	*p*
Input Condition (unsegmented)	−0.02	0.07	.82
Trial type (noun)	0.34	0.06	<.001
Age	0.12	0.07	0.09
Trial number	−0.09	0.06	.14
Condition ×Trial type	−0.24	0.06	<.001

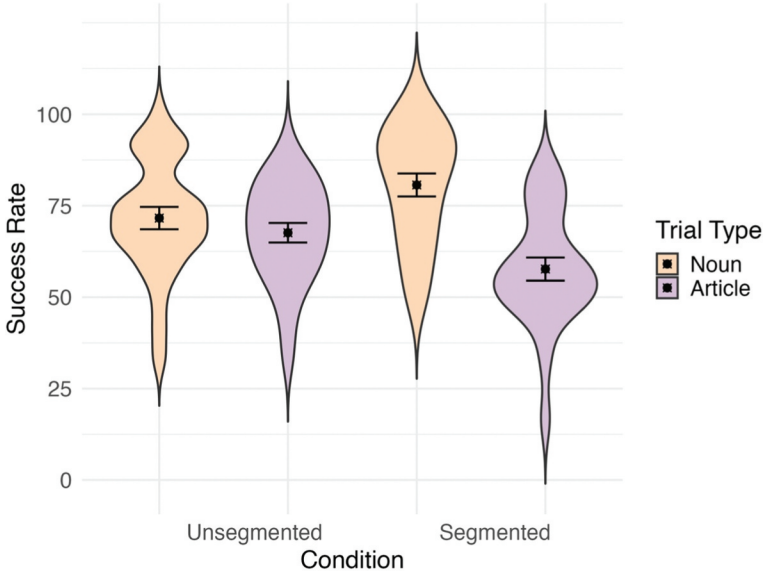

Figure 6. Success rates by trial type across input conditions. Bars represent mean ± standard error.

condition compared to the unsegmented one (β = −.28, SE = 0.13, p = .03). Trial number was not significant in both models (β = −.15, SE = 0.09, p = .08; β = −.03, SE = 0.09, p = .75). As predicted, and as found for adults, exposure to unsegmented-text improved children's learning of article-noun agreement, but not of the nouns themselves. In fact, for children, nouns were learned significantly better in the segmented condition.

Typing trials

We wanted to make sure that participants attended to the written input by asking if their written output matched their input condition. We excluded blank responses, single-word responses, and instances where the article-noun phrase had less than three syllables. This led to the exclusion of 85 responses (25% of all responses). The remaining responses were classified as matching if participants wrote the sentence with a blank space between the article and noun in the segmented condition, and if they wrote the sentence with the article and noun as a single word in the unsegmented condition. The results revealed that 94% of responses corresponded to the input condition, indicating participants were well aware of the written stimuli.

Table 11. Mean accuracy in all input conditions and trial types across studies.

	Study 1		Study 2		Study 3	
	Article	Noun	Article	Noun	Article	Noun
Unsegmented text	.88	.97	.71	.91	.68	.72
Segmented text	.74	.97	.59	.95	.58	.81

Reading test

Reading accuracy was evaluated according to the test guidelines, assigning 1 if the produced form was fully correct and 0 it if was not (this test has a stringent coding criterion where even one mistake renders the response fully incorrect). Three participants were excluded from analysis because their scores were over two standard deviations lower than the mean (these are the ones we also excluded from the analysis of the Greek learning part). Additionally, scores were not recorded for three participants due to sound or recording issues and two participants who requested not to be recorded. The reading scores of the remaining participants ($N = 52$) ranged from 70 to 100 (mean 89.89). There was no difference in reading accuracy between the two input conditions (unsegmented: 91%, segmented: 89% $t(45.89) = -0.84$, $p = .40$), indicating this could not have confounded the results.

In sum, we found that children, like adults, show better learning of article-noun agreement, but not of the nouns themselves, when presented as a single word rather than two separate words (see Table 11).

General discussion

We set out to investigate whether altering the written representation of words will lead to a greater reliance on multiword units, and through that, enhanced learning of the grammatical relations between them. Previous work has shown that starting with multiword units can enhance learning of article-noun agreement: Adults were encouraged to rely on larger units by manipulating the auditory stimuli to be less segmented (e.g., Siegelman & Arnon, 2015). Here, we ask whether subtle orthographic changes can lead to a similar result: we examined how changing the written representation impacts the learning and generalization of grammatical gender agreement in adults and children learning Greek. Our findings indicate that learning article-noun agreement is facilitated when learners see the article-noun sequence written as one word (unsegmented written input), compared to when they see them written as two words (as is the case in actual Greek). Over three studies, adults and literate children showed better learning of article-noun agreement when they saw them written as one word. Our use of typing trials during exposure (Study 3) illustrates that participants were indeed sensitive to the orthographic manipulation, and that this influenced their perception of the article-noun as one word or two. That is, changing the written input impacted the units that participants relied on, and their subsequent learning outcomes. At the same time, exposure to unsegmented written input in learning article-noun agreement did not facilitate generalization of the grammatical gender classes to a novel domain (adjective-noun agreement).

These findings show that changes in orthography, without parallel changes in the auditory stimuli, can encourage reliance on multiword units and lead to better learning outcomes, providing further support for their role in learning grammatical relations (Arnon, 2010, 2021). Related evidence comes from work on how boundary cues influence the learning of dependencies. Morgan et al. (1987) showed that cues marking phrase boundaries facilitate learning of phrase-internal dependencies by indicating to learners which elements belong together. In the present studies, the orthographic manipulation may have functioned in a similar way, providing a visual cue that grouped the article and noun together. In this sense, our findings illustrate another way in which grouping elements into larger units can

facilitate learning of the relations between them, in line with the predictions of the Starting Big approach.

The findings also underscore the idea that learning from multiword units can be beneficial for certain aspects of language, but not others (Siegelman & Arnon, 2015). While learning from unsegmented written input enhanced mastery of article-noun gender agreement, it did not aid in vocabulary learning. Adults learned the noun labels equally well in both input conditions, while children learned them better in the segmented input condition. This pattern suggests that the advantages of learning from multiword units do not apply across the board: While learning from multiword units facilitates the acquisition of semantically opaque dependencies, learning vocabulary, especially for children, may benefit from clearer segmentation into meaning carrying units. A similar pattern was found in Siegelman and Arnon (2015) when learning from unsegmented input facilitated the acquisition of grammatical relations only when the article did not carry semantic information. Findings from child language learning provide further support for the idea that multiword units help with some things, but not others. Early reliance on highly frequent multiword patterns can result in the introduction of errors, as in the case of me-for-I errors (e.g., "me do it"), which arise from frequent multiword sequences in the input such as "let me do it" (Kirjavainen et al., 2009). Similarly, frequent multiword sequences can intrude into inappropriate contexts, yielding errors of commission in children's spontaneous production (McCauley et al., 2021). Together, these findings show that learning from larger units does not uniformly facilitate all aspects of learning and can, in some cases, give rise to systematic errors during the learning process.

More broadly, the effectiveness of learning from multiword units may vary across different aspects of second language learning, depending on whether the relation to be learned is part of a larger unit, and whether the elements to be learned convey independent semantic information. Importantly, the frequency and relevance of such multiword units may differ across languages, depending in part on their morphological structure and on how grammatical relations are expressed (Arnon, 2021). As a result, the usefulness of learning from larger units may differ across linguistic systems and learning contexts, and may extend to other areas of language where adults typically encounter difficulties. Orthographic manipulations similar to those studied here could potentially assist in learning classifier systems and verb-preposition pairings, two domains where the relation to be learned holds between adjacent words, and is semantically opaque, and that adult second language learners struggle with (DeKeyser, 2005; Hansen & Chen, 2001; Kennedy & Miceli, 2001; Koosha & Jafarpour, 2006; Vyatkina, 2016).

Our findings also have implications for second language pedagogy: They point to the potential benefit of adopting a "starting big" approach during teaching, which involves beginning with larger unsegmented units. There is ongoing debate over the optimal approach to teaching a second language (Dos Santos, 2020; Hummel, 2021). There are numerous teaching methods and strategies, each with a different perspective on what should be emphasized. Some advocate focusing on auditory input and prioritizing the development of listening and speaking skills. For instance, the "audio-lingual teaching approach" (Abduh, 2016; Ausubel, 1964; Bidenko & Bespalova, 2017; Wen et al., 2018) and "task-based language teaching" (Gindo & Kawo, 2019; Murad & Smadi, 2009) are prominent methods that emphasize oral communication and learning through practical tasks. Others stress the importance of incorporating written input into the learning process, with findings showing improvement in second language skills when reading is involved in the process (Chio, 2009; Elley & Mangubhai, 1983). Our findings suggest that each method has its advantages. Study 1 showed that adding written input can facilitate learning by providing an additional input modality. However, starting with written forms may hinder the acquisition of certain grammatical relations. An optimal approach may involve combining both methods, by initially using unsegmented auditory input and only later introducing written input. Indeed, such approaches have started to be implemented, for instance, Bidenko and Bespalova (2017) propose a modified implementation of the audio-lingual teaching approach, starting with auditory exercises before introducing reading activities (Bidenko & Bespalova, 2017). Developing teaching methods that leverage the benefits of each technique may result in the most effective learning

outcomes. By starting with oral aspects of language or unsegmented written input and introducing the segmented written input only in later learning stages, learners may achieve better learning outcomes, particularly in learning grammatical relations. This approach acknowledges the complexity of language acquisition and embraces a flexible, adaptive approach to teaching, ultimately fostering greater proficiency and fluency in second language learners.

Our findings suggest that learning article-noun agreement from multiword units improves performance specifically for article-noun agreement but does not necessarily lead to a broader generalization of the gender classes and gender-based agreement patterns. In Study 2, participants in both the segmented and the unsegmented condition did not show evidence of generalizing to adjective-noun agreement (with performance at chance in both groups). One possibility is that the lack of generalization stemmed from participants not fully mastering the article-noun agreement, hindering their ability to generalize the gender classes effectively. Indeed, when looking only at successfully learned article-noun pairings, we found modest evidence of learning, suggesting that the acquisition of article-noun agreements may have contributed to improved performance in noun-adjective agreement tasks. This was found in both input conditions, suggesting that once participants achieved proficiency with article-noun agreement, they could begin to generalize this knowledge to a limited extent. However, it is important to interpret these results in caution, as accuracy remained relatively close to chance. Future studies could focus more directly on this subgroup of learned items to assess whether robust knowledge of article-noun agreement facilitates more reliable generalization. Additionally, future work should explore other types of generalization that may be supported by exposure to multiword units, such as generalization across different determiners (e.g., from definite to indefinite articles or demonstratives), or generalization to novel syntactic contexts.

The relatively short duration of the task (30 minutes of exposure) may have hindered the broader generalization: future experiments incorporating longer exposure periods are needed to gain a deeper understanding of the impact on unsegmented input on generalization. Another question arising from the study concerns the long-term retention of the learned material. Our investigation was limited to a single learning session lasting 30 minutes, followed by immediate testing. However, an important open question pertains to the duration of the effect and how long it will persist beyond this initial timeframe. Future research should examine multiple sessions across time and assess the retention of learned knowledge over extended periods, thereby providing insights into the sustainability of the observed effects. In addition, task complexity may play a key role in moderating the benefits of unsegmented input. It remains unclear whether these benefits will persist under more demanding learning conditions, with increased memory load or less consistency than the fixed frame used here. Future studies should explore whether the observed effects hold in more complex and naturalistic learning environments.

Overall, our findings provide novel support for the advantages of "starting big," suggesting that learning from multiword units can enhance learning of grammatical relations, also in adults and literate children. The findings also open new avenues for adult language teaching methodologies, indicating the potential of incorporating unsegmented input in learning grammatical relations.

Acknowledgments

We wish to the Living Lab in the Bloomfield Science museum, the museum staff, and the children and parents who participated in the studies.

Disclosure statement

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

Funding

This research was funded by Israel Science Foundation grant [420/25] awarded to I. Arnon.

ORCID

Inbal Kimchi http://orcid.org/0000-0002-2553-7117
Inbal Arnon http://orcid.org/0000-0001-8934-718X

Data and code availability

All materials, stimuli, data and analysis are openly available via the Open Science Framework (OSF) website, at: https://osf.io/qnmsv

Ethics approval statement

This study was conducted in accordance with the guidelines of the Department of Psychology at the Hebrew University of Jerusalem Ethics Committee and was approved under approval number 2023-28031. All participants provided informed consent prior to their participation in the study.

References

Abduh, N. B. (2016). The use of audio-lingual method in teaching listening comprehension at the second year students of SMK YAPIP Makassar Sungguminasa. *Exposure*, *5*(1), 43–52. https://doi.org/10.26618/ejpbi.v5i1.808

Ai, J., Pan, Y., & Zhong, W. (2021). The role of motivation in second language acquisition: A review. In *2021 3rd International Conference on Literature, Art and Human Development (ICLAHD 2021)*, Xishuangbanna, China, (pp. 623–627). Atlantis Press.

Arnon, I. (2010). *Starting big - the role of multi-word phrases in language learning and use* [Doctoral dissertation, Stanford University]. Stanford Digital Repository.

Arnon, I. (2015). What can frequency effects tell us about the building blocks and mechanisms of language learning? *Journal of Child Language*, *42*(2), 274–277. https://doi.org/10.1017/S0305000914000610

Arnon, I. (2021). The starting big approach to language learning. *Journal of Child Language*, *48*(5), 937–958. https://doi.org/10.1017/S0305000921000386

Arnon, I., & Christiansen, M. H. (2017). The role of multiword building blocks in explaining L1-L2 differences. *Topics in Cognitive Science*, *9*(3), 621–636. https://doi.org/10.1111/tops.12271

Arnon, I., & Clark, E. V. (2011). Why brush your teeth is better than teeth - children's word production is facilitated in familiar sentence-frames. *Language Learning and Development*, *7*(2), 107–129. https://doi.org/10.1080/15475441.2010.505489

Arnon, I., & Cohen Priva, U. (2013). More than words: The effect of multi-word frequency and constituency on phonetic duration. *Language and Speech*, *56*(3), 349–371. https://doi.org/10.1177/0023830913484891

Arnon, I., & Cohen Priva, U. (2014). Time and again: The changing effect of word and multiword frequency on phonetic duration for highly frequent sequences. *The Mental Lexicon*, *9*(3), 377–400. https://doi.org/10.1075/ml.9.3.01arn

Arnon, I., McCauley, S. M., & Christiansen, M. H. (2017). Digging up the building blocks of language: Age-of-acquisition effects for multiword phrases. *Journal of Memory and Language*, *92*, 265–280. https://doi.org/10.1016/j.jml.2016.07.004

Arnon, I., & Ramscar, M. (2012). Granularity and the acquisition of grammatical gender: How order-of-acquisition affects what gets learned. *Cognition*, *122*(3), 292–305. https://doi.org/10.1016/j.cognition.2011.10.009

Arnon, I., & Snider, N. (2010). More than words: Frequency effects for multi-word phrases. *Journal of Memory and Language*, *62*(1), 67–82. https://doi.org/10.1016/j.jml.2009.09.005

Ausubel, D. P. (1964). Adults versus children in second-language learning: Psychological considerations. *Modern Language Journal*, *48*(7), 420–424. https://doi.org/10.1111/j.1540-4781.1964.tb04523.x

Bannard, C., & Matthews, D. (2008). Stored word sequences in language learning of four-word combinations. *Psychological Science*, *19*(3), 241–248. https://doi.org/10.1111/j.1467-9280.2008.02075.x

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, *68*(3), 255–278. https://doi.org/10.1016/j.jml.2012.11.001

Bates, D. M. (2010). lme4: Mixed-effects modeling with R.

Bidenko, L., & Bespalova, H. (2017). Implementing audio-lingual method to teaching Ukrainian as a foreign language at the initial stage. *Advanced Education*, (7), 23–27. https://doi.org/10.20535/2410-8286.82711

Bordag, D., & Pechmann, T. (2007). Factors influencing L2 gender processing. *Bilingualism: Language and Cognition*, *10* (3), 299–314. https://doi.org/10.1017/S1366728907003082

Bybee, J., & Scheibman, J. (1999). The effect of usage on degrees of constituency: The reduction of don't in English. *Linguistics*, *37*(4). https://doi.org/10.1515/ling.37.4.575

Chio, K. U. (2009). Reading and second language acquisition. *HKBU Papers in Applied Language Studies*, *13*, 153–174.

Clahsen, H., & Felser, C. (2006). Grammatical processing in language learners. *Applied Psycholinguistics*, *27*(1), 3–42. https://doi.org/10.1017/S0142716406060024

Corbett, G. G. (1991). *Gender*. Cambridge University Press.

DeKeyser, R. M. (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

Dos Santos, L. M. (2020). The discussion of communicative language teaching approach in language classrooms. *Journal of Education and e-Learning Research*, *7*(2), 104–109. https://doi.org/10.20448/journal.509.2020.72.104.109

Dussias, P. E., Valdés Kroff, J. R., Guzzardo Tamargo, R. E., & Gerfen, C. (2013). When gender and looking go hand in hand: Grammatical gender processing in L2 Spanish. *Studies in Second Language Acquisition*, *35*(2), 353–387. https://doi.org/10.1017/S0272263112000915

Elley, W. B., & Mangubhai, F. (1983). The impact of reading on second language learning. *Reading Research Quarterly*, *19*(1), 53–67. https://doi.org/10.2307/747337

Elman, J. L. (1993). Learning and development in neural networks: The importance of starting small. *Cognition*, *48*(1), 71–99. https://doi.org/10.1016/0010-0277(93)90058-4

Felser, C., & Clahsen, H. (2009). Grammatical processing of spoken language in child and adult language learners. *Journal of Psycholinguistic Research*, *38*(3), 305–319. https://doi.org/10.1007/s10936-009-9104-8

Fernald, A., & Hurtado, N. (2006). Names in frames: Infants interpret words in sentence frames faster than words in isolation. *Developmental Science*, *9*(3), F33–F40.

Gindo, Z. R., & Kawo, K. N. (2019). The effective of task based instruction in improving learners' speaking skills: The case of Micha Preparatory School. *Journal of Education and Practice*, *10*(34), 9–20. https://doi.org/10.7176/JEP/10-34-02

Goldin-Meadow, S., & Arnon, I. (2025). Whole-to-part development in language creation. *Trends in Cognitive Sciences*, *29*(1), 12–14. https://doi.org/10.1016/j.tics.2024.09.015

Gorter, D. (2013). Linguistic landscapes in a multilingual world. *Annual Review of Applied Linguistics*, *33*, 190–212. https://doi.org/10.1017/S0267190513000020

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

Hansen, L., & Chen, Y. L. (2001). What counts in the acquisition and attrition of numeral classifiers. *JALT Journal*, *23*(1), 83–100. https://doi.org/10.37546/JALTJJ23.1-5%E2%80%8C

Hartshorne, J. K., Tenenbaum, J. B., & Pinker, S. (2018). A critical period for second language acquisition: Evidence from 2/3 million English speakers. *Cognition*, *177*, 263–277. https://doi.org/10.1016/j.cognition.2018.04.007

Havron, N., & Arnon, I. (2016). Reading between the words: The effect of literacy on second language lexical segmentation. *Applied Psycholinguistics*, *38*(1), 1–27. https://doi.org/10.1017/S0142716416000138

Havron, N., & Arnon, I. (2017). Minding the gaps: Literacy enhances lexical segmentation in children learning to read*. *Journal of Child Language*, *44*(6), 1–23. https://doi.org/10.1017/S0305000916000623

Havron, N., & Arnon, I. (2020). Starting big: The effect of unit size on language learning in children and adults. *Journal of Child Language*, *48*(2), 244–260. https://doi.org/10.1017/S0305000920000264

Havron, N., Raviv, L., & Arnon, I. (2018). Literate and pre-literate children show different learning patterns in an artificial language learning task. *Journal of Cultural Cognitive Science*, *2*(1–2), 21–33. https://doi.org/10.1007/s41809-018-0015-9

Hay, J. F., Pelucchi, B., Graf Estes, K., & Saffran, J. R. (2011). Linking sounds to meanings: Infant statistical learning in a natural language. *Cognitive Psychology*, *63*(2), 93–106. https://doi.org/10.1016/j.cogpsych.2011.06.002

Hopp, H. (2012). 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. (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., & Lemmerth, N. (2016). Lexical and syntactic congruency in L2 predictive gender processing. *Studies in Second Language Acquisition*, *40*(1), 1–29.

Hummel, K. M. (2021). Introducing second language acquisition: Perspectives and practices.

Jarvis, S., & Pavlenko, A. (2008). *Crosslinguistic influence in language and cognition*. Routledge.

Johnson, J. S., & Newport, E. L. (1989). Critical period effects in second language learning: The influence of maturational state on the acquisition of English as a second language. *Cognitive Psychology*, *21*(1), 60–99. https://doi.org/10.1016/0010-0285(89)90003-0

Jones, G., Cabiddu, F., & Avila-Varela, D. S. (2020). Two-year-old children's processing of two-word sequences occurring 19 or more times per million and their influence on subsequent word learning. *Journal of Experimental Child Psychology*, *199*, 104922. https://doi.org/10.1016/j.jecp.2020.104922

Kennedy, C., & Miceli, T. (2001). An evaluation of intermediate students' approaches to corpus investigation. *Language Learning & Technology*, *5*(3), 77–90. https://doi.org/10.64152/10125/44567

Kirjavainen, M., Theakston, A., & Lieven, E. (2009). Can input explain children's me-for-I errors? *Journal of Child Language*, *36*(5), 1091–1114. https://doi.org/10.1017/S0305000909009350

Koosha, M., & Jafarpour, A. A. (2006). Data-driven learning and teaching collocation of prepositions: The case of Iranian EFL adult learners. *Asian EFL Journal*, *8*(4), 192–209.

Krashen, S. D., Long, M. A., & Scarcella, R. (1979). Age, rate, and eventual attainment in second language acquisition. *Tesol Quarterly*, *13*(4), 573–582. https://doi.org/10.2307/3586451

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2015). Package 'lmertest'. *R Package Version*, *2* 734.

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

Lew-Williams, C., & Fernald, A. (2007b). How first and second language learners use predictive cues in online sentence interpretation in Spanish and English. In *Proceedings of the 31st annual Boston university conference on language development* (Vol. 2. pp. 382–393). Cascadilla Press.

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

Lieven, E., Salomo, D., & Tomasello, M. (2009). Two-year-old children's production of multiword utterances: A usage-based analysis. *Cognitive Linguistics*, *20*(3), 481–508. https://doi.org/10.1515/COGL.2009.022

Lieven, E. V., Pine, J. M., & Barnes, H. D. (1992). Individual differences in early vocabulary development: Redefining the referential-expressive distinction. *Journal of Child Language*, *19*(2), 287–310. https://doi.org/10.1017/S0305000900011429

McCauley, S. M., Bannard, C., Theakston, A., Davis, M., Cameron-Faulkner, T., & Ambridge, B. (2021). Multiword units lead to errors of commission in children's spontaneous production: "What corpus data can tell us?*". *Developmental Science*, *24*(6), e13125. https://doi.org/10.1111/desc.13125

McCauley, S. M., & Christiansen, M. H. (2014). Acquiring formulaic language: A computational model. *The Mental Lexicon*, *9*(3), 419–436. https://doi.org/10.1075/ml.9.3.03mcc

McCauley, S. M., & Christiansen, M. H. (2019). Language learning as language use: A cross-linguistic model of child language development. *Psychological Review*, *126*(1), 1. https://doi.org/10.1037/rev0000126

McManus, K. (2021). *Crosslinguistic influence and second language learning*. Routledge.

Morgan, J. L., Meier, R. P., & Newport, E. L. (1987). Structural packaging in the input to language learning: Contributions of prosodic and morphological marking of phrases to the acquisition of language. *Cognitive Psychology*, *19*(4), 498–550. https://doi.org/10.1016/0010-0285(87)90017-X

Murad, T. M., & Smadi, O. (2009). The effect of task-based language teaching on developing speaking skills among the Palestinian secondary EFL students in Israel and their attitudes towards English. *The Asian EFL Journal*.

Newport, E. L. (1990). Maturational constraints on language learning. *Cognitive Science*, *14*(1), 11–28. https://doi.org/10.1207/s15516709cog1401_2

Peters, A. M. (1983). *The units of language acquisition* (Vol. 1). CUP Archive.

Pickering, M. J., & Garrod, S. (2013). An integrated theory of language production and comprehension. *Behavioral and Brain Sciences*, *36*(4), 329–347.

Ramscar, M., & Gitcho, N. (2007). Developmental change and the nature of learning in childhood. *Trends in Cognitive Sciences*, *11*(7), 274–279. https://doi.org/10.1016/j.tics.2007.05.007

Reiter, S., Shengross, A., & Shany, M. (2011). *"Reading the world" - a second chance for learning to read: The effects of a literacy intervention program on reading skills and well-being*. University of Haifa.

Sabourin, L., Stowe, L. A., & de Haan, G. J. (2006). Transfer effects in learning a second language grammatical gender system. *Second Language Research*, *22*(1), 1–29. https://doi.org/10.1191/0267658306sr259oa

Scherag, A., Demuth, L., Rösler, F., Neville, H. J., & Röder, B. (2004). The effects of late acquisition of L2 and the consequences of immigration on L1 for semantic and morpho-syntactic language aspects. *Cognition*, *93*(3), B97–108.

Siegelman, N., & Arnon, I. (2015). The advantage of starting big: Learning from unsegmented input facilitates mastery of grammatical gender in an artificial language. *Journal of Memory and Language*, *85*, 60–75. https://doi.org/10.1016/j.jml.2015.07.003

Skarabela, B., Ota, M., O'Connor, R., & Arnon, I. (2021). 'clap your hands' or 'take your hands'? one-year-olds distinguish between frequent and infrequent multiword phrases. *Cognition*, *211*, 104612. https://doi.org/10.1016/j.cognition.2021.104612

Stein-Smith, K. (2021). Multilingualism for global solutions and a better world. *Journal of Language Teaching and Research*, *12*(5), 671–677. https://doi.org/10.17507/jltr.1205.05

Tremblay, A., Derwing, B., Libben, G., & Westbury, C. (2011). Processing advantages of lexical bundles: Evidence from self-paced reading and sentence recall tasks. *Language Learning*, *61*(2), 569–613. https://doi.org/10.1111/j.1467-9922.2010.00622.x

Vyatkina, N. (2016). Data-driven learning for beginners: The case of German verb-preposition collocations. *ReCALL*, *28*(2), 207–226. https://doi.org/10.1017/S0958344015000269

Wen, C. H., Ying, L. C., & Azlan, M. A. K. (2018). The influence of task-based language teaching and audio-lingual teaching approach on Mandarin language learning outcomes. *International Journal of Social Sciences*, *4*(3), 123–141. https://doi.org/10.20319/pijss.2018.43.123141