

# Individual Differences in the Tendency to Use Multiword Information in Natural and Artificial Languages

**Dan Steinhof (dan.steinhof@mail.huji.ac.il)**

Department of Cognitive and Brain Sciences, The Hebrew University of Jerusalem,  
Mount Scopus, Jerusalem 91905, Israel

**Noam Siegelman (noam.siegelman@mail.huji.ac.il)**

Department of Psychology, The Hebrew University of Jerusalem,  
Mount Scopus, Jerusalem 91905, Israel

**Inbal Arnon (inbal.arnon@mail.huji.ac.il)**

Department of Psychology, The Hebrew University of Jerusalem,  
Mount Scopus, Jerusalem 91905, Israel

## Abstract

Work in the last decades showed that learning from multiword units is often beneficial for language learning, impacting mastery of arbitrary linguistic relations and predicting efficient language processing. Much of this work has looked at differences between first (L1) and second language (L2) learning, documenting differences in how children and adults approach language learning, with only a few studies looking at individual differences in reliance on multiword units. Here, we ask whether adults differ in their tendency to draw on multiword units when learning a new language, and if so, whether such differences are related to learning outcomes and to language processing. We used an artificial language with grammatical gender to measure participants' tendency to treat article-noun sequences as one unit during language learning, and a multiword recall task to measure their tendency to benefit from multiword units in their native language (L1). Our findings show that individuals differ in their tendency to chunk information into multiword units when learning an artificial language, with most participants falling into one of two distinct groups, showing a steady pattern of preferences to treat article-noun sequences as either one or more than one unit throughout the task. This tendency was found to be numerically, albeit not significantly, related to how much individuals benefit from multiword information in their L1. These findings document a novel dimension of individual differences – the tendency of a learner to rely on multiword units, which may be related to different aspects of language learning and processing.

**Keywords:** Individual differences; Language learning; Multiword units; Chunking; Psycholinguistics.

## Introduction

While words are traditionally seen as the basic building blocks of language, there is growing evidence that multiword units are also important building blocks for language learning and processing (Arnon, 2021; Christiansen & Arnon, 2017). Such units are believed to be formed via chunking, where words that co-occur together often give rise to a multiword representation (Arnon & Christiansen, 2017). Indeed, infants (e.g., Skarabela, Ota, O'Connor, & Arnon, 2021), children (e.g., Arnon & Clark, 2011; Bannard & Matthews, 2008), and adults (e.g., Arnon & Snider, 2010) are sensitive to multiword information and make use of multiword units in learning and processing. Adults' difficulty in learning second languages

(L2) seems to be related, in part, to their lesser reliance on multiword units (Arnon, 2021; Arnon & Christiansen, 2017). Because adults have prior knowledge of what words are, based on their native language (L1), they are less likely to draw on multiword units during learning.

So far, this literature has focused on group-level effects, showing that both children (Bannard & Matthews, 2008) and adults (Arnon & Snider, 2010) are sensitive to multiword information in their native language, and documenting differences between children and adults in their reliance on multiword units in learning and using a second language (Arnon & Ramscar, 2012; Havron, Raviv & Arnon, 2018; Ellis, 2012). Indeed, when adults are encouraged to extract larger units in lab settings, similar to how children are believed to learn language, they show better learning of certain grammatical relations (e.g., Siegelman & Arnon, 2015). These findings illustrate differences in the use of multiword units between different groups of learners.

However, there are reasons to think that individuals may differ in their tendency to rely on larger units, even in adulthood. Adult learners differ from one another in their ability to learn a second language, and these differences are correlated with individual differences in various cognitive and behavioral variables (e.g., Dörnyei, 2005; Li, Hiver, & Papi, 2022). For example, individual differences in statistical learning correlate with sentence comprehension, literacy skills, and vocabulary development, to name a few (see Siegelman, Bogaerts, Christiansen, & Frost, 2017, for a review). However, no study to date has asked whether adults vary in their tendency to draw on multiword units in learning a new language, and if so, whether such differences are related to sensitivity to multiword information in their L1. This raises the possibility that there are individual differences in the reliance on multiword units, and that those impact language learning and processing.

The handful of studies that did look into individual differences in the reliance on multiword information did so only by examining language processing, with the general finding that better recall of multiword information is correlated with faster sentence processing in L1 and L2. McCauley, Isbilen, & Christiansen (2017) used a multiword

chunking task (MWC) to look at the relation between individual differences in reliance on multiword information and sentence processing. In the MWC task, participants are asked to recall sequences of 12 words, which are composed of four sets of trigrams. The trigrams were either frequently co-occurring word sequences (target sequences, for example: *have to eat good to know don't like them is really nice*) or random word sequences (control sequences, made up of words matched in frequency to the target sequences, for example: *years got don't to game have she mean to them far is*), used to assess participants' baseline working memory (WM). Unsurprisingly, in control sequences, participants recalled about 4-7 elements (Cowan, 2001; Miller, 1956). In contrast, participants had an easier time recalling the target sequences, where the words in the trigrams often co-occur together. In terms of individual differences, the *benefit* a participant had from recalling target compared to control sequences is seen as their degree of reliance on multiword information in language processing. In addition to the MWC task, McCauley et al. (2017) measured participants' sentence processing using a self-paced reading task, focusing on sentences with long-distance number agreement with locally distracting number-marked nouns. They found that participants who benefited more from multiword information in the MWC task also had faster reaction times for processing complex sentences. Isbilen, McCauley, & Christiansen (2022) conducted a similar study using the MWC task alongside an auditory statistical learning task. They found that participants' benefit from recalling multiword information in the MWC task was correlated with better learning of the artificial language, but only when that was measured using a similar recall task. Similarly, Pulido (2021) and Pulido et al. (2024) used the MWC task with L1 English speakers who learn Spanish as L2. In both studies, participants' benefit from multiword sequences in their L1 was predictive of sentence processing in their L2.

While supporting the notion of individual differences in the reliance on multiword units, previous studies did not look at the unit size individuals use during language *learning*. In the current study, we ask whether individuals differ in the size of the units they use when learning a novel language. We then ask whether reliance on multiword units when learning a new language is related to reliance on multiword units in L1 processing, as measured by the MWC task.

To measure unit size during language learning, we adapt the paradigm used by Siegelman & Arnon (2015). This study used an artificial language learning (ALL) paradigm to teach adult participants an artificial language with two grammatical "classes" of nouns, marked by preceding articles. Participants saw pictures and heard sentences describing them in the language. Every few trials, participants were asked to type the sentence they heard. Responses were coded for whether the article-noun sequence was written as one word or two. This was used as an on-line measure of the perceived linguistic unit. Participants were exposed to the language both with and without pauses between the words, differing in the order of exposure (i.e., first with pauses or

first without pauses). Siegelman & Arnon (2015) found that participants treated the article-noun pairings as one word more often when exposed first to unsegmented auditory input. Further, participants who treated the article-noun pairing as one unit learned it better, showing a link between larger initial units and better learning.

## The Current Study

In this paper, we offer a novel perspective on the question of how individuals approach language learning, and how this can affect language processing. We suggest that individuals differ in the unit size they tend to focus on when learning a new language, and when processing their L1. To test this prediction, we look at participants' performance on two tasks: (1) an ALL task, modeled on Siegelman & Arnon, (2015), to assess the units learners use when learning a novel language, and (2) the MWC task, to measure their sensitivity to multiword information in their L1, while controlling for WM. Note that the data reported here is the first session of a two sessions preregistered study, aimed also to estimate the reliability of the tasks reported here.

Our primary hypothesis was that participants would differ in their tendency to extract multiword units in the ALL task, pointing to individual differences in the size of the units segmented from a novel language. Additionally, we wished to explore if participants who tend to draw more on multiword units would better learn the article-noun pairings, as was found in previous studies in a between-subject design. Similar to the ALL task, we hypothesize that participants would show variability in their reliance on multiword information in their L1, as measured by the MWC task. Lastly, we hypothesize that the sensitivity to multiword information in individuals' L1 would be related to the reliance on multiword units in a new language.

## Participants

105 students from the Hebrew University of Jerusalem, all native Hebrew speakers, participated in the study (83 women, 22 men; age:  $M=24.8$ ,  $SD=3.5$ ) for either course credit or monetary compensation. Participants were recruited using the university's systems and participated online using their own computers. Participants were excluded according to predetermined guidelines: 1 participant was excluded before coding the results, for restarting the experiment in the middle; 6 participants due to not having enough codable trials in the ALL task; and 2 participants due to not learning the artificial language. After exclusion, 96 participants were used for the final analysis. See Procedure for the exclusion guidelines.

## Artificial Language Learning Task

The task was modeled after the ones used in Siegelman & Arnon (2015), and Abu-Zhaya & Arnon (2024). The major change from previous studies was that we only used the stimuli without pauses (previously called the unsegmented stimuli) throughout the whole study, and did not expose participants to segmented stimuli. This was done for two main reasons. First, in previous studies, participants were

exposed to both kinds of stimuli, resulting in almost all of them segmenting the language by the end of the study. Because we are interested in the individual tendency to either chunk the language into multiword units or segment it into individual words, we avoided giving participants the segmented input. Second, the unsegmented stimuli are more similar to natural language input, which does not have pauses as reliable segmentation cues (Cole & Jakimik, 1980; Erickson & Thiessen, 2015).

**Materials** The auditory stimuli were taken from Siegelman & Arnon (2015). They consisted of 12 novel nouns, 2 articles (*fo* and *si*), and a carrier phrase, with a constant sentence structure: CARRIER PHRASE – ARTICLE – NOUN (see example 1). The carrier phrase was the same in all sentences. The 12 nouns were divided into two arbitrary “classes”, marked by preceding articles so that each noun appeared only with one article. Six of the nouns appeared with the article ‘*fo*’ and the other six with the article ‘*si*’. There was no semantic or phonological cue to class membership. The only cue was distributional (which article the noun appeared with). Each noun had one semantic meaning of a frequent, concrete, inanimate object (based on Łuniewska et al., 2016). The semantic meanings were presented on the screen using pictures, each meaning with a set of 5 different pictures (e.g., 5 possible different pictures of a pan; pictures from Abu-Zhaya & Arnon, 2024, and Siegelman & Arnon, 2015). Because participants’ L1 does use grammatical gender, we made sure that in each noun class L1’s grammatical gender was matched (i.e., 3 masculine and 3 feminine). Alongside the grammatical sentences, 12 article-foil sentences were used to assess learning of the article-noun pairings. These sentences had the same structure, but the nouns appeared with the article not used in exposure (see example 2).

- |     |                     |                  |              |
|-----|---------------------|------------------|--------------|
| (1) | <b>os ferpel ti</b> | <b>fo</b>        | <b>etkot</b> |
|     | carrier phrase      | exposure article | noun         |
| (2) | <b>os ferpel ti</b> | <b>si</b>        | <b>etkot</b> |
|     | carrier phrase      | wrong article    | noun         |

A female speaker recorded the carrier phrase, the articles, and the nouns separately. The elements were synthesized to a frequency of 170 Hz to remove possible prosodic cues to word boundaries using Praat (Boersma, 2001) and concatenated to create full sentences. All the elements in the sentences were heard without pauses between them, to ensure that the only cue for word boundaries inside each sequence was distributional.

**Procedure** The task consisted of two stages: a learning phase, during which participants were exposed to the language and performed the typing trials, and a test phase. Participants were told they would be learning a novel language by seeing pictures and hearing their descriptions in that language. They were instructed to listen carefully as they would later be tested on the language they just heard. They

were also told they would periodically be asked to type in the last sentence they heard.

In the learning phase, participants were exposed to each sentence 10 times, resulting in 120 learning trials. In each trial, participants saw a picture of an object and heard its description in the language. The order of sentences in the exposure was random for each participant.

As in previous studies (Abu-Zhaya & Arnon, 2024; Siegelman & Arnon, 2015) participants also performed typing trials: every few learning trials a text box appeared, and participants were instructed to type in the last sentence they heard. This served as a method to assess participants’ tendency to chunk the new language into multiword units, specifically, by examining if they perceived the noun and the article as one or more units. Participants performed 40 typing trials, 3 trials in every 10 sentences (i.e., 3 trials between learning trials 1-10, 3 trials between learning trials 11-20, etc.). The position of typing trials within each block was consistent for all participants, but because stimuli presentation was random, each participant typed different sentences in each typing trial. Responses were recorded and subsequently coded by a research assistant for whether the article and noun were typed as one word (1) or several separate words (0). To assure inter-rater reliability, the first author of this manuscript also coded the responses, and Cohen’s kappa was calculated to measure agreement between coders. According to predetermined coding guidelines, responses were not coded if: (1) they did not have a clear article; (2) they did not have a clear noun (because such trials could not be coded for the use of an article-noun pairing). Participants were excluded from the analysis if the number of their uncoded responses was more than 2 SD from the mean of uncoded responses.

After the exposure phase, participants’ knowledge of the language was measured in test trials, using a 2-alternative forced choice (2AFC) task. In each trial, participants saw one picture and heard two sentences by clicking on speaker icons under two choice buttons (labeled “This one is correct”). Each sentence could be played up to two times. Participants were asked to indicate which sentence was correct in the language they learned, by clicking on one of the “This one is correct” buttons, with choice available only after hearing each sentence at least once.

Half of the test trials were noun trials, testing participants’ knowledge of noun–object mapping. In these trials, both sentences contained the correct article, but one sentence had a noun that did not match the presented picture. The other half of the test trials were article trials, testing participants’ knowledge of the article–noun pairings. In these trials, both sentences had the correct noun for the presented object, but one sentence had an incorrect article. Each of the 12 objects appeared twice in an article trial and twice in a noun trial, resulting in 48 2AFC trials. Each of the twelve nouns appeared 8 times during the test: four times in article trials (in two trials, each trial as both a target and a distractor, depending on the attached article), and four times in noun trials (twice as a target and twice as a distractor). The order

of trials and the two options within each trial were randomized for each participant. Participants who scored less than two SD from the mean in the noun trials were excluded from the analysis, for not learning the language.

All participants performed the ALL task first and were encouraged to take a short break before performing the MWC task, with both tasks lasting approximately 30 minutes.

## Multiword Chunking Task

The task was modeled after the ones used in McCauley et al. (2017) and Isbilen et al. (2022). In this task, participants were assessed for the processing benefit they gained from relying on multiword information in their L1, given their baseline WM capacity. Participants were asked to recall sequences of 12 words, which were either composed of four separate word trigrams (target sequences), or the same words, but presented in a random order (control sequences).

## Methods

**Materials** 20 sequences of 12 Hebrew words were created for the task. Stimuli were formed to resemble the task used in previous studies (McCauley et al., 2017; Isbilen et al., 2022). The frequency range of the trigrams used in the target sequences was 0.08-40 per-million (mean frequency 2.24 per-million), measured using trigram counts from the OpenSubtitles corpus (Van Paridon & Thompson, 2020). Control sequences used the same 12 words from each of the 10 target sequences, but in a pseudo-random order while avoiding high-frequency bigrams and trigrams. All words were created individually using Google Text-to-Speech, then combined with 75 ms pauses between each word to create both the target and the control sequences.

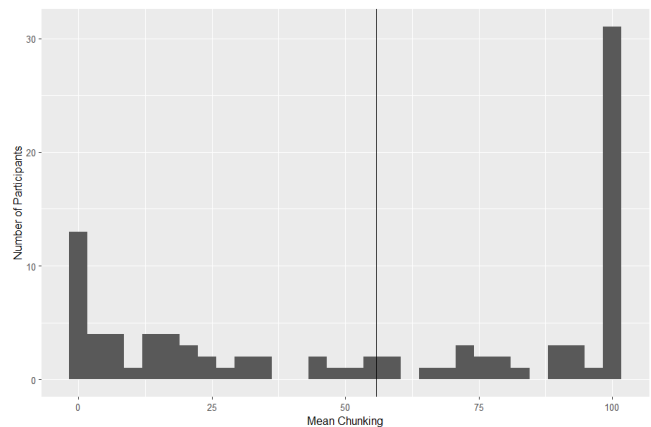
**Procedure** Participants were told they are about to hear sequences of words in their native language. They were instructed to listen to the sequences carefully and type the words they heard when a text box appeared, after the sequence was presented. The order of sequences was random for each participant. Responses were recorded and subsequently coded by a research assistant, for the number of words correctly recalled from the sequence, with possible score range 0-12. To assure inter-rater reliability, the first author of this manuscript also coded the responses, and Cohen's kappa was calculated to measure agreement between coders. Misspellings were corrected and scored as correctly recalled words. Addition or omission of the definite article to the target words was scored as correct recalls, as it is an affix in Hebrew (Wintner, 2000).

## Results

**ALL Typing Trials** The inter-rater reliability in coding the typing trials was high (Cohen's Kappa=0.824). When there was no agreement (10.6% of trials), trials were reviewed and re-coded by the two coders together. Additionally, responses that did not have a clear article or noun were not coded

(10.8%, after exclusion of outliers). As a result, 89.2% of the responses were coded and used in the final analysis.

First, to estimate participants' overall tendency to learn using multiword units, we calculated a mean chunking measure for each participant, by averaging over all the coded typing responses for each participant: the measure indicates the proportion of trials where participants typed the article and noun as one unit. At the group level, participants were more likely to chunk the article and noun into a multiword unit than to segment it to two units, but, importantly, there was large individual variability in this tendency ( $M=55.8\%$ ,  $SD=40.9\%$ ). As can be seen in Figure 1, this large variability is a result of an almost bimodal distribution across participants (although this can be a result of the use of a binary variable; Houghton & Kapatsinski, 2024). That is, whereas some participants treated the article-noun pairings as one word throughout the whole task, many others treated them as two separate words in most of the trials.

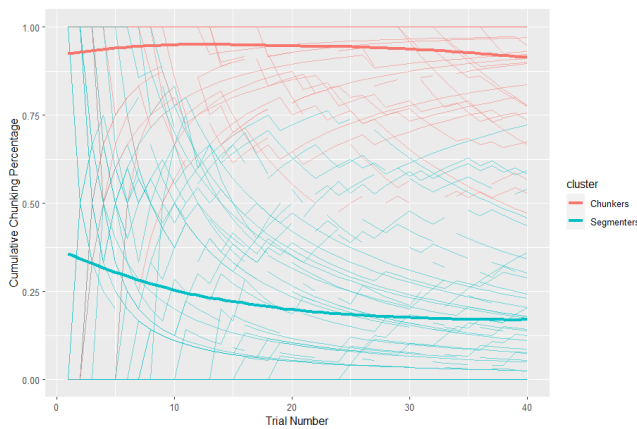


**Figure 1:** Distribution of the mean chunking score in the ALL task: the proportion of trials in which participants typed in the article and noun as one unit during learning phase.

To explore whether participants indeed fall into distinct sub-groups that differ in their behavior, we used time-series clustering to group participants based on their cumulative chunking percentage (proportion of trials treating the article-noun pairing as one word) up to each trial. We calculated this percentage for each trial for each participant by counting the number of article-noun responses typed in as one unit in all trials until trial  $i$ , and dividing it by the number of coded trials for this participant until trial  $i$ . We used the dtwclust package in R (Sarda-Espinosa, 2024), with *dtw\_basic* (dynamic time wrapping) as the distance measure, and *pam* (partition around medoids) as the time-series prototype. To decide how many clusters should be used we calculated the silhouette score for 2 to 10 clusters, using cluster package in R (Maechler, Rousseeuw, Struyf, Hubert, & Hornik, 2024). The result indicated that the optimal number of clusters is 2, with a silhouette score of 0.68. For the sake of clarity, from here onwards we refer to the cluster that includes participants who tended to treat the article and noun as one unit as the *chunkers* cluster, and the cluster with participants who treated the two

units as separate as the *segmenters* cluster. The clustering model resulted in allocating 50 participants to the chunkers cluster, and 46 participants to the segmenters cluster.

Figure 2 shows the cumulative chunking percentage throughout the task for individual participants (thin lines), and the smoothed curve for the two clusters (thick lines). As can be seen, the clustering solution points to two distinct groups of participants, which differ in their tendency to treat the article-noun pairing as one word. Importantly, the smoothed curves of the two clusters (which represent the typical behavior of participants in each cluster) suggest that the behavior of these two clusters is mostly consistent throughout the task. This means that participants' tendency to use large or small units is present early on in the task, after being exposed only to few sentences, and remains mostly steady even when they are exposed to more input.



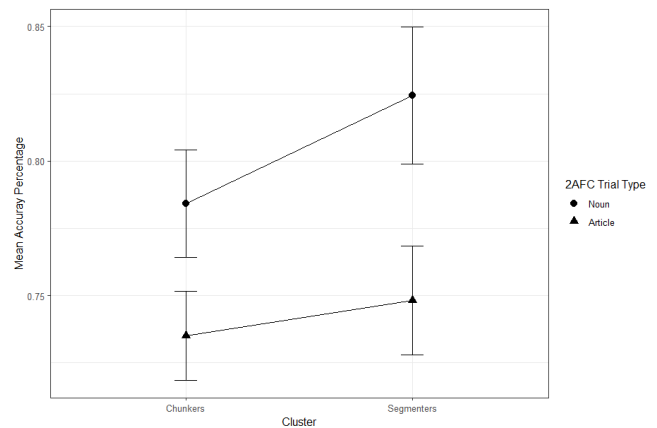
**Figure 2:** Cumulative chunking percentage throughout the ALL learning phase. Each thin line represents the percentage of a single participant in every coded trial throughout the task. Thick lines represent the mean percentage for the two clusters.

**2AFC Task** As expected, participants showed learning of the nouns and their correct articles, as indicated by above chance performance in both types of trials (Noun trials:  $M=80\%$ ,  $SD=15.7\%$ ,  $t(95)=18.89$ ,  $p<0.001$ ; article trials:  $M=74.5\%$ ,  $SD=12.3\%$ ,  $t(95)=18.69$ ,  $p<0.001$ ).

To measure the relation between chunking the article and noun into a multiword unit and learning the article-noun pairings, we used a mixed-effect logistic regression model, using the lmerTest package in R (Kuznetsova, Brockhoff, & Christensen, 2017). Collinearity was small ( $VIF<1.46$ , for all fixed effects). Categorical variables in this model, as well as in the following models in this paper, were effect coded. The dependent variable was accuracy in each trial. The model included clustering allocation (segmenters vs. chunkers)<sup>1</sup>, trial type (noun vs. article), their interaction, and trial number as fixed effects. The random-effect structure included random intercepts for participants and nouns, as well as by-participant random slopes for trial type (the maximal random-

effect structure that converged, Barr, Levy, Scheepers, & Tily, 2013).

Overall, participants did significantly worse in article trials compared to noun trials ( $\beta=-0.32$ ,  $SE=0.06$ ,  $p<0.01$ ), and there was no significant difference in test performance between the clusters ( $\beta=-0.14$ ,  $SE=0.09$ ,  $p=0.12$ ). Unlike our predictions, there was no significant interaction between the cluster and trial type ( $\beta=0.09$ ,  $SE=0.059$ ,  $p=0.122$ ), though the results show a trend towards smaller difference in accuracy between article trials and noun trials for chunkers compared to segmenters, as expected. Specifically, Figure 3 shows that the difference between article trials and noun trials is rather small for chunkers, whereas segmenters performed substantially better on noun trials compared to article trials.

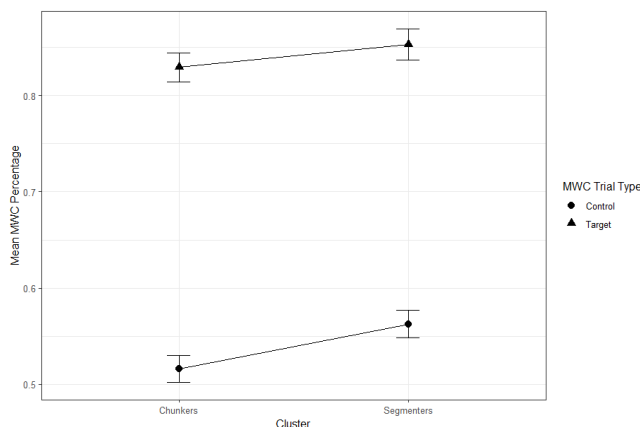


**Figure 3:** Mean accuracy in the 2AFC trials as a function of trial type and the two participant clusters. Error bars depict SE.

**MWC Task** As expected, participants recalled more words in the target sequences ( $M=84.1\%$ ,  $SD=16.4\%$ ) compared to the control sequences ( $M=53.8\%$ ,  $SD=15.3\%$ ). Next, we examined whether the tendency to rely on multiword units in a newly learned language predicts the benefit of recalling multiword sequences in one's L1. We used a mixed-effect regression model, using the lmerTest package in R (Kuznetsova, Brockhoff, & Christensen, 2017). Collinearity was small ( $VIF<1.1$ , for all fixed effects). The dependent variable was the percentage of correctly recalled words in each trial. The model included trial type (control vs. target), the clusters from the ALL task (segmenters vs. chunkers), their interaction, and trial number as fixed effects. The maximal random-effect structure that converged was random intercepts for participants and items, as well as by-participant random slope for trial type. Unsurprisingly, participants recalled significantly more words in the target trials than in control trials ( $\beta=0.15$ ,  $SE=0.009$ ,  $p<0.001$ ). The effect of cluster was marginally significant, with chunkers recalling overall fewer words from the sequences ( $\beta=-0.018$ ,  $SE=0.01$ ,  $p=0.0769$ ). This effect might point to general differences in memory capacity, and we return to this in the general discussion. Lastly, unlike our prediction, there was no

<sup>1</sup> Results were similar when using the numeric mean measure.

interaction between cluster and trial type ( $\beta=0.006$ ,  $SE=0.004$ ,  $p=0.1379$ ), though the results again show a trend towards chunkers benefitting more from recalling target trials, compared to participants in the segmenters cluster (Figure 4).



**Figure 4:** Mean words correctly recalled in the MWC task as a function of trial type and the two participant clusters. Error bars depict SE.

## General Discussion

In the present study, we show that there are individual differences in the reliance on multiword units in language learning and processing. These differences were observed in language learning using an ALL task, where participants differed in how they segmented article-noun sequences. Importantly, participants could be described using two distinct categories: While some participants tended to treat article-noun sequences as one multiword unit, others tended to segment them, treating them as two words (or more). Our findings highlight that much of these differences are present from the first exposure to the language, with participants mostly treating the article-noun pairings as one unit or more already from very early trials and with little change throughout the whole task. Furthermore, albeit not significant, there were numeric trends suggesting that these individual differences relate to the way individuals learn articles vs. nouns in the artificial language, and to the way they process multiword information in their L1.

It is worth highlighting, however, that the relation between reliance on multiword units in the ALL and learning outcomes in this task were different than what we expected. Unlike our prediction, we did not find that participants who tended to treat the article-noun as one multiword unit were better at learning the relation between them. Rather, the difference between the groups had to do with their ability to learn the nouns of the artificial language, with participants who tended to segment the article-noun pairings showing better learning of the nouns. Although this question was not at the center of the current study, these results point to a difference in learning the artificial language based on the realized units of the language, but in a different way than found in previous work (e.g., Siegelman & Arnon, 2015).

However, it is also worth noting that in previous experiments participants were exposed to both segmented and unsegmented input, differing in the order of exposure, while here participants were exposed only to the unsegmented stimuli. Because all participants were exposed only to the unsegmented stimuli, and the only clue for word boundaries was distributional, these differences might point to differences in statistical learning, with segmenters as “better” statistical learners. These participants might have better segmented the nouns from the sentences, leading them to allocate more attention to them while learning, and as a result to better learn them.

Another effect that should be accounted for is the marginal overall worse performance in the MWC task for participants in the chunkers cluster. If indeed replicated in a larger sample, this effect might be an important step in understanding the underlying reason for the individual differences in the tendency to chunk linguistic information into multiword units. In line with chunking theories (e.g., Miller, 1956; Christiansen & Chater, 2016), one possible explanation is that chunkers might have lower memory abilities, leading them to recode more linguistic information into larger units and use more multiword information. These individuals might be prone to chunk linguistic information into larger units when learning, otherwise they might not be able to memorize it and use it later. This can result in the discussed effect: these individuals would be chunking new linguistic information while having a hard time recalling linguistic information that they cannot recode as chunks. This possibility can be examined in future studies by directly testing individual differences in memory abilities.

To summarize, to our knowledge, this study provides the first investigation of individual differences in the linguistic unit size perceived during learning, pointing to an almost dichotomous separation between individuals who attend to larger units and ones who attend to smaller ones, which is present already in the early stages of exposure to a new linguistic input. Moreover, numeric trends in our data suggest that these tendencies relate to how one process their native language: Individuals who tend to chunk linguistic units into multiword units when learning show greater benefit when they rely on multiword units in processing their L1. Taken together, these findings uncover an important dimension of individual differences: The tendency of a language learner and user to rely on multiword units. Additionally, they highlight the need for further investigation of individual differences in unit size used for learning and processing, exploring if these differences are stable over time, and their relation to additional ecological language learning and processing tasks. As a next step to understand the stability of this tendency over time, we are currently performing a test-retest paradigm, to evaluate its reliability.

## Acknowledgments

We thank Tamar Matas for coding the data. Work in this paper was supported by the Israel Science Foundation (ISF)

Grant, project 445/20 (awarded to IA), and project 1034/23 (awarded to NS), and by an Azrieli Early Career Faculty Fellowship (awarded to NS).

## References

- Abu-Zhaya, R. and Arnon, I. (2024) Does early unit size impact the formation of linguistic predictions? *Language Learning*, Published online April 2, 2024.
- Arnon, I. (2021). The Starting Big approach to language learning. *Journal of Child Language*, 48(5), 937-958.
- 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.
- 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.
- 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.
- Arnon, I., & Snider, N. (2010). More than words: Frequency effects for multi-word phrases. *Journal of Memory and Language*, 62(1), 67–82.
- Bannard, C., & Matthews, D. (2008). Stored word sequences in language learning: The effect of familiarity on children's repetition of four-word combinations. *Psychological science*, 19(3), 241-248.
- 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, 255–278.
- Boersma, P. (2001). Praat, a system for doing phonetics by computer. *Glott International*, 5, 341–345.
- Christiansen, M. H., & Arnon, I. (2017). More than words: The role of multiword sequences in language learning and use. *Topics in cognitive science*, 9(3), 542-551.
- Christiansen, M. H., & Chater, N. (2016). The now-or-never bottleneck: A fundamental constraint on language. *Behavioral and brain sciences*, 39, e62.
- Cole, R.A. & Jakimik, J. (1980). A model of speech perception. In R.A. Cole (Ed.), *Perception and production of fluent speech* (pp. 133-163). Mahweh, NJ: Erlbaum.
- Cowan, N. (2001). The magical number 4 in short-term memory: A reconsideration of mental storage capacity. *Behavioral and brain sciences*, 24(1), 87-114.
- Dörnyei, Z. (2014). *The psychology of the language learner: Individual differences in second language acquisition*. Routledge.
- Ellis, N. C. (2012). Formulaic language and second language acquisition: Zipf and the phrasal teddy bear. *Annual review of applied linguistics*, 32, 17-44.
- Erickson, L. C., & Thiessen, E. D. (2015). Statistical learning of language: Theory, validity, and predictions of a statistical learning account of language acquisition. *Developmental Review*, 37, 66-108.
- Havron, N., Raviv, L., & Arnon, I. (2018). Literate and preliterate children show different learning patterns in an artificial language learning task. *Journal of Cultural Cognitive Science*, 2(1), 21-33.
- Houghton, Z. N., & Kapatsinski, V. (2024). Are your random effects normal? A simulation study of methods for estimating whether subjects or items come from more than one population by examining the distribution of random effects in mixed-effects logistic regression. *Behavior Research Methods*, 56(6), 5557-5587.
- Isbilen, E. S., McCauley, S. M., & Christiansen, M. H. (2022). Individual differences in artificial and natural language statistical learning. *Cognition*, 225, 105123.
- Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest package: tests in linear mixed effects models. *Journal of statistical software*, 82(13).
- Li, S., Hiver, P., & Papi, M. (Eds.). (2022). *The Routledge handbook of second language acquisition and individual differences*. New York: Routledge.
- Łuniewska, M., Haman, E., Armon-Lotem, S., Etenkowski, B., Southwood, F., Anđelković, D., ... & Ünal-Logacev, Ö. (2016). Ratings of age of acquisition of 299 words across 25 languages: Is there a cross-linguistic order of words?. *Behavior research methods*, 48, 1154-1177.
- Maechler, M, Rousseeuw, P, Struyf, A, Hubert, M, Hornik, K (2024). *cluster: Cluster Analysis Basics and Extensions*. R package version 2.1.8
- McCauley, S. M., Isbilen, E. S., & Christiansen, M. H. (2017). Chunking ability shapes sentence processing at multiple levels of abstraction. In G. Gunzelmann, A. Howes, T. Tenbrink, & E. J. Davelaar (Eds.), *Proceedings of the 39th Annual Conference of the Cognitive Science Society* (pp. 2681–2686). Austin, TX: Cognitive Science Society.
- Miller, G.A. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review*, 63, 81-97.
- Pulido, M. F. (2021). Individual chunking ability predicts efficient or shallow L2 processing: Eye-tracking evidence from multiword units in relative clauses. *Frontiers in Psychology*, 11, 607621.
- Pulido, M. F., López-Beltrán, P., Dussias, P. E., & Christiansen, M. H. (2024). Individual chunking ability predicts online sensitivity to L2 morphosyntax: Evidence from Spanish gender agreement. *Submitted for Publication*.
- Sarda-Espinosa A (2024). *dtwclust: Time Series Clustering Along with Optimizations for the Dynamic Time Warping Distance*. R package version 6.0.0, <https://CRAN.R-project.org/package=dtwclust>.
- 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.
- Siegelman, N., Bogaerts, L., Christiansen, M. H., & Frost, R. (2017). Towards a theory of individual differences in

- statistical learning. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 372(1711), 20160059.
- 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
- Van Paridon, J., & Thompson, B. (2021). subs2vec: Word embeddings from subtitles in 55 languages. *Behavior research methods*, 53(2), 629-655.
- Wintner, S. (2000). Definiteness in the Hebrew noun phrase. *Journal of Linguistics*, 36(2), 319-363.