INTERSPEECH 2019
September 15–19, 2019, Graz, Austria

Using Prosody to Discover Word Order Alternations in a Novel Language

Anouschka Foltz[1], *Sarah Cooper*[2], *Tamsin M. McKelvey*[2]

[1]Department of English Studies, University of Graz, Austria
[2]School of Languages, Literatures and Linguistics, Bangor University, UK
anouschka.foltz@uni-graz.at, s.cooper@bangor.ac.uk, tamsinmarie94@live.co.uk

Abstract

Native speakers of a language can use prosodic phrasing to disambiguate syntactically ambiguous sentences [1]. The current paper explores whether prosodic phrasing can help learners determine within-clause word order differences in a new language. Unlike many previous studies, we did not train participants in an artificial language, but exploited word order differences that occur in German. Native English speakers with no knowledge of German were trained with simple main clause sentences as well as complex sentences containing a subordinate clause. During training, prosodic phrasing of complex sentences either aligned or did not align with the sentences' clause structure. The results from two experiments showed that the non-aligned prosodic phrasing helps learners discover clause internal word order differences in German, but only if syntactic variability in the test sessions is low. Overall, the results suggest that learners can exploit prosodic structure to learn word order alternations in certain contexts.

Index Terms: prosody, prosodic phrasing, word order, second language acquisition

1. Introduction

Native speakers of a language can use prosodic phrasing to disambiguate syntactically ambiguous sentences [1]. While speakers often do not use prosodic cues to disambiguate syntactically ambiguous sentences in their own productions [2], listeners are quite sensitive to such prosodic cues when comprehending sentences [3]. In addition, adult second language learners can use prosodic phrasing to disambiguate syntactically ambiguous sentences similarly to native speakers [4]. Similar effects of prosodic cues have been found in artificial language learning tasks. For example, native speakers of both Dutch and Korean successfully used final lengthening and, to some extent, a final F0 rise as cues to word segmentation [5]. Listeners can also use prosodic cues to discover syntactic structures. For example, listeners learn the syntax of an artificial language better when listening to sentences with prosodic phrasing that coincides with the sentences' constituent structures compared to prosodic phrasing that either does not coincide with the sentences' constituent structures or to prosody that is monotone [6].

In this study, we explore whether prosodic phrasing can provide monolingual English listeners with word order cues in a novel language. Unlike many previous studies, we do not use an artificial language, but exploit word order differences that occur in German. Specifically, finite verbs in main clauses occur in verb-second position, whereas they occur in verb-final position in subordinate clauses. In contrast, the finite verb in English main and subordinate clauses always precedes the object noun phrase. We explore whether prosodic phrasing can serve as a cue not only to syntactic structure [6], but also to such word order differences *within* clauses. We focus on two kinds of prosodic phrasing to explore which may better support listeners in discovering this verb-second / verb-final alternation in German. The prosodic phrasing of a complex sentence like *Peter denkt, dass der Apfel grün ist* (Peter thinks that the apple is green) can be aligned with the sentence's constituent structure, with a prosodic boundary after *denkt* and main and subordinate clauses each produced in their own prosodic phrase. In contrast, a prosodic boundary after *dass* constitutes prosodic phrasing that is not aligned with the sentence's constituent structure. Instead, the subordinate conjunction *dass* forms one prosodic unit with the main clause of the sentence. As a result, the second prosodic phrase contains the same words (i.e. *...der Apfel grün ist*) as the corresponding main clause sentence *Der Apfel ist grün* (The apple is green), but in a different order. Thus, while the syntactically aligned prosodic phrasing may help listeners discover the main clause / subordinate clause structure of such complex sentences, the non-aligned prosodic phrasing may more easily highlight the fact that the clause-internal word order of the subordinate clause differs from that of a main clause.

2. Experiment 1

Experiment 1 tests how training with syntactically aligned vs. non-aligned prosodic phrasing influences participants' performance in correctly identifying correct sentences and sentences with word-order or lexical errors in a novel language.

2.1. Methods

2.1.1. Participants

30 monolingual English-speaking students at Bangor University (19 female, 11 male, mean age = 22, range 18-26) participated in the experiment. None of the participants reported any knowledge of German.

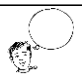
2.1.2. Materials and procedure

The experiment consisted of six alternating training (three trials each) and test (six trials each) phases. We refer to the first three training and test phases as the first half of the experiment and to the last three as the second half of the experiment. During each trial of the training and test phases, participants saw a picture on a computer screen and heard a sentence in German (see Table 1 for examples). A picture showing a boy (Peter) with a thought bubble was always presented with the sentence *Peter thinks*. A picture of a single object always appeared with a simple, main clause sentence, and a picture of a single object in Peter's thought bubble always appeared with a complex sentence containing a subordinate clause. Table 1 shows the target verb *is*, whose position differs in a main vs. subordinate clause, in bold face. All sentences in the training phase were grammatically correct and used appropriate lexical items to describe the

http://dx.doi.org/10.21437/Interspeech.2019-1183

pictures. In addition, all color terms and object names were close cognates with English, allowing participants to identify key lexical items in the speech stream.

Table 1: *Sample pictures and sentences (training).*

sample pictures	sample sentences
	a. Peter denkt. "Peter thinks."
	b. Der Apfel **ist** grün. "The apple is green."
	c. Peter denkt, dass der Apfel grün **ist**. "Peter thinks that the apple is green."

A phonetically-trained native German speaker produced each simple main clause sentence in one prosodic phrase. She also produced two different versions of each complex sentence, one with a prosodic boundary following *denkt* (prosody-syntax aligned) and one with a prosodic boundary following *dass* (prosody-syntax non-aligned). These sentence-medial prosodic boundaries were marked by final lengthening, a continuation rise and followed by an audible pause. All sentences were recorded with a Zoom H1 voice recorder at an overall slow rate as the participants were completely unfamiliar with German.

Table 2: *Sample pictures and sentences (testing).*

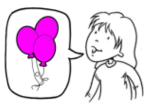
sample pictures	sample sentences for each condition
	*a.* <u>*simple correct:*</u> Die Ballons **sind** pink. "The ballons are pink." *b.* <u>*simple incorrect word order:*</u> Die Ballons pink **sind**. *c.* <u>*simple incorrect lexical item:*</u> Die Ballons **sind** braun.
	*d.* <u>*complex correct:*</u> Anna ruft, dass die Ballons pink **sind**. "Anna says that the ballons are pink." *e.* <u>*complex incorrect word order:*</u> Anna ruft, dass die Ballons **sind** pink. *f.* <u>*complex incorrect lexical item:*</u> Anna ruft, dass die Ballons braun **sind**.

Table 2 shows sample pictures and sample sentences for each experimental condition heard during the test phase. A picture showing several of the same kind of object was combined with a simple, main clause sentence that was either (a) correct, both in terms of word order and lexical items, (b) had incorrect word order, or (c) contained an incorrect color term or object name. A picture showing a girl (Anna) with a speech bubble containing several of the same kind of object was combined with a complex sentence containing a subordinate clause that was either (d) correct, both in terms of word order and lexical items, (e) had incorrect word order, or (f) contained an incorrect color term or object name. Again, the target verb, in this case *are* (bold face in Table 2), differs in terms of position in a main vs. subordinate clause. As before, all color terms and object names were close cognates with English. All test sentences were produced by the same phonetically-trained native German speaker in one prosodic phrase to ensure that test sentences were identical across the participants who had heard syntactically aligned vs. non-aligned prosody during training. The training and test phases also used different verbs (*is* vs. *are*) to ensure that participants were not merely trained to detect the position of the word *is*, and different colors and object names to ensure that any learning occurred for novel sentences.

Participants gave informed consent before completing the experiment in E-Prime. Before each training phase participants were informed that they would see pictures and hear sentences that correctly described the pictures, and that their task was to watch and listen carefully. Before each test phase participants were told that they would see pictures and hear sentences that described the pictures correctly, or contained mistakes, and that their task was to push a button to indicate whether or not they thought that the sentence described the picture correctly. After the experiment, participants filled in a brief language background questionnaire and were debriefed.

2.2. Results

We will first consider the simple, verb-second sentences. Figure 1 shows the proportion of correct responses for the simple sentences. Participants overwhelmingly identified correct sentences as correct and sentences with an incorrect lexical item as incorrect, but only a minority of sentences with incorrect verb-final order as incorrect.

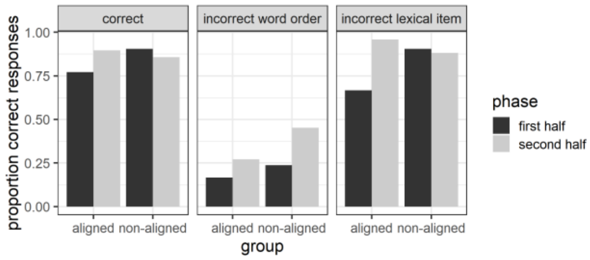


Figure 1: *Results for simple sentences.*

We fit mixed logit models [7], which are appropriate for binary response variables, to the data. The full model had response (correct vs. not correct) as response variable and participant group (aligned vs. non-aligned), sentence condition (correct, incorrect word order, and incorrect lexical item), experiment phase (first vs. second half) and all interactions as predictor variables. Participant and item were added as random effects and random slopes were added for the within-subject predictors sentence condition and experiment phase [8]. Fixed and random effects that did not reliably contribute to model fit were removed to yield the final analysis model. If a model did not converge, the random effects structure was simplified until the model converged. The final model contained participant group, sentence condition, experiment phase, and the sentence condition by experiment phase interaction as fixed effects, and participant and item as random effects. The results from the final model show a significant main effect of group (logit estimate = 0.37, std. error = 0.17, z = 2.19, p < 0.05), with the non-aligned group performing better overall than the aligned group. In addition, there was a marginal sentence condition by experiment phase interaction (logit estimate = 0.78, std. error = 0.4, z = 1.95, p = 0.05). Sentence condition (logit estimate = 0.62, std. error = 0.4, z = 1.54, p = 0.12) and experiment phase (logit estimate = 0.57, std. error = 0.43, z = 1.31, p = 0.19) yielded no significant main effects. We report the marginal and conditional R^2 values for generalized linear mixed effects models to gauge effect sizes [9, 10, 11]. The marginal R^2_{GLMM} value is 0.17 and captures the variance explained by the model's fixed factors, suggesting that little of the variance in correct responses can be explained through the fixed factors in the final model. The conditional R^2_{GLMM} value is 0.62 and captures the variance explained by the model's fixed and random factors, suggesting that the random effects structure contributes more to the variance in correct responses than do the fixed effects.

We now explore the trend towards a sentence condition by experiment phase interaction by looking at all three sentence types separately. To do this, we fit separate mixed logit models for each sentence type as above (except that sentence condition was excluded from the models). The final model for the correct sentences had no fixed or random effects and thus showed no reliable effects of participant group or experiment phase on the proportion of correct responses. That is, the proportion of correct responses for correct sentences was similar across participant groups and across the phases of the experiment. The final model for the incorrect word order sentences also had no fixed effects, but included subject and random slopes for experiment phase for each subject. That is, the proportion of correct responses for incorrect word order sentences was again similar across participant groups and across the phases of the experiment, but there were individual differences in correct responses across the two experiment phases that contributed reliably to model fit. The final model for the incorrect lexical item sentences had no random effects, but showed a significant main effect of experiment phase (logit estimate = 0.59, std. error = 0.27, z = 2.22, $p < 0.05$) suggesting that participants identified sentences with incorrect lexical items as such significantly more frequently in the second than the first half of the experiment. In addition, there was a significant participant group by experiment phase interaction (logit estimate = 0.68, std. error = 0.27, z = 2.55, $p < 0.05$). Follow-up mixed logit models for each group separately showed a significant increase in correct responses from the first to the second half of the experiment for the aligned group (logit estimate = 1.48, std. error = 0.46, z = 3.19, $p < 0.01$), but not for the non-aligned group (final model included no fixed or random effects). Thus, participants in the non-aligned group performed better overall, and participants in the aligned group identified incorrect lexical items better in the second half of the experiment than the first, whereas no such pattern was found for participants in the non-aligned group.

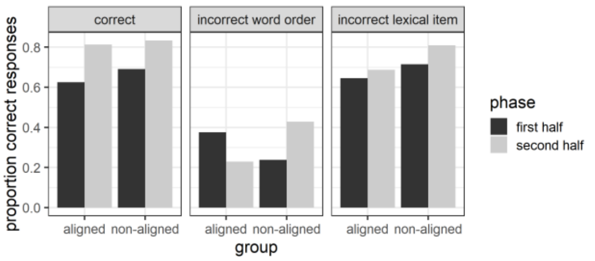


Figure 2: *Results for complex sentences.*

We will now consider the complex sentences, where the verb in the subordinate clause is in final position. Figure 2 shows the proportion of correct responses for the complex sentences. Again, participants overwhelmingly identified correct sentences as correct and sentences with an incorrect lexical item as incorrect but identified only a minority of sentences with incorrect verb-final order as being incorrect. The mixed logit analysis was done exactly as for the simple sentences. The final model included fixed effects of participant group, experiment phase and the participant group by experiment phase interaction, as well as item as random effect. The results from the final model show a significant main effect of participant group (logit estimate = 0.23, std. error = 0.12, z = 1.99, $p < 0.05$), with the non-aligned group again performing better overall than the aligned group. In addition, there was a significant participant group by experiment phase interaction (logit estimate = 0.23, std. error = 0.12, z = 2.01, $p < 0.05$).

We again report the marginal and conditional R^2_{GLMM} to gauge effect sizes. The marginal R^2_{GLMM} value is 0.03, suggesting that only 3% of the variance in correct responses can be explained through the final model's fixed factors. The conditional R^2_{GLMM} value is 0.31, suggesting that the random effects structure again contributes more to the variance in correct responses than do the fixed effects. To further explore the significant interaction, we ran separate mixed logit analyses for each experiment phase. The final model for the first half included no fixed effects and only item as random effect, suggesting that both groups of participants performed equally during the experiment's first half. In contrast, the final model for the second half showed a significant main effect of participant group (logit estimate = 0.48, std. error = 0.19, z = 2.5, $p < 0.05$) in addition to including item as random effect, suggesting that the non-aligned group outperformed the aligned group during the second half of the experiment. Overall, the results for the complex sentences show that the non-aligned group outperforms the aligned group overall, and that both groups start out at the same level, but only the non-aligned group manages to increase their performance over the course of the experiment.

3. Experiment 2

Experiment 2 explores how more syntactic variability during the test session influences participants' performance. Specifically, identifying patterns during the test phase may be more challenging if participants are exposed to more syntactic variety in the correct and incorrect complex sentences.

3.1. Methods

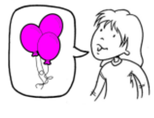
3.1.1. Participants

26 monolingual English-speaking students at Bangor University with no knowledge of German (19 female, 7 male, mean age = 19, range 18-26) participated in the experiment.

3.1.2. Materials and procedure

The materials and procedure were identical to Experiment 1, except that the two additional complex sentence conditions shown in Table 3 were added to the test phases.

Table 3: *Sample picture and sentences of the additional complex sentence conditions (testing).*

sample picture	sample sentences for each condition
	*g.* <u>*complex correct scrambling:*</u> Dass die Ballons pink **sind**, ruft Anna. "Anna says that the ballons are pink." *h.* <u>*complex incorrect scrambling:*</u> Die Ballons pink **sind** ruft Anna dass.

3.2. Results

We will again first consider the simple, verb-second sentences. Figure 3 shows the proportion of correct responses for the simple sentences. As in Experiment 1, participants overwhelmingly identified correct sentences as correct and sentences with an incorrect lexical item as incorrect, but falsely identified sentences with verb-final word order as correct. The mixed logit analysis was done exactly as in Experiment 1. The final model for the simple sentences included only sentence condition as fixed effects and no random effects. The results showed a significant main effect of sentence condition (logit estimate = 0.46, std. error = 0.11, z = 4.08, p < 0.001). A post-hoc test using the `emmeans` package in R [cf. 12] revealed significant differences across all sentence conditions (logit estimate = 3.14, std. error = 0.32, z = 9.83, p < 0.001 for correct vs. incorrect word order, logit estimate = -1.24, std. error = 0.53, z = -2.36, p < 0.05 for correct vs. incorrect lexical item, and logit estimate = -4.37, std. error = 0.49, z = -8.95, p < 0.001 for incorrect word order vs. incorrect lexical item). Thus, participants performed differently across all sentence conditions, but there were no significant differences for participant group or experiment phase. As the final model included only fixed effects, we only report the marginal R^2_{GLMM} value to gauge effect size, which is 0.59, suggesting that over half of the variance in correct responses can be explained through the fixed factor sentence condition.

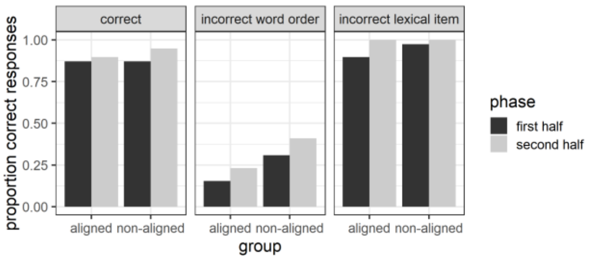


Figure 3: *Results for simple sentences.*

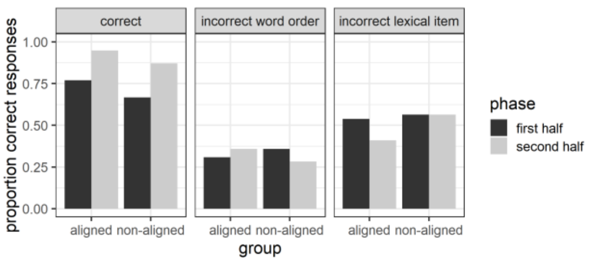


Figure 4: *Results for complex sentences.*

We will now consider the complex sentences, where the verb in the subordinate clause is in final position. Figure 4 shows the proportion of correct responses for the complex sentences. Participants overwhelmingly identified correct sentences as correct, but identified only about half (i.e. around chance level) or fewer sentences with an incorrect lexical item or with incorrect word order as incorrect. The mixed logit analysis was done exactly as in Experiment 1 and the final model contained no fixed effects, but participant and item as random effects and random slopes for sentence condition for each participant. Thus, there were no significant effects of sentence condition, participant group or experiment phase. As the final model includes no fixed effects, we only report the conditional R^2_{GLMM} value to gauge effect size, which is 0.63, suggesting that of 63% the variance in correct responses can be explained through the model's random effects structure. Overall, the results from Experiment 2 suggest that with additional variability in the syntactic structures presented during the test phase, there was no evidence that the prosodic structures heard during training influenced participants' performance. In addition, there was no evidence of learning across the experiment.

4. Discussion and Conclusion

The results from the two experiments show that prosody only seems to support learning of clause-internal word order alternations in a novel language when word order variability during testing is low. Specifically, participants in the non-aligned group performed better than the aligned group for both simple and complex sentences in Experiment 1. Importantly, while both groups in Experiment 1 started out at the same level, only participants in the non-aligned group showed significant learning for the complex sentences in the second half of the experiment. Note that it is the complex sentences, not the simple sentences, that contain clause-internal word order that is unfamiliar (i.e. verb-final) to the native English-speaking participants.

As expected, it is the non-aligned prosodic phrasing that seems to highlight the fact that the clause-internal word order of the subordinate clause differs from that of a main clause. We suggest that the non-aligned prosodic phrasing during training supported detecting the verb-final word order in the subordinate clauses more effectively than the aligned prosodic phrasing because it allowed for easier comparison with the simple main clause sentences. Specifically, the second prosodic phrase in a complex sentence with non-aligned prosodic phrasing contained the same words as the simple main clause, merely in a different order. Hearing exactly the same words packaged together in one prosodic phrase during training seems to facilitate discovering that these same words occur in different orders depending on whether or not there is preceding sentence material.

While non-aligned prosodic phrasing does seem to benefit discovering clause-internal word order differences, it needs to be mentioned that participants overall performed very poorly in the incorrect word order trials. On a majority of trials, participants falsely assumed that the complex sentences with incorrect word order were actually correct sentences in German. One possible explanation for the overall poor performance is that the complex sentences in the incorrect word order trials actually follow English word order in the sense that a word-for-word translation would yield a grammatically correct English sentence. Thus, native English speakers may have transferred their native-language knowledge and falsely assumed a verb-second word order for German subordinate clauses. In contrast, participants performed overall much better in the incorrect lexical item trials, suggesting that they had little trouble identifying cognates in the stream of speech. Interestingly, though, with high variability during testing, participants had substantially more difficulty in detecting cognate words that did not match the picture shown when listening to complex sentences. Increased syntactic variability may have led participants to pay less attention to lexical items.

Overall, the results suggest that prosodic phrasing can highlight clause-internal word order differences at the very early stages of learning if learners experience limited syntactic variability. Future studies should investigate this effect with a larger number of participants and a range of L1 language backgrounds.

5. References

[1] A. J. Schafer, S. R. Speer, P. Warren, and D. White, "Intonational disambiguation in sentence production and comprehension," *Journal of Psycholinguistic Research*, vol. 29, no. 2, pp. 169-182, 2000.

[2] D. W. Allbritton, G. McKoon, and R. Ratcliff, "Reliability of prosodic cues for resolving syntactic ambiguity," *Journal of Experimental Psychology: Learning, Memory, and Cognition*, vol. 22, no. 3, pp. 714-735, 1996.

[3] C. Clifton, K. Carlson, and L. Frazier, "Tracking the what and why of speakers' choices: Prosodic boundaries and the length of constituents," *Psychonomic Bulletin & Review*, vol. 13, no. 5, pp. 854-861, 2006.

[4] H. Hwang, and A. J. Schafer, "Prosodic effects in parsing early vs. late closure sentences by second language learners and native speakers," *Proceedings of the 3rd International Conference on Speech Prosody,* pp. 585-588, 2006.

[5] S. Kim, M. Broersma, and T. Cho, "The use of prosodic cues in learning new words in an unfamiliar language," *Studies in Second Language Acquisition*, vol. 34, no. 3, pp. 415-444, 2012.

[6] J. L. Morgan, R. P. Meier, E. L. Newport, "Structural packaging in the input to language learning: Contributions of prosodic and morphological marking of phrases to the acquisition of language," *Cognitive psychology*, vol. 19, no. 4, pp. 498-550, 1987.

[7] R. H. Baayen, *Analyzing linguistic data: A practical introduction to statistics using R*. Cambridge, UK: Cambridge University Press, 2008.

[8] D. J. Barr, R. Levy, C. Scheepers, and H. J. Tily, "Random effects structure for confirmatory hypothesis testing: Keep it maximal," *Journal of memory and language*, vol. 68, no. 3, pp. 255-278, 2013.

[9] S. Nakagawa, & H. Schielzeth, "A general and simple method for obtaining R2 from generalized linear mixed-effects models," *Methods in Ecology and Evolution*, vol. 4, no. 2, pp. 133-142, 2013.

[10] P. C. Johnson, "Extension of Nakagawa & Schielzeth's R2GLMM to random slopes models," *Methods in Ecology and Evolution*, vol. 5, no. 9, pp. 944-946, 2014.

[11] S. Nakagawa, P. C. Johnson, & H. Schielzeth, "The coefficient of determination R^2 and intra-class correlation coefficient from generalized linear mixed-effects models revisited and expanded," *Journal of the Royal Society: Interface*, vol. 14, no. 134, pp. 1-11, 2017.

[12] S. R. Searle, F. M. Speed, and Milliken, "Population marginal means in the linear model: an alternative to least squares means," *The American Statistician*, vol. 34, no. 4, pp. 216-221, 1980.