

<https://doi.org/10.15388/vu.thesis.984>

<https://orcid.org/0009-0002-0266-4735>

VILNIUS UNIVERSITY

INSTITUTE OF LITHUANIAN LANGUAGE AND FOLKLORE

Richard Udes

Exploring the Effectiveness of Task-Based Language Teaching for Morphologically Complex Languages: The Acquisition of Case in L2 Lithuanian

DOCTORAL DISSERTATION

Humanities,
Philology (H 004)

VILNIUS 2026

The dissertation was prepared between 2021 and 2026 at Vilnius University.

Academic Supervisor – Dr. Rita Juknevičienė (Vilnius University, Humanities, Philology – H 004).

Academic Consultant – Dr. Laura Vilkaitė-Lozdienė (Vilnius University, Humanities, Philology – H 004).

This doctoral dissertation will be defended in a public meeting of the Dissertation Defense Panel:

Chairperson – Prof. Dr. Inesa Šeškauskienė (Vilnius University, Humanities, Philology – H 004).

Members:

Assoc. Prof. Dr. Joan Castellví Vives (University of Barcelona, Humanities, Philology – H 004);

Assoc. Prof. Dr. Roger Gilabert Guerrero (University of Barcelona, Humanities, Philology – H 004);

Assoc. Prof. Dr. Inga Hilbig (Vilnius University, Humanities, Philology – H 004);

Assoc. Prof. Dr. Laura Kamandulytė-Merfeldienė (Vytautas Magnus University, Humanities, Philology – H 004).

The dissertation shall be defended at a public meeting of the Dissertation Defense Panel at 13:00 on October 1, 2026, in Z. Zinkevičiaus Room of the Faculty of Philology, Vilnius University. Address: Universiteto 3, Vilnius, Lithuania. Tel. +370 5 268 7000; e-mail: infor@cr.vu.lt

VILNIAUS UNIVERSITETAS

LIETUVIŲ LITERATŪROS IR TAUTOSAKOS INSTITUTAS

Richard Udes

Užduotimis grindžiamo kalbos mokymo
veiksmingumas mokant
sudėtingos morfologijos kalbų:
lietuvių K2 linksnio kategorijos
įsisavinimas

DAKTARO DISERTACIJA

Humanitariniai mokslai,
Filologija (H 004)

VILNIUS 2026

Disertacija rengta 2021–2026 metais Vilniaus universitete.

Mokslinė vadovė – dr. Rita Juknevičienė (Vilniaus universitetas, humanitariniai mokslai, filologija – H 004).

Mokslinė konsultantė – dr. Laura Vilkaitė-Lozdienė (Vilniaus universitetas, humanitariniai mokslai, filologija – H 004).

Gynimo taryba:

Pirmininkė – prof. dr. Inesa Šeškauskienė (Vilniaus universitetas, humanitariniai mokslai, filologija – H 004)

Nariai:

dr. Joan Castellví Vives (Barselonos universitetas, humanitariniai mokslai, filologija – H 004);

prof. dr. Roger Gilabert Guerrero (Barselonos universitetas, humanitariniai mokslai, filologija – H 004);

dr. Inga Hilbig (Vilniaus universitetas, humanitariniai mokslai, filologija – H 004);

doc. dr. Laura Kamandulytė-Merfeldienė (Vytauto Didžiojo universitetas, humanitariniai mokslai, filologija – H 004).

Disertacija bus ginama viešame Gynimo tarybos posėdyje 2026 m. spalio mėn. 1 d. 13 val. Vilniaus universiteto Filologijos fakulteto Z. Zinkevičiaus auditorijoje. Adresas: Universiteto g. 3, Vilnius, Lietuva, tel. +370 5 268 7000; el. paštas: infor@cr.vu.lt

ABBREVIATIONS

ACC	accusative case
CALF	Complexity, Accuracy, Lexical Variety, Fluency
CH	Cognition Hypothesis
CREED model	Construction-based, Rational, Exemplar-driven, Emergent, Dialectic
DAT	dative case
EOIT	Elicited Oral Imitation Test
GEN	genitive case
HDI	Highest Density Interval
HPD	Highest Posterior Density
INST	instrumental case
IP	Input Processing
L1	first (native) language
L2	second (non-native) language
LAC approach	Limited Attentional Capacity approach
LOC	locative case
NEG	negation
NOM	nominative case
PPP	Present-Production-Practice
<i>pt</i>	Probability of Direction
ROPE	Region of Practical Equivalence
SLA	Second Language Acquisition
SSARC	Stabilize, Simplify, Automatize, Restructure, Complexify
TBLT	Task-Based Language Teaching
TCF	Triadic Componential Framework

TSLT

Task-Supported Language Teaching

VOC

vocative case

CONTENTS

ACKNOWLEDGEMENTS	11
LIST OF TABLES	13
LIST OF FIGURES.....	15
INTRODUCTION.....	17
1 TASK-BASED LANGUAGE TEACHING.....	31
1.1 The central tenets of TBLT.....	31
1.2 Theoretical foundation of TBLT.....	34
1.2.1 Implicit and explicit learning and knowledge.....	34
1.2.2 The cognitive-interactionist model	37
1.2.3 Focus on form: integrating analytic and synthetic teaching approaches	39
1.2.4 Task-supported language teaching	42
1.3 Task design principles.....	44
1.3.1 Tools available for teaching grammar and morphology	44
1.3.1.1 Pre-modified input	44
1.3.1.2 Post-task strategies.....	44
1.3.1.3 Corrective feedback	45
1.3.2 Models designed to assess conceptual task difficulty.....	47
1.3.3 Proposals for designing tasks with morphologically complex L2s	51
1.3.3.1 Gilabert, Castellví – Gradual increase in conceptual task demands with manageable linguistic difficulty.....	53
1.3.3.2 Kogan, Bondarenko – Engage learners with cognitively challenging solo-tasks.....	55
1.3.3.3 Comparing and integrating both proposals.....	56
2 CASE.....	60
2.1 Case in Lithuanian	60
2.2 Second language acquisition of case.....	64

2.2.1	Acquisition of morphosyntactic structures within several theories.....	64
2.2.1.1	The Associative–Cognitive CREED.....	65
2.2.1.2	Input Processing.....	71
2.2.1.3	Processability Theory	73
2.2.1.4	Skill Acquisition Theory.....	78
2.2.2	Former research on L2 acquisition of case	81
2.2.2.1	Case acquisition in L2 Lithuanian	81
2.2.2.2	Case acquisition in other L2s.....	84
2.2.3	Order of acquisition	91
2.2.4	Case acquisition in a morphologically complex language through TBLT and TSLT instruction.....	96
3	STUDY DESIGN AND METHODOLOGY.....	100
3.1	Design and procedure	100
3.2	Participants.....	101
3.3	Target grammatical forms.....	104
3.4	Experimental intervention.....	107
3.4.1	Course and task design	107
3.4.1.1	Selection of the appropriate target tasks	107
3.4.1.2	Alignment of task demands with the CEFR	111
3.4.1.3	Inclusion of the target structures in the tasks.....	113
3.4.1.4	Task sequencing according to the SSARC model	117
3.4.1.5	Implementation of suggestions on adapting TBLT to the teaching of morphologically complex L2s	118
3.4.1.6	Assessment of morphological complexity.....	119
3.4.2	Overview of task contents	122
3.4.2.1	The pre-task stage.....	122
3.4.2.2	The main task stage	124
3.4.2.3	The post-task stage	126
3.4.3	Treatment differences for the TBLT and TSLT groups ...	127

3.5	Instruments.....	130
3.5.1	Rationale behind the test instruments	132
3.5.2	Details of the gap-fill test	135
3.5.3	The elicited oral imitation test	136
3.5.4	Test modifications	138
3.6	Transcription and coding	139
3.6.1	Gap-fill test.....	139
3.6.2	Elicited oral imitation test.....	147
3.7	Statistical methods	153
4	RESULTS	158
4.1	Results of the gap-fill test	158
4.1.1	Progress analysis.....	158
4.1.1.1	Strict binary scoring.....	158
4.1.1.2	Categorical scoring.....	171
4.1.2	Error type analysis	174
4.2	Results of the elicited oral imitation test.....	182
5	DISCUSSION.....	196
5.1	Treatment effects on the acquisition of explicit and implicit case knowledge.....	196
5.2	Lasting explicit and implicit learning gains.....	200
5.3	Different developmental patterns for case acquisition depending on the TBLT and TSLT treatments	203
5.4	Other variables potentially affecting the results	205
5.4.1	The participants' L1.....	205
5.4.2	Item characteristics: syntactic role, case, and paradigm ...	207
5.4.3	Lexical and syntactic load variables	210
5.4.4	Methodological issues	211
5.5	Implications, limitations, and suggestions for further research ..	214
5.5.1	Theoretical implications	214
5.5.2	Practical implications	221

CONCLUSIONS	226
REFERENCES	229
APPENDICES	245
SANTRAUKA	267
ABOUT THE AUTHOR.....	285

ACKNOWLEDGEMENTS

The journey of a PhD student is long and hard, but not necessarily lonely. Over the course of five years, I have had the pleasure of receiving considerable support from my friends, family, and colleagues.

First and foremost, I would like to thank my PhD supervisor, Rita Juknevičienė, for her guidance and continuous support during my PhD studies. No matter the issue, I knew I could always count on her help. Whenever I felt overwhelmed by the magnitude of the work to be done, one call later everything seemed manageable again. I would also like to express my gratitude to my PhD consultant, Laura Vilkaitė-Lozdienė. The fact that I usually referred to her as my second supervisor speaks volumes. I hope we can continue our discussions on statistics in applied linguistics in the future.

This research would not have been possible without the Department of Lithuanian Studies, which helped me set up the administrative support for the courses that served as the instructional treatment for my study. I would like to thank Eglė Gudavičienė and Andrius Apinis, in particular, for their assistance in coordinating the registration process. I would also like to express my gratitude to the VDU Education Academy, which allowed me to pilot my materials and test instruments on several occasions. I am especially indebted to Vilma Leonavičienė for her help.

Speaking of piloting, several of my own students served as guinea pigs for my first task versions. I would like to thank Lena Koronenko, Maksim Makhankou, and Pawel Martchenko for allowing me to try out my teaching materials with them.

Naturally, I would never have succeeded in writing this dissertation without its participants, be it those in the final study, the pilot study, or the control group. While I would like to name each and every one of you, I need to keep you all anonymous, of course. But know that I am very grateful to all of you for your participation in my study.

I would also like to extend my thanks to Eglė Grigaliūnaitė and Jurgis Marčėnas, who recorded the dialogs used as listening texts in the teaching materials.

A special thanks goes out to Danguolė Kapkan, as a conversation at *Alaus biblioteka* six years ago served as the first step in my PhD journey.

I am very fortunate to have had the support of many friends throughout my PhD studies. I am grateful to all of them, but I would especially like to express my thanks to Inga Stirnaitė for her continuous encouragement and Raymond Harder for the valuable discussions.

Naturally, nobody can achieve anything without the support of their family, and I am no exception. I am grateful to my parents, Fred and Wilma, and sister, Melissa, for always caring about my well-being despite the long distance separating us. Last but in no way least, I am immensely grateful to my wife, Paulina Udes, for her loving care, support, and especially patience with my constant busyness. Let us embark on a well-deserved trip when all this PhD stuff is finally behind us!

LIST OF TABLES

Table 1. The make-up of the Task Complexity component with the Triadic Componential Framework for task classification (based on Robinson, 2007a, pp. 15–16).....	49
Table 2. Potential sequences in the SSARC model (<i>Source:</i> Ellis et al., 2019, p. 83).....	52
Table 3. A simplified overview of the nominal declension system of Lithuanian (taken and adopted from Dubasava (2020, p. 64)).....	61
Table 4. Case frequency in adult speech in Lithuanian (Savickienė, 2005)	64
Table 5. Case frequency in written Lithuanian by genre.....	64
Table 6. Possible combinations of cues and outcomes (<i>Source:</i> N. Ellis, 2022, p. 39).....	69
Table 7. Suggested developmental trajectory for questions in L2 English within Processability Theory (taken from Pienemann and Lenzing (2020, p. 160))	74
Table 8. Suggested developmental trajectory of case marking in L2 Russian (<i>Source:</i> Magnani and Artoni (2015, p. 63)).....	76
Table 9. Group and participant details of the study	103
Table 10. Overview of the cases and forms relevant to this study’s target constructions.....	106
Table 11. Overview of steps to be taken for each Main Task.....	110
Table 12. Overview of target structures and other additional grammar points featured in each task	114
Table 13. Implementation of the SSARC model in this study.....	118
Table 14. MC10 values for the reading and listening texts of each task ...	121
Table 15. Number of instances of the target structures in the reading and listening text of each task.....	123
Table 16. List of exercises that were unique to each TBLT or TSLT version of the task	131
Table 17. The target endings of the instruments within the nominal case system in Lithuanian	132
Table 18. Criteria for operationalizing L2 implicit and explicit knowledge constructs relevant to the gap-fill and elicited oral imitation tests.....	133
Table 19. Scoring criteria for participant responses on the gap-fill test	144

Table 20. Scoring criteria for participant responses on the elicited oral imitation test.....	150
Table 21 Overview of models fitted to the data.....	157
Table 22. Descriptive statistics for the binary scoring of responses on the gap-fill test.....	158
Table 23. Bayesian posterior distribution summary for gap-fill test accuracy: Including control group (binary scoring)	159
Table 24. Bayesian posterior distribution summary for gap-fill test accuracy: Two treatment groups (binary scoring)	162
Table 25. Bayesian posterior distribution summary for gap-fill test accuracy: Two treatment groups, including the parameters Case and Paradigm	165
Table 26. Pairwise comparisons for Test Time and Paradigm	168
Table 27. Descriptive statistics for the categorical scoring of responses on the gap-fill test.....	172
Table 28. Bayesian posterior distribution summary for gap-fill test accuracy: Two treatment groups (categorical scoring).....	172
Table 29. Error-type distributions on items requiring subject marking.....	175
Table 30. Error-type distributions on items requiring object marking	176
Table 31. Error-type distributions on items requiring the nominative.....	177
Table 32. Error-type distributions on items requiring the genitive.....	178
Table 33. Error-type distributions on items requiring the accusative.....	179
Table 34. Overview of error-type distributions by subject and object marking and changes across test times	180
Table 35. Overview of error-type distributions by genitive and accusative marking and changes across test times	181
Table 36. Descriptive statistics of the elicited oral imitation tests	183
Table 37. Bayesian posterior distribution summary for elicited oral imitation test accuracy: Including the control group	184
Table 38. Bayesian posterior distribution summary for elicited oral imitation test accuracy: Two treatment groups.....	187
Table 39. Bayesian posterior distribution summary for elicited oral imitation test accuracy: Two treatment groups, including the parameters Case and Paradigm.....	191

LIST OF FIGURES

Figure 1. An overview of the study design in this dissertation	29
Figure 2. The cognitive-interactionist model (adapted from Ellis et al. (2019, p. 31))	38
Figure 3. The Lithuanian case hierarchy (Čižik-Prokaševa, 2003)	63
Figure 4. A sample graph of the power learning curve (adapted in <i>Excel</i> from DeKeyser, 2014, p. 97).....	79
Figure 5. The design of this study	100
Figure 6. Example of a post-task exercise.....	127
Figure 7. Two excerpts from the reading text of Task 3	128
Figure 8. Excerpt from the instructions of the first subtask in the TSLT version of Task 3.....	129
Figure 9. Predicted probabilities of correct responses by Test Time and Treatment Group for the gap-fill model, including the control group.....	161
Figure 10. Predicted probabilities of correct responses by Test Time and Treatment Group for the gap-fill model without the control group	163
Figure 11. Predicted probabilities of correct responses by Test Time and Case for the gap-fill model without the control group.....	167
Figure 12. Predicted probabilities of correct responses by Test Time and Treatment Group for the EOIT model, including the control group	186
Figure 13. Predicted probabilities of correct responses by Test Time and Treatment Group for the EOIT model without the control groups.....	188
Figure 14. Predicted probabilities of correct responses by Test Time and L1 for the EOIT model without the control groups	189
15 paveikslas. Šios disertacijos tyrimo dizaino schema	276
16 paveikslas. Modelio prognozuotos teisingo atsakymo tikimybės pagal testo momentą ir grupę (įrašymo testo duomenys, neįskaitant kontrolinės grupės).....	277
17 paveikslas. Modelio prognozuotos teisingo atsakymo tikimybės pagal testo momentą ir grupę (paskatinto žodinio pakartojimo testo duomenys, įskaitant kontrolinę grupę).....	279

18 paveikslas. Modelio prognozuotos teisingo atsakymo tikimybės pagal testo momentą ir grupę (paskatinto žodinio pakartojimo testo duomenys, neįskaitant kontrolinės grupės).	280
--	-----

INTRODUCTION

This dissertation aims to explore the effectiveness of *Task-Based Language Teaching* (TBLT) for the teaching of a complex grammatical category, namely, grammatical case in Lithuanian as a second language (L2). TBLT is a language teaching methodology in which an L2 is taught through learners actively engaging in real-world communicative tasks. Although TBLT has been recognized as an effective methodology for L2 instruction (R. Ellis et al., 2019), it is often assumed that TBLT is unsuitable for the teaching of morphologically complex L2s (Gilabert & Castellví, 2019; Markina, 2018). Such languages and their complex grammatical categories, such as case, are thought to require extensive form-focused instruction. Therefore, the highly communicative approach of TBLT, in which grammar instruction primarily plays a supportive role, is thought not to suffice. This applies especially to beginner learners (Kogan & Bondarenko, 2022, pp. 78–79). However, to mitigate these challenges, a viable alternative could be *Task-Supported Language Teaching* (TSLT): an instructional approach derived from TBLT that allows for preemptive grammar instruction before task completion. As only limited research has examined the potential applicability of TBLT to teaching morphologically complex L2s and the possible advantages of more extensive grammar instruction through TSLT for such languages, this dissertation aims to challenge these assumptions.

The debate regarding the most effective approach to L2 instruction has persisted since the inception of second language acquisition (SLA) as an academic discipline. TBLT is one such notable approach. It underscores the utilization of the L2 in authentic, real-world communicative contexts, ranging from ordering a cup of coffee at a café to submitting your tax declaration. These real-world tasks are adapted for classroom use into pedagogical tasks. While, in the literature, many definitions for tasks exist, one of the most prevalent ones is that suggested by R. Ellis et al. (2019, p. 10), which specifies that TBLT tasks should include a focus on meaning; contain an information gap, initiating interaction between students, or students and the input; have students rely on their own linguistic and non-linguistic communicative capabilities; and have a clear communicative non-linguistic outcome. Tasks generally follow a clear structure, comprising three stages, ensuring students focus on meaning during task completion (R. Ellis, 2003; J. Willis, 1996). The first stage is the pre-task stage, which includes activities that introduce the task and/or topic, as well as relevant L2 input and activities related to it. Only then is the main task completed, which may be preceded by a planning phase.

Finally, students move on to the post-task stage, in which they discuss task results and, crucially, the grammar relevant to the task.

The four task criteria and the three-stage task structure outlined above have been conceived to achieve TBLT's core aim: fostering implicit L2 acquisition (R. Ellis et al., 2019). Implicit knowledge is the type of knowledge used intuitively by its user, which is often difficult to explain, e.g., that of a person's L1. At the other end of the spectrum lies explicit knowledge, the type of knowledge that is often difficult to use intuitively but a person can explain, e.g., an L2 grammar rule. Traditional teaching methods, such as *Present-Production-Practice* (PPP), have often focused mainly on the acquisition of explicit L2 knowledge (R. Ellis et al., 2019; Long, 2014). Such synthetic teaching approaches divide the target L2 into different units (e.g., collocations, grammar rules), sequence them in a particular order, and present them to learners one by one. It is then the learner's task to synthesize these separate units to communicate effectively in the target L2. Long (2014) calls this a focus on formS, but there has been extensive criticism of these purely form-focused approaches, which foster mainly explicit L2 knowledge. In response to these criticisms, the communicative approach and TBLT emerged to foster not primarily explicit but mainly implicit L2 knowledge (R. Ellis et al., 2019).

According to Long (2014), it is implicit, and not explicit, learning that is the default learning mechanism in adults. Consequently, analytical teaching methods seem far more suitable for L2 teaching, as they have learners engage with L2 input and peers, motivating them to analyze L2 input and output to induce constructions and rules. While implicit L2 learning proceeds considerably more slowly than explicit learning (Long, 2014), it is argued that the resulting knowledge can be applied automatically and is more resistant to forgetting (Paradis, 2009). However, full focus on meaning poses challenges of its own. As Long (2014) argues, while it is the default mechanism, adults are nevertheless constrained in their ability to learn L2s implicitly.

Therefore, Long (2014) suggests another analytic teaching option called 'focus on form'. While the focus of teaching should remain primarily on meaning, it should be complemented by some form of explicit teaching, namely, by a focus on form. Focus on form constitutes brief, form-focused episodes that address communication difficulties learners encountered during meaningful interaction. This can be further complemented by additional focus on formS in the post-task stage, i.e., explicit form-focused exercises, as it allows learners to keep their focus on meaningful interaction during the pre-task and main task stages (Skehan, 1998; D. Willis & Willis, 2007), and subsequent form-focused activities address relevant or attested problems with

form during task completion (Long, 2016). This combination of primarily meaning-focused interaction and supporting explicit form-focused instruction forms the basis of the cognitive-interactionist model that informs TBLT (R. Ellis et al., 2019; Long, 2014). Research within the domain of L2 pedagogy increasingly indicates that TBLT may prove more effective than conventional teaching methods (R. Ellis et al., 2019).

Nevertheless, morphologically complex languages, such as Russian, Finnish, and Lithuanian, continue to pose significant challenges for proponents of TBLT. The acquisition of morphosyntactic constructions is often regarded as one of the most demanding aspects of L2 learning (Slabakova, 2019), particularly in languages with intricate morphological systems, where the large number of constructions makes the process even more complicated (N. C. Ellis, 2022; Savickienė, 2006). For these reasons, many instructors of morphologically complex L2s tend to favor more traditional, explicit methods of teaching forms (Comer, 2012; Gilabert & Castellví, 2019; Markina, 2018). There is even a prevailing assumption that TBLT may be unsuitable for teaching these languages, particularly at beginner levels (Kogan & Bondarenko, 2022).

Among these challenges, grammatical case represents a particularly demanding domain. It is complex and frequently morphologically rich, and is considered one of the biggest challenges for learners of many L2s, such as Lithuanian (Savickienė, 2006). Research on L2 German and Russian has shown that L2 case acquisition does not occur instantly but progresses through a series of stages (Baten, 2013; Cherepovskaia et al., 2022; Magnani & Artoni, 2015). Cherepovskaia et al. (2022) argue that it is both the necessity of a case (i.e., the range of contexts in which a case is useful) and its morphological complexity (i.e., the range of morphological forms attributed to a case) that affect the L2 acquisition process.

This complexity, characterized not only by the wide range of forms but also by their frequent lack of salience, a low degree of contingency, and potential redundancy in relation to learners' L1 (N. C. Ellis, 2022), has led many teachers and instructors of morphologically complex languages to doubt the efficacy of communicative approaches such as TBLT and their primary aim of fostering communicative competence implicitly. Considering the consensus of researchers (N. C. Ellis, 2006, 2009; Frigo & McDonald, 1998) that implicit learning is mainly beneficial for learning adjacent dependencies and frequent chunks, as well as for mastering so-called fuzzy rules, but only after learners have been exposed to large amounts of L2 input, these doubts seem at least partially justified. Instead, such instructors tend to favor more

traditional, explicit methods of teaching forms (Comer, 2012; Gilabert & Castellví, 2019; Markina, 2018), despite the limitations that explicit learning and synthetic syllabi possess (Long, 2014). As learners tend to favor meaningful linguistic markers over functional markers when attending to L2 input (VanPatten, 2014), this might be a valid choice.

Nevertheless, some studies cast doubt on the efficacy of explicit teaching methods. According to Pienemann's Processability Theory (1998), intense explicit grammar instruction cannot be overtly beneficial to the teaching and learning of morphologically complex L2s, as all L2 learners go through the same developmental trajectories. There is no possibility to skip to a subsequent stage if learners are not developmentally ready, regardless of extensive grammar instruction. Supposedly, the same applies when learners' L1s are typologically similar, as L1 similarities can only speed up progression from one stage to the next (Saturno, 2023). Finally, previous research (Kempe & Brooks, 2008; Kempe & MacWhinney, 1998) suggests that learners employ not rule-based but, instead, item-based strategies when learning case, thus effectively suggesting that L2-input-heavy approaches such as TBLT could be more favorable for case acquisition. Therefore, it remains unclear whether explicit teaching approaches are more effective for teaching morphologically complex L2s.

In this context, where there is a potential need for more extensive form-focused instruction for the teaching of morphologically complex L2s, the instructional alternative of *Task-Supported Language Teaching* (TSLT) is particularly relevant. Ellis (2009b) has described the possibility of creating so-called focused tasks, which are pedagogical tasks that adhere to all the aforementioned criteria but target specific grammatical constructions and include preemptive grammar focus in the pre-task stage. TSLT syllabi are mainly comprised of these focused tasks. Nevertheless, critics (e.g., Long, 2014) argue that placing a greater emphasis on grammar throughout the task sequence contrasts with TBLT's primary objective of encouraging implicit learning. Importantly, there have also been concrete suggestions on how to adapt TBLT to the teaching of morphologically complex L2s (Gilabert & Castellví, 2019; Kogan & Bondarenko, 2022; Udes, 2025b). For instance, Gilabert & Castellvi (2019) argue that for the teaching of morphologically complex languages, perhaps form-focused instruction should be present in all task stages, not just the post-task. TSLT seems to align with this suggestion.

Clearly, such alternative approaches to TBLT raise questions about the kind of L2 knowledge they ultimately foster. Importantly, the main question is whether including a greater explicit grammar focus in earlier task stages, e.g., a 'more-TSLT' approach, has a measurable effect on the type of

L2 knowledge acquired, or, more specifically, whether it impedes the acquisition of implicit knowledge in favor of explicit L2 knowledge. The TSLT approach aligns well with Skill Acquisition Theory (DeKeyser, 2014), according to which declarative knowledge (i.e., explicit knowledge of a construction or a rule) may be proceduralized and, perhaps, partially automatized, thereby passing off as implicit knowledge. Conversely, one may ask whether the more conventional TBLT approach supports primarily implicit knowledge acquisition or also enhances explicit knowledge of the L2. Previous research (Kenanidis et al., 2023) has shown that incidental learning of, for instance, case is highly unlikely unless such implicit learning is accompanied by incidental attainment of metalinguistic knowledge. However, this is precisely why TBLT is traditionally equipped with focus-on-form techniques alongside post-task focus-on-forms activities, which means that the more traditional TBLT method might suffice for acquiring even relatively complex grammatical categories.

Despite these concerns, empirical research on the use of TBLT for morphologically complex L2s remains limited. This dissertation aims to address this issue through a study of the teaching and learning of case in L2 Lithuanian after the provision of TBLT- and TSLT-aligned tasks. In this context, it aims to address two key issues: whether TBLT can facilitate the acquisition of a complex grammatical category, such as case, and whether learning gains differ across treatments. In addition, it seeks to ensure lasting progress that transcends immediate learning gains. Markina (2018) showed that L2 Russian students who received TBLT treatment demonstrated immediate learning gains in case acquisition. However, at the delayed posttest, the participants' performance reverted to the baseline. Achieving lasting learning gains for case seems especially challenging in light of Denissenko's (2016) study, which showed that a 150-hour PPP intervention did not yield any significant progress for case marking at four different levels (A1, A2, B1.1, B1.2), thus suggesting that formal instruction alone may not suffice for successful case acquisition. To the best of the author's knowledge, there are few other interventional studies specifically targeting case marking, and none appear to have been conducted for L2 Lithuanian.

In addition to the overall learning outcomes, the dissertation also considers qualitative differences in L2 case development. Denissenko (2016) and Cherepovskaia et al. (2022) have shown that analyses of case error type can shed light on differences in developmental patterns across groups, even when the overall number of errors remains stable. These types of errors are erroneous use of the unmarked base form in contexts where the noun needs to

be marked for case; assignment of an oblique case, i.e., any case besides the nominative base form, in an inappropriate context; case marking on the noun with an ending not belonging to the paradigm of said noun. Dabašinskienė and Čubajevaitė (2009) found that overuse of the nominative, mix-ups of certain oblique cases, and paradigm and noun mismatches are common L2 Lithuanian errors as well. Clearly, error types could explain developmental differences in this study on L2 Lithuanian, which is why they have also been included.

Consequently, the dissertation aims to answer the following research questions:

- RQ1: Does limiting grammar instruction to the post-task stage promote both implicit *and* explicit learning gains related to grammatical case?
- RQ2: Does placing greater emphasis on grammar throughout the task cycle enhance both explicit *and* implicit case-related learning gains?
- RQ3: Can TBLT and TSLT treatments lead to lasting learning gains in grammatical case?
- RQ4: Does a difference in the emphasis on grammar instruction in the treatment lead to any discernible L2 developmental differences concerning grammatical case?

Based on the literature reviewed above, the following hypotheses have been formulated:

- Hypothesis 1a: Participants in both treatment groups will achieve learning gains in explicit knowledge of case.
- Hypothesis 1b: Participants in the TSLT group will outperform those in the TBLT group in terms of explicit knowledge acquisition.

Any form of intervention will most surely lead to certain learning gains. Therefore, while the TBLT group will most likely see some progress in explicit case acquisition, the comparatively limited grammar focus throughout the task cycle will likely make TBLT participants less receptive to the full range of different case forms. By contrast, the TSLT group will benefit from the increased grammar instruction in all task stages and improve more strongly.

- Hypothesis 2a: Participants of both treatment groups will achieve learning gains in implicit knowledge of case.
- Hypothesis 2b: Participants in the TBLT group will outperform those in the TSLT group in terms of implicit knowledge acquisition.

Again, both treatment groups should achieve certain implicit learning gains, regardless of the treatment. However, since TBLT is considered more conducive to implicit L2 acquisition, participants in this group are expected to perform better on tests of implicit knowledge after the treatment. Conversely, the TSLT group may develop a less attuned intuition for case assignment due to the heavier focus on grammar throughout the task cycle, which encourages accuracy over fluency.

- Hypothesis 3: Both treatments lead to significant lasting learning gains for grammatical case.

While the literature provides little evidence for this, achieving lasting learning gains should be possible by combining sufficient formal instruction with a narrow focus on the target structures. Denissenko (2016) examined L2 acquisition of the Russian nominal case system as a whole. Markina (2018) focused on L2 Russian prepositional case government, which involves five different cases. By targeting a specific function and a select number of cases, it may be possible to achieve significant long-term gains, provided the treatment is long enough, and the participants are exposed to ample L2 input to provide sufficient exemplars for implicit processing and potential abstract rule-building.

- Hypothesis 4: While the TBLT group will mainly reduce case assignment errors, the TSLT group will have fewer problems with paradigm-related errors.

This prediction rests on the assumption that the focus on communicative competences of the TBLT treatment may highlight the functional value of the relevant cases during task performance. Furthermore, case assignment rules are generally considered complicated, which is a feature that benefits more from implicit learning with sufficient exposure than from explicit learning, which is more conducive to the acquisition of relatively simple grammar rules. However, minimal attention to morphological forms may lead TBLT

participants to use only the most frequent case paradigms, thereby slowing the acquisition of paradigms. Conversely, since the TSLT group will be more exposed to an explicit focus on morphological forms, they may considerably reduce paradigm mistakes (more on this, see Section 3.3). Nevertheless, this heightened focus on morphological form may impede their progress in case assignment, especially since explicit learning mechanisms are less effective for learning complex grammar rules.

Aim and objectives. The central aim of this dissertation is to establish potential learning gains of L2 course participants on two test instruments measuring case knowledge after receiving two different types of TBLT-aligned treatment. These vary in the placement and extent of explicit grammar instruction, i.e., TBLT (explicit grammar instruction limited to the post-task stage) vs. TSLT (grammar focus present throughout the task cycle). The following tasks have been formulated to achieve this aim:

1. Define the central tenets of the TBLT methodology, specifically, the role of explicit grammar instruction and other task design principles and instruments relevant to the teaching of L2 morphology.
2. Review current research on the acquisition of case in L2 acquisition in order to establish the state of knowledge in this domain.
3. Develop a teaching intervention in the form of a TBLT-aligned L2 Lithuanian course in two versions that differ in the place and extent of explicit grammar instruction.
4. Design test instruments that assess both explicit and implicit L2 knowledge of case.
5. Teach an L2 Lithuanian course with two groups of participants, who will complete the test instruments designed for this study at three specific time intervals to measure their gains.
6. Collect, annotate, and subject the data to appropriate statistical analyses.
7. Interpret the results, refine the hypotheses if necessary, and identify the implications of the findings for the field.

Novelty of the dissertation: To the best of the author's knowledge, this is the first empirical intervention study employing a pretest-posttest-delayed posttest design to investigate adult acquisition of L2 Lithuanian. Previous research has included theoretical discussions of the teaching and acquisition of L2 Lithuanian (Buivydienė & Žukienė, 2007; Pribušauskaitė et al., 2009; Savickienė, 2005, 2006), qualitative error analyses (Dabašinskienė & Čubajevaitė, 2009; Dubasava, 2020), cross-sectional studies on naturally

acquired L2 Lithuanian (Ramonaitė, 2015, 2017, 2020), and test administration following L2 instruction which was not part of the study (Bružaitė-Liseckienė, 2021; Mačiukaite, 2008; Stepšys & Kamandulytė-Merfeldienė, 2024). Furthermore, although previous publications have briefly mentioned the TBLT methodology (Zygmantas, 2011), no empirical studies have yet examined its effectiveness in teaching L2 Lithuanian. The treatment in this dissertation was specifically designed to address the corresponding research questions and incorporates both a between-subjects design (two treatments and a control group) and a repeated-measures design (three testing intervals).

In the broader context of TBLT and the teaching of morphologically complex L2s, a category to which the Lithuanian language is generally assigned, this dissertation represents only the second empirical intervention study. So far, there have been primarily theoretical considerations of applying TBLT to the teaching of such languages (Gilabert & Castellví, 2019; Janowska, 2020; Kogan & Bondarenko, 2022). The only empirical intervention study known to the author is that of Markina (2018) on L2 Russian, which reported only minimal long-term retention for case acquisition. Clearly, empirical research on the application of TBLT to morphologically complex L2s is lacking, and the present study may make a significant contribution to the field. Through a relatively long-term intervention (18 hours) provided to two treatment groups that differ in the amount of explicit grammar instruction, the dissertation seeks to shed further light on the potential efficacy of TBLT for the teaching of morphologically complex L2s, particularly grammatical case acquisition. If successful, the dissertation may have important theoretical and practical implications. Not only would it strengthen the theoretical basis of TBLT, i.e., the cognitive-interactionist model (R. Ellis et al., 2019; Long, 2014), but it would also provide substantial guidance for future task design in TBLT, especially since the study incorporates task design recommendations from Gilabert & Castellví (2019).

Besides TBLT, the dissertation may inform research on case acquisition in morphologically complex L2s more broadly. While numerous studies have examined the acquisition of morphologically complex L2s (see Section 2.2.2 for a selection), more ecologically valid intervention studies on case in L2 Russian (Denissenko, 2016) yielded very limited results. Potential short-term and long-term learning gains of this study may be explained by and also inform relevant SLA theories, such as Input Processing (VanPatten, 2014), the Associative–Cognitive CREED (N. C. Ellis, 2006), Processability

Theory (Pienemann & Lenzing, 2020), and Skill Acquisition Theory (DeKeyser, 2014).

Theses to be defended

1. Restricted focus on form, i.e., brief reminders of form in response to communication difficulties, as well as explicit focus on forms reserved for the post-task stage (i.e., the TBLT treatment) suffices for successful acquisition of explicit case knowledge. Continuous focus on grammar throughout the task cycle (i.e., the TSLT treatment) does not necessarily lead to superior acquisition of explicit case knowledge.
2. By allowing only occasional focus-on-form episodes at the pre-task and main-task stages and explicit focus on forms at the post-task stage (i.e., the TBLT treatment), interaction during task completion remains focused on meaning-making, allowing for implicit case-learning gains in grammatical case. Continuous focus on form throughout the task cycle (i.e., the TSLT treatment) may impede the acquisition of implicit case knowledge.
3. Teaching grammatical case in a morphologically complex L2 by means of TBLT-aligned tasks leads not just to immediate but even to durable learning gains, provided the length of instruction is sufficient, and the target constructions are well-defined.
4. The placement and intensity of explicit grammar instruction in the task cycle have little bearing on developmental patterns in terms of error types.
5. Typologically similar L1s potentially provide only short-term effects on L2 case learning, but, overall, their influence on L2 case acquisition is limited.
6. Teaching grammatical case in a morphologically complex L2 by means of TBLT-aligned tasks leads to implicit learning gains even in adult L2 learners, as hypothesized by the cognitive-interactionist model. Explicit grammar instruction should be limited to the post-task stage.
7. Grammatical case in a morphologically complex L2 can be successfully taught by means of TBLT-aligned tasks, even for low-proficiency learners.

Data and methodology. The main research phase, consisting of test administration and the implementation of the treatment with the experimental groups ($n = 45$), was conducted between November 4, 2024, and February 4, 2025. Additional tests with the control group ($n = 15$) were administered from May to August, 2025. The participants were non-university-affiliated L2

learners of Lithuanian who registered for the course via an application form which was made available on the Vilnius University Department of Lithuanian Studies website. Although participants were not formally tested for their L2 Lithuanian proficiency, the application form specified that the target level of the course was between the A1 and A2 levels according to the *European Framework of Reference for Languages* (CEFR; Council of Europe, 2020). The participants' performance during the course confirmed that they had prior knowledge of Lithuanian and the appropriate language level for this course. The A1–A2 level was chosen for several reasons:

- As learners at this level have already mastered some of the basics, this allowed for freer task development, particularly with regard to spoken interaction.
- Case opposition is thought to not occur immediately in L2 case development, going through certain developmental stages before the nominative base form is juxtaposed to other oblique cases (see Section 2.2.1.3). Therefore, the decision was made to include participants with prior L2 Lithuanian knowledge, while they were still at a level where case oppositions were not yet fully mastered, especially implicitly.
- The hypotheses of this dissertation clearly distinguish between explicit and implicit knowledge. Implicit L2 use by participants with a near-zero proficiency level can hardly be expected, given the very limited exposure to the L2. Therefore, in order to assess these different types of knowledge, participants who already had some prior knowledge of L2 Lithuanian were selected (for more information on the participants, see Section 3.2).

The participants were randomly assigned to the TBLT treatment (i.e., task versions with explicit grammar instruction solely at the post-task stage) and the TSLT treatment (i.e., task versions with grammar focus throughout the task cycle). These groups were further subdivided into four groups in total to make the teaching process more manageable. The treatment length, which coincides with the course length, was 18 hours for each group. The author of this dissertation served as the teacher for all these groups.

The materials used in the treatment, i.e., the L2 Lithuanian course, were created specifically for this dissertation. They consisted of five pedagogical tasks, i.e., real-life tasks turned into three-stage TBLT-aligned tasks for classroom use, deemed communicatively relevant to recent migrants to a new country. Two versions of each task were created, differing in the placement and extent of explicit grammar instruction (for more on the course design, see Section 3.4.1). The materials were tested with several of the

author's own L2 Lithuanian students, as well as students participating in L2 Lithuanian courses at VDU (Vytautas Magnus University) Education Academy in Vilnius. This led to several modifications of the materials to enhance their pedagogical effectiveness.

Test instruments were designed to assess explicit and implicit knowledge. Making this distinction in SLA studies is paramount, since explicit grammar instruction is generally assumed to primarily foster explicit L2 knowledge, which, according to many theoretical perspectives, such as in TBLT and its cognitive-interactionist model, does not automatically transfer into implicit knowledge. TBLT is also said to mainly foster implicit knowledge. Consequently, any study evaluating the efficacy of explicit grammar instruction, particularly in a TBLT context, must include this distinction (for more on the cognitive-interactionist model, see Section 1.2.2).

The participants completed both tests at three time points: before the treatment (pretest; to establish the baseline), immediately after the treatment (immediate posttest; to assess immediate learning gains), and again after a 1.5-month delay (delayed posttest; to assess long-term learning gains). The tests were administered within the same treatment groups, yielding four test administrations per test time (i.e., pretest, immediate posttest, and delayed posttest). The tests were primarily administered by the author, although the tests assessing implicit knowledge were administered with the help of colleagues at Vilnius University. If the participants could not attend a group session for test administration, the author met with them on-site or online to administer the tests. Likewise, test administration with the control group was conducted individually, twice with a 4- to 5-week break, either online or on-site with the author. Figure 1 provides an overview of the entire study design. Participant responses to the test instruments were annotated and entered into *MS Excel*. Statistical analyses on the data were performed in *R* (R Core Team 2013). The data were analyzed primarily by fitting Bayesian mixed-effects models using the *brms* package (Bürkner, 2017). Responses to one of the test instruments were additionally analyzed for error types (for more on study design, see Section 3.1; for the test design, see Section 0; on data annotation, see Section 3.6; on Bayesian mixed-effects models and the statistical methods employed in this study, see Section 3.7).

Initial versions of the test instruments were trialed in a pilot study conducted at VDU Education Academy in Vilnius. It was carried out with two separate groups, both of which completed a pretest and a posttest and received one task, which was used in the final study. One group participated in January 2024 and received a TBLT version of the task, while the other group

participated in May 2024 and received a TSLT version of the task. Based on the results, the test instruments were refined for use in the main study to improve their effectiveness in assessing participants’ explicit and implicit case knowledge.

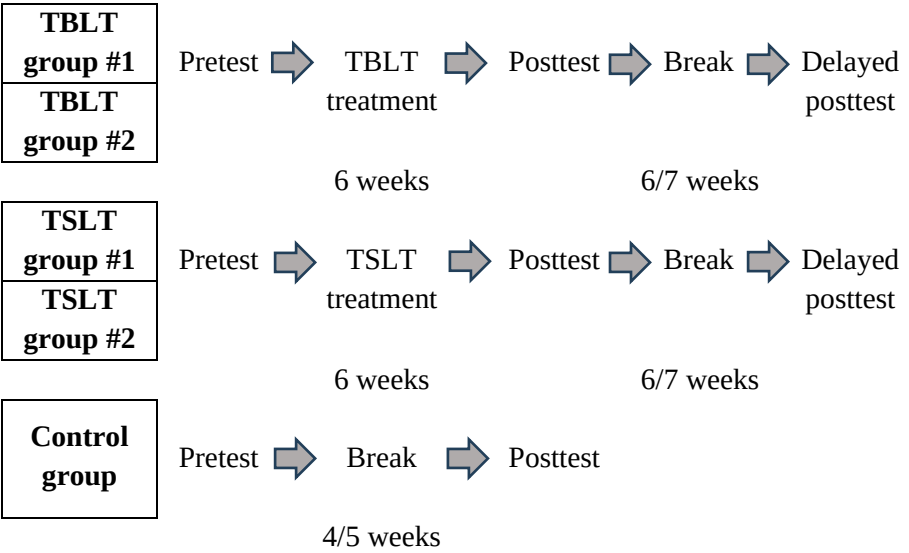


Figure 1. An overview of the study design in this dissertation

Structure of this dissertation. The first two sections provide an overview of the scientific literature pertaining to the research questions and hypotheses of this dissertation. Section 1 focuses on the TBLT methodology, outlining its psycholinguistic underpinnings, task design principles, and considerations relevant to the teaching of morphology. Section 2 addresses case: it provides a brief overview of the Lithuanian case system and discusses theories and findings on its acquisition in the L2. Relevant previous research on L2 acquisition of case is also expanded upon. Section 3 is dedicated to the experiment conducted for this dissertation; it describes the study design, its participants, treatment, test design, transcription procedures, statistical analyses, and the rationale for these. Section 4 presents the results of the statistical and error-type analyses of the data. Finally, Section 5 revisits the hypotheses, discusses the implications of the findings, and provides limitations and suggestions for further research. This is followed by the conclusions, which summarize the study and its primary findings.

Theoretical and practical value of the research. This dissertation's findings have several valuable theoretical and practical implications for the field of SLA. Most importantly, the findings indicate that TBLT can be effective in teaching a morphologically complex L2, particularly in teaching a complex grammatical category such as case. Such courses can lead to even implicit and long-lasting learning gains. These findings will surely inform further research on TBLT in the context of morphologically complex L2s, as well as on SLA in general, on L2 case acquisition. From a purely theoretical perspective, the findings support the cognitive-interactionist model that informs TBLT and have certain implications for other SLA theories. From a practical perspective, the teaching materials of this dissertation can serve as examples of how to teach a morphologically complex L2 through TBLT for other studies and language courses.

Glossing conventions. Interlinear glosses follow the Leipzig Glossing Rules unless otherwise noted. In this dissertation, grammatical information is provided explicitly only for case marking and negation markers. Nominal endings are assumed to be singular unless otherwise indicated. Additional categories are glossed only wherever relevant for the analysis.

Use of AI tools. In writing this dissertation, the author used several AI tools. *Grammarly* (<https://grammarly.com/>) and *ChatGPT* (<https://chatgpt.com/>) were employed to improve written texts for grammar, spelling, and stylistic clarity. *ChatGPT* was also used to help develop and debug code in *R*. All output of the AI tools was thoroughly reviewed and scrutinized, and, wherever necessary, modified by the author. The responsibility for the accuracy of the final text and the code rests solely with the author of this dissertation.

1 TASK-BASED LANGUAGE TEACHING

The first section presents the central tenets of TBLT. It covers not just the central features by which the methodology is considered effective in teaching L2s, but also the disagreements that still exist in the literature, primarily concerning the place of grammar instruction. First, the methodology is briefly introduced. This is followed by a discussion of its theoretical foundation, concerning the target type of L2 knowledge and the role of grammar instruction. An alternative to TBLT is then introduced that more heavily emphasizes grammar: TSLT. The discussion then continues with task design principles. This includes the tools that are typically employed in the design of tasks to integrate grammar instruction, as well as two central models for task design by (preemptively) assessing task complexity. As the role of grammar instruction in the TBLT literature still causes debate, especially in the context of morphologically complex L2s such as Lithuanian, the discussion concludes with suggestions for applying TBLT to the teaching of L2s containing complex morphological systems.

1.1 The central tenets of TBLT

TBLT arose from the need voiced by educators and researchers in the 1980s for a new view on L2 teaching: learners' use of language should not be judged based on its grammatical correctness but on whether the communicative outcome of a task has been achieved (Johnson, 1982). It has its origins in Communicative Language Teaching (Howatt, 1984), which emphasizes communicative intent and fluency over the traditional synthetic approach of foreign language instruction based on an inventory of grammatical structures. Although communicative tasks were also used within Communicative Language Teaching, they never became the central element of the L2 syllabus. From the mid-1980s onward, the first comprehensive descriptions of the TBLT methodology were published (e.g., Long, 1985; Nunan, 1989; Prabhu, 1987), outlining the advantages of teaching an L2 through tasks. According to Nunan (1989), positioning the 'task' as the fundamental unit of instruction can effectively integrate syllabus content (what is taught) with methodology (how it is taught). Additionally, Prabhu (1987, p. 1) argues that tasks create conditions that enable learners to actively engage in communication, which he considers essential for the development of L2 competencies.

TBLT thus promotes realistic use of the target L2 in authentic communicative contexts through specially designed tasks. This creates

pedagogically valid opportunities for learners not merely to learn about the L2 but primarily to learn how to use it. Tasks can range from everyday activities, such as ordering a coffee in a café or visiting a family doctor, to more specific activities tailored to learners' needs, such as writing a summary of an economics article or conducting a routine vehicle check at a border crossing.

However, defining what constitutes a 'task' is not straightforward. Van den Branden (2006) identified no fewer than seventeen different definitions of the term. A definition that can be implemented relatively easily is provided by R. Ellis et al. (2019), who outline four fundamental criteria:

1. Focus on meaning: The task structure should ensure that learners prioritize comprehension and the expression of ideas in the L2.
2. Presence of a 'gap': The task should include an informational 'gap' that motivates learners to communicate, whether to exchange information, to reason, or to express opinions.
3. Reliance on learner resources: While performing the task, learners must draw on their own linguistic and non-linguistic resources. The task itself does not prescribe any specific linguistic forms.
4. Clear communicative goal: The success of the task is measured in terms of achieving the communicative goal rather than linguistic accuracy.

Together, these criteria ensure that any TBLT task involves L2 learners' interaction that arises (relatively) naturally and is judged solely on the achievement of the intended outcome, not on the quality of their L2 output. Importantly, any L2 teaching activity must meet these four criteria to be considered a fully TBLT-aligned task. If not, the activity is not a task, but rather an exercise or, at the most, 'task-like' (R. Ellis et al., 2019, p. 10).

To ensure L2 learners' focus on meaning, TBLT tasks most often follow a standard structure that prioritizes interaction. It typically follows the framework proposed by Willis and Willis (2007), which consists of three stages:

- Pre-task: The topic is introduced, and input is provided through reading and/or listening activities.
- Main task: Learners carry out the main task, which may be preceded by a planning phase to facilitate task performance.
- Post-task: The outcomes of the task and grammar relevant to it are discussed.

Essentially, the format follows the order of preparation, performance, and evaluation. The pre-task primarily prepares learners to achieve the communicative goal of the main task. This ensures that learners are fully

focused on that intended outcome during the first two stages. Only afterward, in the post-task stage, is the performance evaluated, and suggestions are given as to how to improve this L2 performance, and additional practice is provided. Crucially, it is only here that explicit grammar instruction in the task cycle is traditionally situated, so as not to distract learners from the focus on meaning.

As an example, consider the beginner-level L2 Brazilian Portuguese task *Reconhecendo sintomas de gripe* ‘Recognizing the symptoms of influenza’.¹ The communicative goal of this task is for students to describe the symptoms of a foreigner who has arrived in Brazil and is feeling unwell to a healthcare professional. Fellow students then evaluate whether the task has been completed successfully. The structure of this task can be described as follows:

- Pre-task: Students begin by sharing their knowledge of common symptoms of various illnesses, health issues, and personal experiences. This is followed by a vocabulary exercise in which they associate pictures with the names of flu symptoms. Next, they receive input in the form of a short story about people depicted in images and their symptoms. Finally, students independently describe the symptoms associated with the different illnesses presented in another set of images.
- Main task: Each student is randomly given a set of pictures depicting various symptoms, according to which they must describe the symptoms of a foreign visitor in Brazil to a medical professional.
- Post-task: Students discuss the broader context of this kind of situation. They share what they know about flu treatments, describe how the illness is treated in their home country, interview three classmates, and reflect on their own experiences with flu treatment in Brazil. This can be expanded with additional grammar practice comprising the subjunctive and grammar structure in questions (Gurzynski-Weiss, 2023, p. 199).

As can be seen, the task follows the conventional three-stage task structure with a main task that contains a clear communicative outcome and an information gap (i.e., one of the students in each pair did not know the foreign patient’s symptoms). The pre-task and main task stages do not include any explicit form-focused instruction, thus ensuring the expression of meaning

¹ The complete version of this task is available on the Indiana University TBLT Language Learning Task Bank website: <https://tblt.indiana.edu/tasks/details.html?id=43>.

during these stages and reliance on the learners' own L2 resources, which may have been enriched by the L2 input.

Unsurprisingly, TBLT has not been accepted by everyone. With its primary focus on meaning, a common critique on TBLT is that it does not teach 'new language' as there is essentially no room for grammar in the methodology (Sheen, 2003; Swan, 2005). This misconception most likely derives from the standard task format described by Willis (1996). She contends that teachers should not interfere during task performance, arguing that concerns about grammatical accuracy naturally arise as learners share their task outcomes. Hence, relevant grammatical structures should be addressed during the post-task phase. However, it is important to note that task syllabi comprise not one but several tasks and, thus, several post-task stages. Therefore, grammar learning can still occur incrementally with each post-task stage. Importantly, however, the role and importance of grammar instruction are still being debated in the TBLT literature. For instance, in contrast to Willis (1996), Long (2000, pp. 45–46) argues that attention to grammatical forms should be briefly given even during task performance. Additionally, some scholars (e.g., Littlewood, 2007; Shehadeh, 2012) advocate for including grammar instruction in the pre-task stage, which is an approach that Willis (1996) and Long (2014) strongly oppose.

1.2 Theoretical foundation of TBLT

1.2.1 Implicit and explicit learning and knowledge

Regardless of their views on the role of grammar in L2 instruction, the proponents of TBLT generally aim to foster implicit rather than solely explicit knowledge. This distinction is crucial to the theoretical foundation of TBLT, as the central aim is not primarily to learn about the L2 but to learn how to use it. Implicit knowledge is knowledge that people possess but cannot consciously explain or may not even be aware of having. It is deployed automatically, much like a person's knowledge of their L1. Explicit knowledge, on the other hand, is knowledge that people are consciously aware of having and can therefore explain. Learners can often verbalize L2 grammar rules by drawing on their explicit knowledge, while native speakers are often unable to do so because their L1 knowledge is largely implicit. This example illustrates the paradox of explicit knowledge and performance: while L2 learners might be able to explain such rules, this does not necessarily translate into stable performance, whereas the opposite is true for L1 users. Implicit

knowledge is often a product of implicit learning, i.e., learning without awareness of what is being learned. On the other hand, explicit knowledge is a product of explicit learning, i.e., learning with awareness and the conscious search for some underlying patterns or structure when attending to L2 stimuli. Most readers will likely agree that although many L2 textbooks emphasize explicit learning to foster explicit L2 knowledge, most L2 learners would rather achieve the fluent, stable performance that implicit L2 knowledge can provide. This is the primary goal of TBLT (R. Ellis et al., 2019).

Luckily, contrary to popular understanding, Long (2014) argues that implicit learning remains the default mechanism through which L2 learners acquire language. Importantly, studies of non-linguistic learning tasks in adults have shown that performance improvements always precede the ability to verbalize newly acquired knowledge (Berry & Broadbent, 1984; Stanley et al., 1989). Doughty (2003) claims that the same pattern applies to SLA, while N. C. Ellis (2005) maintains that most acquired L2 knowledge, even in adults, is tacit, automatic, and implicit. Moreover, L2 learners mostly struggle with exactly the type of L2 features one would expect to be problematic if implicit learning were the default way of learning, e.g., complex grammatical structures. Artificial L2 grammar learning and complex-systems studies have further shown that learners are capable of detecting regularities and probabilities without awareness, and that generalizations can occur when input is consistent and low in exceptions. As a result, rule-like behavior can emerge from statistical learning (Berry & Broadbent, 1984; Reeder et al., 2009, 2010; Stanley et al., 1989).

These findings clearly support a usage-based view of L2 acquisition. In this perspective, language learning emerges from learners' experience with the target L2 through comprehension, production, and interaction. Through associative learning processes, learners store exemplars (i.e., specific, concrete instances of experience with the L2, such as words or phrases) and gradually abstract recurring patterns across them. These abstracted L2 patterns then become more schematic, almost rule-like, representations or constructions, which learners can apply to novel contexts (N. C. Ellis, 2006, 2008b; Goldberg & Casenhiser, 2008; Tomasello, 2003).

Nevertheless, although implicit learning is possible for adult L2 learners, Long (2014) emphasizes that its capacity is more limited than that of children. Studies by Rebuschat and Williams (2009) and Williams and Kuribara (2008) suggest that adults can implicitly acquire some syntactic and form-meaning regularities and generalize them to novel strings. However, the resulting L2 knowledge they demonstrate often appears fragmentary and

probabilistic, relying substantially on surface chunks (e.g., bigrams and trigrams) rather than fully abstract rule systems. Furthermore, while implicit learning of non-adjacent dependencies appears possible, adult learners experience greater difficulty acquiring such dependencies than adjacent ones (Newport & Aslin, 2004; Wilson et al., 2020). Taken together, these findings suggest that adult implicit learning may be most effective for adjacent dependencies and frequent chunks, such as collocations and local patterns (N. C. Ellis, 2005, 2009; Frigo & McDonald, 1998). At the same time, under conditions of extensive exposure, implicit learning helps infer vague rules from complex material that includes non-salient underlying patterns. Consequently, implicit learning might be less conducive to learning long-distance dependencies, abstract grammatical categories, and morphophonological marking such as gender and, importantly, case, unless L2 learners are exposed to large amounts of L2 input to deduce abstract rules from exemplars and prototypical patterns.

In light of these findings, Long (2014) argues that adult implicit learning needs to be complemented by explicit learning to reshape L2 learners' implicit processing and make it more time-efficient. Importantly, explicit learning should not be the driving force behind L2 acquisition but should aid implicit methods in order for learners to notice low-salient forms and form–meaning connections (R. W. Schmidt, 1990), as well as mismatches between their output and the target L2 (notice-the-gap; R. W. Schmidt & Frota, 1986). Crucially, Long (2014) claims that L2 learners are often entrenched in their automatic L1 processing routines, which makes it difficult for them to attend to L2 input in such a way that learners are sensitive to linguistic features that are irrelevant to the L1. Therefore, Long and N. Ellis claim that explicit knowledge helps re-tune adult L2 learners' 'radios' to a new frequency, after which, following the initial crackling reception, they can return to listening to their broadcast with minimal effort. To quote Long (2014, p. 50) directly:

explicit learning (not necessarily as a result of explicit instruction) involves a new form or form–meaning connection being held in short-term memory long enough for it to be processed, rehearsed, and an initial representation stored in long-term memory, thereafter altering the operation of the way additional exemplars of the item in the input are handled by the default implicit learning process.

In other words, explicit learning can retune which L2 cues a learner's implicit system tracks. As a result, more forms can be noticed and held in working

memory, making it more likely that the input becomes intake and is encoded in long-term memory (Chaudron, 1985; Gass, 1997). Consequently, future input is processed differently. Nevertheless, explicit learning has its own constraints. Studies show that explicit learning is most effective for simple, salient, low-variability patterns and is constrained by learners' developmental readiness (Pienemann, 1998), meaning it cannot foster acquisition that is not aligned with a learner's interlanguage and processing capacity.

In summary, while implicit learning is the default mechanism in L2 acquisition even for adults, it is weaker in adulthood than in childhood, and, on top of that, it is structurally constrained. Therefore, explicit learning is crucial for directing attention to low-salient L2 features and restructuring implicit input processing. This applies especially to such grammatical features as morphology and long-distance dependencies. Hence, instruction needs to primarily facilitate implicit learning while also incorporating explicit processes to enhance implicit processes and more efficiently aid in reaching implicit L2 knowledge.

1.2.2 The cognitive-interactionist model

TBLT aims to achieve this interaction between implicit and explicit learning through the cognitive-interactionist model. The theoretical foundation for this model is Long's (1996) Interaction Hypothesis and the Interaction Approach (S. M. Gass & Mackey, 2014), which was derived from it. While input is crucial for successful L1 and L2 acquisition, learning tends to be more effective and progress more quickly when the input is more closely aligned with the intake, i.e., the subset of that input a learner notices and holds in working memory for further acquisition (Corder, 1967). Long (1981) sought to bridge this gap by proposing an interactional approach, which holds that "conversation is not only a medium of practice; it is also the means by which learning takes place" (Gass, 1997, p. 104). For instance, an L1 speaker or a more advanced L2 speaker can provide more complex, sophisticated L2 input, guiding the learner's attention toward elements that are relevant to the situation. This scaffolding process is essential for language learning, facilitating the learner's engagement not only with vocabulary but also with broader L2 linguistic structures, such as grammar and morphosyntax (N. C. Ellis, 2008a). A large meta-analysis (Keck et al., 2006), spanning across 25 years of research, revealed significant positive effects of interaction on the acquisition of both lexical and grammatical items in both immediate and delayed posttests. Moreover, tasks in which the target linguistic form was

essential for successful completion yielded stronger effects than those in which it was merely useful but not obligatory.

Building on these pedagogical benefits of interaction, TBLT-aligned tasks are designed to stimulate interactions that activate the cognitive mechanisms contributing to successful L2 acquisition (R. Ellis et al., 2019). As discussed above, although adults have a more limited capacity to learn an L2 implicitly compared to children acquiring an L1, Long (2014) argues, based on studies by Rebuschat and Williams (2009) and Leung and Williams (2011), that implicit learning nonetheless remains the primary mode of L2 acquisition for adults as well. While explicit learning in TBLT plays a role, its purpose is comparatively limited. It helps learners notice complex grammatical forms and overcome entrenched L1 routines. Therefore, the objective of explicit instruction is not to transmit explicit knowledge *per se*, but, rather, to reshape the way learners process L2 input, thereby refining the implicit processes that are essential to acquisition. Central to this approach is the role of ‘attention’, and, in particular, the concept of ‘noticing’, i.e., the conscious awareness of linguistic features. For example, in L2 English, a learner interacting through input processing or corrective feedback may recognize that adding ‘s’ to a noun indicates a plural form, thereby noticing the gap in their current interlanguage state (R. W. Schmidt & Frota, 1986). The learner might then deduce the rule that “to indicate ‘more than one’, add an -s to the noun”. Figure 2 illustrates this cognitive-interactionist model of language learning:

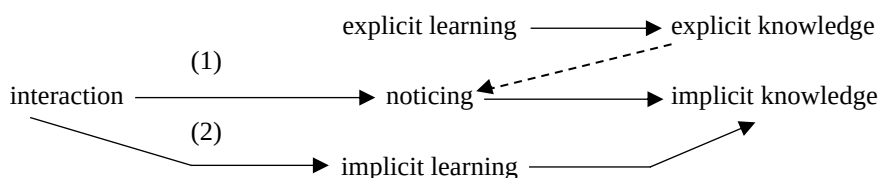


Figure 2. The cognitive-interactionist model (adapted from Ellis et al. (2019, p. 31))

As learners interact and process input, learning proceeds along two ‘pathways’. The first pathway, depicted in Figure 2, leads to implicit knowledge but does so indirectly. Explicit knowledge influences learners’ tendency to notice linguistic features, which, in turn, contributes to the development of implicit L2 knowledge. Importantly, the relationship between explicit knowledge and noticing is not direct (hence the dashed line in Figure 1) and is mediated by various external factors (e.g., teaching strategies) and

internal factors (e.g., learners' capacity to notice grammatical forms). For this reason, explicit instruction alone is insufficient for acquiring implicit knowledge within this model. This mode of learning is referred to as 'incidental learning', and it is argued that adult learners mainly acquire an L2 in this manner (R. Ellis et al., 2019). The second pathway in the figure refers to implicit learning, whereby learning occurs without conscious attention to linguistic features, meaning that implicit knowledge is acquired without 'noticing'. However, as noted earlier, adults' capacity to acquire language in this manner is limited.

1.2.3 Focus on form: integrating analytic and synthetic teaching approaches

Clearly, tasks aim to provide situated language contexts for L2 learners so that they can interact with interlocutors and/or the input, which, in turn, is supposed to lead to implicit L2 knowledge. This may – or may not – involve noticing interlanguage deficits influenced by the learner's acquired explicit knowledge. However, providing opportunities for meaningful interaction is not the only way language is taught in TBLT. This is mostly done through the implementation of 'focus on form', a language-teaching method introduced by Long (1991) to address the shortcomings of other methods.

Long (2014) distinguishes between synthetic and analytic language-teaching approaches, which differ in their presumed roles for the learner. In the case of the former, the target L2 is the point of focus, which is divided into a variety of units (e.g., collocations, grammar rules) and sequenced according to a selection of criteria (e.g., valency, frequency). These linguistic units are then presented to the learner one by one, which makes it the learner's job to synthesize the items for successful communicative use of the L2. *Present-Production-Practice* (PPP) is a clear example of a language teaching method that employs a synthetic approach. Analytic approaches, by contrast, focus on the learner and learning processes. Through exposure to authentic L2 input and gradual engagement in meaningful target L2 production, learners are tasked with analyzing their input and output and, in so doing, inducing rules of grammar and use. Hence, in contrast to synthetic approaches, in which the target L2 is the object of instruction, analytical approaches focus on the message that the learner means to comprehend and/or convey.

Teaching through synthetic syllabi, synthetic teaching materials, etc., Long (2014) refers to as 'focus on formS'. He argues, however, that several of its teaching assumptions do not hold. First, this teaching method assumes

that learners can learn anything deemed important by the teacher or the syllabus whenever it is presented, regardless of the learners' interlanguage. This contrasts with findings on learnability (a learner's ability to learn a linguistic unit), which must guide teachability (the appropriateness of teaching a specific linguistic unit to learners) as described by Pienemann (1998). Second, focus on forms can be considered a process of skill-building, as presented by DeKeyser (2014): first, explicit declarative knowledge is fostered, which, through intensive use, is then restructured into procedural knowledge and stored in long-term memory. Further immense amounts of practice lead to this knowledge being automatized. While such automatization seems possible (Day & Shapson, 1994; Harley, 1998; Lyster, 1994; Muranoi, 2000), the process is time-consuming and thus quite inefficient. Moreover, while automatized procedural knowledge can be mistaken for implicit knowledge, many researchers maintain that explicit and implicit knowledge are separate systems and that explicit declarative knowledge cannot be converted into implicit knowledge (DeKeyser, 2014; Krashen & Scarcella, 1978; Long, 2014; Paradis, 2009). The cognitive-interactionist model underlying TBLT (see Figure 2) generally assumes a weak-interface position: explicit learning and knowledge can affect implicit processing of L2 input and output, but cannot be directly converted into implicit knowledge. Since fostering implicit knowledge is the central goal of TBLT, this poses a problem for teaching through synthetic syllabi.

Language teaching that involves analytic syllabi, analytic teaching materials, etc., Long (2014) calls 'focus on meaning'. Here, language teaching is centered not on linguistic forms but on the message that an interlocutor or a text aims to convey. This approach has a few advantages over 'focus on forms', since it eliminates the need for a strict schedule of linguistic units to be presented to learners and presents them with non-simplified and natural L2 input. Nevertheless, according to Long (2014), 'focus on meaning' is likely insufficient as well. As mentioned above, adults do not have the same capacities as children to learn an L2 solely implicitly. Furthermore, it is time-consuming, lacks negative evidence to motivate learners to notice gaps in their interlanguage, and, in its purest form, includes no explicit instruction, which has been found to be conducive to L2 acquisition.

In order to remedy the problems of the synthetic focus-on-forms and analytic focus-on-meaning approaches, Long (2014) suggests a third language teaching option: focus on form. Although it remains primarily an analytic approach, focus on form additionally emphasizes form in context. It is the reactive use of any type of pedagogic procedure to briefly draw learners'

attention to an L2 difficulty (be it lexical, grammatical, or otherwise) they are facing while carrying out a task. In short, it should be:

- reactive: emerging as a response to learner errors only.
- incidental: occurring during meaning-based or communicative activities.
- brief: so as not to distract from the communicative goal of the task.

This way, while brevity keeps the focus on the communicative message, focus on form also attends to learners' need for negative evidence and their interlanguage readiness. Additionally, as the learner is likely invested in the exchange, they are more motivated to attend to feedback from the teacher or peer, and because they have already considered the meaning to be conveyed, they can devote more attentional resources to the L2 problem in question. As Long (2014, p. 27) himself puts it: "Focus on form capitalizes on a symbiotic relationship between explicit and implicit learning, instruction, and knowledge". While the jury is still out on which language teaching method is ultimately more effective, meta-analyses by Norris and Ortega (2000) and Goo et al. (2015) have found superior L2 acquisition efficacy for explicit instruction over implicit instruction, while Norris and Ortega (2000) showed very comparable treatment effects for 'focus on forms' and 'focus on form', with a slight edge for the 'focus on form' treatment, suggested by Long (1991) and often employed in TBLT.

Nevertheless, a longstanding debate in the literature concerns the extent and place of focus on form within the task cycle. In contrast to Long (2014), Willis and Willis (2007) argue that even brief attention to form during task performance may cause learners to shift their focus away from the task's communicative goal and toward linguistic structures. Skehan (1998) and R. Ellis (2009b) take the opposite approach, arguing for a preemptive focus on form during the pre-task stage. As R. Ellis (2024) puts it, this should not include explicit instruction but rather guided planning for the main-task stage. Notably, the least controversial stage in the task cycle for explicit grammar instruction, i.e., focus on forms, is the post-task stage. This way, learners can concentrate on the expression of meaning at the pre-task and main-task stages (Skehan, 1998; D. Willis & Willis, 2007), and any linguistic problems encountered during these stages can be addressed afterward in the post-task stage, making them relevant to both the task and the L2 learners (Long, 2016).

1.2.4 Task-supported language teaching

Importantly, there exists an alternative approach to TBLT that takes preemptive explicit grammar instruction even further. Preemptive instruction can be integrated into task design, leading to so-called ‘focused tasks’ (R. Ellis et al., 2002). These tasks are designed to elicit specific linguistic constructions while maintaining focus on the intended communicative outcome. Although Long (2016) and Skehan (1998) have argued against such tasks, R. Ellis (2003) maintains that this format enables learners to direct their attention effectively to targeted grammatical forms. This broader interpretation of R. Ellis’s focus on form is referred to as *Task-Supported Language Teaching* (TSLT) and is often regarded as a weaker version of TBLT (R. Ellis, 2003). In response to challenges in applying TBLT to local teaching contexts, TSLT is sometimes advocated as a viable alternative. For instance, Littlewood (2014) argued that communicative language teaching approaches, such as TBLT, do not fit the Chinese culture of learning, leading him to promote the use of TSLT.

TBLT and TSLT differ in the following three key ways (R. Ellis, 2024, pp. 7, 11):

- Syllabus design: In TBLT, syllabus design is centered on the target tasks students are expected to complete; in TSLT, syllabus design is guided by the linguistic units to be taught during the course, and tasks are simply the pedagogical devices employed to achieve L2 learning.
- The place of explicit instruction: TBLT tasks generally do not include preemptive explicit focus on form(s) at the pre-task stage. TSLT tasks, on the other hand, include explicit presentation of target linguistic units at the start of lessons.
- Focused and unfocused tasks: TBLT courses include mainly unfocused tasks (although Ellis allows only limitedly for focused tasks), while TSLT tasks consist solely of focused tasks.

As Ellis notes, the main difference between the two approaches essentially is the implementation of focus on form during task performance: in TSLT, it is directed at the target linguistic items, while in TBLT, it is directed at any linguistic problems arising from interaction.

While limited, these changes reflect very different theoretical positions. As described above, TBLT takes a cognitive-interactionist view that emphasizes holistic and gradual learner development through interaction. TSLT, by contrast, is a return to Skill Acquisition Theory (DeKeyser, 2014). According to this view, declarative knowledge acquired through instruction

proceeds to the procedural stage and, finally, to the automatic stage. For this reason, R. Ellis (2024) claims that TSLT cannot be considered a viable alternative to TBLT, as TSLT will most likely foster explicit L2 knowledge, while TBLT primarily enhances implicit L2 knowledge. This also led him to believe that both approaches cannot be integrated, although there is room for a modular approach.

It might seem that TSLT is simply a return to the traditional PPP method. However, R. Ellis (2024) emphasizes that the practice stage in PPP need not occur in TSLT. Explicit presentation of target forms can immediately be followed by an appropriate focused task that aims to elicit their use, bringing it closer to TBLT than PPP. Furthermore, in a modular approach, target forms for TSLT tasks would be selected based on consistent L2 problems arising from TBLT tasks, making explicit teaching reactive rather than proactive in the PPP tradition. Nevertheless, the present dissertation adopts an interpretation of TSLT that, to a certain degree, integrates the practice stage of PPP into task design. This decision was made to provide a starker contrast between the TBLT method and more traditional teaching methods still often employed by teachers of morphologically complex L2s.

Finally, R. Ellis (2024) mentions that the question remains whether the placement of explicit instruction at the pre-task stage influences subsequent task performance. He argues that explicit focus on form in the initial task stages essentially signals to learners that the lesson's purpose is not communicative but the target structure, which may lead them to focus not on the intended communicative task outcome but on the linguistic items presented in the explicit instruction during task performance. The attention allocated to form may well impact performance, potentially enhancing learners' accuracy but diminishing their fluency or complexity. Despite limited research in this field, Ellis cites Ellis, Li, and Zhu (2019), who did find a significant impact on how the task was performed after prior explicit instruction had been provided. Hence, many TBLT proponents remain somewhat skeptical of the TSLT approach. Long (2014), for instance, dubs it 'lower-case' tblt as opposed to true 'upper-case' TBLT, as it represents a return to synthetic teaching approaches. Nevertheless, even he admits it is "worthy of consideration" (p. 7), as it could serve as a bridge between traditional synthetic syllabi and 'real' TBLT teaching. This seems particularly relevant for the teaching of morphologically complex L2s, for which teachers still often favor explicit teaching methods (Comer, 2012; Gilabert & Castellví, 2019; Markina, 2018).

1.3 Task design principles

1.3.1 Tools available for teaching grammar and morphology

Despite ongoing discussion about the role and extent of grammar instruction, it still has a place in task design. This section provides an overview of tools within the TBLT literature that can be employed to teach grammar in TBLT-aligned tasks.

1.3.1.1 Pre-modified input

The TBLT literature describes several techniques designed to emphasize grammatical forms in the L2 input and enhance their noticeability for learners. These include:

- Input flooding: increasing the frequency of the target grammatical item within the input. The rationale is that the heightened presence of the grammatical element enhances its salience, thereby increasing the likelihood that the learner will notice it. When applying this technique, it is important to keep the input sounding natural and realistic (R. Ellis et al., 2019).
- Input enhancement: visually highlighting grammatical elements within the input, typically through typography such as varying fonts, capitalization, or color (e.g., ‘cars’ ‘carS’). This visual emphasis is intended to draw learners’ attention to the target forms (R. Ellis et al., 2019).

Although some studies have not demonstrated significant effects (e.g., Harley, 1998; Izumi, 2002), many others have shown that these strategies can lead to improved learning outcomes (e.g., Lee & Huang, 2008; Oh, 2001; White, 1998). Still, Ellis (2003) argues that successful L2 acquisition requires both positive and negative input. Nevertheless, since morphological markers are said to be difficult for L2 learners to notice (N. C. Ellis, 2022; VanPatten, 2014), these strategies might be valuable for teaching L2s with rich morphological systems.

1.3.1.2 Post-task strategies

As is most likely clear by now, the post-task stage is often considered the most suitable task stage for addressing grammatical forms in TBLT. The main form-focused strategies for this stage include:

- Consciousness raising: learners are presented with L2 material, from which they collaboratively work out grammatical patterns in pairs or groups. Ideally, this activity should focus on the grammatical elements encountered during the task that proved problematic during performance (R. Ellis, 1997).
- Processing instruction: based on VanPatten's (2014) theory of Input Processing, this strategy aims to foster form-meaning associations. For instance, in the context of L2 Lithuanian, to emphasize the distinction between the accusative and dative cases, learners might be presented with sentences such as:

- (1) *Tėtis duoda knygą dukrai*
dad gives book.ACC daughter.DAT
'Dad gives the book to his daughter.'
- (2) *Tėtis duoda dukrai knygą*
dad gives daughter.ACC book.DAT
'Dad gives his daughter the book.'

Learners then match the sentences to the corresponding images.

- Grammatizing: students fill in missing linguistic forms in a given text (D. Willis & Willis, 2007). The text should already be familiar to the learner, so that they do not have to process new information.

Clearly, these three strategies of grammar instruction differ in their focus on comprehension and production: while Consciousness raising and Instruction processing have learners employ only inductive reasoning to arrive at general grammar rules, Grammatizing also has learners produce forms, potentially leading to comprehension of the grammar rules involved. For this reason, this last strategy seems particularly useful for practicing not just grammar rules but also morphological endings, especially when there are many of them.

1.3.1.3 Corrective feedback

Given the controversy surrounding the role of grammar in the TBLT methodology, the least controversial solution appears to be providing corrective feedback. This feedback can be categorized along two dimensions: 'implicit/explicit' and 'input-providing/output-prompting' (R. Ellis et al., 2019, pp. 52–53). For instance, when a learner makes an error, the L2 teacher can respond in the following ways (R. Ellis et al., 2019, pp. 36–37):

- Clarification request (implicit/output-prompting): The teacher asks a question regarding the incorrect part without specifying where the

error occurred, prompting the learner to reconsider their response. This happens in Example (3), where the student mispronounces the word *kavq* ‘coffee.ACC’.

- (3) *„Aš kasdien geriu gavq.“ → „Kq geri?“
 ‘I drink goffee every day’. → ‘What do you drink?’
- Providing metalinguistic cues (explicit/output-prompting): The teacher offers an invitation to reconsider the utterance, guided by metalinguistic knowledge. This happens in (4), where the noun *pienas* ‘milk’ is used in the nominative but is supposed to be marked with the accusative (*pienq*).
- (4) *„Aš kasdien geriu pienas.“ → „Gerti kq.“
 I every.day drink milk.NOM drink what.ACC
 ‘I drink milk every day.’ → ‘To drink what.’
- Recasts (implicit/input-prompting): The teacher provides a corrected version of the learner’s utterance, such as in (5). The learner can repeat the correct form, though it is not mandatory.
- (5) *„Aš kasdien geriu pienas.“ → „Tu kasdien geri *pienq*.“
 I every.day drink milk.NOM you every.day drink milk.ACC
 ‘I drink milk every day.’ → ‘You drink milk every day.’
- Explicit correction (explicit/input-prompting): The teacher directly corrects the error in the sentence, such as in (6). Again, the learner may repeat the correct sentence, but this is not required.
- (6) *„Aš kasdien geriu pienas.“ → „Ne *pienas*, o *pienq*.“
 I every.day drink milk.NOM not milk.NOM but milk.ACC
 ‘I drink milk every day.’ → ‘Not milk, but milk.’

As for its effectiveness, research suggests that providing corrective feedback is more beneficial than offering none at all (Lyster & Saito, 2010). Both input-providing and output-eliciting forms of corrective feedback contribute positively to L2 acquisition, with the latter generally yielding stronger effects (e.g., Sato & Loewen, 2018; Yang & Lyster, 2010). Additionally, while both implicit and explicit forms of corrective feedback are valuable, explicit feedback tends to be more effective (R. Ellis et al., 2019).

Nevertheless, in TBLT, the type of feedback should also align with task design. This means that, during the pre-task, the teacher should primarily provide implicit, input-providing feedback (e.g., recasts) to keep the focus on meaning and input processing. However, at the main-task stage, the teacher should provide a mix of implicit (e.g., clarification requests) and explicit (e.g., metalinguistic cues) output-prompting feedback, which serve as focus on form (Long, 2014). Finally, the post-task stage is the ideal time for the teacher to provide explicit feedback, both output-prompting and input-prompting (e.g., explicit corrections).

1.3.2 Models designed to assess conceptual task difficulty

When designing tasks and determining their expected complexity for learners, researchers often draw on two influential models: Skehan's (2015) *Limited Attentional Capacity* (LAC) approach, and Robinson's (2007, 2015) *Triadic Componential Framework* (TCF), which is formulated in line with his *Cognition Hypothesis* (CH). Both models aim to assess task complexity by examining how various task components influence L2 learner performance. In the TBLT literature, L2 performance is generally evaluated according to four parameters: structural complexity, i.e., the degree of complexity in the syntactic structures produced by the L2 learner; lexical complexity, i.e., the range and diversity of vocabulary used by the L2 learners, including the extent to which infrequent lexical items are employed; accuracy, i.e., the degree to which L2 utterances are grammatically correct; fluency, i.e., the degree to which L2 learners can express their thoughts smoothly and efficiently, relative to L1 speakers. Skehan and Robinson present different perspectives on how task design and its associated conceptual demands might influence these four performance areas. At a general level, Skehan (2015) asserts that increased task complexity leads to trade-offs between the aforementioned parameters in L2 learners' performance, whereas Robinson (2015) claims that rises in complexity actually lead to improvements across multiple performance dimensions.

More specifically, in Skehan's LAC approach (2015), attentional resources and working memory capacity are limited. As task demands increase, learners must prioritize aspects of performance, leading to trade-offs between fluency and form, as well as within form, between complexity and accuracy. More demanding tasks may therefore improve one area of performance at the expense of another. One of the main aims of the LAC approach is to establish generalizations based on specific task features that can

be employed to predict how learners perform on tasks with those features, thereby enhancing our understanding of the L2 acquisition process (R. Ellis et al., 2019, p. 71). Several such generalizations have been proposed, including that the requirement to organize or integrate complex information within a task raises learners' L2 complexity, and that the need to use infrequent or complex vocabulary within a task significantly reduces both learners' L2 accuracy and fluency. These generalizations have been successfully explained by the model in the LAC approach (Tavakoli & Skehan, 2005; Skehan & Foster, 1997, 1999). Importantly, in each instance, either one or two parameters are raised, but never all simultaneously. However, Skehan (2009) argues that the so-called joint raising of both complexity and accuracy is possible when the model is carefully considered in task design. For instance, this has been shown in narrative tasks (raises complexity) with a clear structure (raises accuracy and fluency). However, from a pedagogical design perspective, while these generalizations can inform task design, they do not provide a coherent framework for creating multiple tasks. Moreover, as noted by R. Ellis et al. (2019) as well, the LAC approach has very little to say about syllabus design and task sequencing.

By contrast, while Skehan's model is valuable for explaining performance trade-offs, Robinson's framework is more directly oriented toward task design and sequencing. As noted earlier, Robinson's perspective on learner resources differs from Skehan's. While Robinson acknowledges that working memory capacity is limited, he argues that attentional resources are not subject to the same constraints. Following Wickens (2007) and Sanders (1998), Robinson maintains that it is more appropriate to conceptualize attention in terms of 'resource pools', and that attentional resources can be 'expanded' to address all necessary aspects when required. Crucially, task requirements should not compete for the same resource pools, so tasks must be carefully designed to avoid such overlap. If successful, the increased task complexity then enhances both the complexity and accuracy of L2 learners' performance, thereby allowing gains across multiple performance parameters simultaneously.

Structurally, the TCF is intended to ensure that tasks are designed with a gradual increase in cognitive complexity while avoiding competition for the same resource pools. Robinson's framework consists of three dimensions that influence complexity:

- Task Complexity: design-related variables that generate cognitive demands.
- Task Conditions: participation-related variables that shape interactional demands.

- Task Difficulty: ability-related variables associated with individual learner differences.

Although the domains of task conditions and task difficulty offer significant opportunities for research (e.g., exploring the relationship between reasoning requirements across task complexity and task difficulty), the process of designing and sequencing tasks is guided primarily by the variables related to task complexity (R. Ellis et al., 2019). For this reason, only task complexity is discussed in more detail here.

Task complexity (see Table 1) determines the cognitive demands of a task and is hypothesized to affect students' L2 performance through the CALF (Complexity, Accuracy, Lexical Variety, and Fluency) parameters. As shown in the table, task complexity variables are divided into two groups: resource-directing and resource-dispersing. Resource-directing variables are believed to prompt learners to focus more closely on the L2 itself, which leads to both accuracy and complexity being raised, as learners' attention shifts toward functional demands, thereby increasing the likelihood of noticing, receiving corrective feedback through interaction, and engaging in negotiation for meaning. As outlined in previous sections, these factors are considered significant in facilitating L2 learning. In contrast, resource-dispersing variables are not directly tied to language or its functional demands and, therefore, do not influence the interplay between accuracy and complexity. While these variables still impact learner performance, their effects are primarily cognitive rather than linguistic (Robinson, 2007a).

Table 1. The make-up of the Task Complexity component with the Triadic Componential Framework for task classification (based on Robinson, 2007a, pp. 15–16)

Task complexity (Cognitive factors)	
Resource-directing variables (making cognitive/conceptual demands)	Resource-dispersing variables (making performative/procedural demands)
+/- here and now	+/- planning time
+/- few elements	+/- single task
+/- spatial reasoning	+/- task structure
+/- causal reasoning	+/- few steps
+/- intentional reasoning	+/- independence of steps
+/- perspective-taking	+/- prior knowledge

It is worth considering some examples of these factors and related research. Starting with a resource-dispersing variable, planning, for instance, is believed to positively impact students' L2 performance by allowing them to consider how to execute a task, i.e., what to say and how to say it (e.g., Damos, 1991; R. Ellis, 2005; Skehan, 2014). This phase helps to free up cognitive resources during task performance, thereby enhancing students' L2 performance in terms of accuracy, fluency, and complexity. However, Robinson notes that simply offering or withholding the option to plan does not, by itself, alter the linguistic output, unlike resource-directing variables, such as causal reasoning. Hence, planning is classified as a resource-dispersing variable within the framework. Another example of a resource-dispersing variable is the requirement to perform two tasks simultaneously rather than just one. Clearly, it is easier for an L2 learner to answer a phone call than to answer a call while also monitoring a television screen for specific information (Damos, 1991). The latter scenario, again, increases task difficulty but does not draw additional attention to linguistic detail. Overall, resource-dispersing variables may result in reduced L2 performance across the CALF parameters (Robinson, 2022).

By comparison, resource-directing variables draw learners' attention to the language itself, resulting over time in more accurate and complex output, albeit often at the expense of fluency. For example, intentional reasoning requires learners to consider the motivations of other agents, which increases the need to attend to complex L2 syntax (e.g., 'he suspected that ...', 'she was not sure if ...' in L2 English; Ishikawa, 2008; Robinson, 2007b, 2012; Tomasello, 2008). Similarly, spatial reasoning provides learners with opportunities to focus on the target L2, specifically on constructions used to describe movement. Such variables can also encourage learners to compare their L1 with the target L2. As Robinson and Ellis (2008) argue, this process encourages L2 learners to "rethink for speaking".

Although R. Ellis et al. (2019) rightfully note that the empirical support for the TCF and CH remains mixed – little evidence has been found for the joint raising of complexity and accuracy – they also acknowledge the framework's value as a systematic basis for task design. By carefully managing the resource-directing and resource-dispersing variables, the intended conceptual complexity of a task can measurably be gauged. This is especially valuable for creating not just one but multiple tasks in a sequence. This sequencing requires the gradual increase of cognitive load and follows two principles (Robinson, 2022). According to the first principle, increases in cognitive load must occur through systematic manipulation of both resource-

directing and resource-dispersing variables. The second principle specifies that resource-dispersing variables should be intensified first, followed by resource-directing variables.

Based on these principles, Robinson derived the 3-step SSARC model for the sequencing of tasks:

- Step 1: SS (Stabilize, Simplify) – The task aims to consolidate learners' existing language skills. At this stage, no additional demands are placed on either resource-dispersing or resource-directing variables.
- Step 2: A (Automatize) – Resource-dispersing variables are increased to promote more efficient use of cognitive resources and foster automatization of L2 processing.
- Step 3: RC (Restructure, Complexify) – Resource-directing variables are introduced to ‘disrupt’ the learner’s existing interlanguage. Task demands should encourage interlanguage restructuring and the development of new form-meaning mappings.

These three steps aim to guide learners from consolidating existing knowledge, through automatization, to reorganizing and increasing the complexity of their interlanguage. Examples of task sequences based on the SSARC model are presented in Table 2. Although the SSARC model is relatively recent and thus lacks extensive empirical validation, its potential to inform research and test hypotheses related to L2 acquisition is promising (Levkina, 2014). Notably, Malicka (2020) reported improved accuracy, fluency, and structural complexity among L2 learners who performed tasks in a sequenced, SSARC-based order, compared to those who completed tasks in isolation. Despite limited research on the SSARC model, it clearly provides a coherent method for creating syllabi with a gradual increase in conceptual complexity.

1.3.3 Proposals for designing tasks with morphologically complex L2s

The reader may have noticed that both the LAC approach and the Triadic Componential Framework focus primarily on the cognitive load imposed by task design and the types of interaction that stem from it, i.e., conceptual task complexity. However, little to no attention is paid to the linguistic challenges posed by the specific constructions featured in the tasks. This lack of linguistic focus is particularly relevant when designing tasks for the teaching of morphologically complex L2s, such as Lithuanian, Russian, and Finnish. It is

Table 2. Potential sequences in the SSARC model (*Source:* Ellis et al., 2019, p. 83)

Step		Possible sequence 1	Possible sequence 2
1 (SS)	Low resource-dispersing demands	Structured	Planned
	Low resource-directing demands	Here-and-now	Few elements
2 (A)	High resource-dispersing demands	Unstructured	Unplanned
	Low resource-directing demands	Here-and-now	Few elements
3 (RC)	High resource-dispersing demands	Unstructured	Planned
	High resource-directing demands	There-and-then	More elements

therefore unsurprising that, despite the TBLT tools for teaching L2 morphology discussed in Section 1.3.1, concerns persist regarding the suitability of TBLT for teaching highly complex morphological L2s, especially at the beginner level (Kogan & Bondarenko, 2022).

Section 2 provides an in-depth overview of the acquisition of L2 morphosyntactic constructions, specifically case, and the inherent difficulties that L2 learners face in acquiring them. Given these severe difficulties, it is understandable that a strongly communicative approach such as TBLT is often met with skepticism. If the acquisition of morphology presents challenges even in analytic languages such as English, these challenges are likely to intensify when learning an L2 characterized by a wide range of morphosyntactic forms. For instance, within the Russian L2 teaching community, there remains a prevailing belief that explicit attention to form is the most effective method for teaching morphologically complex L2s, and that any communicative activity should serve only to reinforce this explicit L2 knowledge (Gilbert & Castellví, 2019). Moreover, certain psycholinguistic research suggests that teaching L2 morphology in languages with complex structures should begin with grammatical mastery, with more frequent and meaningful student interaction introduced later (Comer, 2012). This approach clearly contradicts one of the fundamental principles of TBLT, which emphasizes the acquisition of implicit knowledge through communicative use, even at the beginner stage. Consequently, the application of TBLT to teaching morphologically complex L2s needs further refinement.

The following section presents two proposals for teaching these languages in accordance with TBLT principles: those of Gilabert and Castellví (2019), and of Kogan and Bondarenko (2022). The discussion concludes with the author's comparison of these two models (Udes, 2025b).

1.3.3.1 Gilabert, Castellví – Gradual increase in conceptual task demands with manageable linguistic difficulty

According to Gillabert and Castellví (2019), the complexity of linguistic constructions, i.e., linguistic difficulty, must be a key consideration when designing tasks for morphologically complex L2s. They argue that the lack of attention to linguistic difficulty in the existing literature stems from its relative insignificance in more analytic languages, such as English. However, they assert that this factor cannot be overlooked for the teaching of morphologically complex languages such as Russian, where it significantly increases cognitive processing demands. Accordingly, Gillabert and Castellví have identified the following main points (pp. 534–535):

- Linguistic difficulty competes with conceptual demands, i.e., task design features.
- Linguistic difficulty contributes to the overall cognitive load of the task and directly impacts the successful completion of the task.
- Linguistic difficulty should be assessed not only by the number of grammatical features but also by the relative difficulty of each feature.

Hence, the cognitive load required for task completion is influenced by both linguistic difficulty and conceptual complexity. Consequently, the outcome, namely, task success, depends on learners' ability to manage both effectively. This presents a substantial challenge for task designers working with morphologically complex L2s: excessive linguistic difficulty can hinder learners' ability to perform tasks that would otherwise be manageable for learners of morphologically simpler L2s. To address this issue, Gilabert and Castellví recommend designing tasks that free up as much of learners' attentional and memory resources as possible, enabling them to allocate these resources more effectively to both the conceptual and linguistic aspects of the task.

Based on these theoretical assumptions, Gillabert and Castellví propose several guidelines for designing tasks aimed at teaching morphologically complex L2s (pp. 540–544):

1. Carry out a needs analysis, i.e., extract information about the types of tasks that course participants intend to perform outside the classroom.

According to Long (2014), TBLT curricula must always be based on the real-life tasks, the so-called ‘target’ tasks, that participants aim to perform in the target L2. Pedagogical tasks are created for the course accordingly. Gilabert and Castellví state that, when teaching morphologically complex languages, teachers must pay extra attention to the factors that contribute to task complexity and to the sources of difficulty in terms of the necessary linguistic resources.

2. Anticipate the types of linguistic expressions required, the difficulties these types pose, their underlying causes, and any other factors that contribute to task complexity.
3. Identify the ‘minimal code’ necessary for task completion, that is, the essential morphosyntactic forms without which the task cannot be completed.
4. Achieve a balance between cognitive complexity, conceptual complexity, and linguistic difficulty by modifying the task structure, e.g., by breaking a complex task into smaller, more manageable subtasks. This approach allows learners to build on prior experience, thereby reducing conceptual complexity.
5. Incorporate focus on form at all stages of the task. This should be implemented not only during the post-task stage but also during the pre-task stage, for example, through input enhancement and input flooding (see Section 1.3.1.1).
6. When designing courses for morphologically complex L2s, it is necessary to:
 - a. Gradually increase conceptual complexity while keeping linguistic difficulty at a ‘stable’ and ‘realistic’ level.
 - b. Include opportunities for task repetition so that L2 learners can shift their focus more towards form than meaning when repeating the same task.
 - c. Alternate task modalities (such as oral, written, and hybrid formats) to take advantage of the strengths of each modality (e.g., written tasks may free up more cognitive resources to concentrate on form).

In summary, task design must carefully consider linguistic difficulty to avoid overburdening L2 learners’ attention and memory resources. However, syllabus design should be primarily structured around a gradual increase in conceptual task demands, while ensuring that linguistic difficulty remains realistically manageable.

1.3.3.2 Kogan, Bondarenko – Engage learners with cognitively challenging solo-tasks

The second proposal by Kogan and Bondarenko (2022) primarily focuses on designing tasks for beginner L2 learners and offers a slightly different perspective on managing conceptual complexity. While the authors agree that both conceptual complexity and linguistic difficulty impact learners' task performance, they emphasize the importance of distinguishing between the two: limited linguistic competence does not necessarily reflect limited cognitive abilities in an L2. In fact, research shows that novice L2 learners often demonstrate mature L2 information-processing competencies (Richards, 2015). Therefore, contrary to Gillabert and Castellví's proposal, Kogan and Bondarenko contend that beginning with conceptually simple tasks and gradually increasing their complexity is not the most effective strategy. Instead, they argue that better results can be achieved by designing tasks that align with learners' interests and needs. When a task proves cognitively demanding, the teacher or fellow students can provide additional scaffolding. The researchers strongly believe that if beginners are only given conceptually simple tasks, "we deny our students the intellectual joy of meaningful learning" (p. 83).

To maintain tasks that are conceptually engaging yet manageable for beginner learners, Kogan and Bondarenko suggest revisiting the original definition of a 'task' within the TBLT framework. They point out that, while tasks were initially understood as any form of information processing that leads to a specific outcome (Prabhu, 1987), they gradually became more narrowly defined as verbal interaction (Nunan, 1989). By returning to the earlier definition and removing the requirement for verbal interaction, a wider range of tasks becomes possible, particularly for beginner learners. Solo tasks based on input are particularly suitable, as they encourage L2 learners to process relevant input and interact meaningfully with their environment without requiring verbal output.

Thus, Kogan and Bondarenko propose the following principles for designing beginner-level tasks:

1. Interaction with the environment: Learners should interact purposefully with their linguistic environment by using the L2 meaningfully.
2. Cognitive demands: Tasks should present a manageable challenge based on an information gap, encouraging them to engage their cognitive resources.

3. Teacher as a source of the L2: Instead of relying solely on peer interaction, the teacher can act as an interlocutor, providing both positive input and corrective feedback.
4. Task content based on students' needs: Course content should be guided by learners' interests, preferences, and communicative goals, rather than by the cognitive complexity of the tasks.
5. Measurable output: Each task should produce a measurable outcome, preferably linguistic in nature, such as a solution or a response that requires interaction with the linguistic input.

In summary, task design should focus on tasks that are cognitively challenging yet manageable, while addressing learners' specific L2 needs.

1.3.3.3 Comparing and integrating both proposals

The following discussion is primarily based on Udes (2025b). Clearly, the two proposals share several similarities. First, both Gilabert and Castellví (2019) and Kogan and Bondarenko (2022) emphasize the importance of centering task design on learners' interests. It is necessary to conduct a needs analysis (Long, 2014), as identifying learners' L2 needs makes it possible to anticipate potential conceptual and linguistic difficulties and design pedagogical tasks accordingly (Long, 2014). Both proposals also address incorporating different task modalities. While Kogan and Bondarenko go so far as to suggest that verbal communication can be omitted altogether, Gilabert and Castellví acknowledge that written tasks can free up valuable resources that students can then devote to linguistic form.

Both approaches also discuss task repetition, though their implementation methods differ somewhat. Kogan and Bondarenko suggest designing tasks that naturally prompt students to repeat actions (e.g., a market-related task in which questions about prices are repeatedly asked and answered). By contrast, the subtasks proposed by Gilabert and Castellví could serve a similar purpose, enabling learners to practice elements of the main task incrementally. Gilabert and Castellví also note that allowing students to repeat the same task can be beneficial; however, the author of this dissertation highlights a risk of reduced motivation. Instead of simply repeating the identical task, it is advisable to create a conceptually more challenging version, or a different task that maintains the same procedural approach but differs in content (i.e., procedural repetition; R. Ellis et al., 2019)).

The biggest difference between the two proposals concerns their view on the cognitive demands of tasks. Gilabert and Castellví suggest that

cognitive demands should be minimized at the initial stages, whereas Kogan and Bondarenko argue that cognitive challenges make tasks more engaging for learners. Thus, according to Kogan and Bondarenko, task design for beginner L2 learners should focus on eliminating the need for verbal interaction and providing adequate scaffolding during task performance. A potential solution to reconcile these differing approaches is to integrate both online and offline interaction. Such a course structure could include both tasks that require online interaction (e.g., verbal interaction) and those that rely on offline interaction (e.g., input processing). Because tasks involving online interaction are inherently more cognitively demanding due to the verbal communication required, their conceptual requirements should be minimized. Conversely, offline tasks can be made more conceptually demanding, allowing for greater emphasis on linguistic requirements.

However, it should be emphasized that such a rigid dichotomy need not always be applied. For instance, both online and offline interactions can be integrated within the same task, provided that the cognitive demands are carefully managed. Furthermore, the subtask method proposed by Gilabert and Castellví can effectively reduce cognitive load, thereby enabling beginner learners to perform the main task orally. In addition, the course content should reflect learners' preferences: whether they prefer to perform specific L2 actions (higher cognitive task demands), engage in verbal communication (greater oral interaction), or expand their L2 knowledge (conceptually less demanding input-based tasks with higher linguistic demands).

Another seemingly similar, yet substantively different, aspect concerns the role of linguistic requirements. Gilabert and Castellví focus primarily on identifying the 'minimal code' and assessing linguistic difficulty. They argue that linguistic difficulty is determined not only by the quantity of grammatical elements expected but also by the relative complexity of each grammatical element. They recommend that task design should primarily focus on conceptual complexity while keeping linguistic difficulty manageable. However, Kogan and Bondarenko rightly point out that insufficient knowledge and tools currently exist to accurately determine this relative difficulty. Gilabert and Castellví also acknowledge that little is known about L2 developmental sequences; therefore, it remains unclear which grammatical elements are inherently more challenging than others.

For this reason, Kogan and Bondarenko emphasize the significance of grammatical frequency over difficulty and argue that instruction should focus on the most frequent forms and rules, whether across the target L2 overall or within the specific task context. For instance, in Russian, the dative

case is predominantly used with pronouns (e.g., *mne* ‘1SG.DAT’; *nam* ‘1PL.DAT’) and certain verbs (e.g., *mne nnavitsya* ‘1SG.DAT please.3SG’, ‘I like it’; *mne dali* ‘1SG.DAT give.PST.PL’, ‘they gave to me’). Therefore, it is more practical to start with these common dative uses and later defer to other, less frequent contexts, even when those contexts are considered more prototypical for the case. Despite these differing perspectives, reconciling the two approaches should be relatively straightforward. It is possible to establish a minimum code for the task while simultaneously focusing on the most frequently occurring forms and constructions within it.

Nuss (2022) proposes the *Morphological Complexity Index* (MCI), as defined by Brezina and Pallotti (2019), as a means to assess linguistic difficulty. This index enables the evaluation of morphological diversity in both the input and the expected L1 outputs of tasks, thereby facilitating comparisons across task inputs and allowing predictions of the difficulty of the expected L2 output, particularly with respect to morphosyntactic constructions. However, this approach is limited to determining the difficulty of the grammatical forms and does not address the complexity of their functions. For instance, while it is plausible that encountering three different endings of the genitive case in a text is more challenging for a learner than encountering just two, this assessment provides no insight into the varying functions of the genitive case, the relative difficulty of each function, or how the genitive case compares in difficulty to other cases or grammatical elements from a broader perspective.

Another apparent similarity between the two proposals is the emphasis on incorporating a greater focus on form within the task sequence. Gilabert and Castellví advocate for a more subtle approach, employing form-focused strategies in the input (as discussed in Section 1.3.1.1), providing corrective feedback during the main task (as discussed in Section 1.3.1.3), and reserving direct form-focused strategies for the post-task stage (as discussed in Section 1.3.1.2). In contrast, Kogan and Bondarenko propose a more explicit method of drawing students’ attention to form through the use of ‘inter-tasks’. In the pre-task phase, the teacher may highlight or ask questions about the relevant lexis or grammatical structures pertaining to the task.² In this author’s view, caution is warranted when implementing such inter-tasks. Gilabert and Castellví, likewise, caution that designing subtasks around different grammatical elements risks reverting to an explicitly grammar-

² An example from Kogan and Bondarenko (2022, p. 94): *Focus on the ADJECTIVE + NOUN agreement in plural (a recently introduced pattern)*.

centered curriculum, and the same concern applies to inter-tasks. It remains uncertain whether this more direct form-focused approach could hinder students' acquisition of implicit L2 knowledge. Given the lack of research on this matter, it is advisable to limit inter-tasks to lexical items or useful constructions (templates) and to address grammatical elements only when prompted by students. All other form-related needs can be effectively addressed during the post-task stage.

In summary, integrating the task design approaches proposed by Gilabert and Castellví with those proposed by Kogan and Bondarenko provides a framework for developing tasks to teach morphologically complex L2s while adhering to the fundamental principles of TBLT. The central aspect of this design process is achieving the right balance between the conceptual requirements of a task, task modality, and the linguistic demands. Although accurately determining the difficulty of individual linguistic elements remains challenging, it is feasible to identify the minimal code for a task, introduce the most frequent forms associated with it, and estimate the overall morphological complexity of both the input and the expected output using the MCI. Depending on the learners' needs, tasks can be designed to emphasize more online or offline interaction, or an appropriate balance between the two. While further research is needed to evaluate the effectiveness of these task design strategies in teaching morphologically complex L2s, this approach represents a promising initial step towards integrating TBLT into the teaching of such languages. This dissertation's study, to be described in Section 3, aims to address this research need.

2 CASE

2.1 Case in Lithuanian

Lithuanian is an East Baltic language belonging to the Baltic branch of the Indo-European language family. It is generally considered to be morphologically complex due to its extensive system of inflectional morphology with a wide array of different endings. The nominal inflectional system, in particular, is characterized by a multitude of different endings, divided into five declension patterns and further subdivided into various paradigms, whose marking is determined by the grammatical categories of number, gender, and, importantly, case. Standard Lithuanian has the following seven cases: nominative, genitive, dative, accusative, instrumental, locative, and vocative. Case marking is dependent on the function to be expressed by a noun/pronoun and its agreeing adjective/determiner/participle due to syntactic and/or semantic conditions within a clause. The prototypical function of each case includes:

- Nominative (NOM): subject marking.
- Genitive (GEN): marking of the possessor of the following head noun.
- Dative (DAT): indirect object marking.
- Accusative (ACC): direct object marking.
- Instrumental (INST): marking the means by which the subject achieves an action.
- Locative (LOC): marking a location.
- Vocative (VOC): marking the entity (person, object, etc.) being addressed in a statement.

Importantly, cases can carry more functions than those listed here. While, for instance, the locative and vocative are relatively unifunctional, the genitive, in particular, is notorious for being multifunctional and is involved, for instance, in marking negated subjects and objects. Furthermore, while the accusative is the prototypical case for marking direct objects, certain verbs govern different cases, which can be the genitive (*norėti ko* ‘to want what.GEN’), dative (*padėti kam* ‘to help who.DAT’), and instrumental (*domėtis kuo* ‘to be interested in what.INST’). Prepositions govern the genitive, accusative, and instrumental, which sometimes depend on semantic distinctions within a single preposition (*po pamokos* ‘after class.GEN’, *po stalu* ‘under the table.INST’).

Case is marked through a wide range of inflectional endings (Balčiūnienė et al., 2024). There is no one-to-one relationship between a case

Table 3. A simplified overview of the nominal declension system of Lithuanian (taken and adopted from Dubasava (2020, p. 64))

Case	Declension 1		Declension 2		Declension 3	Declension 4		Declension 5	
	-as	-is (-ys)	-a	-ė	-is	-us	-ius	-uo	sesuo ‘sister’/ duktė ‘daughter’
	Singular								
Nom.	-as	-is	-a	-ė	-is	-us	-ius	-uo	-uo, ė
Gen.	-o	-io	-os	-ės	-ies	-aus	-iaus	-en- s	-er-s
Dat.	-ui	-iui	-ai	-ei	-iai	-ui	-iui	-en- iui	-er-iai
Acc.	-ą	-į	-ą	-ę	-į	-ų	-ių	-en- į	-er-į
Instr.	-u	-iu	-a	-e	-imi	-umi	-iumi	-en- iu	-er-ia
Loc.	-e	-yje	-oje	-ėje	-yje	-uje	-iuje	-en- yje	-er-yje
Voc.	-e!/ -ai!	-i! (-y!)	-a!	-e!	-ie!	-au!	-iau!	-en- ie!	-er-ie!
	Plural								
Nom.	-ai	-iai	-os	-ės	-ys	-ūs	-iai	-en- ys	-er-ys
Gen.	-ų	-ių	-ų	-ių	-ų/-ių	-ų	-ių	-en- ų	-er-ų
Dat.	-	-	-	-	-	-	-	-en- ims	-er-ims
Acc.	-us	-ius	-as	-es	-is	-us	-ius	-en- is	-er-is
Instr.	-	-	-	-	-	-	-	-en- imis	-er-imis
Loc.	-	-	-	-	-	-	-	-en- yse	-er-yse
Voc.	-ai!	-iai	-os!	-ės!	-ys!	-ūs!	-iai!	-en- ys!	-er-ys!

and a single morphological ending, as endings vary not only according to grammatical number (singular vs. plural, e.g., *vyr-as* ‘man-NOM.SG’, *vyr-ai* ‘man-NOM.PL’) but also according to the declension class to which a noun belongs (e.g., *vyr-ai* ‘man-NOM.PL’, *sūn-ūs* ‘son-NOM.PL’). Lithuanian has

five main declensions, subdivided into a total of 14 paradigms. Not all paradigms are equally productive, and some contain only a small number of nouns (e.g., Declension 2, Paradigm 7, which includes only *marti* ‘daughter-in-law’ and *pati* ‘wife’). Nouns also carry gender, although it is primarily relevant for agreement processes such as adjectival marking, since each paradigm is associated with a single gender. As Dubasava (2020) notes, L2 Lithuanian teaching materials generally follow a simplified nominal case system, though it remains relatively multidimensional. A simplified overview of nominal declensional endings is provided in Table 3. Note that the final paradigm contains only two nouns (*sesuo* ‘sister’ and *duktė* ‘daughter’).

Although case is typically regarded as a unified grammatical category, there is substantial variation both across different cases and within individual cases. One important distinction concerns the abstract grammatical and lexical-semantic functions of cases. Traditionally, case has been primarily viewed in syntactic terms, as it indicates grammatical relations between words in a sentence. However, in functional terms, cases may also express more specific semantic relations. To illustrate, in Lithuanian, in a sentence such as *vaikas mato mokyklą* ‘the child sees the school’, the accusative (*mokyklą* ‘school-ACC’) merely indicates the role of a direct object in the context of the verb *matyti* ‘see’. By contrast, in *vaikas yra mokykloje* ‘the child is at school’, the locative (*mokykloje* ‘school-LOC’) does not encode a grammatical relationship but rather expresses a concrete circumstance, namely, the child’s location (Holvoet & Semėnienė, 2004).

Rather than maintaining a strict grammatical-semantic division, Holvoet and Semėnienė (2004) argue that cases should be interpreted according to a grammaticality scale on the obliqueness hierarchy. In the obliqueness hierarchy, direct cases belong more strongly to the nucleus of the sentence; thus, they encode central functions and are obligatory, whereas oblique cases belong to the periphery and, therefore, encode peripheral functions (i.e., circumstances) and are non-obligatory. Holvoet and Semėnienė (2004) attempted to establish the grammaticality/obliqueness of cases in Lithuanian based on a case-government analysis; however, this approach posed some interesting challenges. The first challenge concerns the nominative and accusative, which are generally considered pure ‘structural’ cases, but can alternate with the genitive. For instance, this happens when a partitive interpretation is possible (e.g., *atėjo svečiai* ‘the guests.NOM have arrived’ vs. *atėjo svečių* ‘some guests.GEN have arrived’) or under negation of the verb (e.g., *aš matau vaikus* ‘I see the children.ACC’ vs. *aš nematau vaikų* ‘I don’t see the children.GEN’). This undermines their stability because

it raises the question whether verbs govern only the nominative and the accusative or also the genitive.

The second challenge arises with the dative, instrumental, and especially the locative, which are usually considered semantic cases and are therefore non-obligatory and independent. Nevertheless, in terms of verb government, they seem ‘pleonastic’, as they doubly signify their semantics, both through the verb and the case itself. A typical example is *naudotis kamščiatraukiu* ‘to use a corkscrew.INST’, where both the verb and the case indicate the instrumentality of the object. Furthermore, they are not ‘independent’, since they cannot be replaced by the genitive (e.g., **Aš nesinaudoju kamščiatraukio* ‘I don’t use a corkscrew.GEN’).

To account for these patterns, Holvoet and Semėnienė (2004) draw on the Lithuanian case hierarchy proposed by Čižik (2003) (see Figure 3). Only higher-ranking cases in the hierarchy (i.e., oblique cases) can be replaced by lower-ranking, less oblique cases. Thus, the nominative and accusative can be replaced by the genitive, while the instrumental and locative cases cannot.

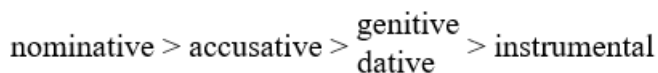


Figure 3. The Lithuanian case hierarchy (Čižik-Prokaševa, 2003)

More broadly, it may be more appropriate to speak of the grammaticality of particular case functions rather than of cases themselves. For instance, although the accusative is generally regarded as a purely structural and grammatical case, it is also used to express temporal reference (*susitinkame pirmadienį* ‘meet.1PL Monday.ACC’, ‘we will meet on Monday’) and duration (*laukėme tris dienas* ‘wait.1PL.PAST three.ACC day.ACC.PL’, ‘we waited for three days’). Conversely, although the instrumental case is considered a strongly semantic case (*važiuoju į darbą autobusu* ‘go.1SG to work.ACC bus.INST’, ‘I go to work by bus’), its usage may also be determined by grammatical factors, such as in predicate constructions (*malonu būti mokytoju* ‘it is nice to be a teacher.INST’).

Furthermore, case frequency is unevenly distributed. For simplicity, the nominative, genitive, and accusative can be considered grammatical cases, while the dative, instrumental, and locative can be considered semantic cases. According to this distinction, grammatical cases are used most frequently, while (more) semantic cases are used least frequently, as shown in Table 4:

Table 4. Case frequency in adult speech in Lithuanian (Savickienė, 2005)³

Nominative	Genitive	Accusative	Dative	Instrumental	Locative
32%	28%	18%	9%	5%	3%

A similar tendency can be observed in written Lithuanian, although distributions vary by genre (Brinkutė, 2018):⁴

Table 5. Case frequency in written Lithuanian by genre

Genre	Nom.	Gen.	Acc.	Dat.	Inst.	Loc.
Administrative	21.02%	50.21%	14.39%	4.94%	5.22%	4.23%
Fiction	30.11%	28.92%	24.08%	2.74%	8.70%	5.45%
Academic	26.35%	42.57%	15.97%	4.27%	6.34%	4.50%
Journalistic	25.80%	40.36%	16.62%	4.38%	6.48%	6.36%

Nevertheless, frequency alone is not sufficient: “the topic of conversation, participants, and situation also determine case distribution” (Savickienė, 2005, p. 64). Case frequency in children’s speech and adult speech directed to children differs from the general tendency. This phenomenon can most likely also be observed when examining the language used by and with L2 speakers. In addition, Savickienė (2005) emphasizes the role of noun semantics in shaping case distribution. For instance, the sample showed that inanimate objects are rarely assigned the dative case. Overall, Lithuanian case usage emerges from the interaction of multiple factors, including grammatical function, lexical semantics, noun class, discourse context, and frequency of use.

2.2 Second language acquisition of case

2.2.1 Acquisition of morphosyntactic structures within several theories

The acquisition of morphosyntactic constructions, such as inflections, is one of the greatest challenges for L2 learners, yet it plays a particularly significant role in second language acquisition (N. C. Ellis, 2022). According to Slabakova’s Bottleneck Hypothesis (2019), difficulties in acquiring

³ However, the study did not consider how the cases were used, e.g., in instances of using the accusative, whether grammatically to mark a direct object, or semantically to express a time circumstance.

⁴ Brinkutė’s (2018) data were recalculated by grouping singular and plural case forms together and discarding the vocative and illative cases.

morphological grammar constitute a substantial obstacle to the overall process of L2 acquisition. This challenge arises because grammatical categories, such as tense, are not conveyed through a single form but rather involve a range of morpho-phonological, syntactic, and semantic features that together shape the meaning and acceptability of sentences. Consequently, full mastery of a specific grammatical category is only possible when an L2 learner has acquired the complete set of associated features. The impact of morphological complexity on L2 acquisition has been demonstrated by Brezina and Pallotti (2019), who argue that learning a morphologically simpler L2 facilitates faster development of inflectional competence compared to learning a morphologically complex L2. Thus, they argue that acquiring inflectional competence takes longer when the target L2 has a more intricate morphological system.

The four theories reviewed below address different aspects of the same broader question: why are morphosyntactic markers difficult to acquire in an L2, and under what conditions can that difficulty be reduced? Although these theories differ in emphasis, they need not be viewed as mutually exclusive. Rather, they operate at partially complementary levels of explanation. Usage-based approaches such as the Associative–Cognitive CREED (N. C. Ellis, 2006) focus on how patterns emerge from experience with input, largely coinciding with the cognitive-interactionist model (see Section 1.2.2.). Input Processing (VanPatten, 2014) specifies how learners allocate attention during comprehension and why some cues are noticed more readily than others. Processability Theory (Pienemann & Lenzing, 2020) addresses developmental constraints on what learners can produce at a given stage. Finally, Skill-acquisition Theory (DeKeyser, 2014), by contrast, explains how explicit knowledge may become proceduralized and automatized through practice, thereby aligning quite well with TSLT (see Section 1.2.4). Taken together, these perspectives offer a multidimensional account of L2 morphological development.

2.2.1.1 The Associative–Cognitive CREED

According to the CREED model proposed by N. Ellis (2006), L2 acquisition follows the same associative and cognitive learning principles that govern all forms of knowledge acquisition. From this perspective, L1 and L2 acquisition are fundamentally similar processes (Wulff & Ellis, 2018). The framework derives its name from its defining characteristics: it is Construction-based, Rational, Exemplar-driven, Emergent, and Dialectic (N. Ellis, 2006). The

CREED model is based on emergentism (N. C. Ellis, 1998) and *Chaos/complexity theory* (Larsen-Freeman, 2011), viewing the L2 system as one in which linguistic patterns arise and develop through dynamic interactions among individuals, brains, societies, and cultures.

The CREED model is described as construction-based since it is grounded in Construction Grammar. Construction Grammar posits that the basic units of human language representation are constructions, defined as symbolic form–meaning pairings. In these pairings, morphological, syntactic, and lexical forms are systematically linked to semantic, pragmatic, and discourse-related functions. Constructions differ in their level of abstraction, ranging from fixed expressions such as *kick the bucket* (i.e., idioms) to highly schematic patterns such as *transitive verb + accusative case*, which function as abstract syntactic templates. According to usage-based approaches, speakers’ linguistic competence emerges from the cumulative experience of all utterances they have encountered and retained, from which frequency-based regularities gradually develop. This principle applies not only to linguistic forms but also to form-meaning mappings. Each encounter with a form, such as in Lithuanian, *lab*, contributes to implicit probabilistic expectations regarding possible continuations (e.g., *labas* ‘hi’, *laba diena* ‘good day’, or *labdara* ‘charity’), resulting in a network of associations, in which activation depends on prior experience and distributional probabilities (N. C. Ellis, 2006).

On this basis, the CREED model maintains that the human language system is rational, conceptualized as a probabilistic network of associations shaped by prior experience. Within this network, contextual information guides language use toward those linguistic constructions most likely to be relevant, thereby facilitating successful comprehension and production. From this perspective, both L1 and L2 acquisition can be understood as forms of statistical construction learning, in which learners infer probability distributions over constructional interpretations, i.e., whether a particular interpretation of a construction follows from a given cue in a specific context.

Usage-based approaches further seek to explain how language users form generalizations from exemplars at different levels of abstraction. For instance, repeated exposure to the expression ‘good morning’ may give rise to a less specific construction such as [good + (time of day)], which can subsequently support the emergence of an even more abstract pattern, [adjective + noun]. Through shared features across specific constructions, linguistic patterns form schemas or prototypes, a principle that also underlies the CREED model. Prototypes are exemplars that are most representative of

their category, most similar to other members of that category, and most distinct from members of other categories, and they play a central role in human cognition. Accordingly, for example, the English ending *-s* is more likely to be categorized as a plural marker (e.g., *sparrows*) than its alternatives such as *-es* (e.g., *albatrosses*) or vowel change (e.g., *geese*) (N. C. Ellis, 2006).

From these exemplar-derived associations, L2 systems are formed. According to chaos/complexity theory, open complex systems emerge spontaneously from apparent chaos through the dynamic interaction of many variables. For phenomena such as weather patterns, ecosystems, or economies to function in relatively stable ways, a system, i.e., a set of regularities, arises from these interactions. As the variables continue to interact, such regularities may also change. However, these changes are driven not by external forces but by gradual development within the system itself. Within the CREED model, this emergentist perspective is also applied to language as a system and to language acquisition (N. C. Ellis, 1998; Larsen-Freeman, 2011). Although grammatical knowledge is often described in terms of rules, it may be more accurate to characterize it as a set of emergent patterns, i.e., relatively stable tendencies that arise from the ongoing interaction of all variables relevant to the language system. This emergentism extends to language acquisition as well: through subconscious statistical computations over constructions, a language system develops, and prototypical as well as more peripheral constructions gradually emerge. An individual's L1 can thus be understood as the outcome of interactions among multiple variables, an approach equally applicable to L2 acquisition in the sense of interlanguage development.

Finally, the CREED model characterizes L2 acquisition as dialectic, i.e., as developing through the productive tension of oppositions. This opposition exists between, on the one hand, the learner's interlanguage at a given moment and, on the other hand, the linguistic target L2 input received from the environment, which may function as either positive or negative evidence. Input from interlocutors enables learners to compare their own interlanguage with other L2 varieties and to notice constructions that diverge from target language patterns.

This process is closely related to the distinction between implicit and explicit learning, which the CREED model assumes to both be integral to L2 acquisition. Explicit instruction is particularly important in the initial presentation of complex constructions, as it helps learners "notice the cue" (R. W. Schmidt, 2001). Implicit learning, by contrast, is crucial in the long term, as it involves statistical computation of constructional usage over time. Compared with the cognitive-interactionist model that informs TBLT, there

appears to be significant overlap. However, the CREED model assumes a considerably stronger potential for interaction between explicit and implicit knowledge (the strong interface position) than is typically assumed in TBLT. The role attributed to explicit learning in the CREED model, however, seems to align quite strongly with TSLT. In summary, the CREED model proposes that L2 acquisition is fundamentally a form of statistical construction learning based on exemplars, emerging through a dialectic process between the L1 and the learners' interlanguage.

At the same time, N. Ellis (2022) notes that morphosyntactic constructions often exhibit properties that hinder their incorporation into the subconscious statistical calculations assumed by the model. He distinguishes three main factors that cause these learning difficulties. The first important factor to consider is salience. Salience depends on the extent to which a linguistic element, such as a construction, stands out from others, as in the way linguistic markers are experienced in the physical world (e.g., louder or more unique sounds) or the potential surprise effects they create for the learner. Research has shown that learners acquire more salient items more quickly than less salient ones. Unfortunately, morphological constructions tend to exhibit low salience. When evaluating the physical salience of a linguistic element based on its phonetic substance, stress, or fixed serial position, it becomes evident that grammatical morphemes, which are often unstressed and depend on other lexical words, are unlikely to exhibit high physical salience. Verb conjugations and noun endings, in particular, are often perceptually vague for L2 learners, and even for L1 users (Herron & Bates, 1997). However, while L1 users can rely on a fully developed L1 system to interpret morphemic cues, L2 learners lack this foundation, making it extremely difficult for them to acquire grammatical morphology from auditory input alone.

Moreover, L2 morphology is not only hard to detect but also contributes less transparently to the language system. To L2 learners, many morphological markers may seem redundant. For instance, the necessity of marking a direct object with the accusative case decreases when its function is already clarified by the object's position in the sentence. Similarly, marking a negated direct object in Lithuanian with the genitive becomes redundant if the prefix *ne-* already denotes negation. Consequently, the psychological salience of morphology is often low, and the frequent occurrence of, for instance, case morphemes in speech ensures that their surprise effect quickly fades.

The second factor that complicates the acquisition of morphology is contingency. According to N. Ellis (2022), both L1 and L2 acquisition are fundamentally statistical processes that rely on the probability distributions of form-meaning mappings. When the contingency of a construction is high (i.e., a specific interpretation consistently follows a particular cue in a given context), the construction is easier to learn. Research on L1 acquisition and natural morphology has shown that children prefer unambiguous endings (Slobin, 1966; Wurzel et al., 1987), and L2 learners are no exception. However, unfortunately, grammatical morphology does not generally conform to this preference.

In Table 6, the letters *a*, *b*, *c*, and *d* represent frequencies. Ideally, the values of *a* and *d* are the highest: when a specific cue is present, the same outcome always follows (*a*), and when it is absent, the same outcome does not occur (*d*). This represents a perfect 1:1 mapping between form and meaning. In Lithuanian, most locative endings (-*yje*, -*oje*, etc.) illustrate this pattern well, since they appear almost exclusively when an adjective and/or a noun is marked in the locative.

Table 6. Possible combinations of cues and outcomes (*Source*: N. Ellis, 2022, p. 39)

	Outcome	No outcome
Cue	<i>a</i>	<i>b</i>
No cue	<i>c</i>	<i>d</i>

However, issues arise when frequencies *c* and *b* are high. Frequency *c* illustrates cases where an interpretation can be expressed by multiple forms. For instance, in Lithuanian, not only do the accusative and genitive cases indicate the direct object or negation, but word order and the prefix *ne-* can also convey these meanings. Frequency *b*, on the other hand, refers to cases where a single form serves multiple functions. A classic example in English is the ending -*s*, which simultaneously marks plural nouns, third-person singular verb forms, and contractions (e.g., *it's*, *she's* (*got*)), leading to ambiguity. Another more abstract example from Lithuanian is the genitive case, which serves multiple functions, such as indicating possession, quantity, and negation. Savickienė (2003) has demonstrated how these different genitive functions are acquired gradually in the L1, beginning with the prototypical function of possession and only later extending to more peripheral uses. This gradual learning process highlights the concept of

prototypicality, which is another important aspect of SLA that will be discussed in more detail below.

The final important factor in this context is the influence of the L1 on L2 acquisition. This is, of course, one of the fundamental distinctions between L1 and L2 acquisition. While linguistic transfer can positively affect interlanguage development by accelerating it through similarities between the L1 and L2 (Ringbom & Jarvis, 2009), it is more relevant here to consider its negative effects, which often result in blocking. Blocking largely arises from the potential redundancy of morphological elements. For instance, if the direct object in the L1 can be indicated through its syntactic position alone, learners may question the necessity of marking it with a case in the L2. Given that the standard word order in Lithuanian is Subject-Verb-Object (SVO), this reasoning might even apply to L2 Lithuanian learners. Some instances of blocking or redundancy are even more widespread. All languages possess lexical means of expressing grammatical tense, such as adverbs (e.g., *yesterday*, *now*), noun phrases (e.g., *pirmadienį* ‘on Monday’, *ateityje* ‘in the future’), and prepositional or postpositional phrases (e.g., *in the morning*, *a year ago*). As a result, any L1 may lead the L2 learner to disregard verbal morphology for tense in the L2 and instead rely on lexical markers, which are easier to process (N. C. Ellis, 2022). However, the morphological properties of a learner’s L1 may impact their attention to L2 morphology, suggesting that greater morphological complexity in the L1 may encourage closer attention to L2 morphology (N. C. Ellis & Sagarra, 2010).

Given these complicating characteristics of L2 morphology, it is worth considering the methods and sequence in which L2 morphological constructions are taught. The question of whether an optimal sequence exists for teaching morphosyntactic forms has long been discussed in SLA research. A usage-based approach suggests that instruction should follow construction prototypicality. Prototypicality is not only a significant factor in L1 acquisition but also plays a crucial role in human cognition more broadly. According to N. Ellis (2022), L2 acquisition proceeds through generalizations based on examples encountered in the input. Through exposure to a variety of L2 examples, learners implicitly develop an understanding of prototypical and peripheral categories of constructions. The more frequent a construction type is, the more readily it becomes ‘accessible’ to learners and adaptable to new contexts. In other words, just as L1 speakers find prototypical constructions easier to process and reproduce, L2 learners should also be able to acquire them more easily. A relevant example of this process is the ‘verb (and prepositional) islands’ (Tomasello, 2003), which children frequently use as

building blocks for acquiring more abstract L1 constructions (N. C. Ellis, 2009).

A comparable strategy may also be effective for L2 instruction. On the basis of research on prototype-rich L2 input (Bybee, 2008; Goldberg et al., 2004), it has been argued that, for optimal acquisition, central category members (i.e., prototypes) should be introduced early and reinforced frequently in L2 curricula (N. C. Ellis, 2009). For instance, in the genitive case in Lithuanian, L2 learners might first receive input rich in examples of the prototypical function of expressing possession, with less frequent or less salient genitive functions introduced gradually over time.

For the present discussion, the CREED's main contribution lies in explaining why case morphology may be difficult to acquire from input alone: many case markers are low in salience, probabilistic in distribution, and partially redundant with lexical or syntactic cues. The framework is therefore especially useful for predicting and explaining differential learnability across case functions.

2.2.1.2 Input Processing

VanPatten's (2014) Input Processing (IP) theory addresses how L2 learners establish form-meaning connections while comprehending L2 input. IP investigates how these connections form, why some are established while others are not, and what psycholinguistic strategies are employed during L2 comprehension. The key claims of IP can be summarized as follows:

- Learning is primarily driven by meaning. When processing input, L2 learners prioritize meaning above all else.
- Cognitive processing and working memory place greater demands on L2 learners than on native speakers. As a result, L2 learners' attention to the input is more limited. In this sense, L2 learners are characterized as 'limited-capacity processors' with reduced ability to process and store information compared with native speakers.
- During comprehension, learners may rely on certain IP universals or on their L1.

These claims form the basis of a set of principles known as the IP principles. The most relevant to the present discussion are the following (VanPatten, 2014):

- *The Primacy of Content Words Principle* (p. 115): learners tend to process content words before anything else in the input.

- *The (Revised) Lexical Preference Principle* (p. 116): if a grammatical form expresses a meaning that can also be encoded lexically, learners will not initially process that grammatical form until they have lexical forms to which they can match it.
- *The Preference for Nonredundancy Principle* (p. 117): learners are more likely to process nonredundant meaningful grammatical markers before redundant ones.
- *The Meaning before Nonmeaning Principle* (p. 118): learners are more likely to process meaningful grammatical markers before nonmeaningful ones.

Clearly, the core concepts of the IP principles are *meaning* and *redundancy*. Learners scan the input for the most meaningful elements and ignore those that carry less, none, or the same meaning as the more meaningful component. For instance, in English, the past-tense marker *-ed* may be redundant when an adverb of time is present. According to the Revised Lexical Preference Principle and the Preference for Nonredundancy Principle, learners are likely to ignore such grammatical marking on verbs in the presence of a lexical item carrying a similar meaning. Eye-tracking studies have shown that beginner and intermediate English L1 learners of Spanish tend to focus on adverbs to determine temporal reference. Conversely, advanced learners and native speakers rely on verb forms. Interestingly, Spanish L1 learners of English show the same tendency to rely on adverbs, suggesting that this strategy is universal and independent of the learners' L1 (VanPatten & Keating, 2007). Conversely, the English ending *-ing* is nonredundant, since it has no lexical equivalent. Therefore, the IP predicts that learners will process and acquire this form more quickly, and research findings corroborate this prediction (VanPatten, 2014). In summary, L2 learners first focus on lexical elements, then on nonredundant grammatical elements, and finally on semantically redundant grammatical markers.

With regard to grammatical case, IP can also generate predictions about how L2 learners of, for example, Lithuanian may process input. According to the theory, semantic cases, such as the locative, dative, and instrumental, are expected to be easier to process and acquire than grammatical cases, such as the nominative, accusative, and genitive. The locative, in particular, appears to be a strong candidate, as it has a clear function of denoting location without any other lexical markers bearing the

same meaning.⁵ In contrast, the nominative and accusative may be more difficult because of their limited semantic content, which may make them seem redundant to learners. Although Lithuanian allows free word order, the basic SVO word order means that subject and object roles can be expressed or interpreted through syntax. Similarly, prepositions carry meaning themselves, which makes subsequent case assignment a purely grammatical operation. Still, L2 learners may pay closer attention to case assignment in situations where two options exist, such as with the preposition *po* in *po savaitės* ‘in a week (GEN)’ and *po stalu* ‘under the table (INST)’. Furthermore, L2 learners might have less difficulty using the accusative for temporal expressions than for marking direct objects. Of course, input semantics alone cannot account for everything in SLA. Frequency, the perceptual salience of grammatical markers in the input, L2 instruction, and production all play important roles (N. C. Ellis, 2022; see below); see section 2.2.1.1).

However, as VanPatten (2014) emphasizes, IP is not intended to be an all-encompassing theory of SLA. Rather, it focuses on how L2 learners make form-meaning connections when processing input, which is highly relevant to this study. Specifically, it predicts that semantically transparent and nonredundant cases should attract attention earlier than structurally assigned or redundant case markers.

2.2.1.3 Processability Theory

Another theory of L2 acquisition, Processability Theory, was developed by Pienemann (1998). The core claim of the theory is that learners can produce and comprehend only those L2 linguistic units for which they are developmentally ready, or, in other words, that their language processor is capable of processing at a given stage. Developmental readiness is assumed to be largely universal across L2 learners, depending on the linguistic constructions in question, allowing the description of developmental trajectories, i.e., the successive stages any L2 learner passes through when acquiring a given L2 construction. Developmental trajectories are predicted based on a set of theoretical frameworks (Bresnan, 2000; Levelt, 1989) and on how grammatical information is transferred within and between phrases

⁵ IP also proposes principles that relate to the likelihood of interpretation through context and lexical semantics. For instance, the sentence *Aš esu Vilniuje* ‘I am in Vilnius’ is unlikely to be interpreted solely through the locative; the context and the word *Vilnius* itself are also significant factors, as the alternative interpretation of ‘I am Vilnius’ is highly improbable.

and sentences. The original processability hierarchy proposed by Pienemann (1998) is as follows:

1. no procedure (e.g., producing isolated words such as ‘yes’).
2. category procedure (e.g., adding a past tense morpheme to a verb, as in ‘talk-**ed**’).
3. noun phrase procedure (e.g., marking plurality as in ‘two kid-**s**’).
4. verb phrase procedure (e.g., moving an adverb out of the verb phrase, as in ‘I went yesterday/yesterday I went’).
5. sentence procedure (e.g., subject–verb agreement, as in ‘Peter see-**s** a dog’).
6. subordinate clause procedure (e.g., use of the subjunctive in subordinate clauses triggered by a main clause, as in ‘The doctor insisted that the patient be quiet’).

It is argued that L2 learners inevitably acquire grammatical structures in this hierarchical order, since (a) each procedure constitutes a prerequisite for the subsequent one, and (b) the hierarchy reflects the course of real-time language processing as hypothesized in the Levelt (1989) model.

Based on this, the following developmental trajectory has been formulated for the acquisition of L2 English questions:

Table 7. Suggested developmental trajectory for questions in L2 English within Processability Theory (taken from Pienemann and Lenzing (2020, p. 160))

Stage	Structure	Example
Stage 1	SVO question	He live here?
Stage 2	WH+SVO	Where he is?
Stage 3	Copula inversion	Where is he?
Stage 4	Aux-second	Where has he been?

Stage 1 is characterized by a canonical declarative SVO order and involves virtually no transfer of grammatical information, placing it firmly in the ‘no (syntactic) procedure’ stage. Stage 2 can be defined as a ‘category procedure’, since a new category, namely a question word, is added in sentence-initial position, although the SVO structure of the sentence remains intact. Stage 3 involves copula inversion, indicating a simple ‘phrasal procedure’. Finally, stage 4 involves auxiliary movement, agreement, and tense/aspect marking across phrases, and thus corresponds to the ‘sentence procedure’ stage (Pienemann & Lenzing, 2020).

Evidently, Processability Theory maintains that L2 learners' acquisition of linguistic structures often follows predictable paths that can be described in terms of developmental trajectories. Importantly, this does not imply that L2 learners' production is always identical, as learners may resolve learning problems in several ways within the developmental stage they are in, the so-called 'hypothesis space'. Nevertheless, because developmental stages cannot be skipped, this implies that factors such as the frequency of a linguistic unit in the input or learners' L1 have only a limited effect on L2 acquisition according to Processability Theory (Pienemann & Lenzing, 2020).

Furthermore, Processability Theory predicts clear limits on the effects of instruction on L2 acquisition (Pienemann & Lenzing, 2020). Instruction can only speed up progression from one stage to the next, but it cannot bypass stages in the developmental trajectory, as each stage is a prerequisite for the next. In this sense, "practice does not make perfect," because output is constrained by processability: learners cannot produce structures they are not yet able to process. Consequently, this forms the basis of a proposal for L2 instruction called the Teachability Hypothesis (Pienemann, 1984), which holds that teachability must adhere to learnability, i.e., the L2 content taught must be constrained by the range of linguistic structures for which L2 learners are developmentally ready, according to the processability hierarchy. Long (2014) argues that the principle of teachability is reflected in TBLT through focus on form due to its responsiveness: as focus on form is provided only in response to learner problems, it thus aligns with learners' internal syllabi and, hence, developmental readiness.

In the context of grammatical case, Magnani and Artoni (2015) proposed a sequence for case instruction. Building on Processability Theory (Pienemann, 1998), they developed a hypothetical framework for acquiring the case system in L2 Russian (see Table 8). As shown in the table, at the first stage, nouns and pronouns are uninflected, appearing exclusively in the (nominative) base form, except in certain set phrases. In the second stage, case marking begins to emerge, though it is largely limited to a binary opposition between the nominative/base form and all other cases. It is not until the third stage that the learner's inflectional competence expands, and case assignment begins to operate not only between nouns and verbs but also with prepositions and adjectives. In the fourth and final stage, the learner is capable of assigning case accurately across all sentence constructions. To test this model, Magnani and Artoni conducted a cross-sectional study with 17 L2 learners of Russian from varied L1 backgrounds. The results supported their proposed case

Table 8. Suggested developmental trajectory of case marking in L2 Russian
(Source: Magnani and Artoni (2015, p. 63))

Stage	Procedure	Structure	Examples
4	Sentence	Object-Verb case agreement	<i>knig-u čitaet mama</i> book-ACC reads mom 'Mom is reading a book.'
		Verb-Object case agreement	<i>čitaju knig-u</i> (I) read book-ACC 'I am reading a book.'
3	Phrasal	Preposition-Noun case agreement	<i>ja s druž'-jami</i> I with friends-INST 'I (am) with friends.'
		Adjective-Noun case agreement	<i>choroš-uju devušku</i> nice-ACC girl-ACC 'A good girl.'
2	Category	Nominative vs. non-nominative	<i>mama</i> vs. <i>mame/mamu</i> 'mom' (NOM vs. non-NOM)
1	Lemma access	No case marking	<i>net urok</i> 'no lesson' <i>menja zovut Lena</i> 'My name is Lena.'

acquisition sequence: all participants reached at least the second stage, 15 advanced to the phrasal stage, and only 9 progressed to the sentence stage.

Based on their results, Magnani and Artoni proposed a more efficient approach to teaching Russian as an L2, grounded in their suggested acquisition sequence. Drawing on Pienemann's Teachability Theory (1998), they argue that introducing grammatical forms beyond a learner's L2 developmental stage is ineffective and may even disrupt the teaching process. This perspective might explain why numerous studies on L2 Russian have reported little progress in case marking despite extended formal instruction (Denissenko, 2016; Henry, 1996; Markina, 2018): the learners were simply not developmentally ready. Magnani and Artoni (2015) recommend teaching inflection in two complementary ways

1. by expanding the inventory of forms that correspond to the learner's current stage and ensuring their correct use.
2. by introducing forms from the subsequent developmental stage.

For instance, during the phrasal stage, once a learner has started to assign case based on verbs, case marking of adjectives and prepositional case assignment can be introduced. Later, as the learner's case-marking skills become stronger, more complex, non-canonical sentence constructions can be introduced, such as *knig-u chitaet* (literally, 'book-ACC reads', meaning 'he is reading a book').

It is also noteworthy that this proposal reflects patterns of L1 acquisition, such as the gradual acquisition of cases through oppositions and the relatively late emergence of case marking with prepositions (Savickienė, 2003). Furthermore, this framework can be adapted to the teaching of L2 Lithuanian. For instance, a common error among both L1 and L2 learners is marking the direct object of a negated verb with the accusative instead of the genitive (Dabašinskienė & Čubajevaitė, 2009). This phenomenon can be explained within the framework: the learner must first grasp that verbs typically govern the less frequent accusative rather than the genitive, and then revise this knowledge once they encounter negation, where the genitive is required. Following Magnani and Artoni's recommendations, the sequence for L2 Lithuanian case acquisition could be structured as follows:

1. *Mama mato berniuk-as.*
mom sees boy-NOM
'Mom sees a boy.'
2. *Mama mato berniuk-o* (and other case endings, e.g., *-q* (ACC))
mom sees boy-GEN
'Mom sees a boy.'
3.
 - a. *Mama mato berniuk-q. / Mama ne-mato berniuk-q.*
mom sees boy-ACC / mom NEG-see boy-ACC
'Mom sees a boy.' / 'Mom doesn't see a boy.'
 - b. *Mama mato berniuk-q. / Mama ne-mato berniuk-o.*
mom sees boy-ACC / mom NEG-see boy-GEN
'Mom sees a boy.' / 'Mom doesn't see a boy.'

While this framework remains hypothetical, the work of Magnani and Artoni highlights the importance of considering developmental sequences in L2 acquisition, as neglecting them may disrupt the overall learning process. Whereas the CREED and IP emphasize input and cue learning, Processability Theory adds a developmental dimension: even frequent or salient forms may remain difficult if the learner's processing system is not yet ready to handle the required grammatical operations. Nevertheless, as noted above,

teachability of forms is incorporated into TBLT, provided it is reactive, that is, provided in response to learner problems encountered during task performance.

2.2.1.4 Skill Acquisition Theory

In contrast to Processability Theory, which assumes that practice does not make perfect, Skill Acquisition Theory (DeKeyser, 2014) assumes that practice *does* make perfect. The theory maintains that learning a wide variety of skills follows a fundamentally similar developmental course and that a set of core principles can be identified that are common to the acquisition of all skills, including L2 learning. Skill Acquisition Theory distinguishes three developmental stages:

- Declarative stage: learners acquire knowledge about a subject, such as an L2 structure, which results in declarative knowledge. This knowledge may be gained through observation of skilled behavior, through verbal explanation by an ‘expert’ (e.g., an L2 teacher), or, most commonly, through a combination of both.
- Procedural stage: declarative knowledge is subsequently proceduralized through attempts to act on this knowledge in order to foster a behavior. The proceduralization process need not be overly time-consuming, provided that the appropriate declarative knowledge is available. A key advantage of procedural knowledge is that learners no longer need to retrieve and assemble separate bits of information from memory; instead, the intended behavior is now available as a ready-made chunk that can be executed when the necessary conditions are met. In L2 acquisition, chunking is often regarded as a crucial learning strategy (N. C. Ellis, 1996), and it also aligns with the principles of Construction Grammar.
- Autonomous stage: achieving robust fluency and accuracy requires extensive practice, which gradually leads to the automatization of previously acquired procedural knowledge. Automaticity, however, can only occur once the relevant procedural knowledge is in place.

In the context of L2 acquisition, these stages can be illustrated by the learning of a grammar rule, such as adding the morpheme *-ed* to verbs in English to mark the past tense. At the declarative stage, learners rely heavily on their working memory, frequently make errors, and must consciously verbalize the rule. At the procedural stage, production becomes more fluent and accurate, with past-tense marking being triggered by contextual cues (e.g. ‘yesterday’) and attention partially shifting from form to meaning. At the autonomous

stage, the rule is applied automatically with minimal cognitive effort, thereby allowing attention to be devoted entirely to meaning.

One central observation in research on Skill Acquisition Theory is the ‘power law of learning’. Reaction times and error rates during the performance of a task decrease with practice, typically declining rapidly at first and then more slowly over time, in accordance with a predictable mathematical curve. This pattern indicates that early practice yields substantial performance gains, whereas later practice yields smaller, incremental gains, as illustrated in Figure 4. It is commonly argued that this pattern reflects the three stages outlined above, with the steep initial section of the curve reflecting proceduralization of declarative knowledge and the following, more level section corresponding to the automatization of procedural knowledge (DeKeyser, 2014).

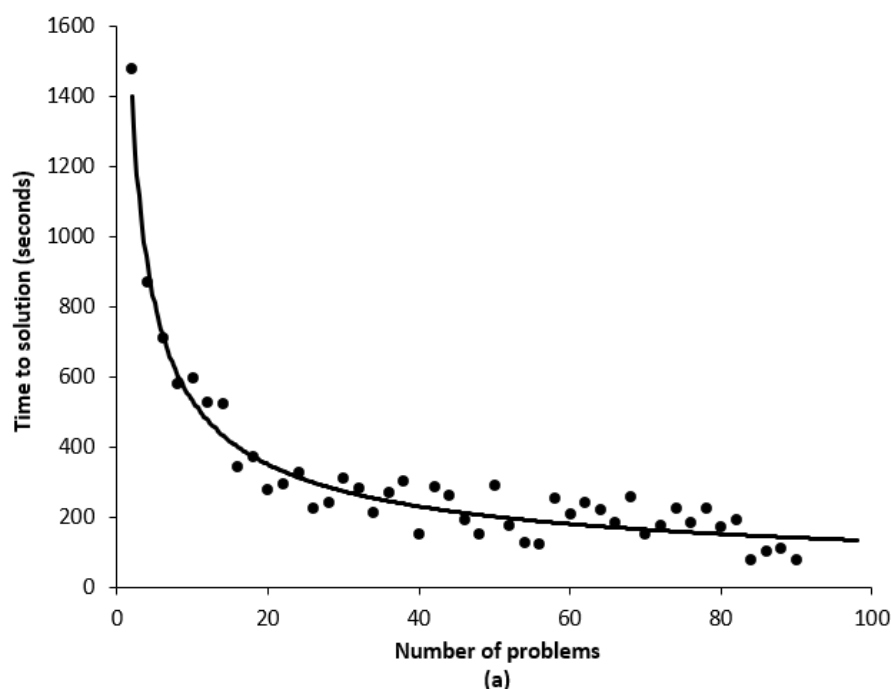


Figure 4. A sample graph of the power learning curve (adapted in *Excel* from DeKeyser, 2014, p. 97)

Granted, the proponents of Skill Acquisition Theory do not claim that it is an all-encompassing theory of L2 acquisition. The theory is assumed to be mostly effective for (a) adult learners who (b) possess an aptitude for explicit language learning when learning (c) relatively simple structures at (d)

fairly early stages of development within (e) instructional contexts. Deviations from these conditions are expected to reduce the effectiveness of declarative knowledge in the initial stages of learning. For this reason, it is argued that learners might benefit from more implicit instructional approaches when dealing with complex grammatical rules. With respect to the explicit-implicit divide, while the declarative-procedural distinction is said not to coincide entirely, declarative knowledge can generally be considered comparable to explicit knowledge. In contrast, procedural knowledge is viewed as closely related to implicit knowledge. Hence, Skill Acquisition Theory adopts a strong interface position, according to which explicit/declarative knowledge can, through processes of proceduralization and automatization, develop into knowledge that is “functionally equivalent to implicit knowledge” (DeKeyser, 2014, p. 106).

Importantly, proponents of Skill Acquisition Theory argue that it should not be viewed as competing with L2 teaching approaches such as TBLT. DeKeyser (2014) even argues that tasks in TBLT essentially provide opportunities for L2 learners to proceduralize and potentially automatize their knowledge, provided that they possess the necessary declarative knowledge. However, such an approach more closely aligns with TSLT. If Skill Acquisition Theory is indeed effective under the conditions described above, TSLT would be expected to be more advantageous than TBLT for teaching relatively simple L2 rules.

This is also particularly relevant when considering L2 acquisition of grammatical case. As semantic cases (e.g., the locative and the instrumental, according to their prototypical semantic functions) carry fairly concrete meaning, which L2 learners tend to focus on when processing input (VanPatten, 2014), the teaching of these cases should be relatively straightforward and therefore benefit from an approach aligned with Skill Acquisition Theory. While the more grammatical cases (e.g., the accusative or the genitive) can be considered more complex due to their wide array of functions, separating these functions into manageable units can reduce their complexity and enhance the potential for teaching them in line with Skill Acquisition Theory. TSLT might then prove exceptionally relevant to the teaching of case, perhaps even more so than TBLT. Nevertheless, cases in L2 Lithuanian are accompanied by a large inventory of morphemes, which likely increases complexity and, consequently, negatively impacts the potential for teaching case according to Skill Acquisition Theory.

Taken together, the four theories converge on the view that L2 morphology is difficult not for a single reason but, rather, because several

constraints interact. Morphological markers may be hard to notice in the input due to their limited salience, low contingency (CREED), and L2 learners' tendency to prioritize meaningful lexical items over functional markers (IP theory). Furthermore, morphological markers may require grammatical operations that emerge only gradually (Processability Theory) and may demand extensive practice before fluent use is possible (Skill Acquisition Theory). At the same time, the theories differ in terms of the weight they assign to input, internal developmental constraints, and explicit instruction. They thus offer distinct but complementary perspectives for interpreting patterns of L2 case acquisition. Furthermore, the role of explicit instruction is especially relevant to the question of what best promotes lasting L2 case acquisition: extensive practice based on explicit, declarative knowledge (aligning with TSLT) or extensive exposure to and use of the L2 supported by limited explicit instruction (as emphasized in TBLT).

2.2.2 Former research on L2 acquisition of case

The following discussion focuses on findings from empirical research on L2 acquisition of grammatical case. First, studies on case acquisition for L2 Lithuanian are presented. Second, several case acquisition studies for other L2s are discussed. The discussion then continues by examining the case acquisition order in L2 German and Russian. Finally, the discussion concludes by examining the only interventional study on case acquisition within a morphologically complex L2, specifically, Russian, through TSLT and TBLT.

2.2.2.1 Case acquisition in L2 Lithuanian

L2 Lithuanian research further highlights the importance of the factors discussed in the previous sections. First, this can be seen in L2 Lithuanian learners' frequent errors in case assignment, as described by Dabašinskienė and Čubajevaitė (2009). The researchers identified five main categories of errors:

1. Overuse of the nominative case.
2. Incorrect case assignment after a pronoun.
3. Incorrect case assignment after a negated verb.
4. Incorrect case marking related to time, place, and direction.
5. Mixing up declension paradigms.

Frequency and salience factors can explain Category 1 errors. The nominative is the most frequently used case in spoken Lithuanian (Savickienė, 2005),

while other case forms and morphemes are generally less salient. Furthermore, it would likely be more appropriate to refer here not to the assignment of the nominative case but to the overuse of the base form. L2 learners likely do not erroneously assign the nominative, but instead produce the canonical base form, as would also be predicted by Magnani and Artoni's (2015) model, in which the first stage is limited to lemma retrieval.

Category 2 and 3 errors can be explained by redundancy and L1 influence. Case marking of a noun phrase following a pronoun is typically a purely grammatical operation; hence, the lexical information of the preposition itself should suffice for students to convey the intended meaning. The same applies to the genitive in negative constructions, where the verb is already marked by the prefix *ne-*. When such grammatical operations are absent in the learners' L1, mastering them in the L2 takes additional time. In this context, it is also noteworthy that the genitive is often used in place of the accusative to mark direct objects. Since the genitive is the second most common case after the nominative, this may reflect possible frequency effects from the input. Category 4 errors occur not only in L2 but also in L1 acquisition, underscoring that L2 learners experience similar difficulties. Here, too, L1 influence is evident. For example, the use of the locative to express time can be attributed to the influence of learners' Polish or Latvian L1 (Dabašinskienė & Čubajevaitė, 2009, p. 56). Finally, Category 5 errors may stem from the low contingency of many morphemes. It is easy to become disoriented within the extensive array of Lithuanian case morphemes, especially since the pairing of a noun with a specific declension paradigm is often arbitrary.

Bružaitė-Liseckienė (2021) examined object marking in Lithuanian by students attending Russian-language high schools. While the participants were adolescents rather than adults and had relatively advanced L2 proficiency in Lithuanian, their language errors exhibited many of the factors already discussed. As the study adopts a usage-based approach, the dissertation also addresses the salience of object markers, noting that case morphemes tend to be less salient and, as a result, are less likely to be noticed by L2 learners. Furthermore, in the context of object marking, the concept of prototypicality is also relevant. Here, this refers to the marking of a single function (in this case, the object) with various cases that differ in prototypicality. Although the accusative is the most prototypical case for object marking (e.g., *skaitau knygą* 'read.1SG book-ACC' ('I am reading a book')), in Lithuanian, the direct object may also appear in the genitive (e.g., *laukiu svečių* 'wait.1SG guest-PL.GEN' ('I am expecting guests')), the dative

(e.g., *vadovauju bank-ui* ‘manage.1SG bank-DAT’ (‘I am managing a bank’)), and the instrumental (e.g., *džiaugiuosi atostogomis* ‘enjoy.1SG vacation-PL.INST’ (‘I am enjoying my vacation’)).

According to usage-based approaches, prototypical categories should be the easiest to acquire. However, Bružaitė-Liseckienė’s (2021) study did not reveal a clear advantage for the accusative case. In fact, a slight preference for the genitive was observed, though this finding was not statistically conclusive. More importantly, the students’ L1 strongly influenced the results: many observed errors could be attributed to discrepancies between Lithuanian and Russian (i.e., negative transfer). Notably, errors tended to involve grammatical cases replacing semantic cases rather than the reverse. This finding strongly supports the case hierarchy proposed by Chizik-Prokaševa (2003).

The significance of verb frequency and the productivity of declension paradigms were especially noteworthy. As noted above, the prototypicality of specific cases for object marking did not influence the students’ performance. However, verb frequency did: the more frequent the verb, the more likely students were to assign the correct case to its object. Although N. Ellis (2022) argues that most studies reveal a stronger impact of specific morpheme features (such as type, frequency, and contingency) over lemma frequency, Bružaitė-Liseckienė’s (2021) results indicate a clear effect of verb lemmas on accurate case assignment. This suggests that accurate case marking depends not only on abstract grammatical knowledge but also on lexical familiarity with individual verbs. The study also identified a formal aspect: the more productive the declension paradigm, the lower the probability of error, and vice versa. Hence, difficulty lies not only in selecting the correct case but also in inflecting the noun according to less productive paradigms. Dabašinskienė and Čubajevaitė (2009), similarly, noted learner problems with declension patterns. Broader language experience and patterns of Lithuanian use also influenced performance. Overall, the study’s results clearly support a usage-based view of SLA, according to which learners are sensitive to the relative frequency of linguistic items.

Another study that illustrates the difficulties adult L2 learners face with declensional paradigms is that of Stepšys and Kamandulytė-Merfeldienė (2024). Although the study also included bilingual children, the following discussion focuses exclusively on the adult L2 Lithuanian participants. The study examined difficulties in adjective agreement (matching adjectives and their head nouns in noun phrases for case, gender, and grammatical number), with the central aim of determining which grammatical categories pose the

greatest problems. Participants were tasked with selecting the correct adjectival ending in a sentence where the noun was already marked for case, gender, and number, and each incorrect option reflected a categorical error (i.e., gender, number, or paradigm).

The results showed few problems for the adult learners in encoding adjectives with the appropriate gender and grammatical number, but revealed considerable difficulty in selecting the correct paradigm. The learners primarily preferred productive, common endings over less frequent ones, a pattern also described by Bružaitė-Liseckienė (2021), and showed transfer of noun inflectional endings to adjectives. Consequently, adult L2 learners struggle more when noun and adjective endings do not align transparently and agreement requires selecting between multiple competing adjectival forms, especially when the target endings are marked. L1 influence was also observed, with learners with analytic L1s seemingly having more trouble marking adjectives than those with more synthetic L1s (e.g., Lithuanian). However, the latter group also demonstrated similar difficulties when faced with adjectival agreement. While the study is not highly informative about case marking *per se* (since the items were already marked for case), it once again illustrates the complexity of not only the nominal but also the adjectival morphological system of Lithuanian that L2 learners face.

Taken together, these three studies suggest that difficulties in L2 Lithuanian case acquisition are systematic rather than random. Across different learner groups and grammatical domains, errors repeatedly reflected the combined influence of frequency, salience, contingency, L1 transfer, and paradigm productivity. Learners tend to rely on more frequent or ‘default’ forms, experience difficulty with low-salience and semantically redundant markers, and struggle when multiple endings must be selected within a complex declensional system. At the same time, lexical factors such as verb-specific case government and broader language experience also shape performance, highlighting the need for extensive L2 input. Overall, the evidence indicates that acquiring L2 Lithuanian case requires learners to master not only abstract grammatical relations but also a highly variable network of form-meaning mappings and inflectional patterns.

2.2.2.2 Case acquisition in other L2s

Moving beyond L2 Lithuanian, Hinz et al.’s (2013) study on L2 Polish is particularly relevant, as it highlights the complex ways in which various factors relevant to L2 acquisition can interact. The participants’ L1 was either

French (a language without a case system) or German (a language with a limited case system). This selection was intentional, as the aim was to account for the potential influence of participants' L1 on their L2 learning process. The participants took part in a two-week intensive L2 Polish course, with a communicative teaching methodology and no explicit grammatical explanations. The study concentrated on the nominative and instrumental cases in the context of professions and nationalities. Depending on the construction, the profession or nationality is marked with the nominative (*To jest lekarz*, 'that is doctor.NOM', 'That is a doctor') or the instrumental (*On jest lekarz-em*, 'he is doctor-INST', 'He is a doctor'). It was predicted that all participants would have more difficulty with the instrumental than the nominative, as the nominative coincides with the base form. Moreover, the participants whose L1 was German were expected to perform better in Situation 2 than those whose L1 was French due to the influence of their respective L1 case systems.

The results, however, were unexpected. Although the French-speaking participants initially underperformed, they eventually caught up with their German-speaking peers, who showed minimal progress between the intermediate and posttests. Interestingly, all participants found instrumental constructions easier to learn than nominative constructions. The authors proposed two potential explanations for this outcome: contingency and salience. For masculine nouns, nominative endings differ considerably (e.g., *lekarz* 'doctor', *tłumacz* 'translator'), whereas the instrumental endings are uniform across nouns (*lekarzem*, *tłumaczem*). Participants seemingly preferred the consistent instrumental endings. For feminine nouns, salience appeared to be the more significant factor. Although the nominative endings for this gender are often consistent (*lekarka* 'female doctor', *tłumaczka* 'female translator'), one of the most distinctive features of Polish for L2 learners is the nasal sounds *q* and *ę*. The ending *-q* serves as the feminine instrumental ending (*lekarką*, *tłumaczką*), which likely made the instrumental case more salient and, consequently, facilitated its acquisition over the nominative. Clearly, multiple positive and negative factors can interact in various ways, thereby significantly shaping L2 learning outcomes.

A valuable finding provided by Hinz et al.'s (2013) study is the lack of L1 transfer that was established for the acquisition of case. This finding was further corroborated by Saturno (2023). Across three studies, Saturno investigated the acquisition of the accusative case in L2 Polish by learners with various Slavic and non-Slavic L1 and L2 backgrounds. The first study examined L2 Polish performance on the accusative case after limited exposure

by L2 Russian learners without another Slavic L1. Despite limited exposure and overuse of uninflected base forms, participants already produced accusative endings similar to those in Russian, indicating L2 transfer. The second study investigated the acquisition of the accusative in L2 Polish by L1 and L2 Russian speakers, again with only minimal Polish input. Both groups demonstrated overuse of the nominative base form, while only the Russian L1 speakers produced accusative endings found in Russian, but in Polish contexts. Study 3 was a large cross-sectional study of 161 participants that compared performance on the accusative case in L2 Polish in an Elicited Imitation Test between learners with (East) Slavic and non-Slavic L1s. The results showed higher accuracy among Slavic L1 speakers, although the acquisition order was largely the same across learners in both groups.

Based on the findings of all three studies, Saturno (2023) argues that the typological proximity of the L1 does not change the underlying developmental trajectory (as per Pienemann's (1998) Processability Theory). However, it can facilitate progress from one stage to the next. Learners in all three studies displayed a universal early stage characterized by overuse of the base form. These studies, however, do provide evidence of transfer in both L1 and L2 Russian users, as evidenced by Russian accusative endings on feminine singular nouns and by declining plural objects according to the Russian system. Furthermore, East Slavic L1 users demonstrated higher accuracy than non-Slavic L1 learners. Nevertheless, Slavic L1 and L2 users at lower levels displayed the same overuse of the base form as other learners, demonstrating limited effects of L1 transfer on the general acquisition process. Hence, while not entirely discarding L1 transfer, the study assigns a much smaller role to L1 influence than do the L2 Lithuanian studies described in the previous section.

Kempe and MacWhinney's (1998) study shows the importance of cue validity when acquiring L2 case. The study focused on the acquisition of morphological case marking in L2 German and L2 Russian by L1 English speakers. The goal was to test two competing explanations of inflectional learning: a rule-based/paradigm-learning approach, according to which learning difficulty stems from paradigm complexity; and an associative, input-driven approach, which predicts that learning depends on the cue validity of individual case forms. German and Russian are supposed to reflect these strategies, as German has a simpler morphological paradigm, with case markers being less reliable and more syncretic (which means that the ease of learning L2 German case would support the rule-based approach). At the same time, Russian has a more complex case system, which means that its case

markers are highly reliable cues for interpreting syntactic roles (suggesting that the ease of learning L2 Russian case would support the associative, input-driven approach). The experiment consisted of a computerized picture-choice comprehension task in which simple SVO sentences (e.g., *Die Mutter sucht den Sohn./Mat' iščet syna.* 'The mother is looking for the son.') were systematically manipulated for case marking, word order, and animacy cues.

The results indicate substantial support for the associative, input-driven learning approach. L2 Russian learners began relying on case endings much earlier than their L2 German peers, despite the system's greater complexity. L2 German learners preferred to interpret sentences based on the animacy of the constituents present. This suggests that learners attend first to strong, reliable cues in the input, and that paradigm complexity does not determine acquisition order. Based on this, Kempe and MacWhinney (1998) argue that the associative learning approach displayed by the L2 learners here is similar to that employed by children in their L1, a position that is at least partially supported by TBLT as well. Clearly, the results align better with the CREED model (advantage of high cue validity) than Skill Acquisition Theory (advantage of applying simpler rules). Furthermore, although morphological endings are considered less salient when presented verbally (N. C. Ellis, 2022), their salience appears heightened when encountered in writing. The results also seem to align with the IP theory (VanPatten, 2014), as syntactic roles are not expressed lexically; they are expressed only syntactically or semantically (animacy). At the same time, 'The Preference for Nonredundancy Principle' provides a strong case for Russian case markers. Nevertheless, the observation that paradigm complexity helps rather than disrupts the acquisition process essentially contradicts the studies discussed previously, although it must be noted that the number of case forms presented in this study was quite limited due to the short sentences in the items.

Another study offering valuable insights into preferred learning approaches for L2 case acquisition is by Kempe and Brooks (2008). They examined whether learners rely primarily on gender-based, case-based, or item-based strategies when acquiring preposition-case combinations in a simplified version of Russian. Crucially, the study contrasted transparent nominative markers of grammatical gender (e.g., *stakan* 'glass', *kniga* 'book') against opaque ones (e.g., *olen* 'deer', *mysh* 'mouse').⁶ Importantly,

⁶ In Russian, similarly to Lithuanian, noun gender can generally be established based on the last sound or letter of the noun. Nouns ending in a consonant are generally masculine, whereas those ending in *-a* or *-ia* are generally feminine gender. Words

however, as gender agreement was not part of the study, it could be argued that the nouns differed not necessarily in (in)transparent gender markers, but rather in (in)transparent paradigms.

Two experiments were conducted. In the first experiment, 43 participants were exposed to nouns with transparent endings and trained on nominative, genitive, and dative forms. The last two were in the presence of an appropriate preposition: *ot* for the genitive ('from'; e.g., *ot stakan-a* 'from the glass-GEN'), and *k* for the dative ('towards'; e.g., *k stakan-u* 'towards the glass-DAT'). In the second experiment, 44 participants learned nouns with opaque endings and were trained on locative and instrumental forms, whose endings vary by gender. Instrumental case followed the preposition *pod* ('under'; e.g., *pod olen-em* 'under the deer-M.INST'; *pod mysh '-yu* 'under the mouse-F.INST') and the locative followed the preposition *na* ('on'; e.g., *na olen-e* 'on the deer-M.LOC'; *na myshi* 'on the mouse-F.LOC'). The input was provided only through listening tasks, without written support.

The first experiment showed clear evidence of gender-based learning. Learners identified gender more accurately than case, and performance on novel nouns remained above chance, suggesting some degree of generalization beyond memorized items. Therefore, Kempe and Brooks (2008) argue that implicit memory-driven processes work alongside attention-based learning, potentially guiding learners toward a more explicit understanding of grammatical rules. However, case marking remained more difficult, indicating that exposure alone may not be sufficient for robust acquisition of inflectional morphology. This suggests that supplementary explicit teaching is necessary, although the amount required may be substantially lower for learning noun gender or, more precisely, inflectional paradigms.

The second experiment revealed a stronger preference for item-based strategies over case-based ones in the absence of transparent endings. This means that when endings were opaque, learners tended to memorize noun-form combinations rather than infer the broader case system. If learners had grasped the case system, errors involving markers of the correct case but the wrong paradigm (such as **pod mysh-em* 'under the mouse-M.INST', with a feminine noun but a masculine instrumental ending) would be expected, but this rarely occurred. Case errors again exceeded gender errors, suggesting that transparent paradigms facilitate pattern abstraction, whereas opaque

ending in a soft consonant or, in writing, the soft sign may be either masculine or feminine, rendering their gender opaque.

paradigms encourage item-specific learning. Nevertheless, as prepositional case assignment is essentially a purely grammatical form of case marking since prepositions themselves convey the semantic meaning, the results could look different when using cases with more semantic functions.

Overall, the outcomes of this study can be interpreted in two ways. They may indicate that learners require stronger explicit instruction in case, consistent with Skill Acquisition Theory (DeKeyser, 2014). Alternatively, they may suggest that the learners in the study had not yet been developmentally ready to acquire the target case forms, as predicted by Pienemann's Learnability Theory (1998). The latter explanation is plausible, given that none of the study participants had prior knowledge of Russian. Furthermore, the results from the second experiment underscore the role of input. Through repeated exposure to dialogs, the participants increasingly succeeded in associating nouns with the appropriate endings according to their grammatical gender. This interpretation echoes the CREED approach (N. C. Ellis, 2006) and the cognitive-interactionist model of TBLT (R. Ellis et al., 2019; Long, 2014).

The next study explores whether incidental learning, as proposed in the previous study and, at least partially, by approaches such as TBLT, is even feasible when acquiring L2 case. Kenanidis et al. (2023) conducted a study in which L1 English and L1 German speakers tried to acquire the inflectional morphology of an artificial L2 (Kepidalo) in a treatment consisting of five (for the L1 English speakers) and six (for the L1 German speakers) training sessions divided over 10 days without explicit grammar instruction. The L1 German group received a longer treatment to test the effects of increased incidental exposure. The results showed significant gains solely for vocabulary learning, with grammatical learning remaining weak and slow for both groups. Participants in both groups performed much better at detecting and using (canonical) word-order cues than at case marking. Both groups showed persistent difficulty with case morphology, relying instead on word order or probabilistic patterns. Despite difficulties, the L1 German group did demonstrate slightly superior performance in case marking than the L1 English group.

Nevertheless, a few caveats need to be highlighted. First, 10 days to acquire a case system, let alone implicitly, is a rather limited timeframe. Furthermore, the superior performance of the L1 German group should be interpreted cautiously, as it may reflect the longer treatment rather than solely L1 transfer. Interestingly, the L1 German participants, who took part in Experiment 2, also completed a metalinguistic awareness questionnaire. The

results showed that greater explicit awareness of grammatical rules was accompanied by superior performance. While this lends support to Skill Acquisition Theory and a more form-focused approach, such as TSLT, it also demonstrates that conscious reflection on form under incidental conditions is possible, as advocated in analytic teaching approaches (see Section 1.2.3). In sum, the authors claim that, while some grammatical knowledge can be acquired incidentally, this mode of learning is slow and limited. Extensive incidental exposure does not seem sufficient for robust acquisition of grammatical case, with metalinguistic awareness seemingly playing a more facilitative role. This suggests that TSLT could be more effective than the purely incidental TBLT for L2 case acquisition.

The final study to be discussed in this section casts doubt on the effectiveness of even extensive L2 instruction for the acquisition of grammatical case. Denissenko (2016) conducted a study in which 83 adult Spanish/Catalan L1 learners of L2 Russian, divided into four proficiency groups (A1-B1), participated in a 150-hour intervention, which was essentially an L2 Russian course at each appropriate level. The course followed a level-based communicative curriculum in an institutional classroom setting, where grammatical case was taught explicitly through a conventional progression of forms and functions. A pretest and posttest were administered before and after the intervention, respectively, and consisted of a written narrative task based on a picture sequence. Performance was measured using the previously described CALF measures, with accuracy analyzed in terms of case form use. The results were striking. In terms of syntactic and lexical complexity, and fluency, only the lowest-level group showed statistically significant progress from pre- to posttest after 150 hours of instruction. No such statistically significant progress was observed in the other groups, indicating a developmental plateau.

Even more striking were the results concerning case use: none of the four learner groups showed statistically significant improvement from pre- to posttest after 150 hours of intervention. This concerned the range of cases used by the learners as well as the quality and frequency of case errors, with case errors being frequent and persistent across all levels. The most frequent problems included overuse of the nominative (i.e., the base form), confusion between the accusative and genitive cases, omission of endings, and misuse after prepositions. Notably, besides the omission of endings, these problems

were also described by Dabašinskienė and Čubajevaitė (2009) for L2 Lithuanian.⁷

Based on her results, Denissenko argues that the classroom context most likely did not provide sufficient input and practice for teaching such a highly complex grammatical system as grammatical case. Communicative strategies of learners seem to stabilize, due to which grammatical refinement does not progress further, which reflects Slabakova's (2019) Bottleneck Hypothesis. Based on these results, it seems that case acquisition requires more than classroom exposure and instruction, even when it is quite extensive. If that is the case, neither TBLT nor TSLT can suffice as approaches to teaching grammatical case and must therefore be complemented by additional exposure and practice outside the classroom.

2.2.3 Order of acquisition

This section covers the order in which case is thought to be acquired. The discussion first considers case acquisition in the L1, as this could be considered the 'natural' order of case acquisition. In Savickienė (2003), L1 Lithuanian case acquisition was documented in the speech of a girl named Rūta. The (simplified) order of case acquisition established for her can be described as follows:

nominative > vocative/genitive > accusative > dative/instrumental/locative⁸

Interestingly, this acquisition order appears to mimic the case frequency found in spoken L1 Lithuanian, as illustrated above, in Table 4. Grammatical cases clearly occur more frequently than semantic ones, and this is generally the order in which Rūta acquired the different cases. That said, children's speech, as well as parents' communication directed to children, differs from 'normal' adult speech. This may be one of the reasons why the vocative appears early in L1 case development, as it is far more frequent in children's speech than in typical adult usage.

Case acquisition occurs gradually, not only between cases but also within individual cases. The 'verb islands' principle mentioned above also

⁷ In Russian, masculine nouns are characterized by zero-endings, making nouns without a morphological ending a typical phenomenon in Russian nominal grammar.

⁸ Both the genitive and accusative cases emerge around the same time (1;7); however, the genitive is much more frequent, with the accusative becoming substantially more frequent only around 1;8.

applied to Rūta, as she initially used the accusative only with verbs related to possession and transfer (e.g., *give*, *have*, etc.). She also showed a preference for clear, unambiguous endings: for instance, when marking a feminine noun in the instrumental, she sometimes replaced the ending *-a* (the correct ending for feminine nouns in both the nominative and the instrumental) with the masculine alternative *-u*. This has also been reported in, for example, L1 Russian (Gvozdev, 1961). Rūta's acquisition of case also occurred gradually. She started to use the genitive quite early, but at first only in its possessive function. The other genitive case functions, such as the partitive and negation, appeared later.

Interestingly, L1 acquisition of case in other languages appears to follow frequency as well. For L1 Russian, Gvozdev (1961) found the following simplified acquisition order: 'nominative > accusative/genitive > dative/prepositional > instrumental', which quite closely follows the case frequencies in L1 Russian speech reported by Zemskaya (1979): nom. 32.6%, acc. 25.3%, gen. 22.0%, prep./loc. 10.1%, instr. 5.0%, dat. 4.1%. The same applies to L1 Polish, for which the following approximate acquisition order has been proposed (Smoczyńska, 1985): 'nominative > accusative/genitive > dative/instrumental/locative', which closely matches the L1 Polish case frequencies found by Laskowski (1989, p. 212): nom. 34.2%, acc. 29.8%, gen. 19.2%, loc. 7.6%, dat. 4.8%, ins. 4.4%). Moreover, reacquisition of case for L1 Russian aphasia patients appears to follow a similar order (Akhutina, 2003; Ruigendijk, 2002). Initially, patients use only the nominative case, but over time they gradually reacquire constructions involving verbs. Once verbs reappear in production, only the nominative and accusative cases are distinguished first, followed by a gradual reintroduction of the genitive and prepositional cases, and eventually the dative and instrumental. This echoes reported L1 Russian case frequencies even more strongly. However, it is important to note that most of these established acquisition orders are heavily simplified, disregarding, for instance, grammatical number. Moreover, Savickienė's (2003) and Gvozdev's (1961) studies were based on the L1 acquisition of one child, making most of the observations mainly informative tendencies.

Returning to case acquisition in the L2, most work has focused on L2 German and Russian. First, Baten (2013) for L2 German and Artoni (2019) for L2 Russian formulated a case-acquisition order based on the Processability Theory (Pienemann, 1998). Both authors hypothesized that, after initial overuse of the nominative base form, learners first start to mark case positionally, i.e., in the canonical post-verb position, which entails the

accusative. Only thereafter are learners developmentally ready to assign case lexically and semantically, which is then followed by grammatical case marking, as can be seen in non-canonical sentence constructions. For L2 Russian, Artoni (2019) focused on the third procedural phrasal stage, as previously described in Magnani and Artoni's (2015) model (see Table 8), as well as verb phrases and prepositional phrases. To test this hypothesized case acquisition order, 15 learners of L2 Russian with diverse L1 backgrounds participated in story-retelling and picture-description tasks.

Based on the results, the following developmental sequence of case within verb phrases was established (Artoni, 2019, p. 126):

- i. Lack of case assignment: nominative.
- ii. Configurational case assignment: non-nominative/accusative in postverbal position.
- iii. Lexical case assignment: instrumental; semantic case assignment: dative.
- iv. Grammatical case assignment: accusatives to objects.

Interestingly, the use of the genitive was hypothesized to be involved in lexical case assignment. However, the only participant who used the genitive was already at the sentence developmental stage, thus indicating that the case was not part of the developmental trajectory of verb phrases here. Nevertheless, development in prepositional phrases displayed a similar pattern (p. 126):

- i. Lack of case assignment: nominative.
- ii. Configurational case assignment: non-nominative/prepositional.
- iii. Lexical case assignment: instrumental, dative, genitive, accusative.

Both these trajectories demonstrate a transition from positional to functional marking, as was hypothesized. The same result was found by Baten (2013) for L2 German. Importantly, however, Artoni (2019) emphasizes that such a cross-linguistic prediction holds only for cases involving grammatical functions. As lexical case assignment (which Artoni defines as cases that are governed lexically by a specific verb, preposition, or adjective) generally depends not on rules but on annotation in the lexicon, acquisition of lexical case is not only highly language-specific but also contingent on the type of input to which learners have been exposed and on learners' lexicon size.

Early evidence for a non-random acquisition order of L2 Russian case was provided by Rubinstein (1995a, 1995b). He interviewed 136 L1 English learners, eliciting oblique case use. The results suggested the order of 'accusative + locative (prepositional) > genitive & instrumental > dative'. As

this order differs somewhat from the L1 acquisition order reported by Gvozdev (1961), Rubinstein argued that this sequence likely reflects an interaction of frequency, formal, and semantic case complexity, and instructional order. Rubinstein (1995b) also performed an error analysis that distinguished three categories: Primary Form, use of the base form instead of accurate oblique case marking; Outside-the-Case, selection of a non-target oblique case; and With-the-Case, selection of the correct oblique case but with an ending belonging to a different declensional paradigm. The analysis showed that learners did not rely heavily on the (nominative) base form alone, but also produced case substitutions and declensional mismatches. This led Rubinstein to conclude that, while the number of case errors varied, the distribution of error types remained unchanged.

A more comprehensive account was later provided by Cherepovskaia et al. (2022). They used data from learners at six different proficiency levels (A1–C1). As some of the data matched Denissenko’s (2016), it included 192 Catalan–Spanish L1 learners of L2 Russian who received a 150-hour treatment. The pre- and posttest consisted of the same written narrative task as in Denissenko (2016), and the responses were analyzed for the distribution of correct case forms, error rates by case, and error types. They observed a clear acquisition order for L2 Russian case:

nominative > locative (prepositional) > accusative > genitive >
instrumental > dative

The locative/prepositional case is acquired early but initially serves only to denote locations. The genitive shows the most complex trajectory, most likely due to its many functions, meaning that progress across these functions is slow. The dative comes last and is the most problematic.

The results of Cherepovskaia et al.’s (2022) study show that case development is gradual and systemic. Learners do not simply reduce errors; instead, over time, the entire distribution of cases changes, through a decrease of the nominative, an increase of oblique cases, of which the overall profile approaches native-speaker usage. Rather than simply achieving higher accuracy with fewer errors, the entire system is restructured. That said, while the error rates for all oblique cases declined significantly over time, even C1-level learners demonstrated performance far removed from native-speaker case use, especially with the genitive, instrumental, and dative. The error analysis carried out in the study employed similar error types as discussed in Rubinstein (1995b). The results showed a strong reliance on the base form among lower-level learners, which gradually decreased with increasing

proficiency, accompanied by an increase in declension errors. Such errors have been described for L2 Lithuanian as well (Dabašinskienė & Čubajevaitė, 2009).

According to Cherepovskaia et al. (2022), the case acquisition described above is determined by two factors. The first factor is the necessity of a case, i.e., how essential it is for communication. This is especially relevant to the accusative, since it marks direct objects, one of the most frequent syntactic roles in sentences, and to the genitive because of its many functions (e.g., subject negation and case government by numerals and adverbs of quantity, as in Lithuanian). On the other hand, the dative is relatively rare in adult discourse, accounting for 4.1% of all case use (Zemskaja, 1979). Nevertheless, in this author's opinion, frequency in the input should not be confused with necessity in L2 output. While syntactic marking of direct objects might seem inevitable and necessary for a native speaker, L2 learners might still rely on syntactic positions to express it instead of the accusative case. Despite allowing for other types of syntactic constructions, this also seems likely in L2 Russian, since the canonical sentence structure is SVO. As noted previously, similar initial positional learning strategies have been observed by Artoni (2019).

The second factor guiding the acquisition order is complexity, both morphological (the number and variability of endings, including syncretism) and semantic (the number and heterogeneity of functions). For instance, the locative appears early not only because it is introduced early in textbooks (although this alone does not guarantee successful acquisition), but mainly because it is relatively simple both morphologically (consisting of essentially one ending in the singular) and semantically (a single salient function). The remaining locative functions and forms appear later. Conversely, the genitive is acquired quite late and slowly since it is exceptionally multifunctional (similar to Lithuanian) and morphologically complex (it is the only case in Russian that includes not only separate endings between paradigms in the singular but also in the plural). Based on this, Cherepovskaia et al. (2022) argue that these two factors create a trade-off effect. In summary, all the results indicate that L2 case acquisition in Russian is not random but follows a systematic order that differs from L1 Russian case acquisition. Moreover, while L2 case is acquired as a system from the beginning of learning, the system matures very slowly.

In summary, both Artoni's (2019) and Cherepovskaia et al.'s (2022) findings suggest that L2 Russian case acquisition begins with marking the locative/prepositional case and/or the accusative. The emergence of the genitive is still under debate, with Artoni including it only in prepositional

phrases, together with the accusative, instrumental, and dative, while Cherepovskaia et al. (2022) argue that it emerges after the acquisition of the accusative, albeit maturing only gradually and slowly. Although in Artoni's developmental sequence the instrumental and dative emerge simultaneously with the genitive in prepositional phrases and even precede it in verb phrases, Cherepovskaia et al. (2022) claim that these cases are acquired only after the emergence of the genitive.

Although these findings might inform L2 Lithuanian case acquisition, one obvious caveat is that the Russian and Lithuanian case systems, while sharing many similarities, remain quite divergent. Key differences include the absence of prepositions to mark the locative in Lithuanian and the lack of animacy marking for direct objects in Lithuanian. In the absence of studies investigating the order of L2 Lithuanian case acquisition, analyses could follow the L1 order; however, the studies above have clearly shown that L1 and L2 case acquisition do not necessarily follow the same developmental path. Frequency of use, on the other hand, might be a better guide for predicting the order of L2 case acquisition, as shown in Table 4 for spoken Lithuanian. If so, L2 case acquisition would follow the order of:

nominative > genitive > accusative > dative > instrumental > locative

Instructional materials for teaching L2 Lithuanian should then try to follow this order, starting with the more grammatical cases (nominative, genitive) and proceeding to the more semantic cases (dative, instrumental, locative). Nevertheless, the necessity of a case or a case function in terms of L2 learners' needs can obviously be at odds with such an order. For instance, L2 learners are likely to want to express where they live quite early when learning Lithuanian, thus necessitating knowledge of the locative, despite its relatively low frequency in L1 speech and its predicted late acquisition according to this hypothesized L2 Lithuanian case acquisition order.

2.2.4 Case acquisition in a morphologically complex language through TBLT and TSLT instruction

TBLT has often been criticized for prioritizing meaning at the expense of grammatical development (R. Ellis et al., 2019). However, discussions on focus on form(s) and the place of explicit grammar instruction in Section 1 show that grammar does have its place in TBLT. This is also illustrated by a large body of research formed over the years. For example, Shintani and R. Ellis (2010) and Shintani (2015) found that young Japanese learners developed knowledge of English singular and plural distinctions, particularly

plural -s, through tasks that required them to interpret these forms accurately in the input. Conversely, semantically redundant forms, such as copula like *be*, proved less receptive to incidental acquisition through tasks. Erlam and Ellis (2018) also found successful learner acquisition of singular and plural distinctions through input-based tasks, but for L2 French. Studies such as those by R. Ellis, Loewen, and Erlam (2006) and Yılmaz (2013) found that corrective feedback provided during or after task completion can enhance grammatical accuracy. Other output-based task research, such as Mackey (1999) and Spada and Lightbown (1999), revealed that learners can progress along developmental acquisition sequences through participation in communicative tasks.

Despite the large research base, research on L2 case acquisition through TBLT remains very limited. To the best of this author's knowledge, Markina (2018) is the only study to date that has investigated the application of both TBLT and TSLT in teaching a morphologically complex L2, Russian, with a (partial) focus on case acquisition. Besides case, the other target construction was verbs of motion, and the results were also analyzed along the CALF parameters described previously. Four L2 Russian courses were conducted, which differed along two dimensions: target construction (grammatical vs. motion verbs) and teaching methodology (TBLT vs. TSLT). All participants completed a pretest, an immediate posttest, and a delayed posttest (one or three months after the course ended). Each test consisted of a writing and speaking task, as well as three grammar tests. Nineteen participants in the TBLT group and twenty participants in the TSLT group took part in the courses focused on cases and prepositions after having already completed 80 hours of L2 Russian instruction. Each course lasted approximately 14 hours and spanned four weeks, during which participants completed seven tasks aimed at fostering case acquisition, such as creating a daily menu for six people and planning an excursion for a group of tourists in Barcelona.

Regarding case acquisition, the results were interesting but somewhat disappointing. Both groups performed better on the written posttest than on the pretest. However, their performance declined substantially on the delayed posttest, returning to approximately pretest levels. Similar patterns were observed for the results of the three grammar tests. The oral tests, however, revealed a different pattern. The TSLT group's performance declined on the immediate posttest but improved again on the delayed posttest, whereas the TBLT group showed consistent improvement, although this was unfortunately not statistically significant. Statistically significant differences were found in

the TBLT group's written pre- and posttest results, as well as in the decline between the immediate and delayed posttests, indicating that performance returned to the baseline pretest level. Moreover, no statistically significant difference was found between the two groups overall. It is also noteworthy that, for the written task, the TBLT group performed better on the posttest than on the pretest and also better than the TSLT group, although this advantage was not maintained. A similar pattern emerged in one of the grammar tests, suggesting that long-term implicit case knowledge was not achieved (Markina, 2018).

Across other parameters, participants showed some progress in almost all areas, but no significant differences emerged between the TBLT and TSLT groups. According to Markina, this discrepancy may stem from the extremely short duration of the courses: the longer the L2 instruction, the more pronounced the difference between teaching methods. She therefore recommends organizing courses that are as long as possible. On the other hand, the TBLT group performed significantly better in two areas: lexical complexity and the use of verbs of motion. To explain why the TBLT group was more successful in mastering verbs of motion than cases, Markina (2018) draws on VanPatten's (2014) IP: verbs of motion are semantically more marked than grammatical cases and were therefore easier for learners to acquire.

Moving back to case acquisition, Markina's study suggests that effects of instruction can be achieved, but only in the short term. Retention was not observed, regardless of the task or the teaching approach, suggesting that any learning gains were likely explicit and therefore short-term rather than implicit, which would be expected to reflect actual acquisition and long-term retention. However, the data did not allow for a clear assessment of the attainment of explicit and implicit knowledge. Nevertheless, tangible learning gains after only 14 hours of treatment are a big difference from the results reported by Denissenko (2016), despite the lack of retention. In terms of treatment, however, both TBLT and TSLT yielded very similar results in case acquisition.

Overall, while Markina's (2018) study demonstrated a short-term instructional effect on L2 case acquisition, which contradicts Denissenko's (2016) findings, the results still raise several questions. It remains unclear whether long-term retention of acquired case knowledge is possible and whether an instructional effect can be discerned. This requires further study and could be achieved in a few ways. First, Markina (2018) herself mentioned the limited duration of the treatment, which should therefore be extended.

Second, Markina (2018) found immediate learning gains after having narrowed down the scope of the target structures, i.e., from the entire case system in Denissenko (2016) to prepositional case government. Finally, as noted above, Markina (2018) did not clearly distinguish between explicit and implicit learning gains of case. Because TBLT is said to primarily foster implicit knowledge and TSLT is thought to mainly promote explicit knowledge, a research design that focuses on acquiring both implicit and explicit knowledge could help uncover instructional differences.

3 STUDY DESIGN AND METHODOLOGY

Having established the theoretical and empirical background, this section describes the experiment conducted in this dissertation. First, the study design and procedure are described. This is followed by information on the participants. Next, the section outlines the specific target items and justifies their selection. Subsequently, a description of the experimental intervention employed and the rationale behind the test instruments used to evaluate the hypotheses is provided. Finally, the relevant transcription, coding, and statistical analyses are laid out.

3.1 Design and procedure

The study in this dissertation employed a *pretest–treatment–posttest–delayed posttest* design (see Figure 5). The study comprised two treatment groups and a control group. The pretest was administered to the treatment groups before the treatment began. After the treatment, which lasted around six weeks, a posttest was administered. After a six- to seven-week break, the delayed posttest was completed by the same participants. The majority of the tests were completed in groups in lecture halls at Vilnius University. The writing portion of each test was conducted in the same classroom under the researcher’s supervision. The oral part was conducted individually with each participant in a separate classroom, under the supervision of the researcher or a colleague. If a participant could not attend the group test sessions, they either completed them online on the *Microsoft Teams* platform or in person in a separate classroom at Vilnius University with the researcher himself. The procedure for the control group differed slightly: only the pre- and posttests were administered, and testing only took place individually, either on location at Vilnius University or online via *Microsoft Teams*. On average, there was a 4- to 5-hour break between the pre- and posttest.

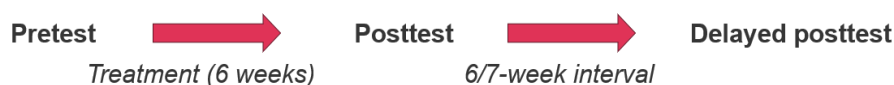


Figure 5. The design of this study

The treatment was essentially an L2 Lithuanian language course comprising twelve 90-minute classes, for a total of 18 hours per group. Participants attended classes twice a week at Vilnius University. To keep class sizes manageable, the TBLT and TSLT groups were each divided into two,

resulting in four instructional groups in total. All groups received the treatment simultaneously.

The researcher served as the teacher for all groups. It should be noted that the researcher preferred the TBLT over the TSLT approach, which may have led to superior teaching performance in the TBLT groups. To minimize this potential source of bias, all task materials were carefully designed with a fairly rigid structure. Additionally, detailed lesson plans were prepared for both task versions to ensure consistent teaching quality across all groups (see Appendix B). Lastly, although the researcher preferred the ‘purer’ TBLT style of teaching, it remained an open question whether the TBLT or TSLT task versions would be more conducive to grammar teaching; hence the present study.

The course originally consisted of five tasks, each lasting four hours. However, only the two TBLT instructional groups completed at least part of the fifth task. The two TSLT instructional groups completed only four of them. Nevertheless, since the treatment duration was the same across both treatment groups, this did not affect the interpretation of each group’s learning gains. While the main tasks were essentially the same for both the TBLT and TSLT groups, the TSLT task versions placed greater emphasis on grammar by highlighting morphological case forms in the input and including additional grammar exercises in the pre-task phase. These treatment differences are elaborated below. Naturally, the control group received no treatment and completed only the pre- and posttests.

Additionally, a pilot study was conducted to assess the suitability of some of the teaching materials and testing instruments. The pilot study involved two separate groups, both already enrolled in an L2 Lithuanian class at VDU Education Academy in Vilnius. Similar to the current final study, one was a TBLT group, while the other was a TSLT group. They completed only one task. Further details and findings of this pilot study are described in Udes (2025a). Additionally, the tests and other materials were further piloted with other students at VDU Education Academy, as well as individual students of the researcher. Piloting of the materials and tests informed several design choices, some of which are described in the sections below.

3.2 Participants

A total of 60 participants took part in this study. They were all adult learners of L2 Lithuanian and participated voluntarily in this experiment. Prior to completing the pretest and starting the course, all participants received an

information sheet detailing the researcher, the study, the course content, and the class and test schedule. The aim of the research was described as exploring the teaching and learning of Lithuanian through tasks; therefore, the target constructions were not mentioned. Participants also signed a consent form (see Appendix A).

The course was advertised as a short, practical Lithuanian language course at the A1–A2 level according to the *European Framework of Reference for Languages* (CEFR; Council of Europe, 2020). No tests were administered to assess the participants' levels, yet all participants appeared to be beginner learners with varying degrees of prior knowledge of Lithuanian. During the course, the materials seemed appropriate for all participants. Participants were randomly assigned to two treatment groups, TBLT (Task-Based Language Teaching; $n = 23$) and TSLT (Task-Supported Language Teaching; $n = 22$), and a control group ($n = 15$). The TBLT group received the treatment with a lower focus on grammar, while the TSLT group's treatment included focus on form(s) at all task stages. The control group received no treatment.

As can be seen in Table 9, the treatment groups were quite comparable. The number of participants, age range, average age, and gender distributions were all very similar. The distribution of L1s in terms of Balto-Slavic and non-Balto-Slavic languages was equally similar. This division between Balto-Slavic and non-Balto-Slavic languages was based on typological similarities not only across Baltic languages but also across Baltic and Slavic languages. Although Baltic and Slavic languages differ considerably in their vocabulary, they share many grammatical similarities, including extensive nominal case systems, with the notable exception of Bulgarian and Macedonian. As for comparisons with the control group, some differences are evident. The control group included fewer participants, a slightly different gender distribution, a lower average age, and fewer participants with a Balto-Slavic L1. These characteristics need to be taken into account when interpreting the data that include the control group. Lastly, only a small number of participants across all groups reported proficiency in typologically similar Slavic L2s. These participants typically had a Slavic L1.

Despite the many similarities, the groups were not entirely equal in prior experience with L2 Lithuanian. The average time spent learning L2 Lithuanian (in months) was 16.21 ($SD = 29.70$) for the TBLT group, 11.05 ($SD = 11.65$) for the TSLT group, and 17.5 ($SD = 14.61$) for the control group. The large standard deviations, however, indicate high variability between participants, especially in the TBLT group. The average length of residence

Table 9. Group and participant details of the study

Course	Task-Based			Task-Supported			Control		
Participants	23	Male: 14	Female: 9	22	Male: 13	Female: 9	15	Male: 6	Female: 9
Age	23–56	Average: 33.1		20–57	Average: 35.8		23–60	Average: 26.3	
L1	8 Balto-Slavic: Belarusian: 1 Bosnian: 1 Latvian: 1 Russian: 4 Ukrainian: 1	15 Non-Balto-Slavic (Catalan, Danish, English, French, Greek, Italian, Spanish, Turkish, Yoruba)		7 Balto-Slavic: Belarusian: 1 Bulgarian: 1 Croatian: 1 Russian: 2 Ukrainian: 2	15 Non-Balto-Slavic (Azerbaijani, Dutch, English, French, German, Italian, Spanish, Tamil, Turkish, Urdu)		3 Balto-Slavic: Russian: 1 Ukrainian: 1 Russian & Ukrainian: 1	12 non-Balto-Slavic (Arabic, Bengali, English, French, Italian, Kazakh, Spanish, Turkish, Vietnamese)	
Level	≈ A1-A2			≈ A1-A2			≈ A1-A2		
Time spent learning Lithuanian	Mean: 16.21 months (SD = 29.70)			Mean: 11.05 months (SD = 11.65)			Mean: 17.5 months (SD = 14.61)		
Residence in Lithuania	Mean: 42.62 months (SD = 33.73)			Mean: 32.71 months (SD = 30.28)			Mean: 36.4 months (SD = 17.94)		

in Lithuania also differed across groups: participants in the TBLT group had lived in Lithuania for an average of 42.62 months ($SD = 33.73$), compared with 32.71 months ($SD = 30.28$) for participants in the TSLT group, and 36.40 months ($SD = 17.94$) for those in the control group. Similarly, standard deviations show a wide range across participants, although it was notably lower in the control group than in the treatment groups. Bayesian one-way ANOVAs were performed in *R* (R Core Team, 2023) using the *BayesFactor* package (Morey & Rouder, 2026) to assess whether the three groups differed substantially in time spent learning Lithuanian and length of residence in Lithuania. The analyses yielded moderate evidence for the null model over the group model, both for time spent learning ($BF_{01} = 4.63$) and length of residence ($BF_{01} = 4.50$) (Wagenmakers et al., 2011). This nevertheless suggests no substantial differences between the groups on either measure.

After completing the posttest, the participants in the control group were asked individually whether they had participated in any formal L2 Lithuanian instruction or studied Lithuanian in any other way between the two test times. Four of the 15 control group participants reported not having deliberately studied any Lithuanian; six stated they had only received passive L2 exposure from their surroundings; and the remaining five acknowledged active L2 Lithuanian learning between meetings, either individually or with a teacher. This should be considered when evaluating the results, as it could have affected post-test performance.

3.3 Target grammatical forms

By narrowing down the nominal case system and selecting a specific function involving only a limited set of cases, not only immediate but also lasting learning gains in L2 case acquisition could be expected. At the same time, the selected function must present a sufficient challenge for self-reported A1–A2 level learners, which led to the decision to make subject and direct-object marking the target structures of this study. Subjects and direct objects are among the most frequent syntactic roles, and marking them involves the three most frequent cases: the nominative, the genitive, and the accusative. Judging from the L2 case acquisition orders described in Section 2.2.3, it makes sense to have low-proficiency learners primarily focus on the most grammatical cases. Furthermore, Dabašinskienė & Čubajevaitė (2009) report frequent mistakes through overuse of the nominative and incorrect case assignment in sentences containing a negated verb.

(7) *But-e* *yra* *balkon-as*.
apartment-LOC is balcony-NOM
'There is a balcony in the apartment.'

(8) *But-e* *nèra* *balkon-ø*.
apartment-LOC NEG.is balcony-GEN
'There is no balcony in the apartment.'

(9) *But-e* *nèra* **balkon-as*.
apartment-LOC NEG.is balcony-NOM
'There is no balcony in the apartment.'

(10) *But-as* *turi* *balikon-q*.
apartment-NOM has balcony-ACC
'The apartment has a balcony.'

(11) *But-as* *ne-turi* *balikon-q*.
apartment-NOM NEG-have balcony-GEN
'The apartment does not have a balcony.'

(12) *But-as* *ne-turi* **balikon-as/q*.
apartment-NOM NEG-have balcony-NOM/ACC
'The apartment does not have a balcony.'

It is worth noting that the word form that L2 learners typically encounter in dictionaries and word lists is the nominative singular (Savickienė, 2005). This form is generally considered unmarked and a word's base form. It is generally assumed that it serves as the foundation for both L1

Table 10. Overview of the cases and forms relevant to this study’s target constructions

Case	Declension 1		Declension 2		Declension 3	Declension 4		Declension 5	
	-as	-is (-ys)	-a	-ė	-is	-us	-ius	-uo	sesuo ‘sister’/ duktė ‘daughter’
Singular									
Nom.	-as	-is	-a	-ė	-is	-us	-ius	-uo	-uo, ė
Gen.	-o	-io	-os	-ės	-ies	-aus	-	-en-s	-er-s
							iaus		
Acc.	-ą	-į	-ą	-ę	-į	-ų	-ių	-en-į	-er-į
Plural									
Nom.	-ai	-iai	-os	-ės	-ys	-ūs	-iai	-en-ys	-er-ys
Gen.	-ų	-ių	-ų	-ių	-ų/-ių	-ų	-ių	-en-ų	-er-ų
Acc.	-us	-ius	-as	-es	-is	-us	-ius	-en-is	-er-is

and L2 acquisition (Gagarina & Voeikova, 2009). Due to its unmarked nature and frequent exposure in L2 learning materials, L2 Lithuanian learners tend to overgeneralize the nominative case in their output (Dabašinskienė & Čubajevaitė, 2009).

Besides selecting the appropriate case, successful case marking requires L2 learners to also identify the declension pattern a noun belongs to. Table 10 gives an overview of the cases and forms relevant to the target structures of this study. While Lithuanian features five main declension patterns, for L2 learners, this typically means acquiring seven distinct paradigms, with several more that overlap. For example, the nouns *balkonas* ‘balcony’ and *brolis* ‘brother’ both belong to the same declensional pattern but to different paradigms, resulting in similar yet different endings (e.g., *balkono* with the genitive ending *-o* and *brolio* with the genitive ending *-io*). In contrast, *diena* ‘day’ is a member of a different declension pattern, including the genitive ending *-os* (*dienos*) and, in this case, the identical accusative ending *-ą* (*dieną*).⁹ Difficulties in correct paradigm and/or declensional pattern selections can lead to errors such as the following:

⁹ The nouns *balkonas* ‘balcony’ and *diena* ‘day’ belong not only to different declension paradigms but also differ in grammatical gender. The noun *balkonas* is masculine, whereas *diena* is feminine. However, when adjective or participle

- (13) *But-as* *ne-turi* **balkon-os*
apartment-NOM NEG-have balcony-GEN
‘The apartment does not have a balcony.’

Both Dabašinskienė and Čubajevaitė (2009) and Bružaitė-Liseckienė (2021) have reported such difficulties experienced by learners.

3.4 Experimental intervention

The following section outlines the experimental treatment the participants received. First, the rationale behind the tasks and their sequencing is explained. Second, an overview of the task contents is provided, clarifying the types of activities in which students engaged during the L2 Lithuanian classes. Finally, the differences in treatment between the TBLT and TSLT groups are described.

3.4.1 Course and task design

3.4.1.1 Selection of the appropriate target tasks

In the TBLT literature on course design, it is commonly argued that the process should start with a needs analysis, which involves asking course participants about their L2 communicative needs. Based on these needs, target tasks are identified: the tasks that course participants aim to perform in the L2. Pedagogical tasks, i.e., simplified versions of real-life tasks, are then created to address the communicative needs reflected in these target tasks. The two proposals discussed previously for adapting TBLT to the teaching of morphologically complex languages likewise emphasize the importance of aligning students’ L2 needs with course content. Nevertheless, for experimental design, it is not feasible to first form a group of participants, perform a needs analysis, and only then develop the experimental treatment materials (Long, 2014). Therefore, a set of tasks was devised that would likely be relevant to most immigrants in a foreign country, even after having already lived in Lithuania for some time.

Another important consideration during the task and course design was the potential for natural use of the target structures. As mentioned above,

agreement and pronominal deixis are not relevant in a context, grammatical gender does not independently influence the case-marking process. Hence, gender can be essentially disregarded under these conditions, as the different endings primarily reflect differences in declension pattern/paradigm.

these included subject marking, which involves the nominative in affirmative sentences and the genitive in negated ones, and direct-object marking, which requires the accusative in affirmative sentences (as in (14)¹⁰) and the genitive in negated ones. While subject and direct-object marking in both affirmative and negated sentences occur frequently in regular everyday speech, the need to produce such sentences would have to emerge naturally from task requirements. Another grammatical challenge concerns the use of the genitive not only after negated verbs but also after adverbs of quantity, as in (15), and after specific verbs governing the genitive case, as in (16).

- (14) *Lietuviai dažnai valgo bulv-es.*
 Lithuanians often eat potato-PL.ACC
 ‘Lithuanians often eat potatoes.’
 [regular accusative marking on the direct object]
- (15) *Lietuviai valgo daug bulv-ių.*
 Lithuanians eat many potatoes-PL.GEN
 ‘Lithuanians eat a lot of potatoes.’
 [genitive marking due to *daug* ‘many’]
- (16) *Lietuviai dažnai nori bulv-ių.*
 Lithuanians often want potatoes-PL.GEN
 ‘Lithuanians often want potatoes.’
 [genitive marking due to the verb *norėti* ‘want’]

Another challenge is the genitive’s partitive function. Both subjects and direct objects can be marked with the genitive when expressing an indefinite quantity or, more specifically, “indefinite nonincremental quantification” (Laugalienė, 2020).

- (17) *Mūsų šaldytuve yra bulv-ių.*
 our refrigerator are potatoes-PL.GEN
 ‘There are (some) potatoes in our refrigerator.’
- (18) *Vakar nusipirkome bulv-ių.*
 yesterday we bought potatoes-PL.GEN
 ‘We bought (some) potatoes yesterday.’

Since these functions of the genitive are conceptually different from solely denoting negation, as in the target structures, high frequencies of these functions in the treatment were deemed undesirable. While such instances of

¹⁰ Examples (14) to (18) were invented by the researcher for explanatory purposes.

the genitive are practically unavoidable, task selection and design were informed by these issues to partially circumvent them.

Based on these considerations, the following target tasks were selected and developed into pedagogical tasks:

1. Expressing your opinion on different residential areas
2. Comparing rental apartments and selecting the most suitable one
3. Deciding what furniture to buy
4. Planning a menu for a dinner party
5. Shopping for groceries for a party

These five tasks were intended to represent the communicative contexts and situations immigrants often encounter when settling into a new country. When searching for accommodation, they would surely consider their preferred neighborhood before choosing a suitable place to live. As apartments and houses may lack certain furniture, they would likely need to purchase it at furniture stores. After having settled in, they might want to invite friends and colleagues to their new home for a housewarming party. In Task 4, this involves organizing a dinner party and setting a menu. Task 5 then expands upon this by requiring learners to agree on a grocery list for their housewarming party. Table 11 presents the steps that the participants followed when performing the main task. The pre-task and post-task phases are described in the next section.

Naturally, the tasks were designed to adhere to the four main criteria defined by R. Ellis et al. (2019). To recap, tasks must (1) focus on meaning, (2) include an informational ‘gap’, (3) require students to rely on their own resources, and (4) have a clear communicative goal. As can likely be inferred from Table 11, criteria 1 and 4 are clearly met. As for criterion 2, informational gaps were ensured by providing each student with different information (e.g., two separate weekly ads in Task 5) or by requiring a joint decision. Regarding criterion 3, although the pre-task stage provided examples and prompts for relevant L2 use, participants were free to use this input during the completion of the main task, thereby meeting this criterion as well.

Table 11. Overview of steps to be taken for each Main Task

Task	Main task steps
<i>1. Expressing your opinion on different residential areas</i>	- In pairs, each student looks at a simplified map of different residential areas; pairs ask and answer questions about it. - Pairs list the advantages and disadvantages of each residential area and express their preference.
<i>2. Comparing rental apartments and selecting the most suitable one</i>	- In pairs, each student looks at a simplified listing of a different apartment; pairs ask and answer questions about them. - Pairs decide which apartment they prefer and provide several reasons for their choice.
<i>3. Deciding what furniture to buy</i>	- Each pair looks at a picture of a living room and discusses what pieces of furniture are needed. - Each student has the inventory list of a different store, which they cannot show to each other. Together, the students orally exchange what items are available in each store. - Each pair can only afford four pieces of furniture. Students discuss and decide which they would like to buy and write down the agreed-upon list of items.
<i>4. Planning a menu for a dinner party</i>	- In pairs, each student looks at the information on the food preferences and allergies of the guests they are planning to invite. They provide this information together with their own food preferences and allergies. - Pairs decide on an appropriate appetizer for their dinner party. They can come up with their own dish or choose one from a list of suggestions. Ingredients can be swapped out if necessary. - Pairs decide on an appropriate entrée for their dinner party. They can come up with their own dish or choose one from a list of suggestions. Ingredients can be swapped out if necessary. - Pairs decide on an appropriate dessert for their dinner party. They can come up with their own dish or choose one from a list of suggestions. Ingredients can be swapped out if necessary.
<i>5. Shopping for groceries for a party</i>	- No sequence of steps specified. - Pairs have information on different food preferences and allergies of some of their guests.

Task	Main task steps
	- Each student receives the weekly advertisement from a different supermarket. - Pairs have a budget of 90 euros. - Pairs have to decide on the following purchases: - o snacks - o drinks - o candy/sweets - o other products

3.4.1.2 Alignment of task demands with the CEFR

As the course's target proficiency level was set at A2, task demands were designed to align with the A2-level requirements outlined in the CEFR (Council of Europe, 2020). Since all tasks primarily require oral interaction between participants, only level descriptors within this dimension were considered, specifically those included in the scales for 'Informal discussion (with friends)', 'Goal-oriented co-operation', and 'Information exchange'. The descriptor scale for 'informal discussion' specifies the following at the A2.2 level (p. 75):

- Can exchange opinions and compare things and people using simple language.
- Can make and respond to suggestions.
- Can agree and disagree with others.

All five tasks involve exchanging opinions and comparing alternatives. For instance, in Task 1, learners compare two residential areas and express their opinions on both. Agreeing and disagreeing become relevant from Task 2 onward, as learners need to come together to decide which apartment to live in. Learners can also make and respond to suggestions in most tasks, although this is particularly relevant from Task 3 onward, since these tasks include more options (such as multiple types of furniture in Task 3).

The 'Goal-oriented co-operation' scale highlights these skills (p. 77):

- Can understand enough to manage simple, routine tasks without undue effort, asking very simply for repetition when they do not understand.
- Can discuss what to do next, making and responding to suggestions, and asking for and giving directions.

According to the fourth criterion, a TBLT task must always include a communicative goal. Because all five tasks require pairs to reach a concrete

outcome (primarily a joint decision on what to choose), learners must negotiate what to do next and coordinate their actions, which closely aligns with the CEFR's 'goal-oriented co-operation' scale. While not all tasks can necessarily be considered routine, unless exchanging opinions and comparing are taken as tasks in an abstract sense, they are confined to two routine thematic domains (living conditions and food), thereby reducing the conceptual load. Furthermore, the pre-task stages are designed to familiarize learners with the target task, thereby scaffolding comprehension.

The scale describing 'information exchange' presents these relevant skills (p. 79):

- Can understand enough to manage simple, routine exchanges without undue effort.
- Can deal with practical everyday demands: finding out and passing on straightforward factual information.
- Can ask and answer questions about plans and intentions.

As per the third task criterion, tasks must include an informational gap, meaning that, in each task, each member of a pair receives slightly different information. Pairs need to share this information to achieve the task's intended communicative goal. The information in these supporting materials is usually quite visual, with larger amounts of text only in later tasks (particularly in Task 5; see Appendix C). Alternatively, explaining this type of visual and written input to a partner can be understood as mediation, which also aligns with the A2-level descriptor (e.g., relaying simple information and clarifying meaning). In this way, reading demands were reduced, and information exchange could take place without undue effort, especially in the initial stages.

As for grammatical accuracy, the CEFR (Council of Europe, 2020, p. 132) specifies the following for the A2-level:

- Uses some simple structures correctly, but still systematically makes basic mistakes; nevertheless, it is usually clear what they are trying to say.

Clearly, messages produced by L2 learners at the A2 Level may be clear, but they are likely to include basic mistakes. While the focus of this study is the acquisition of specific grammatical constructions, this is relevant only to the test instruments designed for the experiment. In TBLT, learners' task performance should always be evaluated according to the achievement of the intended communicative outcome, regardless of grammatical mistakes. This view is consistent with the CEFR's characterization of A2 learners as producing systematic basic errors while still achieving communicative clarity and, thus, a successful task outcome.

3.4.1.3 Inclusion of the target structures in the tasks

Completion of this study's set of tasks was expected to elicit natural use of the target structures. While completing the main task might not necessarily motivate students to use the target structures (since they have not been acquired yet or are not salient enough for students to attempt to use them), ample examples of these constructions can still be incorporated into the input without resorting to extensive input flooding. The constructions could then be addressed during explicit grammar instruction at the post-task stage.

Table 12 shows which tasks include which target constructions, along with other grammar points deemed relevant to each task. This could also be regarded as the minimal code for each task, as described in Section 1.3.3.1. While it was thought that many participants would likely infer from the tests and materials that case was at least one of the main focuses of this study, it was considered essential not to disclose the exact target structures; hence, the inclusion of other relevant grammar topics in each task.¹¹ These additional grammar topics always featured in the input and were therefore thought to be useful for the completion of each task. This also increased the ecological validity of the study, as regular TBLT courses would surely include these grammar points during focus-on-form episodes and post-task activities. It must be noted that the TSLT task versions sometimes featured these additional grammar points in the pre-task stages as well.

The inclusion of such a wide array of grammatical target structures might make it seem that participants were primarily tasked with applying them correctly during task performance, which would violate Ellis et al.'s (2019) first criterion, i.e., focus on meaning. The participants, however, were never explicitly required to use these specific constructions during the main task, and success was assessed solely based on whether they achieved the intended communicative outcome. This ensured that the task structure still prioritized comprehension and the expression of ideas in the L2 and, thus, did not violate Ellis et al.'s (2019) first task criterion. Nevertheless, it could be argued that the design choice to evoke specific target structures required for communication in a given context makes the tasks more focused. As described

¹¹ The participants' responses in a post-course survey actually confirmed these expectations: many participants mentioned grammatical case as one of the grammatical constructions focused on in this study, but none specifically mentioned subject and direct-object marking. Nevertheless, participants most often mentioned the genitive, and two responses noted negation as a point of focus.

Table 12. Overview of target structures and other additional grammar points featured in each task

Task	Target structures featured	Examples of target structures	Additional grammar points
1. <i>Expressing your opinion on different residential areas</i>	- Nominative subject marking. - Nominative subject marking with numerals. - Genitive subject marking in negated sentences.	- <i>Žvėryne yra ligoninė.</i> in Žvėrynas is hospital-NOM 'There is a hospital in the Žvėrynas neighborhood.' - <i>Žvėryne yra trys parduotuvės.</i> in Žvėrynas is three store-NOM.PL 'There are three stores in the Žvėrynas neighborhood.' - <i>Žvėryne nėra bibliotekos.</i> in Žvėrynas NEG.is library-GEN 'There is no library in the Žvėrynas neighborhood.'	- Masculine and feminine numerals. - Present tense conjugation.
2. <i>Comparing rental apartments and selecting the most suitable one</i>	- Nominative subject marking. - Genitive subject marking in negated sentences.	- <i>Bute yra koridorius.</i> in apartment is hallway-NOM 'There is a hallway in the apartment.' - <i>Bute nėra koridoriaus.</i> in apartment NEG.is hallway-GEN 'There is no hallway in the apartment.'	- Nominative and genitive marking after numerals. - Instrumental marking after the preposition <i>su</i> ('with'). - Gender marking and declension of adjectives.
3. <i>Deciding what furniture to buy</i>	- Nominative subject marking.	- <i>Kambaryje yra sofa.</i> in room is couch-NOM 'There is a couch in the room.'	- Adjectival and adverbial comparatives. - Subjunctive mood.

Task	Target structures featured	Examples of target structures	Additional grammar points
	- Nominative subject marking with numerals. - Genitive subject marking in negated sentences. - Accusative direct-object marking. - Genitive direct-object marking in negated sentences.	- <i>Kambaryje yra dvi lemp-os.</i> in room is two lamp-NOM.PL 'There are two lamps in the room.' - <i>Kambaryje nėra fotel-io.</i> in room NEG.is armchair-GEN 'There is no armchair in the room.' - <i>Mes nusipirksime sof-ą.</i> we will.buy couch-ACC 'We will buy a couch.' - <i>Mes ne-pirksime sof-os.</i> we NEG-will.buy couch-GEN 'We won't buy a couch.'	
4. Planning a menu for a dinner party	- Accusative direct-object marking. - Genitive direct-object marking in negated sentences.	- <i>Andrius mėgsta sūr-į.</i> Andrius likes cheese-ACC 'Andrius likes cheese.' - <i>Andrius ne-mėgsta sūr-io.</i> Andrius NEG-likes cheese-GEN 'Andrius does not like cheese.'	- Dative marking in the construction 'be allergic to'. - Accusative and instrumental marking in the construction 'replace sth by/with sth'. - Genitive and accusative marking after a selection of prepositions.

Task	Target structures featured	Examples of target structures	Additional grammar points
5. <i>Shopping for groceries for a party</i>	- Accusative direct-object marking with numerals.	<i>Perkame šešis agurk-us.</i> we.buy six cucumber-ACC.PL 'Let's buy six cucumbers.'	- Genitive direct-object marking due to the partitive. - Subjunctive mood with modal verbs. - Dative marking in the construction 'discount/deal on'.

above, focused tasks adhere to the four task criteria outlined previously but also aim to direct students' attention to a specific grammatical element (R. Ellis, 2003). Clearly, the tasks described above do not target just one or two grammatical elements, which means that they might be categorized as semi-focused.

3.4.1.4 Task sequencing according to the SSARC model

Besides communicative and grammar considerations, course and task design were also guided by Robinson's (2015) Triadic Componential Framework and the SSARC. Preference was given to Robinson's model of task-based performance over Skehan's model, since Robinson's resource-directing and resource-dispersing variables, together with the SSARC model, allow for easier operationalization in experimental course and task design (R. Ellis et al., 2019, pp. 87, 89).

As discussed in Section 1.3.2, course and task design must be guided solely by the resource-directing and resource-dispersing variables. In this study, task design was primarily informed by resource-directing variables (see Table 13). Task 2 was designed to be more difficult than Task 1 in terms of the number of elements under consideration. Task 3 required slightly more complex reasoning: instead of expressing an opinion about a single element (a residential area or an apartment), participants were asked to evaluate multiple pieces of furniture. According to Robinson's framework, this additional step also increased task complexity along the resource-dispersing dimension. Task 4 adds the resource-directing variable of perspective-taking, since participants were required to consider not just their own food preferences and allergies but also those of their guests. Finally, Task 5 was more complex because it lacked an explicit task structure: participants received all necessary information without specific guidance on how to complete the task, increasing its complexity in terms of resource-dispersing variables.

The task sequence of this study was primarily guided by their logical chronological order and did not fully conform to Robinson's (2015) SSARC model, as increased resource-directing demands often preceded or coincided with increased resource-dispersing demands. Nevertheless, task complexity increased incrementally according to the variables defined in Robinson's Triadic Framework.

Table 13. Implementation of the SSARC model in this study

Task	Resource-directing variable	Resource-dispersing variable
1. <i>Expressing your opinion on different residential areas</i>	- few elements - simple reasoning demands	- few steps - structured
2. <i>Comparing rental apartments and selecting the most suitable one</i>	- more elements - simple reasoning demands	- few steps - structured
3. <i>Deciding what furniture to buy</i>	- more elements - more complex reasoning demands	- more steps - structured
4. <i>Planning a menu for a dinner party</i>	- perspective taking - more elements - complex reasoning demands	- more steps - structured
5. <i>Shopping for groceries for a party</i>	- perspective taking - more elements - complex reasoning demands	- more steps - unstructured

3.4.1.5 Implementation of suggestions on adapting TBLT to the teaching of morphologically complex L2s

Attempts were also made to combine Gilabert and Castellvi's (2019) and Kogan and Bondarenko's (2022) proposals for adapting TBLT to the teaching of morphologically complex languages, as proposed by Udes (2025b). This point was outlined in Section 1.3.3. While the five tasks, described in terms of resource-concentrating and resource-dispersing variables, might seem cognitively demanding for low-level L2 learners, it was deemed important not to make the task requirements overly simple or to let the pedagogic versions of these tasks depart too far from their real-life equivalents. This aligns with Kogan and Bondarenko's (2022) argument that tasks should provide cognitively stimulating challenges for L2 learners. Admittedly, they proposed achieving this mainly through solo- or input-based tasks, which the tasks of this study clearly are not.

Although due care was taken to provide sufficient input to familiarize students with the language relevant to each task, all main tasks required spoken interaction. In order to account for this heightened cognitive complexity, this challenge was mitigated by providing pre-task activities that primarily involved offline interaction with written and oral input, as well as subtasks before the main-task stage, as proposed by Gilabert and Castellví (2019). Task 3, for instance, was broken down into a set of smaller skills, those being ‘establishing what is present and missing’ and ‘expressing a preference between two objects’. Subtasks were then created around these skills to prepare learners for the main task, which provided less guidance on how to complete it.

Post-task grammar instruction was designed to adhere to Kogan and Bondarenko’s advice to limit instruction to the most important forms and functions. Grammatical post-task exercises were limited to the target structures laid out in Table 12 and, in the context of noun declension, covered the paradigms encountered in the input. Example sentences were almost always taken from the input; sentences included in the exercises were either taken from the reading and/or listening texts or only slightly adapted to reduce their linguistic difficulty and facilitate comprehension of the grammatical constructions involved.

3.4.1.6 Assessment of morphological complexity

Linguistic difficulty was essentially established *post factum*, after the tasks were created based on the conceptual task demands. It was operationalized here primarily in morphological terms due to the target structures involving primarily case marking. This does not imply that other dimensions of linguistic difficulty, such as lexical or syntactic complexity, are irrelevant, as such factors may also increase comprehension difficulties of the input and shape task performance. However, including them would have introduced additional sources of variance that are less directly related to the central research questions, which led to the decision to focus primarily on morphological complexity.

Morphological complexity of the input was established by means of computing the Morphological Complexity Index (MCI; Brezina & Pallotti, 2019), previously mentioned in Section 1.3.3.3. The MCI was designed as an operational and replicable measure of the morphological complexity of L2 texts and/or L2 learner performance. Essentially, it identifies any morphological form or change that marks a grammatical category in a word,

which they call ‘exponences’. In Brezina and Pallotti (2019), the method was applied to verbal inflections, but it can naturally be extended to nominal inflections as well. These exponences are then sampled into standardized subsets, from which the average within-sample variety and between-sample diversity can then be calculated.

The MCI is designed to reflect the average inflectional diversity of a single word class. As the target constructions in this study revolve around nominal inflections, calculations of the MCI for the task input were based on nouns. Therefore, all types of nominal inflections were counted as separate exponences (see Table 3 for the full list of possible exponences). All nouns within the reading and listening texts were coded according to their endings (e.g., *faktaj* – ai; *miesto* – o; *rajonus* – us, etc.). For each text, 100 random samples of 10 randomly selected tokens were drawn. For each sample, the within-sample variety is calculated, which is the number of unique exponences within the same sample set. For instance, in the following samples:

Sample 1:

mėnes-j ‘month-ACC’ (-j); *stal-as* ‘table-NOM’ (-as); *technik-a* ‘equipment-NOM’ (-a); *viet-a* ‘place-NOM’ (-a); *automobile-j* ‘car-ACC’ (-j); *įrengim-as* ‘installation-NOM’ (-as); *kambar-ys* ‘room-NOM’ (-ys); *but-ui* ‘apartment-DAT’ (-ui); *lov-a* ‘bed-NOM’ (-a); *rūs-ys* ‘basement-NOM’ (-ys).

Sample 2:

šildym-as ‘heating-NOM’ (-as); *met-ai* ‘year-NOM.PL’ (-ai); *skalvim-o* ‘laundry-GEN’ (-o); *telefon-u* ‘telephone-INST’ (-u); *kain-a* ‘price-NOM’ (-a); *grind-ys* ‘floor-NOM.PL’ (-ys); *but-as* ‘apartment-NOM’ (-as); *bokšt-as* ‘tower-NOM’ (-as); *automobil-iui* ‘car-DAT’ (-iui); *eur-ų* ‘euro-GEN.PL’ (-ų).

the number of unique exponences for sample 1 is 5, whereas for sample 2, it is 8, making the average of just these two samples 6. The average of unique exponences within each sample is then calculated across all 100 samples. Sample pairs are then compared to compute between-sample diversity, defined as the number of exponences unique to one of the two samples. For example, when comparing Samples 1 and 2, Sample 1 has two unique exponences (-j and -ui), whereas Sample 2 has five (-ai, -iui, -o, -u, -ų). This yields a between-sample diversity of 3.5. The average between-sample diversity is then calculated over all pairs. These averages are then taken to compute the MCI with sample sizes of 10 (MC10), which is as follows:

$$\text{MCI} = (\text{within-subset variety} + \text{between-subset diversity}/2) - 1$$

In the case of the previous two example samples, this amounts to:

$$\text{MC10} = (6 + 3.5/2) - 1 = 6.75$$

The theoretical range for MC10 is 0–19 (Brezina & Pallotti, 2019, p. 9). MC10s for the reading and listening texts of each task were calculated in *R* (R Core Team 2013) and are listed in Table 14. Task 5 does not include a value for the reading text, since authentic supermarket weekly ads were used, which participants did not have to read in full.

Table 14. MC10 values for the reading and listening texts of each task

Task	Modality	Noun MC10
Task 1	Reading	11.33
	Listening	11.36
Task 2	Reading	10.63
	Listening	11.55
Task 3	Reading	9.43
	Listening	9.88
Task 4	Reading	10.93
	Listening	11.82
Task 5	Reading	NA
	Listening	11.96

As shown in the table, morphological noun complexity was not entirely stable across the tasks and did not increase incrementally, as would be preferred when designing materials. Values for the reading texts are high in Task 1, decrease gradually until Task 3, and then increase again in Task 4. Values for the listening texts are more stable, although Task 3 again contains the simplest listening text in terms of inflectional complexity. While an incremental increase in the linguistic difficulty of the L2 input across the tasks would have been preferable, this does not necessarily cause problems. If VanPatten’s (2014) IP theory is correct, L2 learners will essentially filter out most morphological markers, as they carry little semantic content and their functions can often be inferred from the context (e.g., pronouns for verbal marking, syntax for subject and direct object marking).¹² Furthermore, the

¹² On the other hand, studies by Kempe & MacWhinney (1998) as well as by Kempe & Brooks (2008) have shown learner preference for item-based learning, providing

MCI accounts only for the variety of morphological endings and not for the variety of functions that these endings encode. The variety of functions was likely more limited within each task, as each task required the completion of specific linguistic acts. Moreover, while the input may have included a wide variety of morphological endings, this does not necessarily hold for the required output of the main task, especially since it was learner-dependent (as per the third TBLT-task criterion). These observations notwithstanding, future task design should perhaps include a gradual increase in the MCI for L2 input, provided that it does not compromise the authenticity and relevance of the input texts.

3.4.2 Overview of task contents

All five tasks consisted of a pre-task, a main task, and a post-task stage. This section provides further detail on what the participants were expected to do in each stage of each task. Importantly, while the task contents overlap to a considerable degree, only the task versions created for the TBLT group will be discussed here. Differences in treatment will be discussed in the subsequent Section 3.4.3.

3.4.2.1 The pre-task stage

The pre-task in the TBLT versions of each task consisted of the following parts:

- an introduction to the topic
- a reading text with exercises
- a listening text with exercises
- a set of subtasks

These parts were always sequenced in this order, with the exception of Task 4, where the reading and listening parts were reversed. Each new task began with a short ‘warming-up’ session designed to elicit previously learned vocabulary and prime it for subsequent task performance. The reading texts served different purposes: they provided background information relevant to the task topic (Task 1), provided examples of relevant language (Tasks 2 and 5), or served as examples of task performance (Tasks 3 and 4). The listening texts usually served as examples of task performance, except for Task 4, where the listening text mainly provided relevant background information.

evidence for at least subconscious tallying of combinations of lexical words and specific morphological endings and their functions.

The goal of the reading and listening texts was not just to provide relevant information and examples but also to expose participants to the target structures. Nevertheless, no input flooding strategies were employed, as that could have diminished the authenticity of the language used in the input. The number of instances of target constructions (subject marking in existential clauses and direct-object marking) is displayed in Table 15.

Table 15. Number of instances of the target structures in the reading and listening text of each task

Task	Text	Subject		Object	
		Nominative	Genitive	Accusative	Genitive
Task 1	Reading	10	0	15	0
	Listening	4	1	8	0
Task 2	Reading	16	0	5	0
	Listening	10	1	11	1
Task 3	Reading	0	1	32	3
	Listening	0	1	29	2
Task 4	Reading	0	0	18	5
	Listening	0	0	28	4
Task 5	Reading	4	0	22	1

Each noun phrase (including adjectives marked for case) or pronoun was counted as a separate instance of (existential) subject marking or direct object marking, provided that it did not include the partitive genitive. The verbal form *yra* ‘there is/are’ could be omitted from existential clauses, which is common in Lithuanian. As shown in Table 15, instances of subject marking were numerous in the first two tasks, after which they became quite rare. This was expected given the nature of the target tasks. Another observation is that negation was quite rare in the reading and listening texts. These texts were primarily developed by adapting authentic sources to promote certain target structures, which likely contributed to the rarity of negated subject and direct object marking. Importantly, however, while planned exposure to negated subjects and direct objects was quite limited, performance in the subtasks (as described below) and main tasks likely led to more production of and/or exposure to these forms.

The reading texts were written or edited in different ways: one was based solely on the researcher’s imagination and experience (Task 4), while others were loosely based on existing online texts and discussions (Tasks 1 and 3), a simplified version of authentic materials (Task 2), or real authentic

materials themselves (Task 5). The listening texts, by contrast, were all scripted by the researcher, and professional actors were hired to record the conversations for them. The original text was mainly intended as a guideline, and the actors were asked to improvise to produce a naturally sounding conversation.

Lastly, as mentioned previously, the subtasks were designed to prepare students cognitively for the main task by breaking it down into smaller sub-skills, each targeted by a separate subtask. Many of these subtasks also included grammatical operations relevant to this study's target structure. The second subtask in Task 1 had participants analyze a map of a city district in Vilnius and answer questions about which facilities were present and which were absent. This subtask naturally elicited affirmative and negated subject marking. Task 2 included a similar subtask in which participants compared two apartments and established what was present in both and what was absent from one of them. In Task 3, the first subtask had participants each describe a picture of a living room that they could not show to each other and establish how they differed, i.e., what was present in both or absent in one of them. Besides subject marking, this subtask also naturally elicited object marking through verbs such as *turėti* 'to have' and/or *matyti* 'to see'. Both subtasks in Task 4 elicited object marking: participants expressed which foods they liked and disliked, and then established which foods two friends disliked and which ingredients needed to be changed in three dishes based on those preferences. Finally, all three subtasks in Task 5 involved object marking: participants determined how much (fictional) money they had as a group, added items to a grocery list, and agreed on a grocery list within a 10-euro budget based on a selection of discounts.

Appendix B provides a detailed overview of the pre-task content for each task.

3.4.2.2 The main task stage

The steps required to complete each main task were outlined previously in Table 11. In addition to the instructions, participants were given supporting materials. In Task 1, these included small, simplified maps of two city districts (i.e., neighborhoods), *Šnipiškės* and *Naujininkai*, as well as a list of symbols introduced during the pre-task activities to facilitate information exchange. Similarly, in Task 2, each pair member received the layout of one of two apartments, along with additional information such as its location and rental terms. They were also provided with a property listing-style checklist, which

they completed to facilitate information exchange. Task 3 provided a picture of a shared living room in the instructions, and each pair member was given one of two store catalogs containing six items. In Task 4, each participant received the food preferences and allergies of two different guests, and each pair was also given a list of suggested dishes to prepare for the dinner party. Finally, each participant received the weekly ads of one of two supermarkets to peruse for completing Task 5. The main tasks for both treatment groups were essentially similar; only the supporting materials differed somewhat. All these materials are presented in Appendix C.

As for this study's target constructions, each task was designed to elicit appropriate subject and direct-object marking in a natural manner. In Task 1, by listing the advantages and disadvantages of each residential area, the participants were expected to mention which facilities are present and which are absent, potentially leading to affirmative and negated subject marking. Similarly, in Task 2, discussions about which apartment to choose, as well as the reasons for that choice, were expected to mainly elicit affirmative and negated subject marking. Task 3 first required participants to identify what was absent, thereby potentially eliciting negated subject marking. Afterward, participants were tasked with describing the furniture they saw in the stores they were visiting and with suggesting and choosing furniture to purchase, thereby eliciting direct-object marking. Task 4 mainly involved direct-object marking, as participants discussed their own likes and dislikes regarding food, as well as those of some fictional friends. Based on these food preferences, the participants chose dishes to prepare for a dinner party and, where necessary, the ingredients to modify, providing further opportunities for direct-object marking. Task 5 gave participants the most freedom, as they had to make a grocery list while adhering to constraints such as a budget. Discussions about what to purchase and what was not necessary to buy surely provided ample opportunity for direct-object marking, although in this task it often involved the partitive genitive or the genitive after quantity words.

While the main tasks aimed to elicit these target constructions, participants were never actively encouraged to use them, as per the third criterion for TBLT tasks, which has learners rely on their own resources (R. Ellis et al., 2019). Hence, if the participants managed to express their thoughts and complete the target tasks without accurately marking subjects and direct objects, they were not reminded to do so. Nevertheless, the teacher sometimes provided incidental focus on form, which, according to Long (2014), is useful in meaning-focused contexts such as these, provided that the focus is brief and

reactive (see Section 1.2.3). These instances of focus on form differed in their implicitness and explicitness (see Section 1.3.1.3) but mostly consisted of recasts and explicit error corrections. Only when participants explicitly requested a grammatical explanation was the relevant rule briefly explained. In all other cases, corrective feedback was restricted to brief recasts and comments on accuracy.

3.4.2.3 The post-task stage

The post-task stage for each task was identical for both treatment groups. It included an exercise in which participants identified at least four phrases from the reading and listening texts that they found useful for their own L2 use. Tasks 1, 4, and 5 also included a classroom discussion exercise in which the participants shared their task results, while Task 2 included a writing exercise.

In addition to these content-based exercises, a substantial part of each post-task stage was devoted to explicit grammar exercises, which covered all the structures described in Table 12. Most of these exercises highlighted relevant grammatical constructions within sentences from the pre-task input and, in some cases, from the main-task stage. They were generally designed to provide relatively brief opportunities for explicit focus on forms. While these exercises were mostly carried out individually, they served as consciousness-raising activities (see Section 1.3.1.2), as participants had to apply grammatical rules based on the input and performance encountered during the pre-task and the main task stages. After the completion of the grammar exercises on nominal case, tables showing the case endings according to the paradigms relevant to the nouns encountered in the input were shown in class and sent afterward by email. Both treatment groups were shown these tables, which were identical to the ones provided in the pre-task to the TSLT group (see Section 3.4.3), but completed.

An example of a post-task grammar exercise from Task 1 is given in Figure 6. The instructions read: “Compare two maps. Write down what is there and what is not.” The figure shows two maps (Map 1, Map 2), with the words *yra* ‘there is’ and *nëra* ‘there is no’ displayed below them. The first map already contains two example answers, namely *parkas* ‘park-NOM’ and *stotelès/stotelių* ‘bus stop.GEN.SG/.PL’. The target structure of this exercise was subject marking in affirmative (*yra* ‘there is’ + nominative) and negative (*nëra* ‘there is no’ + genitive) existential sentences. It required participants to write down which places and services were present in one map but absent from

4. Palyginkite du žemėlapius. Parašykite, kas juose yra ir ko nėra.

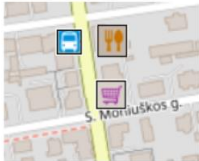
1 žemėlapis	2 žemėlapis
	
yra:	yra:
_____ <i>parkas</i> _____	_____ _____
_____	_____
_____	_____
nėra:	nėra:
_____ <i>stotelės/stotelių</i> _____	_____ _____
_____	_____
_____	_____

Figure 6. Example of a post-task exercise

the other. Essentially, this was a repetition of what the learners had already done during the main task stage, but which now served as a brief instructional prompt to highlight nominative and genitive case marking in the context of indicating the presence or absence of, for instance, services in a location.

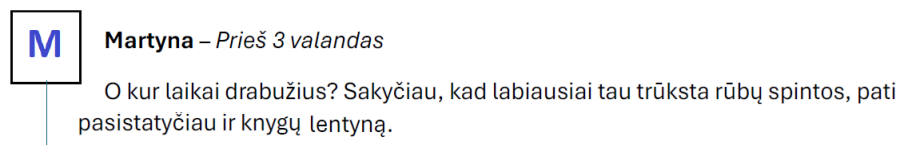
Task repetitions were considered and originally intended to be included in all tasks. The rationale was that task repetition would provide participants with the opportunity to shift their attention more towards form than towards meaning, as argued by Gilabert and Castellví (2019). Rather than simply repeating the same tasks, they were supposed to be cognitively more challenging versions, usually including more elements. For instance, in Task 3, store inventories were supposed to include eight items each, and participants were supposed to furnish two different rooms instead of one. However, due to the already considerable length of each task, the decision was made to exclude task repetitions from the post-task stages.

3.4.3 Treatment differences for the TBLT and TSLT groups

The task descriptions in the preceding sections mainly focused on the TBLT versions of each task. The TSLT group received essentially the same tasks but

with several important differences. These modifications served to implement Gilabert and Castellví's (2019) suggestion to incorporate focus on form at all stages of the task. First, the input was enhanced to make grammatical forms, particularly case endings, more salient. This was done not only in the reading texts and listening transcripts but also in the instructions for the subtasks (e.g., examples and prompts) and the supporting materials for the main task. Figure 7 provides an example of input enhancement from the reading text of Task 3.

TBLT version:



TSLT version:

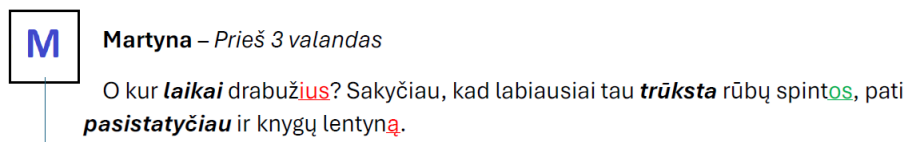


Figure 7. Two excerpts from the reading text of Task 3

The English translation in both versions reads: ‘Where do you keep your clothes? I’d say that what you need most is a wardrobe, I’d also put a bookshelf’. In the TSLT version, the case endings for the words *drabužius* ‘clothes’, *spintos* ‘wardrobe’, and *lentyna* ‘shelf’ were colored and underlined. The colors distinguished the different cases; in this example, red represented the accusative and green the genitive. Since cases carry a wide range of functions, coloring and underlining all case forms (in this task, the accusative and the genitive) risked overwhelming the participants, which would have made input enhancement counterproductive, as the target construction (direct-object marking) could have been obscured by the abundance of identical case markings serving different functions (e.g., prepositional government). Therefore, words governing a case according to a certain function were additionally bolded. These did not always involve the target structures of this study but also included other relevant grammatical constructions listed in Table 12, such as the locative in Task 1 and the preposition *su* ‘with’, which governs the instrumental case, in Task 2. It is unclear, however, whether these forms of input enhancement affected participants’ case acquisition. Participants were never explicitly told what the colors and other forms of input

enhancement represented in order to keep it as unobtrusive as possible, and questions about this aspect were rare.

Importantly, input enhancement in the examples and prompts within the subtask instructions sometimes altered the nature of these exercises. Below is an excerpt from the instructions of the first subtask in Task 3:

Pavyzdžiui:

A: Mano kambaryje yra fotelis.

A: Aš turiu kilimą.

B: Mano kambaryje taip pat!

B: O aš neturiu kilimo.

Naudingos frazės:

- Čia yra (*fotelis*)/nėra (*fotelio*).
- Mano/šiamė kambaryje yra (*sofa*)/nėra (*sofos*).
- Aš turiu (*paveikslą*)/neturiu (*paveikslo*).

Figure 8. Excerpt from the instructions of the first subtask in the TSLT version of Task 3

The upper translation reads “For example: A: There is an armchair in my room.; B: In my room as well!; A: I have a carpet.; B: I do not have a carpet.” The lower translation is the following: “Useful phrases: – There is (an armchair)/there is no (armchair).; There is/There is no (couch) in my/this room.; I have (a painting)/I do not have (a painting).” Case endings are underscored and colored according to the specific case (blue = nominative, green = genitive, red = accusative).

In this subtask, participants worked in pairs. Each pair received a picture of a living room and was required to describe the furniture they saw to their interlocutor and identify the similarities and differences between the rooms by discussing it together. While the TBLT group appeared to interpret the task as such, the TSLT group seemed to focus more strongly on the grammar involved, specifically on nominative and genitive marking for subjects in positive and negated sentences, respectively. This difference can likely be explained by the input enhancement in the examples and prompts. Consequently, what was largely a communicative task for the TBLT group turned into a fairly explicit focus-on-form exercise for the TSLT group. This effect was observable across many subtasks in other tasks, potentially turning them into practice activities that align closely with the PPP method. Nevertheless, since no direct feedback was collected from the participants after completing this or other subtasks with similar input enhancement, no

firm claims can be made about the specific effects of this type of intervention at this time.

Another important difference between the TBLT and TSLT task versions was the replacement of certain communicative exercises with explicit form-focused exercises. In Task 2, one exercise was shortened rather than replaced. The decision to replace communicative exercises rather than add grammar-focused exercises was intended to keep the completion time of both task versions roughly similar. For the listening activity, instead of filling in blanks with potentially useful words and phrases, participants were required to supply missing grammatical forms, most commonly case endings. Essentially, this adhered to the grammatizing strategy previously outlined in Section 1.3.1.2, but was used here in the pre-task rather than the post-task. Similarly, communicative exercises in the reading section were frequently replaced with exercises that used a table of case endings. Participants were then tasked to complete the table based on the words and endings found in the reading text. Essentially, these exercises worked like the inter-tasks proposed by Kogan and Bondarenko (2022), as their aim was to highlight relevant grammatical structures for each task. Alternatively, these exercises could be interpreted as a TSLT approach that aligns more closely with PPP, with the case tables for the reading texts representing the Presentation stage and the gap-fill exercises for the listening text representing the Practice stage.

Table 16 presents which exercises in the TBLT versions were replaced with grammar exercises and specifies the targeted cases. It should be noted that the grammar exercises in Tasks 4 and 5 targeted only the dative case, which was relevant to these specific tasks but not one of the target cases in this study. Importantly, however, both TSLT groups did not complete Task 5 due to time constraints.

3.5 Instruments

To assess participants' knowledge of grammatical case in subject and direct-object marking, two instruments were employed: a gap-fill test and an *Elicited Oral Imitation Test* (EOIT). The order in which these instruments were presented to the participants was randomized. As described above, the target structures were subject marking (nominative, genitive) and direct-object marking (accusative, genitive) in both affirmative and negated sentences. To operationalize the measurement of participants' performance more effectively, the intended target items were kept in the singular (although

Table 16. List of exercises that were unique to each TBLT or TSLT version of the task

Task	TBLT exercise	TSLT table(s)
1.	Write an interesting fact about your hometown.	Singular and plural locative, plural genitive endings for six paradigms.
2.	Provide three advantages and disadvantages for each apartment.	Provide two advantages and two disadvantages for each apartment. Singular genitive, singular and plural instrumental endings for six paradigms.
3.	Write a comment with advice for the topic starter on the 'Reddit' forum.	Singular and plural genitive and accusative endings for five paradigms.
4.	In pairs, discuss and write down what other dishes or ingredients, mentioned in the dialog, could be changed or replaced.	Singular and plural dative endings for six paradigms.
5.	In pairs, write down and discuss which products you would like to get deals on and what kind of deals they should be.	Singular and plural dative endings for six paradigms.

participants were allowed to use the plural if they wanted), and paradigms that were not among the four most frequent were excluded. Thus, all target words belonged to the paradigms ending in *-as*, *-is/-ys*, *-ė*, and *-a*.¹³ Since plural endings are typically introduced later in L2 Lithuanian courses and the excluded paradigms are quite rare, these endings would likely have placed additional strain on participants' performance. The updated Table 17 shows the exact target endings for the instruments to follow.

¹³ These paradigms were selected based on the frequencies established by Bružaitė-Liseckienė, (2021, p. 58) in the *Lietuvių kalbos žodynas: elektroninė versija* 'Lithuanian Dictionary: Electronic Version', compiled by A. Smetona. In percentages, the share of each paradigm over all nouns in the dictionary is as follows: *-as* (34.5%), *-is/-ys* (27.1%), *-ė* (25.3%), and *-a* (9.6%). However, it should be added that these percentages reflect the frequency of endings in the dictionary, and not in L1 Lithuanian usage as a whole. Unfortunately, such information is, as of yet, unavailable, hence the decision to rely on dictionary frequencies of paradigms.

Table 17. The target endings of the instruments within the nominal case system in Lithuanian

Case	Declension 1		Declension 2	
	-as	-is (-ys)	-a	-ė
	Singular			
Nom.	-as	-is/-ys	-a	-ė
Gen.	-o	-io	-os	-ės
Acc.	-ą	-į	-ą	-ę

Target nouns were checked for suitability along two criteria. First, they were supposed to belong to the 2,500 most frequent lemmas according to the *Wordlist of lemmas from the joint corpus of Lithuanian* (Dadurkevičius, 2020).¹⁴ If they were not, they had to be included in one of the most frequently used textbooks of L2 Lithuanian: *Nė dienos be lietuvių kalbos* (2018), intended for absolute beginners progressing to CEFR level A2, or *Sėkmės!* (2022), intended for absolute beginners progressing to CEFR level A1. As chunking and subsequent emergence of abstracted rules are considered important processes in L2 learning, valency must be considered as an important factor for the acquisition of case marking; hence, the same two criteria were applied to the verbs accompanying the target nouns. The verb *užduoti* ‘give/set’ was the only verb that did not meet these criteria, but was nevertheless accepted as it was used in the collocational context of *užduoti klausimą* ‘ask a question’. L2 students were thought to encounter this collocation regularly. The full list of target nouns and verbs for the gap-fill tests and EOITs, as well as the full list of items, is provided in Appendix D.

3.5.1 Rationale behind the test instruments

As mentioned above, this study employed two instruments: a gap-fill test and an elicited oral imitation test. The gap-fill test was administered to assess participants’ explicit knowledge of case. As its name suggests, this test presented participants with sentences containing a gap and asked them to fill it in. Each gap was accompanied by the target word in its base form (i.e., the nominative singular) in brackets, as shown in the example below:

¹⁴ The 2500 lemma range is taken from Milton, (2010, p. 224)).

- (19) *Nežinau, kiek mokėti, nes nėra _____ (sąskaita).*
 ‘I do not know how much I have to pay because there is no _____
 (check).’

The EOIT was selected to assess participants’ implicit knowledge of case. EOITs have participants listen to a statement, mark a response (often ‘agree’ or ‘disagree’), and then repeat the sentence. The rationale is that, while focused on the content and meaning of the statement, especially under time constraints, participants disregard its grammar, leading to the use of implicit knowledge during repetition. Crucially, a proportion of these statements is, deliberately, grammatically incorrect. The reasoning behind this is that if participants’ implicit knowledge of a grammatical category is sufficiently robust, they will spontaneously correct the error in the statement when repeating it. Below are examples of both a grammatically correct and an incorrect statement:

- (20) *Čia ant stalo yra mano telefonas.* ‘My phone is here on the table.’
 (21) **Čia ant stalo nėra sūris.* ‘There is no cheese on the table.’ (nominative instead of genitive marking in a negated sentence)

As mentioned above, the test instruments were designed to assess different types of case knowledge: the gap-fill test assessed participants’ explicit case knowledge, and the EOIT assessed their implicit case knowledge. The rationale behind this strategy draws on the criteria for operationalizing L2 implicit and explicit knowledge constructs laid out in Han and Ellis (1998) and Ellis (2009a). The criteria most relevant for these tests are the following: ‘degree of awareness’, ‘time available’, ‘focus of attention’, and ‘utility of metalanguage’. Table 18 summarizes how these criteria are applied to the instruments in this study (Ellis, 2009a).

Table 18. Criteria for operationalizing L2 implicit and explicit knowledge constructs relevant to the gap-fill and elicited oral imitation tests

Criterion	Gap-fill test	EOIT
Degree of awareness	Rule	Feel
Time available	Unpressured	Pressured
Focus of attention	Form	Meaning
Utility of metalanguage	Yes	No

The ‘degree of awareness’ criterion concerns how a participant attends to the answer, whether they base their response on a rule or on

intuition. Rule-based decision making is considered an explicit way of responding, while intuition, supposedly, taps into implicit knowledge. The presence or absence of time constraints to solve a problem either allows test-takers to use their explicit knowledge (absent) or forces them to resort to their intuition (present). Thus, time-constrained tests are intended to assess participants' implicit knowledge. Test design can also be manipulated by shifting participants' focus to form or meaning, which leads to assessments of explicit and implicit knowledge, respectively. Similarly, tasks can motivate participants to use their metalinguistic knowledge, or not, leading to the assessment of explicit or implicit knowledge. In this study, the gap-fill test most likely prompted participants to respond according to rules (if they knew them), was not time-constrained, and directed participants' attention to form, specifically by asking for correct endings in the test examples, which encouraged participants to use their metalinguistic knowledge. Conversely, the EOIT presumably required participants to rely on their L2 intuition while being pushed by time constraints, and especially to focus on meaning, as they were instructed to respond to the content of the statement before repeating it. Supposedly, this approach led them to avoid relying on their metalinguistic knowledge.

A few disclaimers are in order here. First, the prediction that participants responded based on either rules or feeling is exactly that: a *prediction*. For instance, participants who had a poor grasp of the subject or direct-object marking rules probably relied more heavily on intuition. Furthermore, those participants who realized that grammatically incorrect sentences were included in the EOIT might have relied more on their metalinguistic rule knowledge. These caveats need to be considered when analyzing the results.

Furthermore, time constraints for each test need to be expanded upon. First of all, the gap-fill test required participants to complete it within 20 minutes, but this seemed more than sufficient for most participants, hence the claim that time pressure was minimal. It is also worth noting that, in EOITs, administrators usually move on to the next item after the first repetition of a statement. In this study, it was decided to allow participants to revise their repeated sentences if they wished, as it was thought that this would not immediately lead to the use of explicit grammar knowledge. The participants often questioned not the grammar, but their memory of the original sentence. Additionally, memory constraints were deemed sufficient as a time constraint for tapping into participants' implicit knowledge of case. The longer

participants waited to repeat a sentence, the harder it became to do so accurately.

3.5.2 Details of the gap-fill test

Test administration was preceded each time by a short introduction focusing on the instructions and two examples. The target items for each test time (i.e., pretest, posttest, delayed posttest) consisted of eight subject-marking items and 11 object-marking items, for a total of 19 target items. In the subject-marking items, participants had to fill in the subject marked for the appropriate case (which could be the nominative base form or the genitive form when negated, as in the example above), while the object-marking sentences required participants to fill in the object in its correct case form (the accusative in affirmative sentences and the genitive in negations). The decision was made to include more direct-object marking items, as they were deemed more complex and, thus, insightful, since a correct response to non-negated subject-marking items would simply require participants to copy the target word presented in brackets. In addition to the 19 target items, the tests at each test time also included seven filler items to reduce the likelihood that participants would understand the exact target structures of the study. These fillers resembled the target items in that participants were tasked with providing the correct nominal forms, but in a non-target case (nominative in non-existential sentences, dative, instrumental, and locative).

Regarding the target paradigms, each gap-fill test included seven *-as* nouns (Subject: 3, Object: 4), five *-is* nouns (Subject: 2, Object: 3), five *-ė* nouns (Subject: 2, Object: 3), and two *-a* nouns (Subject: 1, Object: 1). These numbers were meant to reflect the relative distribution of these endings in natural language use, as found by Bružaitė-Liseckienė (2021, p. 58). The distribution of target cases varied slightly across the pre-, post-, and delayed posttests. While subject marking always included three items targeting the nominative and four targeting the genitive, direct-object marking in the pre- and delayed posttests targeted the accusative six times and the genitive five times; the posttest included five target items for the accusative and six for the genitive.

Each test consisted of two versions. The order of target items and fillers was randomized but adhered to a select number of principles:

- Each test started with two filler items.
- No two filler items should follow each other thereafter.

- A maximum of two subject or direct-object marking items could follow each other.
- A maximum of two positive or negated sentences could follow each other.

These principles were designed to distract participants from the target constructions. By starting each test with two fillers, spreading the remaining fillers across the test, and distributing the other two target structure conditions (subject/direct object; affirmation/negation) evenly, it was assumed that participants would be less likely to identify the target construction. While the participants surely understood that grammatical case was the point of focus, the aim was not to overtly emphasize the marking of subjects and direct objects in affirmative and negative statements.

3.5.3 The elicited oral imitation test

Before each test, participants completed three training items, one of which was also written down. Each test was administered individually by either the researcher himself or a colleague from Vilnius University. Data collection involved playing a recording of the statements while participants' production was recorded with a voice recorder. Statements were played separately and only once; there was no possibility to repeat any of the statements. The subsequent statement was played when participants gave an intonational or behavioral cue that they would not further attempt to repeat the statement. Test administrators tried to ensure that participants did not repeat the sentence until they had marked their response on their answer sheet, so as not to repeat it immediately.

The statements were quite simple to make them suitable for lower-level learners. Consequently, the 'agree/disagree' response format usually employed in EOITs did not fit, which led to the following four response options: *teisingai* 'correct', *neteisingai* 'incorrect', *nežinau* 'I do not know', *nesuprantu* 'I do not understand'. The *nežinau* 'I do not know' option was available in case participants did not know whether the statement was true. While different, the true/false options were thought to nevertheless encourage participants to focus on meaning first, with form receding into the background. This was the only purpose of the response options; hence, the responses were not analyzed further. All statements related to participants' own habits or opinions, their immediate surroundings during the test, or Lithuania in general.

The syllable length of items ranged from six to 13, with close mean values for the pretest (8.8), posttest (9.2), and delayed posttest (8.9). Such syllable lengths seem on par with EOITs performed by Hamayan et al. (1977; a mean syllable length of 9) and Spada et al. (2015; a mean syllable length of 10) but small compared to those designed by Erlam (2009; a mean syllable length of 13.53) and Sarandi (2015; a mean syllable length of 12.27). Earlier versions of the EOIT designed for this study were tested in the pilot study described in Udes (2025a) with the TBLT group, but the longer sentences included therein proved too difficult for most students to process, leading to data loss. Therefore, the problematic items were either changed or slightly simplified for further trials. The TSLT group in the pilot study received these updated items and had fewer problems processing them. Many of these items ended up on the final item list for this study.

Erlam (2009) repeats Gallimore and Tharp's (1981) claim that the position of the target structure in the sentence matters in terms of a heightened cognitive load when repeating a sentence, with target structures positioned at the beginning of the sentence being the easiest to process, followed by sentence-final placement, while target structures placed in the middle are deemed the most difficult to imitate. In this study, almost all target structures included sentence-final placement of the target structure except for one sentence in the pretest, two in the posttest, and one in the delayed posttest that featured mid-sentence placement of the target structure.

Each test consisted of 16 statements: eight subject-marking and eight direct-object-marking target items. Each of these eight target items included four erroneous statements in terms of subject- or direct-object-marking, with the exception of the pretest, in which the subject-marking items consisted of five correct and only three erroneous sentences. This happened due to an oversight by the researcher. The test did not include any fillers because the exercise was designed to have participants focus on meaning rather than form, which inherently rendered the target constructions less salient. In terms of the target paradigms, each test included six *-as* nouns (Subject: 3, Object: 3), four *-is* nouns (Subject: 2, Object: 2), four *-é* nouns (Subject: 2, Object: 2), and two *-a* nouns (Subject: 1, Object 1).

Each test comprised two versions containing the same target items in different orders to control for possible order effects. The order of the target items was randomized but adhered to a select number of principles:

- The first item needs to be grammatically correct.
- A maximum of two subject or direct-object marking items can follow each other.

- A maximum of two grammatically correct or incorrect sentences can follow each other.¹⁵

The first principle was adopted to ease participants into the test and to avoid immediately drawing attention to erroneous items. As for the remaining principles, they were, again, intended to distract participants from the target constructions as much as possible.

3.5.4 Test modifications

Piloting of the test instruments informed a number of their design features. The biggest change was the exclusion of target sentences containing the verb *patikti* ‘to like’ from the final study, due to problems described in Udes (2025a). Constructions with this verb were included only as filler items. Another important modification was the decision to have the verbs preceding the target nouns appear only in the present tense. The original versions of certain target items used imperative or past-tense forms, which could have been less easily recognized by participants, despite the high frequency of the verbs. While certain other noun-related challenges could have been the cause of some of the problematic items (for instance, nouns referring to locations were often marked erroneously in the locative), which were considered harder to control for, it was deemed important to at least control for the verbal forms in all the items. Certain items were also shortened where appropriate, and items containing low-frequency target nouns (e.g., *virduļys* ‘kettle’) that did not seem particularly relevant to the participants were substituted with easier ones.

Regarding the EOITs administered to both groups of the pilot study, piloting proved essential. For many participants in the TBLT group, around nine out of 20 items proved too difficult in terms of syllable length and/or certain nouns or verbs used in the statements. Problems arose due to specific target nouns (*galimybē* ‘possibility’, *ligonīnē* ‘hospital’, and *paplūdimys* ‘beach’ proved problematic) or not understanding the sentence as a whole, which led to some participants being unable to repeat certain statements. As a result, syllable length was shortened, and the target nouns and constructions were simplified. Additionally, a sentence related to participants’ immediate surroundings was piloted (**Čia ant stalo yra peilio* ‘There is a knife on the table’) to evaluate its efficacy for checking subject-marking knowledge. Since

¹⁵ Due to the aforementioned oversight resulting in a superfluous grammatically correct statement in the posttest, one version of the pretest had a sequence of three grammatically correct sentences.

it proved successful with this pilot group, the decision was made to substitute problematic items with this new type of statement. Additionally, in order to mitigate the risk of losing items due to comprehension difficulties, the decision was made to increase the number of items per test from 10 to 16 for the final study.

3.6 Transcription and coding

3.6.1 Gap-fill test

Assessment and scoring of the data were based on the accuracy of participants' responses, i.e., correct case marking for the subject or direct-object role in sentences with a positive or negated verb. Scoring included four categories to establish response accuracy:

- 1: a completely accurate response.
 - e.g., *nëra parduotuv-ēs* 'NEG.is store-GEN' ('there is no store')
- 1a: a semi-accurate response due to minimal spelling errors.
 - e.g., **nëra parduotuv-es* 'NEG.is store-GEN(?)' ('there is no store') [e instead of é, leaving room for interpretation]
- 0a: a semi-accurate response because of paradigm errors or incorrect assignment of the genitive for a plural noun.
 - e.g., **nëra parduotuv-os* 'NEG.is store-GEN' ('there is no store') [wrong paradigm]
- 0: a completely incorrect response.
 - e.g., **nëra parduotuv-ē* 'NEG.is store-NOM' ('there is no store') [nominative instead of genitive case]

Table 19 further elaborates on the criteria for assigning one of the four categories, together with additional examples based on the target items.

Some of the scoring decisions require additional clarification. As mentioned above, all target items were designed to elicit responses containing a noun in the singular. Nevertheless, participants occasionally preferred a plural noun instead. When the plural noun was marked with the correct case, the response was scored as correct for case marking, even if the use of the plural created contextual problems. In some cases, plural interpretations were incompatible with cues provided elsewhere in the item, such as relative pronouns (e.g., **kavin-ēs, kur-ioje*, 'café-NOM.PL which-LOC.SG', 'cafés in which', where *kurioje* is singular). This also applied to some of the target nouns themselves, for which plural responses were less natural (e.g., *jogurtų*, 'yoghurt-GEN.PL', 'yoghurts'). However, interpreting such context cues was

not part of the target structures under investigation and requires grammatical and lexical sensitivity beyond the study's primary focus. Furthermore, these cues were not controlled for in the test design. It was therefore decided to score only the accuracy of case marking and not penalize contextual mismatches of this kind.

In the case of errors involving the soft sign *-i*, it was decided to categorize these mistakes as spelling errors, rather than paradigm errors, since the intended case and paradigm are still easily identifiable. Moreover, endings such as *-o* and *-io* belong to the same declension pattern (recall the example of *balkonas* 'balcony' (gen. *balkono*) and *brolis* 'brother' (gen. *brolio*)). Responses including no diacritics were also categorized as spelling mistakes. While many Lithuanians do not use diacritics in their everyday chat correspondence, making such answers acceptable to some, it was nevertheless decided not to accept them as fully accurate responses. This still led to some challenges, however, since certain case endings without diacritics can be interpreted in several ways. This is visible in the following response from the data sample:

- (22) *Jeigu mėgsti saul-e, atostogauk rugpjūtį!*
 if you like sun-NOM/ACC/INST(?) go on vacation in August
 'If you like the sun, go on vacation in August!'

While the fully correct response *saulė* is easily interpretable as accusative marking, the response *saule* can carry a number of different case interpretations, namely the nominative (*saulė*), accusative (*saulę*), and instrumental (*saule*) cases. Despite this, since such a spelling of the accusative is quite common in informal written Lithuanian and is readily understood by most L1 speakers, the decision was made to mark these responses as '1a'. An exception was made for endings potentially belonging to different paradigms because, besides the multitude of possible readings, this would not be deemed acceptable by native speakers.¹⁶

The '0a' category was defined to account for erroneous responses that nevertheless reflect some developmental progress. The literature shows that students of L2 Lithuanian often struggle not only with assigning the correct case but also with marking it according to the correct paradigm (Dabašinskienė & Čubajevaitė, 2009). This even applies to bilingual Russian–

¹⁶ For instance, **savaitgala* 'weekend' (target response: *savaitgalį*) could be interpreted as the nominative, accusative, or instrumental. If we allow for switches between the vowels *a* and *o*, as the data occasionally suggest, the word could also have been marked genitive.

Lithuanian high school students (Bružaitė-Liseckienė, 2021). The decision also included adding incorrect genitive marking for plural direct objects. As described above, direct objects are marked genitive in the case of a negated verb. However, certain readings and contexts can also allow for, or even require, partitive genitive marking after a non-negated verb, such as in the following target items:

- (23) *Butas taip pat turi rūš-ių . su*
 apartment also has basement-PL.GEN with
metalinėmis durimis
 metal doors
 ‘The apartment also has some basements with metal doors.’
- (24) *Ekskursijos pabaigoje gidė visada užduoda*
 at the end of the excursion guide always asks
turistams klausim-ų apie Vilnių.
 tourists question-PL.GEN about Vilnius
 ‘At the end of the excursion, the guide always asks tourists some questions about Vilnius.’

However, certain verbs or contexts render genitive marking impossible:

- (25) *Jeigu lyja, leidžiu *savaitgal-ių prie televizoriaus.*
 when it rains I spend weekend-PL.GEN near television
 ‘When it is raining, I spend (some) weekends in front of the television.’
- (26) *Sveiki, pirmą kartą planuojame *kelion-ių į Kroatiją*
 hello first time we plan trips-PL.GEN to Croatia
automobiliu, prašome patarimų.
 by car we ask for advice
 ‘Hello, we are planning a trip to Croatia for the first time, please give us some advice.’

Acquiring such subtle distinctions in genitive plural case marking with non-negated verbs is most likely very difficult and more typical of at least intermediate-level L2 Lithuanian proficiency. At the same time, such erroneous genitive marking does seem to reflect certain developmental progress, as participants did not default to the base form or another unrelated case, but selected the genitive, which is grammatical in related constructions, such as (18). However, unlike the contextual mismatches discussed above,

these responses contained a clear grammatical error within the target predicate-noun relationship itself. For this reason, such responses were scored ‘0a’.

Lastly, it must be noted that most analyses of the results did not include the categorical distinction among the four response categories described above. Instead, they used a strict binary scoring method. Under this method, only fully accurate responses were accepted, i.e., only those responses scored ‘1’. All other response types, i.e., ‘1a’, ‘0a’, ‘0’, were scored ‘0’.

Errors were additionally analyzed by type. These did not include ‘1a’ errors involving the absence of diacritics. They could be case assignment errors, when the selected case was unsuitable for the given context:

- (27) *Jis neis ten, kur nėra *eil-ę, nes*
 he will.not.go there where NEG.is line-ACC because
tai reiškia, kad tame restorane kažkas negerai.
 it means that in.that restaurant something wrong
 ‘He will not go there where there is no line, because it means that
 there is something wrong in that restaurant.’

A related, however, separately categorized error involved the base form. These were not considered case assignment errors since, essentially, they included no change to the ending. While the base form coincides with the nominative singular, chances are low that errors include specifically nominative case marking:

- (28) *Jis neis ten, kur nėra *eil-ė, nes*
 he will.not.go there where NEG.is line-NOM because
tai reiškia, kad tame restorane kažkas negerai.
 it means that in.that restaurant something wrong
 ‘He will not go there where there is no line, because it means that
 there is something wrong in that restaurant.’

The final error type involves the aforementioned 0a category, in which the selected form belongs to the correct case, but to a different declensional paradigm:

- (29) *Jis neis ten, kur nėra *eil-io, nes*
 he will.not.go there where NEG.is line-GEN because
tai reiškia, kad tame restorane kažkas negerai.
 it means that in.that restaurant something wrong
 ‘He will not go there where there is no line, because it means that
 there is something wrong in that restaurant.’
 [While the noun carries the genitive ending *-io*, it belongs to the
 wrong paradigm, as the form should be *eilės*, with the genitive ending
-ės.]

If an ending was non-existent or it was impossible to identify which case the erroneous case belonged to, those endings were marked as unidentifiable (?).

- (30) *Šalia yra *kavin-ov, kurioje skanus maistas ir*
 next.to is café-?, in.which tasty food and
malonus aptarnavimas.
 pleasant service
 ‘There is a café next door which offers tasty food and pleasant
 service.’
- (31) *Ne-skaitau *straipsn-a, kai iš nuotraukos*
 NEG-I.read article-NOM/ACC/INST? when from picture
viskas aišku.
 everything clear
 ‘I do not read the article when everything is clear from the pictures.’

All the data were transcribed and coded by the author. To ensure coding reliability, two additional coders received brief training and independently coded a small subset of the data comprising 100 items (about 4% of all items). The additional coders were fellow PhD students at the Faculty of Philology of Vilnius University. Agreement was assessed using Fleiss’s Kappa, which allows direct comparisons among three raters. Interrater reliability was excellent, with Fleiss’s Kappa indicating almost perfect agreement ($\kappa = .918$; Landis & Koch, 1977). Any discrepancies in coding were analyzed afterward, and the coding scheme was refined.

Table 19. Scoring criteria for participant responses on the gap-fill test

Code	Conditions	Description	Example
1	Correct response	Correct case, correct case paradigm, singular ending	<i>draugas ne-organizuoja švent-ės</i> friend NEG-organizes celebration-GEN 'The friend does not organize a celebration.'
		Only the grammatical ending was assessed; any errors present in the root word, prefixes, etc., were disregarded.	<i>draugas ne-organizuoja *svənt-ės</i> friend NEG-organizes celebration-GEN 'The friend does not organize a celebration.'
	Correct alternative response	Correct case, correct case paradigm, plural ending	<i>draugas ne-organizuoja švenč-ių</i> friend NEG-organizes celebration-GEN.PL 'The friend does not organize celebrations.'
		Certain contextual mismatches for plural marking were disregarded and, thus, were scored 1: - pronouns outside the target predicate-noun relationship. - less lexical-semantic naturalness of the noun in plural form	<i>aš turiu mokinius/-ių,</i> I have student-ACC.PL/-GEN.PL <i>*kuris (...).</i> who-NOM.SG 'I have students who (...).' <i>ne-valgau jogurt-ų</i> NEG-I.eat yoghurt-GEN.PL 'I do not eat yoghurts.'
1a	Correct response with minimal spelling errors	The ending contains a superfluous soft sign (<i>i</i>) or lacks it.	<i>*savaitgal-o</i> [-o should be -io] weekend-GEN <i>*savaitgal-ų</i> [-ų should be -ių] weekend-GEN.PL
		The ending lacks diacritics.	<i>*parduotuves</i> [-es should be -ės] store-GEN <i>*parduotuviiu</i> [-iu should be -ių]

Code	Conditions	Description	Example
			store-GEN.PL
		The ending includes two spelling errors: a missing soft sign and a missing diacritic.	* <i>parduotuvu</i> [-u should be -i <u>ų</u>] store-GEN.PL
		This category also applies to alternative (plural) responses.	<i>gidė visada užduoda klausim-u</i> [-u should be - <i>ų</i>] guide always ask question-GEN.PL 'The guide always asks questions.'
0a	Correct case assignment, incorrect paradigm marking	The ending assigned by the participant belongs to the correct case, but to the incorrect paradigm.	<i>nėra</i> * <i>virtuv-os</i> [-os should be - <i>ės</i>] NEG.is kitchen-GEN 'There is no kitchen.'
			<i>nėra</i> * <i>virtuv-o</i> [-o should be - <i>ės</i>] NEG.is kitchen-GEN 'There is no kitchen.'
	Incorrect assignment of the genitive in plural direct objects.	Plural genitive is a grammatical mistake within the predicate-noun relationship itself.	<i>leidžiu</i> * <i>savaitgalių</i> I.spend weekend-GEN.PL 'I spend weekends.'
			<i>planuojame</i> * <i>kelionių</i> we.plan trip-GEN.PL 'We plan trips.'
		Spelling inconsistencies outlined in 1a apply to the latter 0a category.	<i>savaitga-liu</i> [-iu should be -i <u>ų</u>] weekend-GEN.PL
			<i>kelion-iu</i> [-iu should be -i <u>ų</u>] trip-GEN.PL
0	Incorrect response	Incorrect case assignment	<i>draugas neorganizuoja</i> *švent- <i>ė</i> / friend NEG-organizes celebration-NOM/

Code	Conditions	Description	Example
			<i>*švent-ę</i> celebration-ACC ‘The friend does not organize a celebration.’ [requires genitive marking because of negation]
		Unidentifiable ending	<i>nėra *laik-asa</i> NEG.is time-? ‘There is no time.’
Incorrect nominative ending		The 0a category was not applied to the nominative target structures, since the base form was included with each gap, bracketed.	<i>pasaulyje yra miest-a</i> [-a should be -as] in.world is city ‘In the world there is a city.’
Empty gap		Lack of a response was interpreted as evidence of inability to mark the noun for case.	
Correct case, incorrect paradigm, but with (potential) spelling errors.		The ending assigned potentially belonged to the correct case, but to the incorrect paradigm. Due to spelling error(s), this is hard to determine.	<i>pokalbia</i> [-ia should be -io] conversation-GEN? <i>savaitgala</i> [-a should be -į] weekend-ACC? <i>savaitgales</i> [-es should be -io] weekend-GEN? <i>kelionas</i> [-as should be -es] trip-ACC.PL?

3.6.2 Elicited oral imitation test

Data coding was based on participants' accuracy in case marking, i.e., correct case marking of the subject or direct object in sentences with an affirmative or a negated verb. Scoring included three categories of established response accuracy:

- 1: a completely accurate response
 - o e.g., *nėra sūr-io*, 'NEG.is cheese-GEN' ('there is no cheese')
- 0: a completely incorrect response.
 - o e.g., *nėra *sūr-is*, 'NEG.is cheese-NOM' ('there is no cheese')
- NA: a response lacking the target structure or not meeting one of the other necessary structural requirements.
 - o e.g., *nėra *something*, 'NEG.is ?' ('there is no *something*')

Table 20 further elaborates on the criteria for assigning one of the three categories, together with examples based on the target items.

Some of the scoring decisions require additional explanation. First, regarding sentence structure, the choice was made to evaluate the target response even when the subject or the predicate was missing. Lithuanian is a pro-drop language, which often results in sentences lacking a subject. Furthermore, for lower-proficiency learners, even short prompts may create substantial processing demands, as participants were required to first attend to meaning before formulating a response. This was likely even more difficult if, for instance, the verb was unknown to the participant. As a result, participants may have failed to retrieve or produce the full predicate while still demonstrating knowledge of correct subject or direct-object marking by marking the correct case on the noun. Therefore, responses such as in (32) were accepted:

- (32) *Vakare *pause* dažnai aš *pause* film-q*
in.evening often I movie-ACC
'In the evening I often movie.'

Second, while the target response for grammatically incorrect statements involved correcting the case assigned to the target noun, responses that included changes to the verb's negation or non-negation were also accepted, since these resulted in a grammatically correct response. See the following Example (33):

- (33) Original sentence: *Čia ant stalo yra *butel-io.*
[incorrect genitive here on table is bottle-GEN
marking] ‘There is a bottle on the table.’
Target response: *Čia ant stalo yra butel-is.*
here on table is bottle-NOM
‘There is a bottle on the table.’
Alternative response: *Čia ant stalo nėra butel-io.*
here on table NEG.is bottle-GEN
‘There is no bottle on the table.’

Although participants were instructed to repeat the same sentence without changing its content, they occasionally misremembered it, seemingly relying on the target noun's case ending for repetition.

Third, the sound quality of the vowels in the case endings uttered by the participants often posed challenges to their accurate evaluation. Regularly, it proved impossible to determine with confidence exactly which vowel was pronounced. Since reduced sound quality is often considered a common characteristic for grammatical markers, which, as discussed, makes them more difficult to acquire (N. C. Ellis, 2022), the decision was made to disregard the reduced vowel quality and accept such responses. Only in cases where it was possible to determine with confidence that an erroneous vowel was uttered would the response be considered incorrect. Fourth, for the same reason, it was decided not to include paradigm errors in the assessment process, since it would often have been nearly impossible to establish whether the correct or an alternative paradigm-ending had been produced. Therefore, EOIT responses were scored using only three codes, rather than the four used for the gap-fill test.

Fifth, nonsensical responses were coded slightly differently depending on whether the target noun was identifiable or the sentence as a whole made little sense. While slight deviations were deemed acceptable, the target noun still needed to be identifiable by a native speaker in order to be evaluated and not coded ‘NA’. Granted, in case the target noun was replaced by an alternative yet grammatically correct noun, the response was accepted. Likewise, responses were scored as correct when participants altered a statement in a way that made it semantically illogical but remained grammatically well-formed. This decision was motivated by the nature of the test items, which were relatively unauthentic. In order to remain comprehensible to lower-proficiency learners while still allowing for truth-value judgments, the stimuli had to consist of relatively simple sentences,

which often rendered them unnatural. Furthermore, the object under study was grammatical case, making the meaning conveyed less of a priority.

Finally, because it was extremely challenging to assess correct endings accurately, let alone errors, no error analyses were performed on the data from the EOIT.

All the data were transcribed and coded by the author. To establish interrater reliability, the same two alternative coders received brief training and independently coded a small subset of the data comprising 100 items (about 5% of all items). Agreement was assessed using Fleiss's Kappa, which allows direct comparisons among three raters. Fleiss's kappa was again high ($\kappa = .893$), thus indicating substantial agreement (Landis & Koch, 1977). Any discrepancies in the coding were analyzed afterward, and the coding scheme was refined.

Table 20. Scoring criteria for participant responses on the elicited oral imitation test

Conditions	Code	Examples
<i>Procedure</i>		
Only the target word and its grammatical ending were evaluated.		
Only the last sentence repetition was evaluated.		
The participant did not repeat the sentence.	NA	
The participant successfully repeated the sentence, despite not understanding the meaning.	1	
<i>Structure</i>		
The response included at least a verb and the target noun.	1	<i>Néra sūr-io.</i> NEG.is cheese-GEN 'There is no cheese.'
The response included at least a subject and the target noun.	1	<i>Aš futbol-q.</i> I soccer-ACC 'I soccer.'
The response included only the target noun.	NA	<i>Sūr-is.</i> cheese-NOM 'Cheese.'
The response lacked the target noun.	NA	<i>Čia ant stalo nėra</i> *"something". here on table NEG.is ? 'There is something on the table.'
Not the case, but the verb was changed to match the case marking in the erroneous sentence.	1	Original sentence: <i>Čia ant stalo yra</i> * <i>butel-io.</i> [incorrect genitive here on table is bottle-GEN marking] 'There is a bottle on the table.' Alternative response: <i>Čia ant stalo nėra butel-io.</i> here on table NEG.is bottle-GEN 'There is no bottle on the table.'

Conditions	Code	Examples
In case the prefix <i>ne-</i> is present next to the verb, the genitive is required, even in the case when the verb is unidentifiable.	1	<i>Aš dažnai žodžiu futbol-q.</i> I often ? soccer-ACC ‘I often ? soccer.’ <i>Aš dažnai ne-rau futbol-o.</i> I often NEG-? soccer-GEN. ‘I do not often ? soccer.’
<i>Endings</i>		
Schwa sound (ə)/unclear vowel	1	<i>nėra (*)banan-ə</i> , NEG.is banana-GEN(?), (‘there is no banana’)
Alternative correct plural ending	1	<i>nėra banan-ų</i> , NEG.is banana-GEN.PL, (‘there are no bananas’)
Clearly incorrect vowel	0	<i>nėra *banan-a</i> , NEG.is banana-ACC(?), (‘there is no banana’)
Missing ending	0	<i>nėra *banan-</i> , NEG.is banana-Ø, (‘there is no banan’)
Clearly incorrect paradigm	0	<i>nėra *sof-o</i> , NEG.is couch-GEN, (‘there is no couch’) [-o should be -os]
Incorrect ending due to a change of the verb in the response.	0	Original sentence: <i>Aš žiūriu film-q.</i> I watch movie-ACC ‘I watch a movie.’ Response: <i>Aš *yra *film-q.</i> I is movie-ACC ‘I is a movie.’
<i>Noun</i>		
The ending is correct, but the noun is unidentifiable.	NA	<i>traukin-io</i> , train-GEN → <i>trauk-io</i> , ?-GEN
The ending is correct, but the noun is slightly divergent from the norm.	1	<i>savait-ę</i>, week-ACC → <i>savan-e</i>, ?-ACC <i>fotelis</i>, armchair → <i>sotelis</i>, <i>fortelis</i> <i>aikštė</i>, ‘square’ → <i>ėkštė</i>

Conditions	Code	Examples
The ending is correct, but the noun is different.	1	<i>savait-ę</i> , ‘week-ACC’ → <i>savaitgal-į</i> , ‘weekend-ACC’ <i>lang-as</i> , ‘window-NOM’ → <i>lemp-a</i> , ‘lamp-NOM’
<i>Meaning</i>		
The ending is correct, but the sentence is nonsensical.	1	<i>Aš nevalgau restoran-ų prie namų.</i> I NEG.eat restaurant-GEN.PL near home ‘I do not eat restaurants near my home.’

3.7 Statistical methods

Statistical analyses were conducted by employing Bayesian mixed-effects models via the *brms* package in *R* (Bürkner, 2017). Mixed-effects models were selected because the data had a repeated-measures hierarchical structure, with multiple observations nested within participants across several test times (pretest, posttest, delayed posttest). Furthermore, such models account for the non-independence of repeated observations while also estimating individual differences in baseline performance at pretest and developmental trajectories. Additional random effects accounted for item-level variability. Most importantly, mixed-effects models allow examination of interactions between time and group, thereby enabling comparisons of learning gains over time across treatments.

The choice for Bayesian models requires some additional explanation. Bayesian models and more traditional frequentist logistic mixed-effects models differ in two fundamental ways: (1) the information sources used for parameter estimation and (2) the way the results are interpreted. In frequentist models, estimates rely solely on data from a single study. Bayesian models, on the other hand, incorporate not only the study's data but also prior distributions of parameter values, which are data known or assumed beforehand. By combining these two sources through Bayes' theorem, a posterior distribution is generated that represents the updated range of plausible values for each parameter, incorporating both the priors and the collected data. In frequentist models, the p -value is used to decide whether to accept or reject the alternative hypothesis. No such dichotomous rigidity exists in Bayesian models, which are rooted in probabilities. Instead of focusing solely on a p -value, Bayesian model interpretations assess the strength of evidence for an effect or interaction. For this study, Bayesian models were favored over frequentist logistic mixed-effects models, as the former are noted to perform better with smaller sample sizes and also allow for more flexible interpretations of effects and interactions (Wong, 2025).

To analyze and present the findings from the Bayesian models, three decision rules were used, along with associated criteria for interpretation, all adopted from Wong (2025). This decision was made because Wong (2025) makes suggestions for implementing Bayesian statistics specifically in L2 interventional research. First, *Highest Density Intervals* (HDIs) were utilized to summarize posterior uncertainty. As noted by Wong (2025), HDIs are particularly effective when working with small sample sizes. They represent a range of values from a Bayesian posterior distribution that includes the most

credible parameter values. In this study, the HDI was set at 89%, thus indicating that the interval encompasses all values that account for 89% of the most likely parameter estimates. In other words, one could say, “there is an 89% probability that the true value lies within the reported interval”. In case the HDI’s range is quite narrow and does not include zero, this usually indicates strong evidence for an effect or interaction.

The second decision rule is the *Region of Practical Equivalence* (ROPE), which evaluates the practical significance of the results. A ROPE represents a narrow range of values around zero and is typically applied to determine whether an effect is negligible. ‘Negligible’ is understood here not from a statistical perspective but rather from a practical perspective. In the following analyses, the ROPE was set at ± 0.18 on the log-odds scale. When substantial portions of the posterior distribution fall within the ROPE, this suggests that the effect is likely practically negligible. While no strict threshold was employed, posterior proportions of about 30% or more within the ROPE were interpreted as indicating less clear effects.

Lastly, the Probability of Direction (*pd*) was calculated to indicate the level of certainty that an effect is either positive or negative. A *pd* value nearing 100% reflects strong evidence in favor of a positive or negative direction.

Most of the settings for the Bayesian mixed-effects models were taken from Wong (2025) as well. The response variable in most models was binary, necessitating an analysis using the Bernoulli family with a logit function. One model included all four possible coding outcomes from the gap-fill test, leading to the use of the categorical distribution family with a logit function for the computation of that particular model. Weakly informative priors were utilized: fixed effects were specified with a $\text{normal}(0, 2)$ distribution, while random effect standard deviations were allocated a $\text{Cauchy}(0, 1)$ distribution. Model estimation was performed with four *Markov Chain Monte Carlo* (MCMC) chains, each comprising 6000 iterations and a warm-up period of 2000. The control parameter *adapt_delta* was set to 0.99, and successful convergence was confirmed by verifying that *Rhat* equaled 1.00.

Inclusion of (slightly) more informed priors was explored. One approach that this dissertation considered was to extract priors from the data in Udes (2025a). The mean and standard deviation of each relevant variable and interaction were taken, and each standard deviation was doubled to avoid the priors restricting the model too much. However, certain key differences between the trial study described in Udes (2025a) and the final study had to be addressed for prior computation. First, the trial study included only a gap-

fill test; there was no EOIT. Hence, it was decided to include trial study priors solely for the Syntactic Marking variable, as the effect could hypothetically be similar in both test modalities. Second, the gap-fill test did not include a control group, thus forcing alterations to potential intercept and group priors. This was further necessitated by the fact that, in the trial study, a more lenient scoring method was applied, which accepted responses including spelling errors (in this study, 1a errors). The intercept prior was adjusted accordingly.

Another prior extraction method explored was to use this study's data to center the intercept. This was done by logit-transforming the grand mean of the observed success rate across all conditions. The priors for the remaining variables and the standard deviation were left the same as described above (normal(0, 2), Cauchy(0, 1)). The *loo_compare* function was used to compare the models with weak priors with those with slightly more informed priors. It revealed no advantages for the gap-fill models and only minimal advantages for the elicited oral imitation models with more informed priors. To ensure continuity across all models, it was decided to abandon the idea of extracting more informed priors. As a result, the priors' specification described above and adopted from Wong (2025) was used instead. The priors used for model testing are available in Appendix E.

Random effects always included Items and Participants. Accuracy was mostly represented as a binary outcome variable (1 = correct, 0 = incorrect). For the gap-fill test, three models employed a strict scoring system, meaning only answers marked 1 were deemed correct, whereas responses scored as 1a, 0a, and 0 were recoded as errors and scored 0. This allowed assessment of clear potential learning gains without any caveats. A fourth model was fitted to the same data, using a categorical outcome variable that retained all possible response types (i.e., 1, 1a, 0a, 0), as described in Table 19.

Since the control group completed only the pretest and the posttest, the model arguments differed depending on which test data were analyzed and whether the control group was included. The main purpose of the models that included the control group was to determine whether the treatment groups showed greater gains from pre- to posttest than the group that received no treatment. These models included the fixed effects Test Time (Pretest, Posttest), Treatment Group (TSLT, TBLT, Control), Syntactic Marking (Subject, Object), L1 (Balto-Slavic, non-Balto-Slavic), and the interaction between Test Time and Treatment Group. In the EOIT models, Correctness (Grammatical Item, Ungrammatical Item) was added as an additional fixed effect.

Models excluding the control group were used to examine treatment differences and longer-term development across three test times. These models included Test Time (Pretest, Posttest, Delayed Posttest), Treatment Group (TSLT, TBLT), Syntactic Marking (Subject, Object), L1 (Balto-Slavic, non-Balto-Slavic), and the interactions between Test Time and Treatment Group, and between Test Time and L1. Again, the EOIT models also included Correctness (Grammatical Item, Ungrammatical Item). Additional models were fitted for both tests in which Syntactic Marking was replaced by Case (Nominative, Genitive, Accusative) and Paradigm (-as, -is, -é, -a). An overview of these models can be found in Table 21.

Different kinds of contrast coding were used for the categorical variables depending on the variables in question. Treatment contrast was employed for variables with a clear reference level, i.e., the pretest for Test Time and the control group for Group. This decision was made because these reference levels constitute a clear baseline condition against which the other variable levels (Posttest, Delayed Posttest; TBLT, TSLT) could be compared. Conversely, for variables without a reference level, sum contrast (-0.5, 0.5) was applied, since none of the levels could serve as a logical baseline condition. Sum contrasts assess statistical differences relative to the overall mean across all levels (Schad et al., 2020). This means that, under sum coding, the model no longer tests direct comparisons at the predictor level, and the main-effect coefficients represent the average effect across all levels of the predictor. Interaction terms, on the other hand, represent how individual levels deviate from that average.

Because interaction effects are difficult to interpret directly from the model output, pairwise comparisons were performed to explore these effects. These comparisons primarily concerned within-group improvement from pre- to posttest, or between-group performance at a given Test Time. Pairwise comparisons were computed by using the *emmeans()* and *contrast()* functions from the *emmeans* package. For each comparison, the 89% *Highest Posterior Density* (HPD) is reported, which is essentially equivalent to the HDI in both application and interpretation. As with the HDI, HPD intervals that lie entirely above or below 0 indicate strong evidence of an effect. Visualizations based on the pairwise comparisons were created by using the *ggplot()* function from the *ggplot2* package.

Table 21 Overview of models fitted to the data

Model	Data	Outcome Variable	Test Times	Groups	Random effects	Main Fixed Effects	Additional Predictors
1 Table 23	Gap-fill	Binary (strict)	Pretest, Posttest	TSLT, TBLT, Control	Participants Items	Time, Group, Time x Group	Syntactic Marking, L1
2 Table 24	Gap-fill	Binary (strict)	Pretest, Posttest, Delayed posttest	TSLT, TBLT	Participants Items	Time, Group, Time x Group	Syntactic Marking, L1, Time x L1
3 Table 25	Gap-fill	Binary (strict)	Pretest, Posttest, Delayed posttest	TSLT, TBLT	Participants Items	Time, Group, Case, Paradigm, Time x Case, Time × Paradigm	Syntactic Marking, L1
4 Table 28	Gap-fill	Categorical (1,1a,0a,0)	Pretest, Posttest, Delayed posttest	TSLT, TBLT	Participants Items	Time, Group, Time x Group	Syntactic Marking
5 Table 37	EOIT	Binary	Pretest, Posttest	TSLT, TBLT, Control	Participants Items	Time, Group, Time x Group	Syntactic Marking, L1, Correctness
6 Table 38	EOIT	Binary	Pretest, Posttest, Delayed posttest	TSLT, TBLT	Participants Items	Time, Group, Time x Group	Syntactic Marking, L1, Correctness Time x L1
7 Table 39	EOIT	Binary	Pretest, Posttest, Delayed posttest	TSLT, TBLT	Participants Items	Time, Group, Case, Paradigm, Time x Case, Time x Paradigm	Syntactic Marking, L1, Correctness

4 RESULTS

4.1 Results of the gap-fill test

This subsection covers the gap-fill test results. First, the accuracy scores are analyzed to determine whether there were any significant learning gains. Afterward, the focus shifts to error types and whether they reveal differences in developmental patterns between the two treatment groups.

4.1.1 Progress analysis

4.1.1.1 Strict binary scoring

The overview of the gap-fill test results will begin with the strict scoring assessment of the responses. This means that a ‘1’ was assigned to a perfectly correct response and a ‘0’ was given to an erroneous response. The latter included the categories ‘0a’ and ‘1a’, formerly described in Table 19. This approach led to the descriptive statistics shown in Table 22.

Table 22. Descriptive statistics for the binary scoring of responses on the gap-fill test

Group	Test	Mean (max = 19)	Gains from pretest	SD
TBLT (n=23)	Pretest	8.04		5.72
	Posttest	12.3	+4.26	5.30
	Delayed posttest	12.3	+4.26	4.96
TSLT (n=22)	Pretest	5.59		4.53
	Posttest	9.05	+3.46	4.87
	Delayed posttest	8.68	+3.09	5.05
	(n=19)			
Control (n=15)	Pretest	7.33		5.87
	Posttest	8.33	+1.00	5.51

The table displays the mean performance of all the participants in each group at each test time. The results clearly show quite sizeable learning gains for both treatment groups. While the TSLT group seemed to have been weaker at the outset, as demonstrated by the lower average pretest score (5.59 points out of 19, compared to 8.04 for the TBLT group), learning gains in terms of average accuracy were similar across both groups, with 4.26 for the TBLT

group and 3.46 for the TSLT group from pre- to posttest. The control group also showed modest learning gains, with an average increase of 1 point in accuracy across all participants, thus indicating potential training effects. Another interesting finding is the stable test results observed in the delayed posttest. The TBLT group demonstrated the same average accuracy (12.3) on the post- and delayed posttests, whereas the TSLT group showed only a small decrease from 9.05 to 8.68.

The first model, as presented below in Table 23, was fitted to establish whether the treatment groups made greater learning gains than the control group.¹⁷

Table 23. Bayesian posterior distribution summary for gap-fill test accuracy: Including control group (binary scoring)

Parameter	Median	HDI (89%) Low	HDI (89%) High	<i>pd</i>	% in ROPE (±0.18)
Intercept	-0.49	-1.33	0.34	82.55	18.32
<i>Test Time: Posttest</i>	<i>0.38</i>	<i>-0.17</i>	<i>0.97</i>	<i>86.55</i>	<i>24.24</i>
Treatment: TSLT	-0.57	-1.49	0.37	83.37	15.61
Treatment: TBLT	0.23	-0.72	1.17	64.83	23.21
Syntactic Role: Object	-0.99	-1.48	-0.50	99.92	0.00
L1: Balto-Slavic	0.40	-0.41	1.20	78.48	21.68
Test Time x Treatment:	0.76	0.32	1.22	99.65	0.00
TSLT					
Test Time x Treatment:	1.17	0.71	1.64	100.00	0.00
TBLT					

Note. The reference levels for each variable were as follows: Test Time: Pretest; Treatment: Control group; Syntactic Role: Subject; L1: non-Balto-Slavic.

Abbreviations: HDI = Highest Density Interval; *pd* = Probability of Direction; ROPE = Region of Practical Equivalence.

Strongly supported parameters are bolded. Moderately supported parameters are italicized.

¹⁷ The reader is reminded that the strength of evidence of an effect or interaction within Bayesian modeling can be based on the following three parameters: the 89% HDI, whether it is relatively narrow and does not include 0; the *pd*, values close to 100 indicate a strong positive or negative tendency, depending on the median; the percentage of the posterior distribution that falls within the ROPE, with zero to low percentages indicating practical significance of the results.

Regarding the control group's performance, while there does seem to be a positive effect as indicated by the Test Time coefficient for the posttest (Median = 0.38), it is supported by weak evidence, since the 89% HDI [-0.17, 0.97] includes zero and is quite wide, and the *pd* (86.55%) is not very strong, as it is quite far removed from 100. Nevertheless, the share of posterior estimates within the ROPE (24.24%) is not very large. As both Treatment Group and Test Time were part of an interaction, a pairwise comparison using *emmeans()* and *contrast()* was performed to check for a lack of a clear effect. It also showed weakly supported evidence of pretest-posttest gains for the control group (Median = 0.38, 89% HPD [-0.19, 0.95]). Consequently, although the descriptive statistics showed improved performance in the control group at the posttest, the evidence is insufficient to claim significant learning gains.

As for group differences, there appear to be no marked differences in performance at pretest between the control and the TSLT group (Median = -0.57) as the 89% HDI [-1.49, 0.37] is wide and includes zero, the *pd* (83.37%) is not very strong, although the percentage of posterior estimates that lie in the ROPE (15.61%) is not extremely large. Similarly, there is essentially no evidence for a difference in pretest performance between the control and the TBLT groups (Median = 0.23), since the 89% HDI [-0.72, 1.17] is again wide and includes zero, and the *pd* (64.83%) is closer to 50 than 100, meaning there is little certainty of the positivity of the posterior median. Nevertheless, once more, the percentage of posterior estimates that lie in the ROPE (23.21%) is not extremely large. Because the variable Treatment Group was part of an interaction, these interpretations had to be checked with pairwise comparisons. Nearly identical results were found for both the control group and the TSLT group (Median = -0.57, 89% HPD [-1.48, 0.38]) and the control group and TBLT group comparisons (Median = 0.23, 89% HPD [-0.68, 1.20]). Figure 9 visualizes the interaction between Test Time and Treatment Group using Bayesian logistic regression and shows the observed tendencies.

As Syntactic Marking and L1 were sum-coded and not part of an interaction, their coefficients reflect the average adjusted effect of these predictors across all levels of the other variables in the model. Overall, participants had more difficulty with Object than Subject items, which is shown by the large negative effect (Median = -0.99) and strong evidence, since the 89% HDI [-1.48, -0.50] does not include zero, the *pd* (99.92%) is extremely close to 100, and none of the posterior estimates fall within the ROPE (0%), indicating full practical significance of the results. However, L1

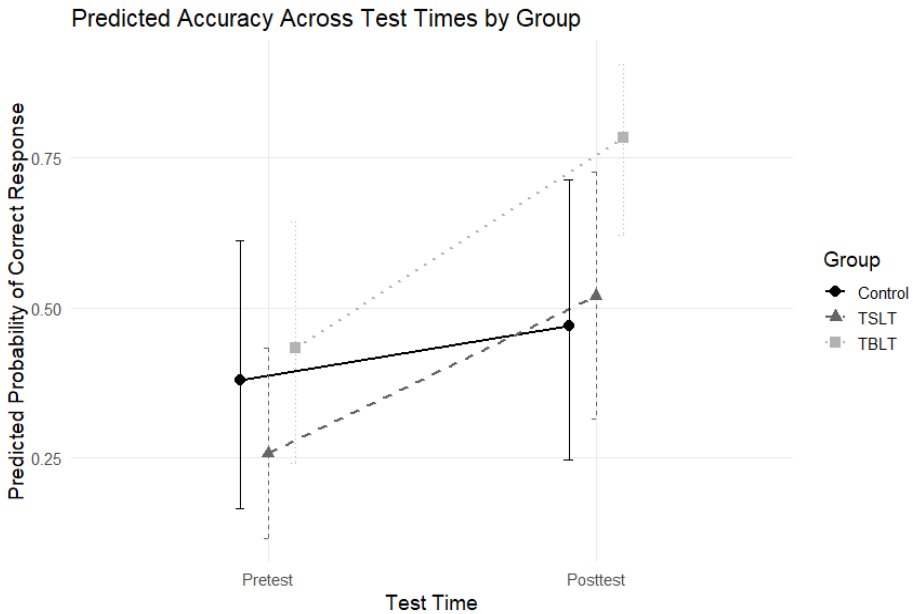


Figure 9. Predicted probabilities of correct responses by Test Time and Treatment Group for the gap-fill model, including the control group. Error bars represent the 89% HDIs

did not appear to be a relevant effect overall (Median = 0.40, 89% HDI [-0.41, 1.20], $pd = 78.48\%$, ROPE = 21.68%).

More relevant, however, are the Test Time and Treatment interactions, as they indicate whether changes in performance from pretest to posttest differed between the treatment groups and the control group. As shown in Table 23, the interaction between Test Time and the TSLT treatment was positive (Median = 0.76) and supported by very strong evidence (89% HDI [0.32, 1.22], $pd = 99.65\%$, ROPE = 0.00%). This indicates that the TSLT group improved more from pre- to posttest than the control group. Likewise, the interaction between Test Time and the TBLT treatment group showed a positive effect (Median = 1.17), again with extremely strong evidence (89% HDI [0.71, 1.64], $pd = 100.00\%$, ROPE = 0.00%), indicating greater gains for the TBLT group relative to the control group. The steeper slopes for the TSLT and TBLT groups shown in Figure 9 corroborate this, thereby allowing for further analyses of the treatment differences across all three test times, including the delayed posttest.

The Bayesian model, fitted to data containing only the treatment groups and the delayed posttest responses, corroborates the descriptive

statistics’ observation that performance was stable across the two posttests.

Table 24. Bayesian posterior distribution summary for gap-fill test accuracy: Two treatment groups (binary scoring)

Parameter	Median	HDI (89%) Low	HDI (89%) High	<i>pd</i>	% in ROPE (±0.18)
Intercept	-0.70	-1.28	-0.11	97.02	5.69
Test Time: Posttest	1.48	0.98	2.00	100.00	0.00
Test Time:	1.35	0.84	1.86	100.00	0.00
Delayed posttest					
<i>Treatment: TBLT</i>	<i>0.82</i>	<i>-0.05</i>	<i>1.72</i>	<i>93.47</i>	<i>9.08</i>
<i>L1: Balto-Slavic</i>	<i>0.02</i>	<i>-0.91</i>	<i>0.92</i>	<i>51.68</i>	<i>26.16</i>
Syntactic Role:	-1.19	-1.61	-0.79	100.00	0.00
Object					
Test Time: Posttest x	0.43	0.01	0.85	94.82	15.40
Treatment: TBLT					
<i>Test Time:</i>	<i>0.39</i>	<i>-0.03</i>	<i>0.82</i>	<i>93.24</i>	<i>20.23</i>
<i>Delayed posttest x</i>					
<i>Treatment: TBLT</i>					
Test Time: Posttest x	0.78	0.33	1.23	99.77	0.00
L1: Balto-Slavic					
<i>Test Time: Delayed</i>	<i>0.18</i>	<i>-0.26</i>	<i>0.64</i>	<i>73.62</i>	<i>42.43</i>
<i>posttest x</i>					
<i>L1: Balto-Slavic</i>					

Note. The reference levels for each variable were as follows: Test Time: Pretest; Treatment: TSLT; L1: non-Balto-Slavic; Syntactic Role: Subject.

HDI = Highest Density Interval; *pd* = Probability of Direction; ROPE = Region of Practical Equivalence.

Strongly supported parameters are bolded. Moderately supported parameters are italicized.

As shown in Table 24, not only is there a strongly supported positive effect for the performance of both groups from the pre- to posttest (Median = 1.48, 89% HDI [0.98, 2.00], *pd* = 100.00%, ROPE = 0.00%), but also from the pre- to delayed posttest (Median = 1.35, 89% HDI [0.84, 1.86], *pd* = 100.00%, ROPE = 0.00%). This indicates that lasting learning gains were achieved, seemingly by both treatment groups. Pairwise comparisons confirm this for both the TSLT group (pretest–posttest: Median = 1.27, 89% HPD [0.73, 1.83]; pretest–delayed posttest: Median = 1.16, 89% HPD [0.59, 1.71]) and the TBLT group (pretest–posttest: Median = 1.70, 89% HPD [1.14, 2.25]; pretest–

delayed posttest: Median = 1.55, 89% HPD [0.99, 2.09]). Figure 10 visualizes the interaction between Test Time and Treatment Group based on Bayesian logistic regression and shows the gains for both groups.

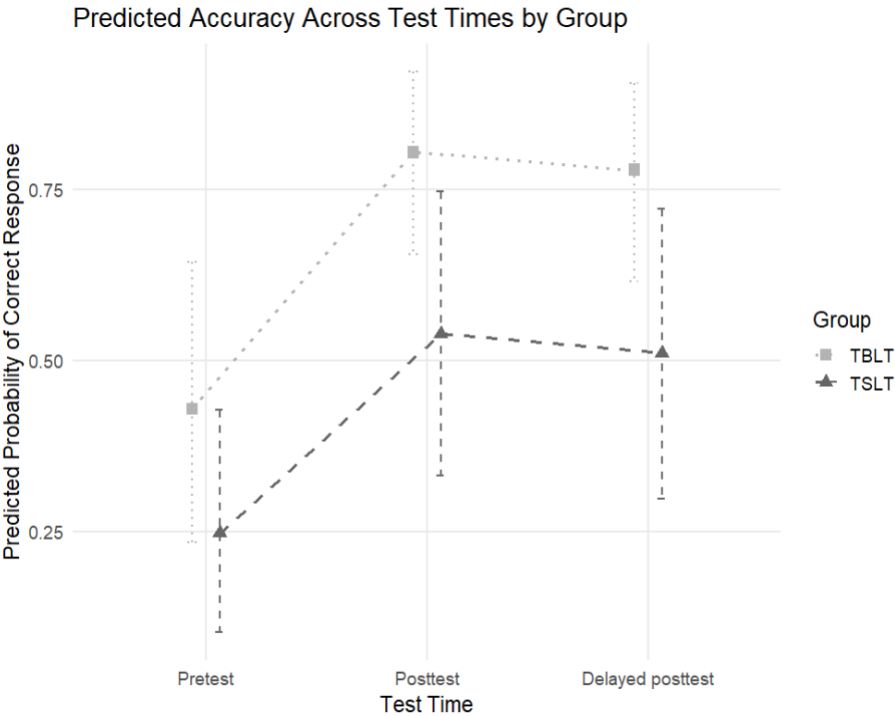


Figure 10. Predicted probabilities of correct responses by Test Time and Treatment Group for the gap-fill model without the control group. Error bars represent the 89% HDIs

However, the Test Time x Treatment Group interaction suggests greater improvement in the TBLT group than in the TSLT group. The posttest and TBLT group interactions display particularly strong evidence of superior improvement (Median = 0.43, 89% HDI [0.01, 0.85], *pd* = 94.82%, ROPE = 15.40%), while the interaction at the delayed posttest for the TBLT group is weaker but still moderate (Median = 0.39, 89% HDI [-0.03, 0.82], *pd* = 93.24%, ROPE = 20.23%). In Figure 10, the slope for the TBLT group from the pretest to the posttest seems slightly steeper than for the TSLT group. This seems surprising, as the TSLT treatment was hypothesized to be more conducive to explicit learning, which the gap-fill test aimed to assess.

Nevertheless, potential differences between the treatment groups at the pretest stage need to be taken into account. While participant selection for each group was random, the model indicates moderate evidence for TBLT

performance superiority over the TSLT group at the pretest level (Median = 0.82, 89% HDI [-0.05, 1.72], $pd = 93.47\%$, ROPE = 9.08%). A pairwise comparison using *emmeans()* and *contrast()* showed a very similar result (Median = 0.82, 89% HDI [-0.08, 1.68]). While the data are inconclusive, superior TBLT improvement at the immediate and delayed posttests might stem from superior L2 knowledge at the outset.

Again, the target syntactic role proved an important effect across all tests, with object marking being much more problematic for both groups (Median = -1.19, 89% HDI [-1.61, -0.79], $pd = 100.00\%$, ROPE = 0.00%). Interactions between Syntactic Role and Treatment Group and Test Time were explored but did not yield any notable findings (see Appendix F). Additionally, while there seemed to be little evidence for positive Balto-Slavic L1 effects on response accuracy at the pretest, there was a markedly strong positive interaction for L1 with the posttest (Median = 0.78, 89% HDI [0.33, 1.23], $pd = 99.77\%$, ROPE = 0.00%), although this interaction was no longer present at the delayed posttest (Median = 0.18, 89% HDI [-0.26, 0.64], $pd = 73.62\%$, ROPE = 42.43%). This might indicate a short-term advantage of having a Balto-Slavic L1 for achieving learning gains of explicit case knowledge. A three-way interaction between Test Time, L1, and Treatment Group was explored but yielded no significant findings (see Appendix F).

Because the type of syntactic marking proved to have such a strong effect, another model was fitted to test for specific case effects (nominative, genitive, accusative), including interactions with Test Time and Treatment Group, and potential influences of the target noun paradigms ('-as', '-is/-ys', '-è', '-a'). The posterior distribution summary of this model is presented in Table 25.

Disregarding the effects previously discussed, there appears to be moderately strong evidence for inferior performance of both groups concerning the genitive in comparison to the nominative and the accusative at the baseline (Median = -0.59, 89% HDI [-1.20, 0.01], $pd = 94.10\%$, ROPE = 11.96%) and exceptionally strong evidence of even more inferior performance with the accusative at the outset (Median = Median = -1.03, 89% HDI [-1.75, -0.30], $pd = 98.54\%$, ROPE = 0.74%). Regarding the paradigms, while evidence of an effect for *-is* does not seem to exist, participants almost surely had more trouble successfully declining nouns belonging to *-è* at the pretest (Median = -1.67, 89% HDI [-2.43, -0.92], $pd = 99.94\%$, ROPE = 0%), while nouns belonging to *-a* gave them less difficulty (Median = 1.09, 89% HDI [0.17, 2.00], $pd = 97.11\%$, ROPE = 3.32%).

Table 25. Bayesian posterior distribution summary for gap-fill test accuracy:
Two treatment groups, including the parameters Case and Paradigm

Parameter	Median	HDI (89%) Low	HDI (89%) High	<i>pd</i>	% in ROPE (±0.18)
Intercept	-0.56	-1.08	-0.04	95.61%	10.52%
Test Time: Posttest	0.93	0.56	1.31	99.99%	0%
Test Time:	1.40	1.02	1.79	100%	0%
Delayed posttest					
<i>Case: Genitive</i>	-0.59	-1.20	0.01	94.10%	11.96%
Case: Accusative	-1.03	-1.75	-0.30	98.54%	0.74%
Paradigm: -is/-ys	0.37	-1.07	0.35	79.59%	24.06%
Paradigm: -è	-1.67	-2.43	-0.92	99.94%	0%
Paradigm: -a	1.09	0.17	2.00	97.11%	3.32%
<i>Treatment: TBLT</i>	0.80	-0.04	1.67	93.61%	8.94%
Syntactic Marking:	-0.40	-0.76	-0.05	96.76%	13.30%
Object					
L1: Balto-Slavic	0.30	-0.57	1.20	71.36%	23.39%
Test Time: Posttest x	2.17	1.33	3.0	100%	0%
Case: Genitive					
<i>Test Time: Delayed</i>	0.81	-0.06	1.65	93.44%	8.88%
<i>posttest x Case: Genitive</i>					
Test Time: Posttest x	-1.37	-2.33	-0.43	98.88%	0%
Case: Accusative					
Test Time: Delayed	-1.27	-2.18	-0.35	98.52%	0.70%
posttest x					
Case: Accusative					
Test Time: Posttest x	0.01	-1.01	1.03	50.68%	23.37%
Paradigm: -is/-ys					
Test Time: Delayed	0.28	-0.72	1.27	67.45%	22.01%
posttest x					
Paradigm: -is/-ys					
Test Time: Posttest x	0.51	-0.51	1.51	79.04%	17.20%
Paradigm: -è					
Test Time: Delayed	-0.03	-1.05	0.99	51.54%	23.65%
posttest x Paradigm: -è					
Test Time: Posttest x	-0.78	-2.05	0.58	82.41%	11.56%
Paradigm: -a					
Test Time: Delayed	0.34	-0.97	1.64	65.71%	16.70%
posttest x Paradigm: -a					

Parameter	Median	HDI (89%) Low	HDI (89%) High	<i>pd</i>	% in ROPE (±0.18)
Case: Genitive x Treatment: TBLT	0.89	0.42	1.36	99.89%	0%
Case: Accusative x Treatment: TBLT	0.71	0.15	1.26	97.85%	4.40%
Paradigm: -is/-ys x Treatment: TBLT	-0.15	-0.74	0.46	64.88%	35.76%
Paradigm: -ě x Treatment: TBLT	0.69	0.08	1.3	96.46%	7.18%
Paradigm: -a x Treatment: TBLT	-0.89	-1.67	-0.11	96.38%	5.13%

Note. The reference levels for each variable were as follows: Test Time: Pretest; Treatment: TSLT; Case: Nominative; Paradigm: -as; Syntactic Marking: Object; L1: non-Balto-Slavic.

HDI = Highest Density Interval; *pd* = Probability of Direction; ROPE = Region of Practical Equivalence.

Strongly supported parameters are bolded. Moderately supported parameters are italicized.

Interactions between Test Time and Case all displayed moderate to strong evidence. Clearly, participants in both groups improved markedly more in marking the genitive case at the posttest than in marking the nominative or the accusative cases (Median = 2.17, 89% HDI [1.33, 3.00], *pd* = 100%, ROPE = 0%). At the same time, such a superior improvement was moderately supported at the delayed posttest (Median = 0.81, 89% HDI [-0.06, 1.65], *pd* = 93.44%, ROPE = 8.88%). Strong evidence was found for inferior learning gains from pre- to posttest concerning the accusative case in contrast to the nominative and genitive both at the posttest (Median = -1.37, 89% HDI [-2.33, -0.43], *pd* = 98.88%, ROPE = 0%), as well as the delayed posttest (Median = -1.27, 89% HDI [-2.18, -0.35], *pd* = 98.52%, ROPE = 0.70%).

Since these results could indicate not only poorer accusative case performance but even a decline in case knowledge, pairwise comparisons were computed to explore variable-specific changes over Test Time. Interestingly, while no evidence was found for pre- to posttest improvement for marking the accusative (Median = 0.25, 89% HPD [-0.35, 0.86]), pre- to delayed posttest learning gains (Median = 0.76, 89% HPD [0.18, 1.35]) did occur, while there is even some suggestive evidence of learning gains between

the immediate post- to delayed posttest (Median = 0.52, 89% HPD [-0.13, 1.17]), suggesting improvements in the time slot when the treatment had already finished. Surprisingly, performance on marking the nominative case appears to be similar, with no pre- to posttest improvement (Median = 0.54, 89% HPD [-0.22, 1.31]), while pre- to delayed posttest (Median = 1.63, 89% HPD [0.84, 2.41]) as well as immediate post- to delayed posttest (Median = 1.09, 89% HPD [0.30, 1.87]) improvement does seem to exist. Figure 11 visualizes these interactions using Bayesian logistic regression estimates and confirms the observed patterns.

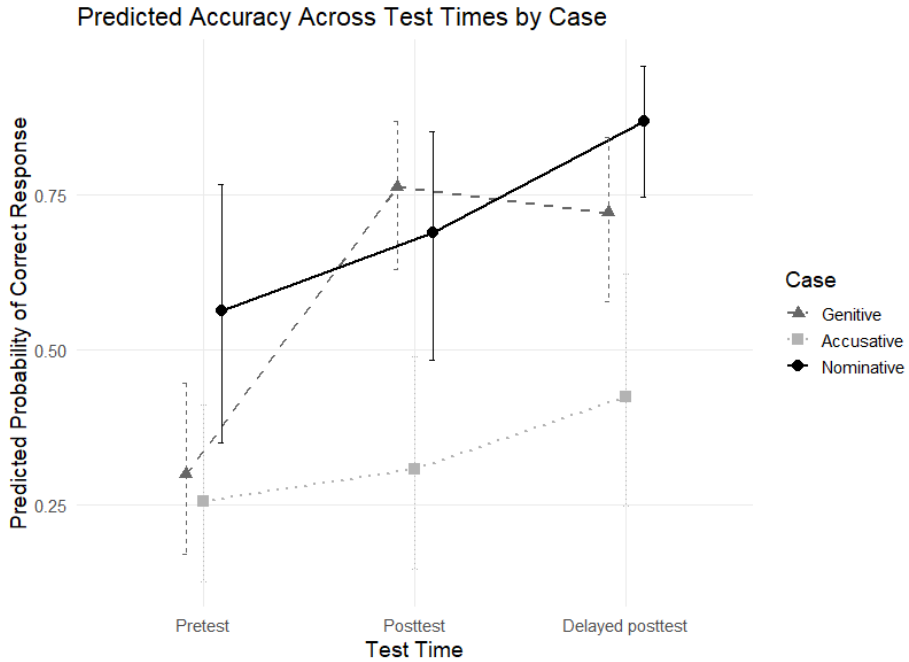


Figure 11. Predicted probabilities of correct responses by Test Time and Case for the gap-fill model without the control group. Error bars represent the 89% HDIs

The model provided no strong evidence of interactions between Test Time and Paradigm, indicating that learning gains did not differ substantially across paradigms. In order to explore changes over Test Time within each separate paradigm, the *emmeans()* and *contrast()* functions were utilized to compute pairwise comparisons. These results are summarized in Table 26.

Table 26. Pairwise comparisons for Test Time and Paradigm

Interaction	Median	HPD (89%)	
		Low	High
Pretest – Posttest: -is/-ys	0.94	0.33	1.55
Pretest – Delayed posttest: is/-ys	1.54	0.96	2.11
<i>Posttest – Delayed posttest: -is/-ys</i>	<i>0.59</i>	<i>-0.02</i>	<i>1.22</i>
Pretest – Posttest: -è	1.19	0.55	1.82
Pretest – Delayed posttest: -è	1.39	0.75	2.03
<i>Posttest – Delayed posttest: -è</i>	<i>0.20</i>	<i>-0.40</i>	<i>0.82</i>
<i>Pretest – Posttest: -a</i>	<i>0.55</i>	<i>-0.28</i>	<i>1.42</i>
Pretest – Delayed posttest: -a	1.56	0.70	2.45
Posttest – Delayed posttest: -a	1.02	0.05	1.97
Pretest – Posttest: -as	1.06	0.54	1.58
Pretest – Delayed posttest: -as	1.10	0.59	1.63
<i>Posttest – Delayed posttest: -as</i>	<i>0.05</i>	<i>-0.50</i>	<i>0.60</i>

Note. HPD = Highest Posterior Density.

Strongly supported parameters are bolded. Moderately supported parameters are italicized.

These results show pre- to posttest and pre- to delayed posttest improvement across essentially every paradigm, suggesting that the paradigms were not an obstacle to achieving case-related learning gains. The only exception seems to be the *-a* paradigm, which does not display any evidence for pre- to posttest learning gains (Median = 0.55, 89% HPD [-0.28, 1.42]), while pre- to delayed posttest (Median = 1.56, 89% HPD [0.70, 2.45]) and even immediate post- to delayed posttest improvement (Median = 1.02, 89% HPD [0.05, 1.97]) do display strong evidence. However, the reader is reminded that the number of *-a* type items in each test was only two, which makes it hard to draw any firm conclusions.

Going back to the results provided in Table 25, the interactions between Case and Treatment Group as well as Paradigm and Treatment Group averaged over all test times reveal higher odds of a correct response for the TBLT group when marking the genitive (Median = 0.89, 89% HDI [0.42, 1.36], *pd* = 99.89%, ROPE = 0%) and the accusative (Median = 0.71, 89% HDI [0.15, 1.26], *pd* = 97.85%, ROPE = 4.40%), as well as declining *-è* paradigm nouns (Median = 0.69, 89% HDI [0.08, 1.30], *pd* = 96.46%, ROPE = 7.18%). However, they demonstrated inferior performance when declining *-a* paradigm nouns in comparison with the TSLT group (Median = -0.89, 89% HDI [-1.67, -0.11], *pd* = 96.38%, ROPE = 5.13%). Three-way interactions for Test Time, Treatment Group, and Case/Paradigm were also explored, but

yielded no noteworthy results. This suggests that the TBLT group performed better overall than the TSLT group, but there is no clear evidence of differential learning gains between the groups across test times (see Appendix F).

In order to check for potential negative item outliers, the model presented in Table 24 was refitted so that item effects were estimated separately within each group, which was done by nesting the random intercept of Item within the variable Treatment Group. This allowed the random intercepts to be visualized in a caterpillar plot (see Appendix G). It shows model-estimated probabilities for each item, faceted by Treatment Group and Test Time. Dots represent the estimated probability of a correct response of a single item, with horizontal intervals indicating the corresponding 95% credible intervals. Dots further to the left indicate more difficult items, whereas dots further to the right indicate easier items. The items are ordered from most to least challenging for each test phase.

Negative outliers were identified based on a posterior mean estimate below 0.375 on the probability scale. By this metric, at the pretest, both treatment groups saw difficulties responding to items 9 (target word: *parduotuvė* ‘shop’; target case: genitive; syntactic role: subject), 8 (*virtuvė* ‘kitchen’, genitive, subject), 6 (*traukinys* ‘train’, genitive, subject), and 14 (*pokalbis* ‘conversation’, genitive, object). The TBLT group demonstrated inferior performance on item 18 (*šventė* ‘holiday/celebration’, genitive, object), while the TSLT group was exceptionally challenged by items 12 (*klausimas* ‘question’, accusative, object), 19 (*kokybė* ‘quality’, genitive, object), and 7 (*sąskaita* ‘check/account’, genitive, subject). Unsurprisingly, almost none of these belong to the most frequent *-as* paradigm. The only item with an *-as* noun led mainly to oblique-case assignment errors, suggesting that participants were primarily challenged by the item’s underlying logic. This could be the case, at least in part, due to the verb *užduoti*, which did not meet the requirements outlined in Section 0. Similar problems were observed in Udes (2025a); however, it was thought that including the present tense form, rather than the past tense form in Udes (2025a), could lead to better results. Yet, this does not seem to be the case. As for the other items, the vast majority of errors are related to the base form, suggesting that participants were either unaware of a context necessitating a certain case form (most of the problematic items required the genitive, where the expected answer involved negation) or unsure how to inflect nouns ending in *-ė/-is/-ys*, or both.

Regarding the posttest, both treatment groups faced difficulties with items 40 (*Panevėžys*, a city in Lithuania; accusative, object), 38 (*vaikas*

‘child’, accusative, object), and 46 (*saulė* ‘sun’, accusative, object), while mainly the TBLT group was challenged by items 34 (*kavinė* ‘café’, nominative, subject) and 33 (*knyga* ‘book’, nominative, subject) and the TSLT group by items 41 (*rūsys* ‘basement’, accusative, object) and 35 (*eilė* ‘line/queue’, genitive, subject). At the delayed posttest, both groups found items 61 (*sostinė* ‘capital’, genitive, object) and 63 (*vakaras* ‘evening’, accusative, object) particularly difficult, while only the TBLT group had difficulties responding to items 68 (*renginys* ‘event’, accusative, object) and 71 (*katė* ‘cat’, accusative, object).

Because both treatment groups improved across the tests, the total number of problematic items was lower on the immediate posttest and delayed posttest. The endings *-ė/-ys* seemed particularly difficult. Beyond endings, the nature of most difficulties seemed to change somewhat from the pretest to both posttests. During the pretest, the majority of errors stemmed from using the base form, while items 40 and 46 for the TSLT group and items 33, 34 (posttest), and 63 (delayed posttest) for both groups involved an erroneous oblique case. Most problems in item 61 were either base-form or paradigm errors. This suggests that, rather than simply leaving the base nominative form, participants were more inclined to experiment with case usage. This might even have been detrimental to the participants’ performance, judging from problems that arose with items 33 and 34, which required the nominative. In the pretest, most problems arose from marking the genitive, whereas in the subsequent posttests, more problems emerged with the items requiring the accusative.

Although these problems might be attributed to the endings or case, potential issues arising from the lexical elements in the items must not be disregarded. The nouns *rūsys* ‘basement’ and *renginys* ‘event’ are not necessarily likely to feature often in learner contexts, while the verbs used in items 63 (*leisti* ‘spend’) and 71 (*auginti* ‘raise, take care of’) may be frequent but did not feature in the textbooks consulted for this study. Similarly, while most case errors with the target noun *knyga* ‘book’ involved the accusative, the noun *kavinė* ‘café’ was often erroneously marked in the locative, a frequently encountered form for nouns denoting locations.

Overall, the analysis suggests that learner difficulty with certain items was systematically related to morphological form, the target case, and the lexical context. At the pretest, the greatest challenges arose from less frequent declensional paradigms and contexts requiring the genitive under negation. The number of problematic items decreased substantially over the subsequent test times, indicating overall learning gains. At the same time, errors shifted

from reliance on the base form towards greater experimentation with oblique case marking, thus suggesting developing but still unstable knowledge of morphological endings.

4.1.1.2 Categorical scoring

Moving on to the categorical response scores, the descriptive statistics already allow for some interesting observations (see Table 27). Firstly, when focusing solely on the ‘1’ (correct response) and ‘0’ (incorrect response) categories, the results again show large learning gains for both treatment groups from pre- to posttest. Secondly, the total number of ‘1a’ (a response including spelling errors) and ‘0a’ (a response including paradigm errors or an incorrect assignment of the genitive for a plural noun; for a full overview, see Section 3.6.1) scores seems relatively small, casting doubt on the need for such a detailed analysis.

Nevertheless, they do reveal some interesting tendencies worth exploring. While the distribution of responses appears quite similar at the delayed posttest for the TSLT group, the TBLT group shows subtle changes. The number of fully correct responses remained stable, but the number of completely incorrect responses decreased and was replaced by ‘1a’ and ‘0a’ responses. This indicates further positive L2 developments for these participants, as the responses are no longer ‘fully’ incorrect but now include accurate case assignment, with only minor errors in spelling or paradigm selection. These data were also fitted with a Bayesian model, which can be found below, in Table 28.

Despite the low number of alternative ‘1a’ and ‘0a’ responses, they provide noteworthy insights into developmental patterns. The most significant evidence can be found for category ‘1’ (correct response). Since these responses correspond to the same responses assigned ‘1’ in previous analyses, these significant effects will not be presented again. However, interestingly, while in this model there seems to be slightly stronger evidence for a positive interaction involving the delayed posttest and the TBLT treatment than in the previous binary model (Median = 0.45, 89% HDI [0.00, 0.90], $pd = 94.57\%$, ROPE = 15.15%), evidence for such an interaction at the posttest in this model is essentially non-existent (Median = 0.15, 89% HDI [-0.28, 0.60], $pd = 70.12\%$, ROPE = 45.03%).

Table 27. Descriptive statistics for the categorical scoring of responses on the gap-fill test

Group	Test	Code							
		1		1a		0a		0	
TBLT (n=23)	Pretest	185	(42%)	38	(9%)	13	(3%)	201	(46%)
	Posttest	284	(65%)	17	(4%)	12	(3%)	124	(28%)
	Delayed posttest	283	(65%)	28	(6%)	28	(6%)	98	(22%)
TSLT (n=22)	Pretest	123	(29%)	43	(10%)	6	(1%)	246	(59%)
	Posttest	199	(48%)	48	(11%)	22	(5%)	149	(36%)
	Delayed posttest (n=19)	165	(46%)	41	(11%)	20	(6%)	135	(37%)

Note. Percentages may not sum to 100% due to rounding.

Table 28. Bayesian posterior distribution summary for gap-fill test accuracy: Two treatment groups (categorical scoring)

	Parameter	Median	HDI	HDI	<i>pd</i>	% in ROPE (±0.18)
			(89%) Low	(89%) High		
0a	Intercept 0a	-3.81	-4.52	-3.18	100.00	0.00
	Intercept 1	-0.49	-1.06	0.07	91.81	16.57
	Intercept 1a	-2.28	-2.84	-1.77	100.00	0.00
	Test Time: Posttest	1.13	0.36	1.92	98.81	0.17
	Test Time:	1.70	0.93	2.47	99.94	0.00
	Delayed posttest					
	<i>Treatment: TBLT</i>	0.81	-0.03	1.68	93.76	9.07
	Syntactic Role: Object	-0.18	-0.79	0.47	67.36	33.21
	Test Time: Posttest x	-1.32	-2.28	-0.38	98.78	0.14
	Treatment: TBLT					
	Test Time:	-0.23	-1.15	0.67	66.09	24.30
	Delayed posttest x Treatment: TBLT					
	Test Time: Posttest	1.46	0.95	1.97	100.00	0.00
1	Test Time:	1.56	1.05	2.06	100.00	0.00
	Delayed posttest					
	<i>Treatment: TBLT</i>	0.84	-0.03	1.70	93.76	8.68
	Syntactic Role:	-1.16	-1.59	-0.74	100.00	0.00
	Object					

Parameter		Median	HDI (89%) Low	HDI (89%) High	<i>pd</i>	% in ROPE (±0.18)
1a	Test Time: Posttest x Treatment: TBLT	0.15	-0.28	0.60	70.12	45.03
	Test Time: Delayed posttest x Treatment: TBLT	0.45	0.00	0.90	94.57	15.15
	Test Time: Posttest	0.28	-0.39	0.94	75.33	28.24
	<i>Test Time:</i>	<i>0.53</i>	<i>-0.12</i>	<i>1.19</i>	<i>90.46</i>	<i>16.12</i>
	<i>Delayed posttest</i>					
	Treatment: TBLT	-0.07	-0.67	0.51	57.16	38.89
	Syntactic Role: Object	0.34	-0.21	0.93	83.97	26.31
	Test Time: Posttest x Treatment: TBLT	-0.95	-1.61	-0.31	99.09	0.20
	Test Time:	-0.06	-0.69	0.55	56.26	37.68
	Delayed posttest x Treatment: TBLT					

Note. The reference levels for each variable were as follows: Test Time: Pretest; Treatment: TSLT; Syntactic Role: Subject.

HDI = Highest Density Interval; *pd* = Probability of Direction; ROPE = Region of Practical Equivalence.

Strongly supported parameters are bolded. Moderately supported parameters are italicized.

In contrast to the fully correct responses, which show an advantage of the TBLT treatment, the less accurate ‘1a’ and ‘0a’ responses paint a different picture. While strong positive effects were found for ‘0a’ (paradigm or genitive plural error) responses in the post- (Median = 1.13, 89% HDI [0.36, 1.92], *pd* = 98.81%, ROPE = 0.17%) and delayed posttest (Median = 1.70, 89% HDI [0.93, 2.47], *pd* = 99.94%, ROPE = 0%) for both treatment groups, strong support was found for a negative interaction between the posttest and the TBLT group (Median = -1.32, 89% HDI [-2.28, -0.38], *pd* = 98.78%, ROPE = 0.14%). This finding quite accurately reflects the descriptive statistics, which showed a fairly low number of ‘0a’ responses in the TBLT group at the posttest compared with the TSLT group, after which the number more than doubled at the delayed posttest. This indicates a sharp increase in ‘0a’ responses compared to ‘0’ responses in the TBLT group, relative to the TSLT group. Furthermore, a strongly supported negative interaction can be observed for ‘1a’ (spelling error) responses at the posttest for the TBLT group (Median = -0.95, 89% HDI [-1.61, -0.31], *pd* = 99.09%, ROPE = 0.20%),

although not at the delayed posttest (Median = -0.06, 89% HDI [-0.69, 0.55], $pd = 56.26\%$, ROPE = 37.68%).

While the TBLT group demonstrated an increase at the delayed posttest in '1a' errors, this increase still paled in comparison with the rise in such responses for the TSLT group. Taken together, the TSLT group's advantages for '0a' and '1a' responses and the TBLT group's advantage for fully correct '1' responses suggest different developmental patterns. While the TSLT group appears to improve case marking gradually, the TBLT group appears to progress more directly from incorrect to fully correct responses, largely bypassing spelling and paradigm errors. Nevertheless, stating this point firmly seems impossible due to the possibility of 0a- (Median = 0.81, 89% HDI [-0.03, 1.68], $pd = 93.76\%$, ROPE = 9.07%) and 1-type (Median = 0.84, 89% HDI [-0.03, 1.70], $pd = 93.76\%$, ROPE = 8.68%) response advantages for the TBLT group at pretest, as well as the absence of strong evidence for the interaction on 1-type responses between Posttest and TBLT group.

4.1.2 Error type analysis

As could be gathered from the previous models, different response categories revealed interesting developmental patterns that warrant further analysis of the errors. Below, the tables first display errors by target syntactic role and treatment group, followed by analyses of errors across different cases and treatment groups. As a reminder, the following error types were distinguished: erroneous assignment of any of the oblique cases in Lithuanian; erroneous use of the nominative base form, which suggests no evidence of case assignment; accurate case assignment, but with marking of an ending that belongs to a different paradigm than that of the target noun; an unidentifiable case or ending (?); absence of any response (NA). Percentages of error types in each test relate only to the oblique case, base form, and paradigm errors, as these are most informative. In the following, whenever changes in error rates from one test to another are discussed and the tests differed in the number of target items or participants, error rates were normalized by calculating them separately for each error type and test as the number of errors divided by the total possible number of responses.

Table 29. Error-type distributions on items requiring subject marking

Group	Test	Oblique case errors	Base form errors	Paradigm errors	?	NA	Total errors
TBLT (n=23)	Pretest	38 (54.0%)	23 (32.0%)	10 (14.0%)	12	7	90
	Posttest	33 (75.0%)	8 (18.2%)	3 (6.8%)	1	0	45
	Delayed posttest	16 (48.5%)	6 (18.2%)	11 (33.3%)	1	0	34
TSLT (n=22)	Pretest	32 (39.5%)	46 (56.8%)	3 (3.7%)	10	5	96
	Posttest	30 (75.0%)	4 (10.0%)	6 (15.0%)	6	3	49
	Delayed posttest (n=19)	23 (62.2%)	7 (18.9%)	7 (18.9%)	9	1	47

Note. ? = unidentifiable error; NA = no response; Percentages of error types relate only to the oblique case, base form, and paradigm errors.

Table 29 shows error types for the subject-marking items at each test time by treatment group. As a reminder, subject marking involved the nominative and genitive cases. For subject marking, participants in the TBLT group showed a sizeable decrease in errors from pre- to posttest (-50%). This, however, involved mainly the base form (-65.2%), the paradigm (-70%), and unidentifiable (-91.7%) or missing responses (-100%), as case assignment errors decreased only slightly (-13.2%). Interestingly, however, while the number of errors further decreased at the delayed posttest (-24%), paradigm errors grew considerably (+267%), back to baseline. As this coincided with a sharp decrease in case assignment errors (-52%), this seems to suggest that the TBLT group began to assign cases more accurately, but by marking endings that corresponded to non-target paradigms.

As for the TSLT group, the total number of errors similarly decreased markedly from pre- to posttest (-49%), largely due to a decrease in base form errors (-91.3%). Most other error types, especially case assignment (-6.3%), remained largely stable. This implies that participants more frequently began declining nouns and, apparently, quite successfully so. Interestingly, while the number of errors remained relatively stable at the delayed posttest, the weighted difference showed a slight increase in total errors compared to the immediate posttest for this group (+11.1%). This was observed despite a

Table 30. Error-type distributions on items requiring object marking

Group	Test	Oblique case errors	Base form errors	Paradigm errors	?	NA	Total errors
TBLT (n=23)	Pretest	46 (49.5%)	42 (45.2%)	5 (5.4%)	11	20	124
	Posttest	41 (49.4%)	34 (41.0%)	8 (9.6%)	7	2	92
	Delayed posttest	33 (42.3%)	33 (42.3%)	12 (15.4%)	5	10	93
TSLT (n=22)	Pretest	48 (40.3%)	66 (55.5%)	5 (4.2%)	28	9	156
	Posttest	63 (60.6%)	31 (29.8%)	10 (9.6%)	11	7	122
	Delayed posttest	41 (44.6%)	42 (45.7%)	9 (9.8%)	13	2	107

Note. ? = unidentifiable error; NA = no response; Percentages of error types relate only to the oblique case, base form, and paradigm errors.

further decrease in case-marking errors, which might suggest quite stable retention of explicit knowledge of case assignment but a decrease in knowledge of case endings and the appropriate paradigms.

Table 30 shows a similar distribution, but for object-marking items. Object marking involved the genitive and accusative cases. As predicted by the models described above, which demonstrated strong evidence for participants' difficulties with this syntactic role, these results show somewhat different trajectories than those for subject marking. The total number of errors decreases for both treatment groups at the posttest (TBLT: -26%; TSLT: -21.8%); however, they then remain stable at the delayed posttest (TBLT: +1.1%; TSLT: 1.6%). For the TBLT group, the decrease in skipped answers contributed to a decrease in errors (-90%), indicating a slight increase in accuracy. In terms of error type distributions, however, they remain largely stable over all tests, save for paradigm errors, which gradually increase (posttest: +60%; delayed posttest: +50%), albeit in small total numbers. Nevertheless, again, since this coincides with a decrease in oblique case assignment errors (posttest: -11%; delayed posttest: -19.5%) and base form errors (posttest: -19%; delayed posttest: -2.9%), this suggests that case assignment for the TBLT participants improved but did not fully coincide with paradigm accuracy. Regarding the TSLT group, superior posttest performance was mainly due to a decrease in unidentifiable (-60.7%) and base form errors (-53). Nevertheless, this led to a rise in case assignment errors (+31.3%),

which implies that participants began to decline nouns more frequently, though not necessarily accurately. As for the error type distribution at the delayed posttest, it seems to reflect a return to baseline, with a decline in case assignment errors (-24.6%) but a rise in base form errors (+56.9%).

Table 31 covers errors concerning the nominative. Interestingly, the TBLT group demonstrated gradual improvement across tests, with case assignment errors remaining stable at the posttest but decreasing substantially at the delayed posttest (-54%). The TSLT group seems to exhibit a similar developmental pattern, although it reduced its case assignment errors more gradually over all tests (Posttest: -28.6%; Delayed posttest: 46%). Since the nominative base forms were given in brackets, many of the error types besides those involving oblique cases were unlikely. Hence, further analysis of nominative errors is not particularly insightful.

Table 31. Error-type distributions on items requiring the nominative

Group	Test	Oblique case errors	Base form errors	Paradigm errors	?	NA	Total errors
TBLT (n=23)	Pretest	24 (100.0%)	0 (0.0%)	0 (0.0%)	4	2	30
	Posttest	24 (100.0%)	0 (0.0%)	0 (0.0%)	0	0	24
	Delayed posttest	11 (91.7%)	0 (0.0%)	1 (8.3%)	0	0	12
TSLT (n=22)	Pretest	21 (100.0%)	0 (0.0%)	0 (0.0%)	3	2	26
	Posttest	15 (100.0%)	0 (0.0%)	0 (0.0%)	3	1	19
	Delayed posttest	7 (100.0%)	0 (0.0%)	0 (0.0%)	5	1	13

Note. ? = unidentifiable error; NA = no response; Percentages of error types relate only to the oblique case, base form, and paradigm errors.

Table 32 deals with the genitive, making it the only table that covers both subject and object marking. Both treatment groups show improvements in genitive case performance from pre- to posttest (TBLT: -50%; TSLT: -52.9%), which remain stable at the delayed posttest (TBLT: +6%; TSLT: +7.9%). However, while error distributions at the delayed posttest for the TSLT group show an increase in case assignment (18.9%), paradigm (10.4%),

Table 32. Error-type distributions on items requiring the genitive

Group	Test	Oblique case errors	Base form errors	Paradigm errors	?	NA	Total errors
TBLT (n=23)	Pretest (n=10)	34 (38.2%)	42 (47.2%)	13 (14.6%)	14	7	110
	Posttest (n=11)	21 (42.0%)	19 (38.0%)	10 (20.0%)	4	1	55
	Delayed posttest (n=10)	9 (18.8%)	16 (33.3%)	23 (47.9%)	2	3	53
TSLT (n=22)	Pretest (n=10)	23 (20.5%)	84 (75.0%)	5 (4.5%)	20	7	139
	Posttest (n=11)	30 (54.5%)	10 (18.2%)	15 (27.3%)	9	8	72
	Delayed posttest (n=10)	28 (52.8%)	12 (22.6%)	13 (24.5%)	8	0	61

Note. ? = unidentifiable error; NA = no response; Percentages of error types relate only to the oblique case, base form, and paradigm errors.

and especially base-form errors (22.6%), the TBLT group showed a strong decrease in case assignment errors (-52.9%) and a marked increase in paradigm errors (+153%). Although the number of cases is relatively small, a similar trend was observed during the pilot study, specifically, Udes (2025a). This suggests that, by the delayed posttest, the TBLT participants had become more proficient at accurate case assignment but still lacked the skills to assign the corresponding paradigm. The TSLT group, on the other hand, seemed to demonstrate partial attrition in their case-marking skills, as indicated by the relative increase in base-form errors from post- to delayed posttest.

Table 33 includes errors concerning the accusative case. Evidently, despite certain minimal fluctuations, participants' performance in marking the accusative remained fairly stable over all tests. Surprisingly, however, both groups demonstrated a slight increase in errors at the posttest (TBLT: +7.1%; TSLT: +10.3%) and a decrease at the delayed posttest (TBLT: -11%; TSLT: -19%). Nevertheless, these findings seem to indicate little learning gains for this type of case marking. Error distributions do show some subtle dynamics, including slight increases in case assignment (+33.8%) and base-form errors (+20%) for the TBLT group at the posttest, which then largely diminish at the delayed posttest (oblique case errors: -17%; base-form errors: -20%).

Table 33. Error-type distributions on items requiring the accusative

Group	Test	Oblique case errors	Base form errors	Paradigm errors	?	NA	Total errors
TBLT (n=23)	Pretest	26	23	2	5	9	65
	(n=6)	(51.0%)	(45.1%)	(3.9%)			
	Posttest	29	23	1	4	1	58
	(n=5)	(54.7%)	(43.4%)	(1.9%)			
	Delayed posttest	29	22	0	4	7	62
	(n=6)	(56.9%)	(43.1%)	(0.0%)			
TSLT (n=22)	Pretest	36	28	3	15	5	87
	(n=6)	(53.7%)	(41.8%)	(4.5%)			
	Posttest	48	25	1	5	1	80
	(n=5)	(64.9%)	(33.8%)	(1.4%)			
(n=19)	Delayed posttest	29	26	3	8	1	67
	(n=6)	(50.0%)	(44.8%)	(5.2%)			

Note. ? = unidentifiable error; NA = no response; Percentages of error types relate only to the oblique case, base form, and paradigm errors.

The TSLT group, similarly, demonstrated a large increase in case assignment errors at the posttest (+60%), which then decreased at the delayed posttest (-42%). Paradigm error percentages are quite large, but due to their low absolute numbers they are not very informative.

Finally, Table 34 and Table 35 summarize the patterns observed in these error-type analyses. Changes in error types over test time are additionally described in words. Table 35 does not include the nominative, as it is less informative given the limited variety of errors.

Table 34. Overview of error-type distributions by subject and object marking and changes across test times

Group	Test	Subject marking			Object marking		
		Error type	Change in %	Change description	Error type	Change in %	Change description
TBLT	Posttest	Oblique case	-13.2	decrease	Oblique case	-10.9	decrease
		Base form	-65.2	strong decrease	Base form	-19.0	decrease
		Paradigm	-70.0	strong decrease	Paradigm	60.0	increase
	Delayed posttest	Oblique case	-51.5	strong decrease	Oblique case	-19.5	decrease
		Base form	-25.0	decrease	Base form	-2.9	stable
		Paradigm	266.7	strong increase	Paradigm	50.0	increase
TSLT	Posttest	Oblique case	-6.3	stable	Oblique case	31.3	increase
		Base form	-91.3	strong decrease	Base form	-53.0	strong decrease
		Paradigm	100.0	strong increase	Paradigm	100.0	strong increase
	Delayed posttest	Oblique case	-11.2	decrease	Oblique case	-24.6	decrease
		Base form	102.6	strong increase	Base form	56.9	strong increase
		Paradigm	35.1	increase	Paradigm	4.2	stable

Note. Where tests differed in the number of target items and/or participants, error rates were normalized by dividing the number of errors for each error type by the total possible number of responses and expressing the result as a percentage.

Table 35. Overview of error-type distributions by genitive and accusative marking and changes across test times

Group	Test	Genitive			Accusative		
		Error type	Change in %	Change description	Error type	Change in %	Change description
TBLT	Posttest	Oblique case	-44.0	strong decrease	Oblique case	33.8	increase
		Base form	-59.0	strong decrease	Base form	20.0	increase
		Paradigm	-30.0	decrease	Paradigm	-40.0	strong decrease
	Delayed posttest	Oblique case	-53.0	strong decrease	Oblique case	-17.0	decrease
		Base form	-7.4	stable/decrease	Base form	-20.0	decrease
		Paradigm	153.0	strong increase	Paradigm	-100.0	strong decrease
TSLT	Posttest	Oblique case	18.6	increase	Oblique case	60.0	strong increase
		Base form	-89.0	strong decrease	Base form	7.1	stable
		Paradigm	173.0	strong increase	Paradigm	-60.0	strong decrease
	Delayed posttest	Oblique case	18.9	increase	Oblique case	-42.0	strong decrease
		Base form	52.8	strong increase	Base form	0.4	stable
		Paradigm	10.4	increase	Paradigm	189.0	strong increase

Note. Where tests differed in the number of target items and/or participants, error counts were normalized by dividing the number of errors for each error type by the total possible number of responses and expressing the result as a percentage.

4.2 Results of the elicited oral imitation test

This subsection covers the results of the EOIT. A factor that needs to be considered before presenting the results is some changes to the test design, as outlined in Section 3.5.3. First of all, due to an oversight, the pretest included nine grammatically correct and, consequently, seven grammatically incorrect sentences, instead of the originally balanced eight/eight design. Furthermore, one grammatically incorrect posttest item was discarded because most participants interpreted the target noun (*prezidentë* ‘president (female)’) as masculine rather than the intended feminine form, making their production difficult to evaluate.

With that in mind, Table 36 presents the results for the EOIT, divided not only by treatment group but also by statement category (i.e., grammatical correctness). To ensure full transparency, the table includes not only the mean accuracy per participant and the standard deviation but also the percentage of correct responses for each test and statement category after omitting NA responses, the total number of NAs for each test and statement category, and the percentage of NAs within their respective test and statement category. As a reminder, NAs mainly consisted of failures to provide a response or responses lacking the target structure.

The following analyses are based on the accuracy percentages excluding NAs (% correct (-NA)) from the descriptive statistics. For the grammatically correct statements (+), the TBLT group demonstrated modest gains that were maintained at the delayed posttest, whereas the TSLT group displayed stable performance across all tests. The control group showed a slight decline at posttest. As for the grammatically incorrect statements (-), both treatment groups improved at posttest, although the gains were larger and more sustained for the TBLT group. The TSLT group showed some decline at the delayed posttest. Surprisingly, the control group showed substantial improvement in performance despite receiving no treatment. These results seem to indicate that both treatment groups demonstrated learning gains in implicit knowledge of case. Confoundingly, however, the control group also showed some improvement. Clearly, statistical modeling is required to determine whether the changes in the treatment groups exceed those observed in the control group and thus reflect treatment effects.

Table 36. Descriptive statistics of the elicited oral imitation tests

Group	Test	Mean (max = 8)	SD	% correct (-NA)	NA	% NA
TBLT	Pretest (max = 9)	7.78	1.25	91.3	11	5.3
	+ Posttest	6.91	1.28	94.6	16	8.7
	Delayed posttest	7.30	1.12	94.9	7	3.8
	Pretest (max = 7)	0.26	0.61	4.0	12	7.5
	- Posttest	1.48	1.81	21.7	4	2.5
	Delayed posttest	1.87	2.49	24.2	6	3.3
TSLT	Pretest (max = 9)	7.45	1.41	90.6	17	8.6
	+ Posttest	6.27	1.71	89.0	21	11.9
	Delayed posttest	6.63	1.35	90.0	12	7.9
	Pretest (max = 7)	0.18	0.49	3.0	22	14.3
	- Posttest	1.00	1.00	14.9	6	3.9
	Delayed posttest	0.84	0.99	11.1	8	5.3
Control group	Pretest (max = 9)	7.60	1.21	88.4	6	4.4
	+ Posttest	5.53	1.78	84.7	22	18.3
	Pretest (max = 7)	0.40	0.63	7.1	14	13.3
	- Posttest	1.33	1.19	19.4	2	1.9

Note. + : grammatically correct input items; - = grammatically incorrect input items; % correct (-NA) = the percentage of correct responses for each test and statement category after omitting NA-cases; % NA = percentage of responses at test time scored NA; due to an oversight, max scores at the pretest were different than at the other test times.

Table 37 contains the Bayesian model fitted to the pre- and posttest data across the two treatment groups and the control group.

Table 37. Bayesian posterior distribution summary for elicited oral imitation test accuracy: Including the control group

Parameter	Median	HDI (89%) Low	HDI (89%) High	<i>pd</i>	% in ROPE (±0.18)
Intercept	-0.24	-0.88	0.40	73.34	30.78
<i>Test Time: Posttest</i>	<i>0.74</i>	<i>-0.05</i>	<i>1.54</i>	<i>93.26</i>	<i>10.18</i>
Treatment: TSLT	-0.04	-0.64	0.57	53.85	38.49
Treatment: TBLT	0.07	-0.52	0.67	57.15	38.57
Correctness:	-5.16	-5.87	-4.50	100.00	0.00
Incorrect (-)					
Syntactic Role: Object	-0.12	-0.76	0.54	61.88	35.48
L1: Balto-Slavic	0.27	-0.14	0.68	85.19	33.36
Test Time: Posttest x Treatment: TSLT	-0.02	-0.72	0.65	52.44	34.57
<i>Test Time: Test Time x Treatment: TBLT</i>	<i>0.48</i>	<i>-0.19</i>	<i>1.15</i>	<i>87.34</i>	<i>18.52</i>

Note. The reference levels for each variable were as follows: Test Time: Pretest; Treatment: TSLT; Correctness: Correct (+); L1: non-Balto-Slavic; Syntactic Role: Subject.

HDI = Highest Density Interval; *pd* = Probability of Direction; ROPE = Region of Practical Equivalence.

Strongly supported parameters are bolded. Moderately supported parameters are italicized.

While inconclusive, there was moderate evidence for an improvement from pretest to posttest in the control group (Median = 0.74, 89% HDI [-0.05, 1.54], *pd* = 93.26%, ROPE = 10.18%). Even more concerning are the interaction effects: although there is weak to moderate evidence that the TBLT group improved more than the control group (Median = 0.48, 89% HDI [-0.19, 1.15], *pd* = 87.34%, ROPE = 18.52%), the estimated improvement in the TSLT group is virtually identical to that of the control group (Median = -0.02, 89% HDI [-0.72, 0.65], *pd* = 52.44%, ROPE = 34.57%). A three-way interaction for Test Time, Treatment Group, and Correctness was explored, but this yielded no strong results (see Appendix F). Naturally, this lack of significant performance differences will be addressed in the discussion section.

Regarding the remaining variables included in the model, there appeared to be no substantial differences between groups at the outset. Pairwise comparisons confirmed this finding (Control – TSLT: Median = 0.04, 89% HPD [-0.57, 0.65]; Control – TBLT: Median = -0.07, 89% HPD [-

0.67, 0.53]). Unsurprisingly, participants had most trouble accurately responding to grammatically incorrect sentences (Median = -5.16, 89% HDI [-5.87, -4.50], $pd = 100.00\%$, ROPE = 0.00%). Finally, there were no clear effects for the target syntactic role of the items (Median = -0.12, 89% HDI [-0.76, 0.54], $pd = 61.88\%$, ROPE = 35.48%) and there was some relatively weak evidence for an L1 effect, but the ROPE is quite high (Median = 0.27, 89% HDI [-0.14, 0.68], $pd = 85.19\%$, ROPE = 33.36%).

As the model provided only suggestive evidence of a TBLT treatment effect relative to the control group, several follow-up analyses were conducted to examine whether further comparisons were warranted. First, pairwise comparisons using the *emmeans()* and *contrast()* functions were computed to assess pretest–posttest gains for each group. Although both the control group and the TSLT group display only suggestive evidence of learning gains at the posttest (Control: Median = 0.74, 89% HPD [-0.08, 1.51]; TSLT: Median = -0.71, 89% HPD [-0.02, 1.50]), the TBLT group shows strongly supported improvement (Median = 1.22, 89% HPD [0.45, 1.96]). Furthermore, pairwise comparisons showed no group differences at the pretest. At the posttest, however, moderate to strong evidence was found for the TBLT group outperforming the control group (Median = 0.55, 89% HPD [-0.01, 1.10]) and even stronger performance for the TBLT group outperforming the TSLT group (Median = 0.61, 89% HPD [0.12, 1.12]).

Overall, these results suggest that only the TBLT group’s improvement was clearly supported and substantial. The steeper slope for the TBLT group observed in the graph with Bayesian logistic regression estimates for the elicited oral imitation test displayed in Figure 12 provides even further evidence of this. Taken together, this provided sufficient motivation to continue comparing the two treatment groups across all test times, including the delayed posttest.

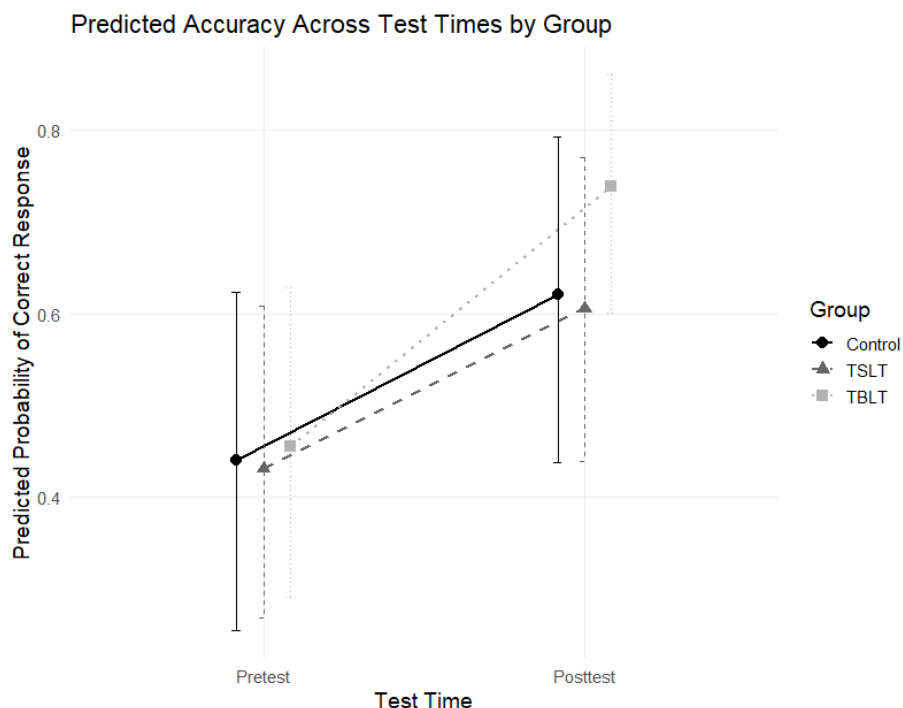


Figure 12. Predicted probabilities of correct responses by Test Time and Treatment Group for the EOIT model, including the control group. Error bars represent the 89% HDIs

This leads to the model presented in Table 38. This model yields several interesting findings. Firstly, the average increase in performance across both treatment groups from pre- to posttest (Median = 0.74, 89% HDI [0.10, 1.38], $pd = 96.81\%$, ROPE = 5.67%) as well as from pre- to delayed posttest (Median = 0.99, 89% HDI [0.35, 1.62], $pd = 99.11\%$, ROPE = 0.00%) receives strong evidence. However, pairwise comparisons reveal that only learning gains by the TBLT group are strongly supported (posttest: Median = 1.01, 89% HPD [0.28, 1.69]; delayed posttest: Median = 1.38, 89% HPD [0.70, 2.09]) and not those by the TSLT group (posttest: Median = 0.48, 89% HPD [-0.26, 1.19]; delayed posttest: Median = 0.58, 89% HPD [-0.16, 1.31]). Interestingly, a difference in increase in performance from pre- to posttest between the two treatment groups is only weakly to moderately supported by the evidence (Median = 0.52, 89% HDI [-0.13, 1.20], $pd = 90.01\%$, ROPE = 17.10%), while superior learning gains achieved by the TBLT group at the delayed posttest (Median = 0.79, 89% HDI [0.13, 1.46], $pd = 97.31\%$, ROPE = 4.59%) are strongly supported. Figure 13, which shows the interaction between Test Time and Treatment Group based on Bayesian logistic

regression estimates from the model, presents this dynamic. Further pairwise comparisons of both treatment groups at each Test Time reveal fairly strong evidence for performance differences at the posttest (Median = 0.64, 89% HPD [0.02, 1.30]) and solid evidence at the delayed posttest (Median = 0.92, 89% HPD [0.27, 1.55]), while, at pretest (Median = 0.11, 89% HPD [-0.54, 0.79]), there was no evidence for group differences. While there is the caveat of unclear pre- to posttest differences between the control and TBLT groups, these results do seem to suggest superior improvement in implicit knowledge of case for the TBLT group compared with the TSLT group.

Table 38. Bayesian posterior distribution summary for elicited oral imitation test accuracy: Two treatment groups

Parameter	Median	HDI (89%) Low	HDI (89%) High	<i>pd</i>	% in ROPE (±0.18)
Intercept	-0.25	-0.77	0.27	78.64	33.84
Test Time: Posttest	0.74	0.10	1.38	96.81	5.67
Test Time:	0.99	0.35	1.62	99.11	0.00
Delayed posttest					
Treatment: TBLT	0.11	-0.54	0.79	61.14	34.05
Correctness:					
Incorrect (-)	-5.52	-6.12	-4.99	100.00	0.00
Syntactic Role: Object	-0.67	-1.18	-0.17	98.16	3.69
<i>L1: Balto-Slavic</i>	<i>0.63</i>	<i>-0.08</i>	<i>1.35</i>	<i>92.04</i>	<i>12.98</i>
<i>Test Time: Posttest x</i>	<i>0.52</i>	<i>-0.13</i>	<i>1.20</i>	<i>90.01</i>	<i>17.10</i>
<i>Treatment: TSLT</i>					
Test Time: Delayed posttest					
x	0.79	0.13	1.46	97.31	4.59
Treatment: TBLT					
<i>Test Time: Posttest x</i>	<i>-0.69</i>	<i>-1.36</i>	<i>0.00</i>	<i>94.61</i>	<i>9.93</i>
<i>L1: Balto-Slavic</i>					
<i>Test Time: Delayed posttest x</i>	<i>-0.36</i>	<i>-1.05</i>	<i>0.34</i>	<i>79.21</i>	<i>24.43</i>
<i>L1: Balto-Slavic</i>					

Note. The reference levels for each variable were as follows: Test Time: Pretest; Treatment: TSLT; Correctness: Correct (+); L1: non-Balto-Slavic; Syntactic Role: Subject.

HDI = Highest Density Interval; *pd* = Probability of Direction; ROPE = Region of Practical Equivalence.

Strongly supported parameters are bolded. Moderately supported parameters are italicized.

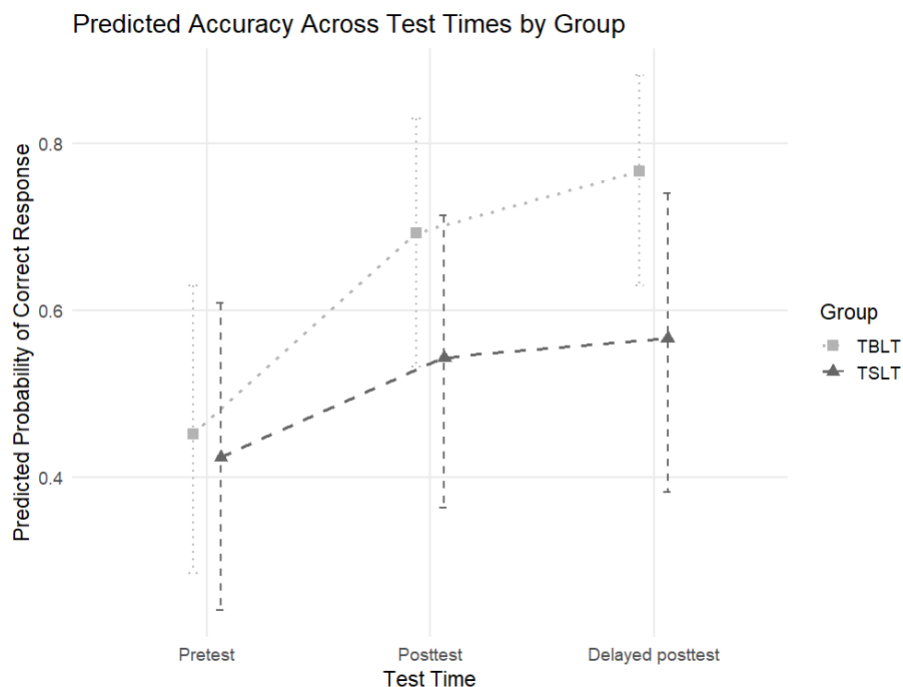


Figure 13. Predicted probabilities of correct responses by Test Time and Treatment Group for the EOIT model without the control group. Error bars represent the 89% HDIs

As for the other effects, sentence correctness, once again, heavily influenced the results (Median = -5.52, 89% HDI [-6.12, -4.99], $pd = 100.00\%$, ROPE = 0.00%); surprisingly, Syntactic Role exhibited a fairly strong negative effect for objects at baseline (Median = -0.67, 89% HDI [-1.18, -0.17], $pd = 98.16\%$, ROPE = 3.69%). Interactions involving these parameters were explored. Sentence correctness by test time revealed positive effects, but they were only weakly supported. Syntactic role by test time revealed a positive effect at the posttest and a negative effect at the delayed posttest, both of which, again, had only weak evidence. This implies no, or only minimal, progress in performance on grammatically incorrect and object-marking items (see Appendix F).

Finally, the interaction between Test Time and L1 was explored. Interestingly, a Balto-Slavic L1 advantage was moderately supported at pretest (Median = 0.63, 89% HDI [-0.08, 1.35], $pd = 92.04\%$, ROPE = 12.98%). Interactions, however, point towards a worse increase in performance for Balto-Slavic L1 speakers at the posttest (Median = -0.69, 89% HDI [-1.36, 0.00], $pd = 94.61\%$, ROPE = 9.93%), while evidence for

such a disadvantage was no longer present at the delayed posttest (Median = 0.36, 89% HDI [-1.05, 0.34], $pd = 79.21\%$, ROPE = 24.43%).

Pairwise comparisons reveal why this was the case. At the pretest, there is only suggestive evidence for a difference in performance between Balto-Slavic L1 speakers and L1 speakers of different languages (Median = 0.63, 89% HPD [-0.09, 1.33]). Nevertheless, pretest–posttest (Median = 1.08, 89% HPD [0.40, 1.75]) and pretest–delayed posttest (Median = 1.16, 89% HPD [0.48, 1.84]) gains were far more substantial for the non-Balto-Slavic L1 speakers than for the Balto-Slavic L1 speakers (posttest: Median = 0.40, 89% HPD [-0.37, 1.16]; delayed posttest: Median = 0.81, 89% HPD [0.03, 1.57]). This dynamic is illustrated in Figure 14. It shows that, at the posttest, the non-Balto-Slavic L1 group caught up with the Balto-Slavic L1 group. Their performance then stabilized, while the Balto-Slavic L1 group nevertheless improved slightly. This suggests that the potential advantages of an L1 for acquiring implicit grammar knowledge in typologically similar L2s might be more subtle than often realized.

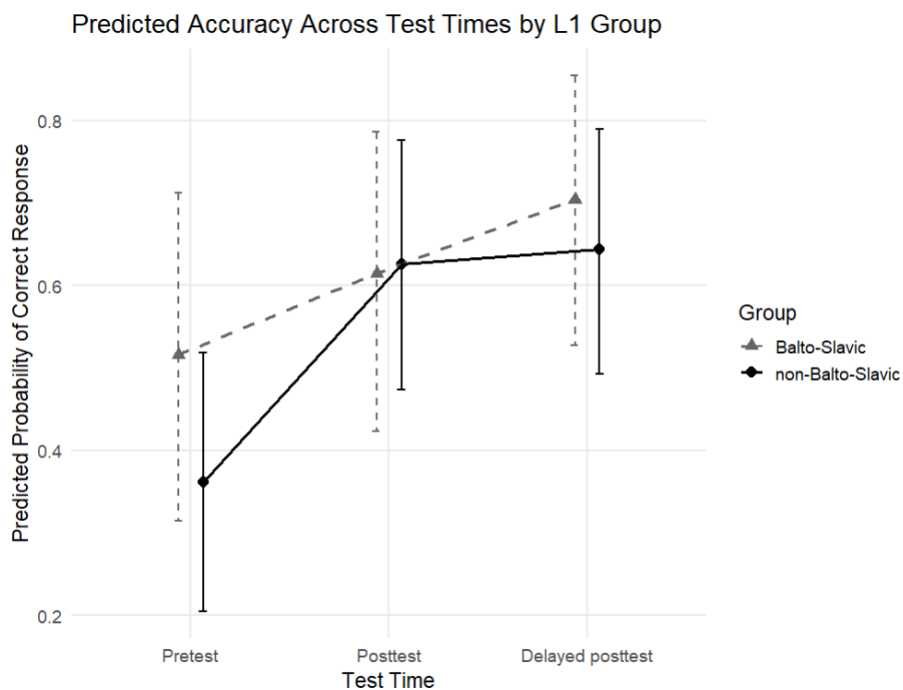


Figure 14. Predicted probabilities of correct responses by Test Time and L1 for the EOIT model without the control group. Error bars represent the 89% HDIs

Since the last model demonstrated significant syntactic role effects, it was decided to fit another model that included Case and Paradigm, as shown in Table 39 below. At the pretest, moderate to strong evidence suggests that participants of both treatment groups had most difficulty with items targeting the genitive (Median = -1.12, 89% HDI [-2.21, 0.01], $pd = 94.29\%$, ROPE = 6.12%) and the accusative (Median = -1.19, 89% HDI [-2.50, 0.12], $pd = 92.54\%$, ROPE = 7.03%). The paradigms the target nouns belonged to seemed not to matter at the outset. Of the two-way interactions, only Test Time and Case reveal a strong improvement in genitive marking from pre- to delayed posttest for both groups (Median = 1.81, 89% HDI [0.31, 3.33], $pd = 97.06\%$, ROPE = 1.90%). Pairwise comparisons provide strong support for improved genitive marking from pretest to both posttests (immediate posttest: Median = 1.07, 89% HPD [0.14, 1.97]; delayed posttest: Median = 1.72, 89% HPD [0.86, 2.67]). No evidence of improvement for the nominative and the accusative could be established.

Out of the three-way interactions, only the interaction between posttest, TBLT group, and accusative case was strongly supported (Median = 2.45, 89% HDI [0.71, 4.19], $pd = 98.84\%$, ROPE = 0%), suggesting that the TBLT group showed superior accusative case marking gains compared to the TSLT group at posttest, although this advantage was not sustained at the delayed posttest. The three-way interaction between delayed posttest, TBLT group, and the ‘-is/-ys’ paradigm displays moderate evidence (Median = 1.91, 89% HDI [-0.01, 3.83], $pd = 94.39\%$, ROPE = 3.46%), suggesting superior pre- to delayed posttest gains for the TBLT group in the implicit declension of ‘-is/-ys’-type nouns. However, because solid evidence for meaningful interactions is scarce, and the interactions that are strongly supported do not carry over across tests, these findings do not appear to be highly informative.

It was decided to check for negative outliers here as well. Similar to the gap-fill test, the model presented in Table 38 was refitted by nesting the random intercept of Item within Treatment Group, and a caterpillar plot was created to visualize the random intercept Item (see Appendix G). It includes model-estimated probabilities for each item, faceted by Treatment Group, Test Time, and Correctness. Again, items are ordered from most challenging to least challenging, and negative outliers were identified as those with a posterior mean estimate below 0.375 on the probability scale. Regarding grammatically correct sentences at the pretest, the TBLT group struggled with item 11 (target noun: *futbol* ‘soccer’; target case: accusative; target syntactic role: object), while the TSLT group faced problems with item 1

Table 39. Bayesian posterior distribution summary for elicited oral imitation test accuracy: Two treatment groups, including the parameters Case and Paradigm

Parameter	Median	HDI (89%) Low	HDI (89%) High	<i>pd</i>	% in ROPE (± 0.18)
Intercept	-0.01	-0.59	0.58	51.31	39.91
Test Time: Posttest	0.76	0.04	1.50	95.46	7.47
Test Time: Delayed posttest	0.82	0.10	1.55	96.55	5.49
Treatment: TBLT	0.24	-0.52	1.00	69.01	27.20
<i>Case: Genitive</i>	<i>-1.12</i>	<i>-2.21</i>	<i>0.01</i>	<i>94.29</i>	<i>6.12</i>
<i>Case: Accusative</i>	<i>-1.19</i>	<i>-2.50</i>	<i>0.12</i>	<i>92.54</i>	<i>7.03</i>
Paradigm: -is/-ys	-0.18	-1.43	1.09	59.41	18.55
Paradigm: -è	0.26	-1.02	1.49	63.38	19.01
Paradigm: -a	0.90	-0.53	2.35	84.41	9.97
Correctness: Incorrect (-)	-5.73	-6.35	-5.16	100.00	0.00
Syntactic marking: Object	0.09	-0.69	0.83	57.49	30.74
L1: Balto-Slavic	0.27	-0.30	0.85	77.86	31.06
Test Time: Posttest x Treatment: TBLT	0.55	-0.26	1.34	85.91	16.96
<i>Test Time: Delayed posttest x Treatment: TBLT</i>	<i>0.66</i>	<i>-0.15</i>	<i>1.45</i>	<i>90.66</i>	<i>12.62</i>
Test Time: Posttest x Case: Genitive	0.61	-0.95	2.12	73.86	12.80
Test Time: Delayed posttest x Case: Genitive	1.81	0.31	3.33	97.06	1.90
Test Time: Posttest x Case: Accusative	-0.31	-1.87	1.27	62.38	14.52
Test Time: Delayed posttest x Case: Accusative	-0.76	-2.32	0.87	77.61	10.83
Case: Genitive x Treatment: TBLT	-0.47	-1.73	0.78	72.06	16.01
Case: Accusative x Treatment: TBLT	-0.42	-1.63	0.78	71.46	17.03
Test Time: Posttest x Paradigm: -is/-ys	-0.60	-2.31	1.08	71.82	12.60

Parameter	Median	HDI (89%) Low	HDI (89%) High	<i>pd</i>	% in ROPE (±0.18)
Test Time: Delayed posttest x Paradigm: -is/-ys	-0.22	-1.96	1.54	58.11	13.78
Test Time: Posttest x Paradigm: -è	-0.74	-2.48	1.08	74.52	10.78
Test Time: Delayed posttest x Paradigm: -è	-0.65	-2.29	1.04	73.96	11.78
Test Time: Posttest x Paradigm: -a	0.76	-1.22	2.74	73.24	10.30
Test Time: Delayed posttest x Paradigm: -a	0.35	-1.62	2.34	60.91	11.33
Paradigm: -is/-ys x Treatment: TBLT	0.34	-1.06	1.68	65.16	16.24
Paradigm: -è x Treatment: TBLT	-1.08	-2.48	0.32	88.70	8.57
Paradigm: -a x Treatment: TBLT	0.48	-1.07	2.00	69.05	13.49
Test Time: Posttest x Treatment: TBLT x Case: Genitive	-0.83	-2.48	0.77	79.88	10.84
Test Time: Delayed posttest x Treatment: TBLT x Case: Genitive	0.53	-1.17	2.23	69.04	12.71
Test Time: Posttest x Treatment: TBLT x Case: Accusative	2.45	0.71	4.19	98.84	0.00
Test Time: Delayed posttest x Treatment: TBLT x Case: Accusative	1.38	-0.23	3.02	91.63	5.62
Test Time: Posttest x Treatment: TBLT x Paradigm: -is/-ys	0.71	-1.10	2.49	73.22	11.47
<i>Test Time: Delayed posttest x Treatment: TBLT x Paradigm: -is/-ys</i>	<i>1.91</i>	<i>-0.01</i>	<i>3.83</i>	<i>94.39</i>	<i>3.46</i>
Test Time: Posttest x Treatment: TBLT x Paradigm: -è	-0.81	-2.80	1.17	74.18	9.83

Parameter	Median	HDI (89%) Low	HDI (89%) High	<i>pd</i>	% in ROPE (±0.18)
Test Time: Delayed posttest x					
Treatment: TBLT x	-0.83	-2.60	0.95	77.33	10.69
Paradigm: -ė					
Test Time: Posttest x					
Treatment: TBLT x	0.43	-1.54	2.41	63.71	11.34
Paradigm: -a					
Test Time: Delayed posttest x					
Treatment: TBLT x	-1.36	-3.43	0.71	85.21	6.74
Paradigm: -a					

Note. The reference levels for each variable were as follows: Test Time: Pretest; Treatment: TSLT; Case: Nominative; Paradigm: -as, Correctness: Correct (+); L1: non-Balto-Slavic; Syntactic Role: Subject.

HDI = Highest Density Interval; *pd* = Probability of Direction; ROPE = Region of Practical Equivalence.

Strongly supported parameters are bolded. Moderately supported parameters are italicized.

(*bankas* ‘bank’, genitive, subject). Problems seemed to mainly have arisen from the verb: item 11 included the verb *žaisti* ‘play’, which was conjugated in the first person *aš žaidžiu* ‘I play’ and was difficult for some participants to understand, and while item 1 included simply *nėra* ‘there is no’, the most frequent mistake while repeating the sentence was to substitute the verb with the form *yra* ‘there is’, making the original genitive marking on the noun *bankas* ‘bank’ incorrect. As for grammatically incorrect sentences, both groups experienced difficulties repeating item 4 (*sūris* ‘cheese’, genitive, subject), the TBLT group found items 8 (*upė* ‘river’, genitive, subject) and 13 (*alkoholis* ‘alcohol’, genitive, object) the most challenging, while this was the case with item 2 (*langas* ‘window’, nominative, subject) for the TSLT group. Clearly, the most problematic items targeted the genitive case, while other tendencies that cause these challenges, for instance, in terms of the paradigm or the accompanying verb, do not seem to exist.

Moving on to grammatically correct sentences at the posttest, both groups faced most difficulties correctly repeating item 19 (*bananas* ‘banana’, genitive, subject), while item 23 (*aikštė* ‘square’, nominative, subject) proved difficult for the TBLT group and item 27 (*filmas* ‘movie’, accusative, object) created difficulties for the TSLT group. Items 19 and 23 seemed to have caused problems due to the target nouns themselves. The genitive *banano* was often pronounced *banana* as in English, which resembles the accusative case and is thus a mistake. The noun *aikštė* ‘square’ seemed to have been difficult

for many participants to recognize, despite its frequency and assumed relevance in daily-life contexts. The problem with item 27 might have been the accompanying verb *žiūrėti* ‘watch’, which was not always recognized and was difficult for many participants to pronounce. The presence of the adverb *dažnai* ‘often’ might have made accurate perception and repetition even more challenging. Attention to the verb might have distracted from accurate marking on the target noun. As for grammatically incorrect sentences, both groups struggled with item 24 (*katė* ‘cat’, genitive, subject), while item 28 (*kompiuteris* ‘computer’, accusative, object) proved exceptionally difficult for the TSLT group. Both nouns are assumed to be very frequent in L2 learner contexts, making it difficult to determine the specific problems with these items.

Lastly, out of the grammatically correct sentences at the delayed posttest, item 43 (*jogurtas* ‘yoghurt’, accusative, object) was difficult for both groups, while the TBLT group struggled with items 48 (*kelionė* ‘trip’, accusative, object) and 39 (*virtuvė* ‘kitchen’, nominative, subject). Items 43 and 48 likely proved difficult due to additional contextual information. For example, item 43 included *retai* ‘rarely’, which some participants did not recognize and which may have been distracting. Out of the grammatically incorrect sentences, the TSLT group had more difficulty correctly repeating item 45 (*numeris* ‘number’, genitive, object). Again, the context might have been the issue, as some participants seemed initially puzzled by the words *namas* ‘house’ and *numeris* ‘number’ in the same sentence.

Finally, failure to repeat a sentence could indicate lexical problems. Such instances were coded NA. Consequently, item outliers with a high number of NAs were identified as well. At the pretest, participants of both groups more often failed to repeat items 2 (*langas* ‘window’, incorrect sentence, nominative, subject; 10 NAs), 10 (*stalas* ‘table’, incorrect sentence, accusative, object; 9 NAs), 15 (*varškė* ‘curd’, incorrect sentence, genitive, object; 6 NAs), and 16 (*savaitė* ‘week’, correct sentence, accusative, object; 12 NAs). Problems with items 2, 15, and 16 were likely directly related to the target nouns themselves: *langas* ‘window’ was presented in the sentence as *lango* ‘window.GEN’ and, due to the genitive form, was often not recognized; *varškė* ‘curd’ is a popular food in Lithuanian culture but not necessarily in others, and is perhaps even disliked by certain individuals, leading to it not being as readily recognized; *savaitė* ‘week’ was often confused with *savaitgalis* ‘weekend’, which was accepted as an answer provided that the case form was correct, but this led to issues nonetheless. Item 10 was simply too difficult due to a longer sentence including too much context and a double

negation, which is grammatically required in Lithuanian. At the posttest, most problems arose with items 21 (*traukinys* ‘train’, correct sentence, genitive, subject; 10 NAs) and 29 (*krepšinis* ‘basketball’, correct sentence, genitive, object; 7 NAs). The main problem seemed to be the target nouns themselves, which were often not recognized. The fact that they were both presented in the genitive case (*traukinio* ‘train.GEN’; *krepšinio* ‘basketball.GEN’) likely did not help. At the delayed posttest, some participants had a particular problem solely with item 48 (*kelionė* ‘trip’, correct sentence, accusative, object; 13 NAs). As mentioned above, this item most likely proved too difficult because of the added context.

Rather than reflecting certain developmental tendencies depending on the treatment, the negative outliers mainly revealed potential design issues that need to be addressed in the future. While the possibility always exists that certain target nouns are less frequent for certain individuals, the fact that presenting them in other case forms may diminish their comprehensibility is something to consider in the future. Additionally, added context or difficult verbs may have distracted participants from accurately assigning case to target nouns. Naturally, a certain level of distraction is required for the test; hence, these features do not automatically constitute a design problem. Nevertheless, the level of distraction that is necessary for an EOIT should somehow be gauged.

5 DISCUSSION

This final section attempts to provide answers to the research questions formulated in the Introduction and address the hypotheses that were proposed in that section. Each is considered separately, together with the relevant hypotheses formulated in the Introduction. The discussion then continues with additional potentially confounding factors that may have influenced the results. Subsequently, this study's conclusions are presented, followed by their implications, the study's limitations, and ideas for further research.

5.1 Treatment effects on the acquisition of explicit and implicit case knowledge

The following two research questions were proposed:

- RQ1: Does limiting grammar instruction to the post-task stage promote both implicit *and* explicit learning gains related to grammatical case?
- RQ2: Does placing greater emphasis on grammar throughout the task cycle enhance both explicit *and* implicit case-related learning gains?

Clearly, the TBLT treatment group corresponds to the type of grammar instruction described in RQ1, while the TSLT treatment group adheres to the type of grammar instruction described in RQ2. Both research questions will be discussed in the same subsection, but divided by their corresponding hypotheses. The first question to be addressed pertains to the acquisition of *explicit* knowledge:

Hypothesis 1a: Participants of both treatment groups will achieve some learning gains in explicit knowledge of case.

Hypothesis 1b: Participants in the TSLT group will outperform those in the TBLT group in terms of explicit knowledge acquisition.

Since both hypotheses relate to explicit knowledge acquisition, the following discussion covers the gap-fill test results. Judging from the results, Hypothesis 1a seems to have been confirmed. Both the TBLT and TSLT groups achieved significant learning gains on the posttest and delayed posttest. However, surprisingly, the results do not support Hypothesis 1b. There appear to be no superior learning gains for the TSLT group at either posttest or delayed posttest. Moreover, the Bayesian model provided strong support for superior

pre- to posttest improvement in the TBLT group and even a potential advantage at the delayed posttest.

Several potential explanations can be formulated. The first relates to a combination of test characteristics and treatment effects. While the gap-fill test was designed to measure explicit case knowledge, such research instruments rarely reflect rigid dichotomies. Thus, when participants were unsure about which grammatical rules to apply to a given item, they were likely to rely on their intuition, leading to an answer closer to the implicit end of the spectrum. Additionally, the focus that the TSLT treatment placed on not just case assignment and marking, but especially paradigms, might have weakened the noticing bridge portrayed in Figure 2. Consequently, explicit instruction for the TSLT group largely remained explicit knowledge, while the more limited grammar instruction given to the TBLT group led to more noticing (i.e., sensitivity to exemplars) and, therefore, stronger implicit knowledge of case. In other words, the TBLT group made more successful implicit guesses. The TBLT group's apparent tendency to skip spelling and/or paradigm errors and immediately provide fully accurate endings, as suggested by the categorical model, seems to corroborate this idea. Importantly, this interpretation could serve as strong support for the cognitive-interactionist model, which is essentially the principal theoretical foundation of TBLT (see Section 1.2.2).

Another explanation relates to group differences. While there seemed to have been no substantial differences in pretest performance between the control and the TBLT group on the one hand, and the control and the TSLT group on the other, the model presented in Table 24 *did* show moderate support for pretest differences between the two treatment groups. Given that the participants in the TBLT group, on average, had spent more time learning Lithuanian and living in Lithuania, this is not particularly surprising. If the TBLT group was indeed at a higher average developmental L2 proficiency level than the TSLT group, this could, in theory, have had important effects. Based on VanPatten's (2014) IP, it could be argued that slightly more proficient participants could free up more cognitive resources for constructions with less semantic meaning, making it easier for the TBLT group participants to acquire novel constructions. Based on Pienemann's (1998) Processability Theory, a similar claim could be made: the TBLT participants were developmentally more ready to acquire novel constructions than the TSLT participants were. A more usage-based perspective, motivated by the CREED model, could suggest that the TSLT participants were less sensitive to exemplars.

Switching the focus specifically to case, evidence for differences in developmental readiness seems mixed at best. Error analyses (see Table 31, Table 32, Table 33) showed that both groups made the most progress in marking the nominative and the genitive, while performance in marking the accusative remained largely stable. Likewise, the model depicted in Table 25 showed strong support for gains in marking the genitive from pre- to posttest and moderate evidence from pre- to delayed posttest, while negative interactions involving the accusative were strongly supported. Nevertheless, pairwise comparisons revealed relatively strong support for learning gains in accusative marking from pre- to delayed posttest, albeit for both groups. Moreover, although the model indicated superior performance of the TBLT group in marking both the genitive and the accusative, the absence of evidence for three-way interactions among Test Time, Treatment, and Case suggests no difference between the two treatment groups in learning gains for either case. This suggests that the treatment groups were developmentally similar in terms of their receptivity to case.

In the end, it should be emphasized that the pairwise comparisons provided strong support for learning gains in both groups, and the potential pretest difference between the treatment groups affects only statements regarding the potential superiority of the TBLT treatment for acquiring explicit knowledge. While the data *do* indicate this, moderate evidence for pretest group differences does not allow for firm assertions on that account.

The second pair of hypotheses pertains to the acquisition of *implicit* case knowledge and, thus, to the EOIT.

Hypothesis 2a: Participants of both treatment groups will achieve some learning gains in implicit knowledge of case.

Hypothesis 2b: Participants in the TBLT group will outperform those in the TSLT group in terms of implicit knowledge acquisition.

As can be gathered from the test results, insights into this question are much less straightforward. The initial Bayesian model, including the control group, as illustrated in Table 37, showed moderate evidence of pre- to posttest improvement for the control group and a weakly supported difference for the TBLT group. The TSLT group appeared to perform almost identically to the control group. Pairwise comparisons revealed suggestive evidence of pre- to posttest improvements in the control and TSLT groups, and strongly supported evidence of such an improvement in the TBLT group. This was further supported by pairwise comparisons showing strongly supported

performance differences at the posttest between the TBLT group and both the TSLT and control groups, which were not present at the pretest. While this allowed for further analyses of the data, including the delayed posttest, the potential improvement in the control group must be considered further.

The subsequent model, including the delayed posttest, as depicted in Table 38, revealed strong support for learning gains from pre- to posttest and pre- to delayed posttest for both groups. While the interaction between Test Time and Treatment Group strongly supported differences in learning gains only at the delayed posttest in favor of the TBLT group, pairwise comparisons showed strong support for pre- to posttest improvement only in the TBLT group. Corresponding evidence for the TSLT group was much weaker. This suggests that while there is some evidence of learning gains for the TBLT group, the TSLT group did not achieve substantial gains, thus failing to support Hypothesis 2a. Based on this, it can already be established that Hypothesis 2b was, indeed, correct, with the TBLT group clearly outperforming the TSLT group on this test. The strong support for performance differences between the treatment groups at the post- and delayed posttests, which were not found at the pretest, only further confirms this finding.

Again, an attempt will be made to provide some possible explanations for these findings. Firstly, the superior learning gains of the TBLT group were expected under this hypothesis. As per the cognitive-interactionist framework (R. Ellis et al., 2019), greater implicit interaction with fellow participants and input led to stronger gains in implicit case knowledge. Nevertheless, the TSLT group's lack of significant progress is striking. Based on the same framework, it could again be argued that the greater amount of explicit instruction strengthened only the TSLT group's explicit knowledge but may somehow have hampered their noticing skills. Alternatively, preemptive focus on form may have shifted task completion away from meaning-focused interaction toward exercises targeting correct grammatical constructions, as TBLT researchers such as Long (2014) and Willis and Willis (2007) have warned against. Several design choices, as discussed in Section 3.4.3, which brought the TSLT task versions closer to PPP in certain respects, may have further contributed to this effect. This must surely have hampered the acquisition of implicit case knowledge. Besides providing additional evidence for the cognitive-interactionist model, these results also seem to support the weak interface position on the interaction between explicit and implicit knowledge, which is frequently adopted in TBLT. At any rate, there was no evidence of

any group differences at the pretest, indicating that any potential developmental advantages for the TBLT group were likely not present.

The final finding worth discussing is that the TBLT group showed only suggestive evidence of superior pre- to posttest learning gains compared to the control and TSLT groups; however, it solidified its gains and improved at the delayed posttest. This suggests more lasting TBLT treatment effects. Less explicit grammar instruction could have led participants to attend to the input and their output differently, while focusing more on meaning and, therefore, more effectively noticing which L2 constructions are necessary in both. In N. Ellis's (2006) CREED terms, this might have aided the detectors involved with the target structures, which proved more resilient after a period without treatment or, possibly, benefited more from exposure to L2 Lithuanian in the participants' everyday surroundings.

Nevertheless, certain issues with the results cannot be ignored. While pairwise comparisons showed strong support for learning gains from the pretest to both test times in the TBLT group, strong support for treatment and test interactions was found only at the delayed posttest. Combined with the moderate support for pre- to posttest improvement in the control group, this casts serious doubt on the initial learning gains achieved by both treatment groups. Improved performance might not have been due to treatment but, instead, due to other potential confounding variables. Evidently, these issues raise some methodological questions, which will be discussed in Section 5.4.4.

5.2 Lasting explicit and implicit learning gains

The next research question and hypothesis concern lasting learning gains, which were assessed by using the delayed posttest.

RQ3: Can TBLT and TSLT treatments lead to lasting learning gains in grammatical case?

Hypothesis 3: Both treatments lead to significant lasting learning gains for grammatical case.

As discussed above, some lasting learning gains were, indeed, achieved. The gap-fill test showed significant improvements in explicit knowledge of case for both groups. The EOIT, however, yielded significant, lasting learning gains only for the TBLT group, but not the TSLT group, thus providing only partial support for the hypothesis. Treatment differences have already been discussed and will not be repeated here. More importantly,

though, is the simple fact that lasting learning gains can be observed in the first place. In Markina's (2018) study, the participants demonstrated learning gains for case in a gap-fill test at the posttest, which then returned to baseline at the delayed posttest. Denissenko's study (2016) even found no significant learning gains for case at any level, even after an extensive treatment lasting 150 hours.

Before continuing, it is important to highlight some differences between this and the other studies mentioned. The first, most obvious difference is the different target L2s: L2 Lithuanian in this study and L2 Russian in Markina's (2018) and Denissenko's (2016). The areas of case marking that were researched also differ. Markina's study examined prepositional case government, while this study centers on subject and direct-object marking. These may seem like two very different linguistic operations, but both are similarly grammatical and carry little semantic content. In terms of VanPatten's (2014) IP, this lack of clear meaning most probably posed an obstacle for L2 students in attending to case marking. In prepositional phrases, prepositions carry most of the semantic load, while case marking serves mainly a grammatical function. A similar dynamic emerges in the case of negation, where the verb *nėra* 'is not' and the prefix *ne-* 'no' convey the negation, and the genitive case marking only strengthens this meaning. This might also apply to non-negated direct-object marking, since the verb itself and the frequent postverbal position of direct objects denote the appropriate grammatical role, with accusative case marking again confirming the object status of the postverbal noun. Moreover, while prototypical subject and direct-object marking in Lithuanian involves only three cases, prepositions in Russian govern up to five cases (genitive, dative, accusative, instrumental, and locative), which may make subject and direct-object marking conceptually simpler in Lithuanian.¹⁸ Nevertheless, the study design was similar in that it included a gap-fill test administered as a pre-, post-, and delayed posttest, but with a shorter 14-hour treatment.

Denissenko's (2016) study focused on all cases, without a specific functional focus. Moreover, this study's test involved writing a narrative based on a set of images and creating an ending for it. Given that students had 30 minutes to write the first story and an additional 10 minutes to invent an ending, these participants most surely drew on a combination of explicit and

¹⁸ As discussed previously in the literature review, besides the prototypical verb-governed accusative, verbs can also govern the genitive, dative, and instrumental cases. These verbs do not allow case alternation under negation. However, these types of objects were not part of the target structures in this study.

implicit case knowledge. Additionally, the study included only an immediate posttest, not a delayed posttest, which was administered after a lengthy 150-hour treatment.

The question then arises as to why the participants in this study achieved not only immediate learning gains but also lasting learning gains for case, at least in terms of explicit knowledge revealed by the gap-fill test, with possible long-term implicit gains by the TBLT group, as revealed in the EOIT. The reasoning for thinking that the treatment groups in this study had a greater chance of long-term case acquisition was the clearer focus in terms of the target constructions. Denissenko's (2016) study examined case use in its entirety, which entails a large array of case functions and paradigms. Moreover, treatment consisted of regular formal L2 Russian instruction and was not adjusted to the research goals. Markina's (2018) study, admittedly, applied a narrower scope by focusing on prepositional case government and providing tasks focused on this kind of case government. Nevertheless, prepositions cover a wide range of semantic meanings and involve five cases in Russian. In comparison, the target structures of this study seem conceptually more coherent (subject marking, direct-object marking, and negation therein) and cover only three possible cases (nominative, genitive, and accusative).

The SLA literature offers possible explanations for why this might have benefited long-term case acquisition gains. First, it is worth recalling Magnani and Artoni's (2015) hypothetical framework for the acquisition of L2 Russian case (see Table 8). Accurate production of the target structures requires a development from the second 'Category' stage to the third 'Phrasal' stage. 'Verb-object case agreement' is mentioned as one of three prerequisites for reaching this stage. However, the present study narrowed the target structures further by focusing solely on direct-object marking and excluding verbs that govern only the genitive, dative, or instrumental cases. This focus on a specific subcategory at this stage could have facilitated participants' development toward this stage.¹⁹ From the perspective of VanPatten's (2014) IP theory, however, the target structures do not appear to provide any advantages over those in the other studies. As discussed above, they carry little semantic meaning and primarily serve grammatical functions in the presence of more semantic cues. This kind of redundancy is also an important factor in N. Ellis's (2006) CREED framework; however, alternative explanations may

¹⁹ In the same subsection, however, a potential developmental order for negated direct-objects in L2 Lithuanian was proposed, allowing for an alternative interpretation. See Section 5.4.2.

apply here. While these case endings seem redundant and carry little salience, the narrow scope of the target constructions made them very frequent. Furthermore, the focus on negation also likely increased the contingency of the genitive: negation equals genitive marking. This heightened contingency might even have led to participants' superior performance in marking the genitive compared to the nominative and, especially, the accusative. Such high contingency likely facilitated implicit acquisition processes, potentially leading to more durable learning gains, as reflected in the present findings.

5.3 Different developmental patterns for case acquisition depending on the TBLT and TSLT treatments

Error analyses on the gap-fill test responses were performed in order to answer the fourth research question and test Hypothesis 4.

RQ4: Does a difference in the emphasis on grammar instruction in the treatment lead to any discernible L2 developmental differences concerning grammatical case?

Hypothesis 4: While the TBLT group will mainly reduce case assignment errors, the TSLT group will have fewer problems with paradigm-related errors.

To reiterate, the lowered focus on grammar would lead the TBLT group to focus more on meaning when interacting with the input and in dyads, which could potentially highlight case functions. This would be further amplified at the post-task stage, where relevant grammatical constructions were presented according to their appropriate functions in the context, with little regard for paradigms. Consequently, the TBLT group would assign the appropriate case but would have greater difficulty marking it, struggling to select the appropriate paradigm for the target noun. A similar trend was observed during the pilot study (Udes, 2025a), which prompted further investigation. Conversely, the TSLT group's treatment focused on paradigms and endings throughout the task, facilitating development in this regard. However, this might have distracted their attention during interaction and at the post-task stage from the functionality of grammar constructions, including the target constructions. Essentially, the TBLT group in each test was hypothesized to show a decrease in case assignment errors, along with stable or worse performance with respect to paradigm errors. The TSLT group was supposed

to show the opposite: an increase in case assignment errors and a decrease in paradigm errors.

Starting with the error analyses for Subject and Object marking, these trends were observed in fewer than half of the cases (see Section 4.1.2, Table 34). The trend predicted in Hypothesis 4 was clearly visible only in the TBLT group's posttest results for object marking and its delayed posttest results for both subject- and direct-object marking items. The TSLT group showed no instances of the predicted trend; instead, it displayed steady increases in paradigm errors and frequent decreases in case-assignment errors. From the perspective of case (see Table 35), only the TBLT group results at the delayed posttest for genitive marking and the TSLT group results at the posttest for accusative marking seem to depict something akin to what is described in Hypothesis 4. Error types of the nominative are ignored here since the base form was always given in brackets with the appropriate ending. All other instances in terms of performance with genitive or accusative marking for both groups did not confirm the hypothesis.

Taken together, only the TBLT group's performance, based specifically on the subject and direct-object marking dichotomy, adheres mostly to the developmental patterns described in Hypothesis 4. It could be argued that the 1.5-month break mainly impeded paradigm knowledge, rather than case assignment knowledge, as expected under the hypothesis. This is further supported by the claim that learning gains from implicit treatments are more durable, which would be an advantage to the TBLT group and the case-marking knowledge that the group was thought to be more receptive to (Long, 2016, p. 15). Posttest results by the TBLT group could have differed because the input to which the participants were exposed provided sufficient instances of the four target paradigms to foster not only case assignment but also case marking according to the appropriate paradigms, at least for subjects. Nevertheless, as could be judged from the absolute numbers presented in Table 29 to Table 33, paradigm errors were relatively rare; hence, changes in percentages are not particularly informative.

Error type differences by grammatical case reveal a number of distinct patterns. For genitive marking, the TBLT group's performance initially improves on all accounts, only to somewhat decline in terms of paradigm errors at the delayed posttest. By contrast, the TSLT group appeared to experiment more with declining nouns, as evidenced by the sharp drop in base-form errors and the corresponding increase in both case-assignment and paradigm errors, although all three types of errors increased again at the delayed posttest. For accusative marking, the TBLT group displayed the

complete opposite pattern to that predicted by the hypothesis, showing virtually no problems with paradigms and only improving case assignment at the delayed posttest. Meanwhile, the TSLT group performed as predicted by the hypothesis at the posttest but returned to baseline at the delayed posttest. These observations do not seem to point to any clear developmental patterns at the case level.

Importantly, the categorical model presented in Table 28 *did* indicate a difference in developmental patterns between the two groups. Judging from the negative interactions between the TBLT group and the posttests for both 0a and 1a responses, participants in the TSLT group seemed to have undergone gradual improvement in case knowledge by progressing from fully inaccurate responses to semi-accurate ones, including spelling, paradigm, and contextually faulty case-assignment errors. Furthermore, the strongly supported positive interaction between the delayed posttest and the TBLT group suggests superior retention of not only case-assignment knowledge but also spelling and paradigms. This may even indicate that the TBLT participants largely skipped 1a and 0a errors, advancing directly from fully incorrect 0-type to fully accurate 1-type responses. However, the interaction for 1-type responses at the posttest with the TBLT group was not strongly supported, and the model showed potential advantages for the TBLT group at the pretest on 0a- and 1-type responses. Consequently, instead of the developmental pattern laid out in Hypothesis 4, there does seem to be some evidence for different developmental patterns depending on TSLT or TBLT treatment, but the evidence so far is mixed.

5.4 Other variables potentially affecting the results

This subsection aims to identify all other confounding variables not included in the main hypotheses of this study. As far as possible, an attempt is made to explain these effects and to establish the extent to which these variables influence not only the results but also the interpretations in light of the hypotheses.

5.4.1 The participants' L1

The first potentially confounding variable to be discussed is the participants' L1. As noted, L1s were grouped into Balto-Slavic on the one hand and non-Balto-Slavic on the other. Latvian, the only other living Baltic language today, shares many lexical and structural similarities with Lithuanian, including a similarly rich case system, which may benefit L1 Latvian speakers when

learning L2 Lithuanian. Slavic languages share considerably fewer lexical commonalities with Lithuanian but do include a sizeable number of structural similarities, especially in their case systems.²⁰ Given that the target constructions in this study involve case marking, these typological similarities could likewise prove advantageous to speakers of a Slavic L1.

In the SLA literature, L1 effects are often established. According to N. Ellis's CREED (2006) and VanPatten's IP (2014), the presence of a nominal case system in the L1 should make case endings less redundant and, thus, more salient in the L2. Dabašinskienė and Čubajevaitė (2009) described certain frequent L2 Lithuanian case errors that can be attributed to the L1. Similarly, Stepšys and Kamandulytė-Merfeldienė (2024) found evidence of performance differences on adjectival case marking between adult L2 learners whose L1 was analytic (e.g., English) as opposed to synthetic (e.g., Russian). Conversely, Dubasava (2020) described very similar case errors made by her students despite their Russian L1 background. Likewise, Saturno (2023) found Russian L1 learners of L2 Polish making nearly identical mistakes to learners with other L1s, the only difference being that the Russian L1 learners progressed more quickly from one developmental stage to the next. Taken together, these findings suggest that the jury is still out on the role of the L1 in the L2 acquisition of nominal case marking.

The results reported in this dissertation seem to reflect this point. Regarding the gap-fill test, participants' L1 background yielded no clear advantage in case-marking performance at the pretest. At the posttest, however, L1 speakers of Baltic or Slavic languages appeared to have a significant advantage. Interestingly, however, this advantage then diminished at the delayed posttest. This might indicate that a typologically similar L1 is beneficial only immediately after treatment. Likewise, the EOIT showed very limited signs of beneficial L1 effects. Although there was some suggestive evidence that the Balto-Slavic L1 speakers had superior implicit case knowledge at pretest, the non-Balto-Slavic L1 speakers caught up completely at the posttest, at which point performance no longer differed. At the delayed posttest, the performance of the non-Balto-Slavic L1 speakers remained stable, whereas that of the Balto-Slavic L1 speakers slightly increased, but there was no evidence of a substantial difference in gains. Essentially, these results indicate that typologically similar L1s do not have an effect on improvement in implicit case acquisition. Nevertheless, clearly, based on the

²⁰ Notable exceptions are Bulgarian and Macedonian, which lack nominal case systems. One participant in the TSLT group had Bulgarian as their L1 and was encoded as Slavic accordingly, despite this absence of nominal case.

results of this study, no firm conclusions can be drawn regarding the effect of a typologically similar L1 on the acquisition of L2 case. Regarding the hypotheses, it is important to note that the number of Baltic and/or Slavic L1 speakers was very similar across both groups, thus rendering this effect moot for treatment comparisons.

5.4.2 Item characteristics: syntactic role, case, and paradigm

The next set of potentially confounding variables is item-related: syntactic role, case, and target noun paradigm. Regarding the gap-fill test, syntactic role proved relevant: object-marking items were consistently more problematic than subject-marking items, and this pattern remained stable across test times. Moreover, in contrast to the pilot study (Udes, 2025a), no substantial differences were observed between the treatment groups for either subject- or object-marking. The fact that subject-marking items were generally easier for participants than object-marking items could be explained by the way the target nouns were presented – specifically, in their base form. Subject-marking items requiring the nominative merely asked participants to copy the base form presented in brackets. As for the EOIT, a similar pattern was established. Object-marking items were more difficult for the participants in both treatment groups and remained stable across test times. The reason behind this might not be the nouns but rather the verbs that accompany them – while in subject-marking items, these could only be *yra* ‘there is’/ *néra* ‘there is no’, object-marking items involved a much wider array of verbs. Given that implicit learning is thought to facilitate chunking and the subsequent abstraction of grammar rules, this appears to lend weight to this interpretation.

Analyses based on grammatical case, however, seem to depict a somewhat different picture. On the gap-fill test, pretest performance showed slight support for lower accuracy in genitive marking (covering both subject and direct-object marking) compared with the overall performance average for case. Vastly inferior performance at marking the accusative in contrast to that same average, however, was strongly supported. Over both posttests, performance on the accusative remained markedly inferior in comparison to the marking of the two other cases. At the same time, strong evidence emerged for superior genitive marking performance at the posttest along with moderate evidence at the delayed posttest. In terms of learning gains for both groups, pairwise comparisons primarily revealed pre- to posttest and pre- to delayed posttest learning gains for genitive marking. Interestingly, learning gains in marking the nominative and accusative were only established at the delayed posttest. In terms of group differences, superior TBLT group performance was

observed only at the pretest for both genitive and accusative marking; however, the lack of evidence for further three-way interactions across test times indicates no further changes in this dynamic.

Following this, it seems that genitive marking was the most susceptible to learning gains. Superior performance when marking the genitive over both other cases, as well as seemingly stronger learning gains, suggest that the concept of negation, as marked by either *nėra* ‘there is no’ or the prefix *ne-* ‘NEG-’, might have allowed for a high level of contingency, which could surely have aided participants in appropriately marking negated subjects in the genitive. Alternatively, it might have provided sufficient adjacency for successful chunk learning and subsequent rule abstraction to take place. This is all in addition to the high frequency of the genitive case in the Lithuanian language in general, although it was shown in Table 15 that the number of instances of subject and/or object negation in the L2 input was rather limited. Perhaps the activities included in the subtasks, main tasks, and post-task stages, which often focused on the presence or absence of certain elements, made use of the genitive under negation more salient. Moving past negation alone, incorporating the partitive function of the genitive (as discussed in Section 3.4.1) would diminish the high contingency described, presumably leading to differing learning trajectories.

As for the EOIT, similar trends were observed. At the pretest, performance in marking both the genitive and the accusative showed a clear negative effect. However, pairwise comparisons revealed robust evidence of improvement from pre- to posttest and from pre- to delayed posttest in marking the genitive. No such evidence of improvement was found for nominative or accusative marking. Three-way interactions show strong evidence of superior improvement in accusative marking for the TBLT group at the posttest, but this effect is no longer present at the delayed posttest. One way to account for this observation might be the developmental cycle proposed for L2 Lithuanian object-marking in Section 2.2.1.3. It was argued that L2 Lithuanian students at Stage 2 would begin marking any case other than the nominative, most likely the genitive, given its high frequency. At Stage 3a, learners start marking objects in the accusative in both negated and non-negated contexts, while only at Stage 3b do learners mark objects in the accusative or genitive in the appropriate contexts. The participants in the TSLT group may have been predominantly at Stage 2, leading to more frequent genitive marking, which proved mostly accurate. In contrast, the TBLT group had already advanced to Stage 3a, leading to more accurate accusative marking for direct objects, but not in negated contexts.

Nevertheless, the absence of a significant interaction at the delayed posttest indicates that this developmental advantage was not sufficiently retained after 1.5 months.

Noun paradigms revealed no particularly important findings. Starting again with the gap-fill test, at the pretest, no clear differences between the *-is* and *-as* paradigms could be established; however, participants clearly had more trouble declining *-ė*-paradigm nouns, while the *-a*-paradigm ones presented less of a challenge. There seemed to be little change over test times, save for a decrease in performance when declining nouns in the *-a* paradigm. As for treatment, evidence emerged for superior performance at the pretest in declining the *-ė* paradigm for the TBLT group, while the TSLT group demonstrated superior skills in declining *-a*-paradigm nouns. Across test times, however, little changed, as evidenced by the lack of evidence for the corresponding three-way interactions. Results for the EOIT were no more informative. At pretest, groups showed no significant differences in performance, either together or separately, across all paradigms. Furthermore, no effect of test time was found. One moderately supported three-way interaction was established for the delayed posttest, the TBLT group, and the *-is* paradigm, showing greater improvement in nouns containing this ending in comparison to the TSLT group. In the absence of strongly supported interactions that form a coherent picture, these effects do not allow for any clear insights.

Nevertheless, since these item characteristics showed several significant effects and interactions, it was deemed important to check for outlier items. Starting with the gap-fill test, item outliers seem to confirm the established difficulties in declining the *-ė* paradigm. Across all tests, this specific paradigm seems the most problematic. A possible explanation for this is the fact that, no matter the case, endings belonging to this paradigm always include a diacritic sign (nominative: *-ė*; genitive: *-ės*; accusative: *-ę*). According to the rigid ‘all-or-nothing’ evaluation approach, failure to add such a sign leads to the response being marked incorrect. Therefore, unsurprisingly, most 1a responses belong to the *-ė* paradigm (53%). Another interesting finding is that, while *-is* and *-ys* endings are considered to belong to the same paradigm, the L2 participants in this study had considerably more difficulty declining *-ys* endings than *-is* endings, judging from the fact that four out of six target nouns ending with *-ys* appeared in the list of negative outliers (those being *traukinys* ‘train’, *Panevėžys* ‘Panevėžys’, a city in Lithuania’, *rūsų* ‘basement’, and *renginys* ‘event’). The city name *Panevėžys*, in particular, seemed to cause problems, most likely because it is a proper

noun. Thus, while the (masculine) *-is* and *-ys* endings are often considered to belong to the same paradigm, L2 learners seem to respond to them differently.

From the perspective of the target cases, the decrease in problematic genitive-marking items across testing sessions seems to indicate an improvement in marking this specific case. Accusative-marking items increased accordingly. Surprisingly, the TBLT group seemed to have had exceptional trouble marking the nominative in Items 33 and 34, especially the noun *knyga* ‘book’. Possibly, the TBLT group was more strongly primed to change the base form, leading to superfluous case marking. A final insight is that the number of negative outliers is considerably smaller for the delayed posttest results. This might indicate that items were more evenly distributed in terms of difficulty.

Moving on to the EOIT, it appears somewhat more complicated to draw coherent insights. Concerning the grammatically correct sentences, the presence of the target nouns *bananas* ‘banana’ and *jogurtas* ‘yoghurt’ can be explained by the fact that the genitive target form *bananO* was often pronounced *bananA* according to the English translation, whereas the target accusative form *jogurtA* was often pronounced *jogurtO*, corresponding to the genitive form, which is quite common with uncountable food items. Further patterns could not be established. As for the grammatically incorrect sentences, most of the problematic items targeted the genitive case, while the target paradigms were mostly *-ė* and *-is*. The decrease in outlier items targeting the genitive seems to corroborate the improvements observed in this case compared to tests in Table 39. The same table also showed inferior performance for the *-is* and *-ė* paradigms compared to *-as*. Clearly, paradigms affect L2 learner performance, although it is surprising that the supposedly less frequent *-a* paradigm causes fewer difficulties than the more frequent *-ė* and *-is* paradigms.²¹

5.4.3 Lexical and syntactic load variables

As mentioned above, a high rate of failures to repeat target items might indicate processing issues, which is why outliers in NA counts were identified. It is difficult to provide coherent insights, but it will be argued that a combination of factors influenced performance: lower target noun frequency,

²¹ As noted previously, these frequencies were based on dictionary frequencies rather than on L1 Lithuanian usage frequencies. It is entirely possible that the *-a* paradigm is much more frequent in L1 Lithuanian usage than the *-ė* or *-is* paradigms.

non-base form presentation, less frequent words in the context, and sentence length. Consider the following examples:

- (34) *Aš ne-žaidžiu krepšin-io.*
 I NEG-play basketball-GEN
 ‘I do not play basketball.’
- (35) *Aš visada planuoju savo savait-ę.*
 I always plan my week-ACC
 ‘I always plan my week.’

In Example (34), sentence length was likely not a factor. However, the combination of a lower-frequency verb, the target noun, and genitive marking (*krepšinio* instead of *krepšinis* ‘basketball’) might have made it difficult for the participants to repeat the sentence, even though it was presented as grammatically correct. As for Example (35), sentence length likely impeded repetition. Furthermore, a substantial proportion of the participants had trouble understanding and, thus, repeating the word *visada* ‘always’ and/or replaced the target noun *savaitė* ‘week’ with *savaitgalis* ‘weekend’, which might be a more frequent noun in the daily context of L2 participants. While lexical challenges can never be fully avoided, the four aspects mentioned can inform further test design to create new versions of EOITs. For instance, lower-proficiency participants may be able to repeat target nouns in non-base forms, as long as the test design controls for sentence length, target noun frequency, and the frequency of surrounding contextual vocabulary.

5.4.4 Methodological issues

Finally, some possible methodological issues related to the EOIT need to be discussed. Certain design choices were made to accommodate the participants’ low proficiency levels. However, these choices might have compromised the nature of the case knowledge tested. To reiterate, EOITs are deemed instruments for the measurement of implicit grammar knowledge of L2 participants (R. Ellis, 2009a). Because participants must respond to the statement before repeating it, they supposedly attend to meaning first. The need to repeat the statement afterward, before forgetting it, imposes a semi-natural time constraint on production, thus ensuring the use of implicit grammar knowledge. Direct formulation of the original statement is deemed unimportant in this regard, since sentence form is generally forgotten before its intended meaning (Erlam, 2009, p. 66). Corrected ungrammatical items provide even stronger evidence of the application of participants’ implicit

knowledge, since they involve the accurate reconstruction of the statement, ostensibly purely based on meaning. A meta-analysis (Kostromitina & Plonsky, 2022) of different studies that included the EOIT found that it is indeed a good measure of L2 proficiency.

One design choice that may have influenced the cognitive processes at work during this test is sentence length. The pilot study revealed issues that those participants had with approximately half of the items, several of which stemmed from lengthy sentences. Based on this observation, statements were shortened. Generally, while it is thought that sentence length does not affect participants' production (Erlam, 2009; Spada et al., 2015), it is advisable to provide statements of various lengths (Bley-Vroman & Chaudron, 1994, p. 253). Kostromitina & Plonsky (2022), however, argue that increased sentence length provides a more accurate measure of L2 proficiency. As mentioned prior, while the average syllable length of items in this study (8.97) showed a huge difference from, for example, Erlam's (2009) study (13.53), it was closer to Spada et al.'s (2015) study (9.95).

Nevertheless, word counts might be more relevant than syllable counts. As frequently mentioned before, morphological markers in oral speech are considered less salient (N. C. Ellis, 2022), meaning L2 learners attend more to separate words than to endings. Taking the items from Spada et al.'s (2015) study, their average word count was somewhat higher (6.86) than in this study (4.25). When focusing specifically on content words, the difference between their average (4.24) and this study's (2.56) remains considerable. Theoretically, this is important given the cognitive processes involved in participants' processing of test statements. While absolute sentence length in syllables might matter, a lower density of content words may reduce processing load for participants, thereby reducing the delay between stimulus and response, particularly when both truth value and sentence repetition are involved. Consequently, participants may be more likely to recall the original formulation of the statement, leading to direct repetition or greater attention to potential grammatical inconsistencies. Crucially, this is an important methodological consideration when designing EOITs to measure participants' proficiency in morphologically complex L2s.

Another design choice that potentially influenced the results was the decision to change the type of statements that this study's participants had to respond to. According to R. Ellis (2009a), statements should be belief statements that participants must respond to before repeating them, so the focus falls on meaning. While the current study included belief statements, they were, importantly, not designed for responses along the 'agree/disagree'

paradigm, but rather along the ‘true/false’ paradigm. During instrument design, it proved very difficult to invent short statements that participants could agree or disagree with while avoiding confounding grammatical variables such as adjectives (which also carry case, leading to double targets) and plural nouns (which often allow partitive use of the genitive). The problem with these true/false belief statements is that they might be easier to interpret. Processing a statement based on its truth value seems simpler than deciding whether to agree or disagree. Interestingly, however, Spada et al. (2015) included mainly statements that elicited responses according to their truth value. Thus, although it remains unclear whether participants attend to items differently when responding with true/false or agree/disagree, it is crucial to specify this in study design descriptions.

Another issue is whether the test truly taps into participants’ knowledge. While the test adheres to the criteria laid out in Table 18, and previous studies have found credible evidence for results regarding their participants’ implicit grammar knowledge (Erlam, 2009; Spada et al., 2015), problems arise when participants realize that some of the items are grammatically incorrect. As in Erlam (2009), this study’s participants were never explicitly told that they would hear grammatically incorrect items. Instead, they were instructed to repeat the sentences in “correct Lithuanian”. Nevertheless, in this study, several participants from both treatment groups and the control group reported noticing mistakes or asked whether this was the case. In both cases, the participants were either told that no comments could be made on the matter and/or the instructions were repeated. Such participants either started correcting the sentences or repeated them erroneously as heard in the audio, despite the instructions. In case of the former, it is highly doubtful that these participants used implicit knowledge when accurately repeating the sentences, since they no longer did so subconsciously with a focus on meaning.

This leads us to the next problem, which concerns potential training effects over test times. While the results were inconclusive regarding significant superior performance by the control group at the posttest, the fact that the group came so close without any treatment, while only the TBLT treatment showed some potential for a significant difference in learning gains at the posttest, is concerning. One of the main problems seems to be that recognizing the presence of grammatically incorrect items not only changes how participants attend to them. The additional danger that arises is twofold: (1) at pretest, participants might not immediately realize that some of the statements are grammatically incorrect, leading them to start correcting

sentences when several items have already passed, while in subsequent tests this awareness is already present from the start; (2) with each additional test, the chances of participants noticing the presence of grammatically incorrect sentences increase. Suga & Loewen (2023) found the latter in their study. Nevertheless, such training effects can be mitigated if there is a significant difference in performance between the control and treatment groups, or within the treatment groups, for which there was some evidence at the posttest and quite strong evidence at the delayed posttest.

5.5 Implications, limitations, and suggestions for further research

5.5.1 Theoretical implications

The study may have numerous theoretical and practical implications. Beginning with its theoretical implications, it seems paramount to consider its significance for TBLT's fundamental principles, namely the cognitive-interactionist model. To recap, at its core stands the Interaction Hypothesis (Long, 1996) and, therefore, interaction as the chief motor behind effective and efficient L2 acquisition. As Long (2014) and N. Ellis (2005, p. 306) claim, even for adults, L2 learning mainly relies on implicit learning mechanisms. However, adults' implicit learning capacities are not as effective as children's, at least in part because these implicit processes are mainly tuned to the learners' L1, thus rendering many important L2 specifics (e.g., vocabulary, syntax, or morphology) less noticeable. This is at least one reason why morphological markers, such as case markers, are 'fragile features' that are learned relatively late, or not at all. Hence, explicit teaching and/or learning is necessary, but one should not overdo it. Explicit instruction is valuable mainly as a facilitator of noticing, not as a mechanism of acquisition. Essentially, explicit learning yields explicit knowledge, which does not convert directly into implicit knowledge, but rather aids learners' implicit L2 learning processes.

The interaction hypothesis claims that interaction creates instances of 'negotiation for meaning', which allow for the most efficient means of hybrid explicit and implicit learning. While attending to meaning during task completion, negative (but also positive!) feedback and communication problems arising from learner output can trigger these episodes of negotiation for meaning, leading learners' attention to briefly shift from meaning to the linguistic code. Such feedback and so-called interactional adjustments by the interlocutor are said to facilitate acquisition as "it connects input, internal learner capacities, particularly selective attention, and output in productive

ways” (Long, 1996, pp. 451–452). These brief nudges toward the linguistic code in response to communication problems are called ‘focus on form’, which aims to have learners briefly notice and consider grammatical issues in their output, after which attention returns to the task at hand. These short instances of focus on form are intended to provide essential explicit L2 knowledge that informs implicit processing of L2 input and, in so doing, allows incidental L2 acquisition and, consequently, implicit L2 knowledge. TBLT as a methodology primarily adheres to this cognitive-interactionist model. While it could be argued that the TSLT version of TBLT is at least partially informed by the same model, its focus on form is no longer reactive or brief. Because TSLT allows for preemptive focus on form at all task stages (i.e., before potential communication problems arise), it takes a much stronger skill-acquisition approach (DeKeyser, 2014) that may lead to different learning pathways.

The results of the TBLT group in this study strongly support the cognitive-interactionist model. Implicit learning of a complex grammatical construction was achieved through meaning-focused interaction, which provided opportunities for negotiation of meaning. Furthermore, while it might not be entirely considered ‘reactive’ focus on form, brief grammatical instruction limited to the post-task stage supposedly drew participants’ attention to relevant grammatical forms and constructions. This led not just to the priming of subsequent implicit processing but also to the formation of explicit case knowledge. Surprisingly, this was even possible for such a fragile category as case, with its low salience and often ambiguous endings. Apparently, the combination of meaningful task-induced interaction together with post-task instruction was sufficient to retune participants’ L1-biased processing. Furthermore, the input seemed to have provided sufficient exemplars for learners to form accurate form-meaning mappings.

Results for the TSLT group further support the cognitive-interactionist model and call into question the need for an alternative TSLT approach. As the cognitive-interactionist model illustrates, accumulated explicit knowledge cannot be automatically converted into implicit knowledge. The TSLT group showed just that: while it showed successful improvement in explicit knowledge of case, learners in this group did not appear to have achieved any significant implicit learning gains. This once again emphasizes that explicit knowledge, rather than being the goal, should have merely a supporting role for implicit learning mechanisms. It could be argued that the additional preemptive focus on form in the pre-task stages encouraged the participants to attend to the input and the subsequent tasks in

a more metalinguistic, problem-solving manner, thereby inhibiting implicit, usage-based learning. Additional noticing of linguistic form may have reduced the attentional resources available to learners during task performance for processing form-meaning mappings.

Nevertheless, it is important to note that the TSLT treatment interpretation employed in this dissertation is only one of many. Here, the interpretation aligned fairly closely with the PPP method due to the partial inclusion of a Practice stage through the gap-fill exercises after each listening text and, possibly, the subtasks, which, in the TSLT versions, might have functioned more like grammar exercises. Hence, alternative implementations of TSLT might yield different results. Judging from the results of this dissertation, especially those that align more closely with TBLT.

Regarding the long-term retention of case knowledge, both methodologies appear effective. While the TBLT group showed strong evidence of both sustained explicit and, possibly, implicit learning gains, the TSLT group also achieved durable improvement in explicit case knowledge. Nevertheless, despite this difference, the categorical results might illustrate some further developmental differences in learning. As described in Section 1, it is generally believed that explicit learning is most effective for teaching and learning relatively simple rules, whereas implicit learning is more effective for chunk learning and mastering so-called fuzzy rules, including non-salient underlying patterns. The rules behind the target constructions are relatively straightforward (i.e., direct objects are marked with the accusative; negated subjects and objects are marked with the genitive), although the marking through a wide array of endings is not. Moreover, sentences often do not consist solely of a subject, a predicate, and an object, which means that context can make the case-marking process even more complicated. Assuming that the TSLT group attended to the L2 through more metalinguistic problem-solving, while the TBLT group employed a more implicit usage-based learning approach, it could be argued that the TSLT group was successful at primarily dissecting the rules behind the target constructions, whereas the TBLT group relied more heavily on exemplars, leading to the formation of a system based on the fuzzy rules of less salient case marking paradigms. This system seemed particularly stable: while the categorical results strongly supported superior growth of 0a (paradigm errors and semi-appropriate oblique case assignment) and 1a (spelling errors) responses for the TSLT group at both posttests, it was at the delayed posttest that similarly strong evidence supported superior growth of 1 (fully accurate) responses for the TBLT group. Once again, 0a- and 1a-type responses were quite rare, and

the error analyses did not show clear developmental trends, warranting caution in drawing strong conclusions. Nevertheless, it can be strongly argued that the TBLT and TSLT groups in this study learned quite differently, with outcomes that could be largely predicted by the cognitive-interactionist model.

The cognitive-interactionist model coincides in many ways with N. Ellis' (2006) CREED, which means that most of the results of this study similarly strengthen the CREED framework. Essentially, the CREED posits that L2 learning occurs through statistical learning of constructions from meaningful exemplars. Like Long, N. Ellis argues that implicit learning is the default mechanism for adult L2 learners, emphasizing the emergence of functioning L2 systems through use. The TBLT group's performance demonstrated that it is possible to abstract constructions, i.e., form-meaning mappings, from usage events without heavy pre-task metalinguistic explanation. Moreover, since the TSLT group did not show any measurable implicit learning gains, it could be argued that the preemptive form-focused instruction led these participants to treat case endings as isolated symbols rather than as parts of functional constructions. This is further corroborated by the fact that the available exemplars in the input were identical for both groups and that only the TSLT group received these exemplars with visually enhanced endings.

Nevertheless, in the context of the superior TBLT group improvement – even in explicit case knowledge – it is important to consider the potential pretest advantage they had. Within the CREED framework, this can be explained in terms of better initial cue weighting from prior L2 exposure, which enabled them to more accurately interpret and organize case endings from the outset, thus allowing them to further, and potentially more quickly, enhance their emerging L2 case systems. A rather odd finding of this study within the CREED framework is the possible immediate jump from fully inaccurate to accurate responses, while skipping paradigm and/or spelling errors, as illustrated in the categorical model. In an emerging system, chunk learning would be expected to be followed by grasping functional distinctions, after which further morphological refinement would follow. Despite these details, the results do seem to support the CREED framework's understanding of L2 learning processes, in that durable knowledge of case can emerge from the interaction of comprehension, production, and feedback across tasks, rather than from explicit pre-task work alone. Salience alone, such as through input enhancement, is insufficient without communicative contingency.

In the context of VanPatten's (2014) Input Processing (IP) theory, the results are particularly interesting, as the theory predicts serious challenges for

learners in acquiring grammatical case, especially the target structures. As a reminder, the theory posits that L2 learning is primarily driven by meaning, leading learners to prioritize it over other factors when processing L2 input. According to a set of principles that further detail L2 input processing, the likelihood of L2 learners focusing on the target structures of this study is minimal. First, they are merely morphological endings rather than content words. Second, the ‘meanings’ of negation and (negated) direct object status are expressed not just morphologically but also (semi)lexically (*nèra* ‘there is no’; *turiu* ‘I have’). Consequently, these genitive inflections convey abstract (and largely redundant) grammatical functions rather than transparent semantic meaning.

Despite these predictions of IP, the study found strong evidence of learning gains in both treatment groups. One explanation for this might be the role of explicit instruction that is envisioned in the cognitive-interactionist model: it aims to retune implicit processing. While the extent of explicit focus on form differed between the TBLT and TSLT groups, it apparently sufficed for both to retune the originally L1-based implicit processing of the input, yielding strong improvement in case knowledge. Importantly, though, input enhancement did not seem to have made an impact. Judging from the ‘0a’ and ‘1a’ responses in the categorical model and the error-type analyses, the TSLT group did not perform better at selecting correct paradigms or writing them accurately. Even more so, the TBLT group demonstrated a potential advantage over the TSLT group in this regard. This essentially corroborates the IP theory in that, even despite input enhancement, input is still mainly processed for meaning. Lastly, the “Meaning before Nonmeaning Principle” might explain the weaker learning gains in the accusative case. While this might be a matter of perspective, it could be argued that ‘negation’ is closer to a meaning than the grammatical role of the direct object. Disregarding other potentially influential factors such as frequency, this could explain the superior learning gains for the genitive, especially for negated direct objects. Essentially, they are not marked as negated direct objects, but simply as negated nouns.

Likewise, the results of this study seem to largely support the predictions established based on Pienemann’s (1998) Processability Theory. As discussed in Section 2.2.1.3, the theory posits that developmental trajectories are universal across languages and learners, regardless of the learners’ L1, and that these trajectories can be predicted based on a processability hierarchy. This processability hierarchy is based mainly on the nature of the transfer of grammatical information within and between phrases and sentences, with each stage being a prerequisite for the next. Nevertheless,

this does not mean all L2 learners perform identically. Learners demonstrate variation within a hypothesis space, which allows for creative use of the L2 but, as long as the learner is not developmentally ready, does not, in itself, transcend the procedural stages proposed by the theory. Although learners cannot skip any stage in the processability hierarchy, L1s with similar grammatical characteristics can speed up acquisition, i.e., the transition from one stage to the next, as Saturno (2023) shows.

The results of the present study appear to support the theory in two ways. First of all, learners' typologically similar L1s seemed to have had only a limited effect on learning gains in L2 Lithuanian case. For the gap-fill test, Balto-Slavic L1s appeared to have yielded superior L2 Lithuanian case improvement at the posttest; however, this advantage was diminished by the delayed posttest, thereby indicating no long-term L1 benefits for the acquisition of explicit L2 case knowledge. Naturally, if instruction had continued, this effect might have persisted and even grown incrementally. Conversely, for the EOIT, there was some moderate evidence of a negative effect of having a typologically similar Balto-Slavic L1. However, this effect was found only at the posttest, thus rendering any long-term retention nonexistent. Nevertheless, it is interesting that typologically similar L1s did not seem to help learners achieve short-term gains in either explicit or implicit knowledge. This suggests that the effects of similar L1 knowledge are limited to more effective noticing of grammatical forms but do not automatically transfer to superior unconscious case assignment, both of which the cognitive-interactionist model predicts. Crucially, typologically similar L1s showed no evidence that learners could skip developmental stages, at least in terms of long-term performance.

The second aspect supporting the Processability Theory is the developmental behavior demonstrated by the treatment groups. Magnani and Artoni's (2015) hypothetical framework for acquiring the case system in L2 Russian comprises four stages, with Stages 2 and 3 being the most crucial for this study. The second categorical stage predicts an emerging competition between marking the base form and non-nominative forms, typically representing the most frequent case inflections. Direct object marking, which was one of the target constructions in this study, belongs to the third phrasal stage, which involves the exchange of phrasal information. Error-type analyses revealed that, while the TBLT group demonstrated a decrease in both oblique case assignment and base form errors from pre- to posttest, most of the TSLT group's gains seemed to stem from a decrease in base form errors. Furthermore, at the delayed posttest, the TSLT group demonstrated a strong

increase in base-form errors, while the TBLT group showed a further decrease in both error types. This suggests that most learners in the TSLT group are still at the binary-categorical stage, seemingly still figuring out in which contexts nouns need to be assigned a different ending. Conversely, the TBLT group appears to have largely moved on to the third phrasal stage, judging not only from the further decrease in errors from post- to delayed posttest but also from the strongly supported interaction observed in the results of the EOIT between the TBLT group, posttest, and improvement in marking the accusative case. While the evidence is not overwhelmingly robust, there are indications that the TBLT group is developmentally slightly more advanced than the TSLT group, which can be determined within the framework of Pienemann's (1998) Processability Theory and Magnani and Artoni's (2015) hypothetical framework.

DeKeyser's (2014) Skill Acquisition Theory was highly relevant to this study, as it provided a solid theoretical framework for the TSLT angle of TBLT. Skill Acquisition Theory's central claim is that "practice makes perfect" and that explicit knowledge of a subject, such as an L2 structure, develops through practice and goes through three developmental stages: firstly, the declarative stage (solely knowledge about the subject); subsequently, the procedural stage (the knowledge is now available as a ready-made chunk); and, finally, the autonomous stage (fluent application of the knowledge). Acquisition of initial declarative knowledge is a prerequisite for advancing to the subsequent stages. And while proceduralization can occur relatively quickly, automatization requires extensive practice. Hence, expecting automatization of knowledge in an L2 intervention study such as this is unlikely; however, proceduralization of that knowledge seems feasible. Nevertheless, the feasibility of acquiring an L2 as a skill depends not only on the learners' characteristics and/or the instructional context but also on the target structure. As skill learning is considered mostly conducive to learning relatively simple rules, the target constructions in this study do not seem ideal. While the two main functions (negation and direct object marking) seem relatively straightforward but rather abstract, the multitude of endings used for case marking most surely presents a difficult challenge.

Evidently, while the TSLT group achieved explicit learning gains, it showed no evidence of implicit case acquisition, rendering the TSLT approach less effective than the TBLT treatment. If declarative knowledge is considered explicit and procedural knowledge is deemed implicit, it could be argued that the TSLT group's case knowledge remained declarative and that the treatment did not contribute to its proceduralization. While it might be argued that this

did indeed occur for the TBLT group, it must be emphasized that this treatment placed much less emphasis on metalinguistic knowledge of the target constructions, which is considered a prerequisite for advancement to the proceduralization and automatization stages. This calls into question the necessity of mastering metalinguistic knowledge of a complex grammatical category as a prerequisite for further acquisition. Studies such as those by Kenanidis et al. (2023) found this to be a prerequisite for stronger incidental learning, suggesting that the limited focus on form(s) in the TBLT treatment was enough to facilitate implicit learning processes. As mentioned previously, certain characteristics may have impeded the potential effectiveness of the TSLT treatment if it is equated with skill acquisition. However, this raises doubts about the necessity of such an approach if it is less effective than standard TBLT overall.

5.5.2 Practical implications

Beyond theoretical implications, this study might offer important guidance for both task design in educational contexts and for further research on task design in the context of case acquisition and the teaching of morphologically complex L2s, not just L2 Lithuanian. First and foremost, the study shows that it is indeed possible to teach a complex grammatical category, such as case, through the performance of TBLT-aligned tasks, resulting not just in immediate explicit and (potentially) implicit learning gains, but also in long-term improvement in both forms of knowledge. In other words, the acquisition of explicit case knowledge does not necessitate the inclusion of explicit focus on form through all task stages. This may substantially inform further research on task design. For instance, the results raise doubt about the need for focused tasks and a TSLT approach, at least one that more closely resembles PPP. While it has often been argued that teaching morphologically complex languages and/or complex grammatical structures requires more explicit instruction than usually provided for in strict TBLT settings, this study's results have shown that the TBLT methodology might be just as conducive to the acquisition of explicit case knowledge, while also fostering implicit case knowledge. Naturally, the participants in this study were not complete beginners and, most likely, already possessed some knowledge of subject and direct-object marking from the start. However, such clear signs of successful attainment are nonetheless incredibly informative.

Second, A1–A2-level L2 Lithuanian learners managed to perform relatively complex tasks involving speaking. While Kogan and Bondarenko (2022) argue for minimizing speaking requirements when designing tasks for

beginner learners, favoring solo tasks that involve mainly input, Gilabert and Castellví's (2019) proposed subtasks appear to have adequately prepared learners to perform the final main tasks. This design choice can be implemented in courses and tasks designed for teaching L2 Lithuanian and other morphologically complex languages.

Further TBLT research on the teaching of morphologically complex languages can be guided by these findings. First, while all tasks provided focus on form, and not just on the target constructions of this study, subject and direct object marking were the only categories consistently featured in each task and the only categories tested in the test instruments. This is rather unrealistic in actual language courses, where the focus needs to be on a broader range of important grammatical structures. The effectiveness of TBLT should be researched in such a context as well. It might be that the TSLT approach is more effective in such an all-encompassing L2 course, or that a combination of focused and unfocused tasks is warranted. Second, potential follow-up studies might even include a 'third' TBLT approach devoid of any explicit focus-on-form instruction to determine whether incidental learning of complex grammatical structures is possible without explicit awareness after all. If the results of such a group proved unsuccessful, this would provide additional evidence for the cognitive-interactionist model (R. Ellis et al., 2019; Long, 2014).

Third, subtasks were found to be a valuable design tool; however, it is unclear how task performance would have developed without them. Thus, further research might compare similar TBLT tasks that differ in whether they include subtasks to determine whether subtasks enhance performance. Alternatively, it might be explored whether input enhancement in instructions and prompts for the subtasks actually coerces learners into treating them as grammar tasks, as suspected in the TSLT materials. Subsequent differences in results as to the acquisition of explicit and implicit knowledge should then also be established.

Finally, while the participants in this study were low-proficiency learners, they were not absolute L2 Lithuanian beginners. Although successful case marking is important for absolute beginner learners, it might be worthwhile to investigate the design and long-term effectiveness of solo tasks, as well as comparisons between solo tasks and tasks involving speaking requirements, with and without subtasks, at higher levels.

This study's results may also inform SLA research in general. Although previous studies showed that L2 instruction minimally affected learners' performance in case acquisition, the present study demonstrated

robust evidence of durable case acquisition resulting from treatment. Long-term acquisition of explicit knowledge was even observed following both treatments. The reason for this difference might lie in a narrower scope within the selected target constructions. While Denissenko's (2016) study researched gains in L2 Russian case performance concerning the entire case category, the present study focused solely on (existential) subject and (prototypical) direct object marking. Most likely, for this reason, Markina (2018) did find immediate learning gains in prepositional case marking. Nevertheless, neither treatment group demonstrated long-term retention of prepositional case-marking knowledge.

The cause for this might be that some case functions are more susceptible to L2 instruction than others. While it might be argued that both prepositional case marking and subject and direct object marking for case are similarly grammatical operations, the former appears to be more random (prepositions govern a specific case or a specific set of cases which simply need to be learned by heart) than the latter (i.e., subject = nominative, direct object = accusative, negation = genitive). Therefore, it might be worthwhile in future SLA studies to investigate not only the acquisition of the whole case system or specific cases but also separate case functions and the extent to which they are affected by L2 learning. One way to operationalize this in the context of this study's target constructions might be to focus solely on direct-object marking and include non-prototypical direct-object marking, i.e., verbs that govern the genitive, dative, and instrumental, as they essentially mimic the 'randomness' argued for prepositional case marking. This might then further inform the design of lesson materials for the teaching of L2s featuring case.

Another important insight from this study concerns the role of metalinguistic knowledge in successful case acquisition. Although differences in metalinguistic awareness between the two treatment groups were not tested, the TSLT group was more likely to have higher levels of metalinguistic knowledge of the target constructions than the TBLT group, due to treatment differences. Previous research, such as that by Kenanidis et al. (2023), suggests that metalinguistic knowledge enhances incidental learning, which is a point also predicted by the cognitive-interactionist model (R. Ellis et al., 2019; Long, 2014). While this might suggest a TSLT advantage, this study's results indicate that the explicit focus on the forms provided in the post-task may have been sufficient for successful explicit and incidental acquisition of case. Kempe and MacWhinney's (1998) study showed that L2 learning is not necessarily rule-based but rather item-based, a finding further corroborated by

Brooks and Kempe (2008). This seems to echo the above-outlined interpretation of long-term retention of case knowledge, which holds that the TSLT group employed a more metalinguistic problem-solving approach. In contrast, the TBLT group processed the L2 in a rather usage-based manner through exemplars, leading to stronger item-based learning. The previously suggested third approach, in which post-task focus on form(s) would be absent, could potentially illuminate whether purely usage-based L2 acquisition of a complex grammatical structure is possible at all. In summary, metalinguistic knowledge seems important for (incidental) learning, although the desired extent remains open to debate.

A related issue in this regard is the apparent ineffectiveness of input enhancement. As a reminder, the reading and listening texts in the TSLT tasks included visual enhancements mainly of case forms. While these visual enhancements did not always align with the study's target constructions, they would likely have made case endings in general more salient. Nevertheless, both treatment groups produced paradigm errors, and the TBLT group even showed some indication of abrupt developmental shifts from fully inaccurate to fully accurate responses, suggesting less difficulty with spelling and case endings. Previous studies have shown mixed effects of pre-modified input (see Section 1.3.1.1) and the current study seems to indicate small effect sizes, if any. Nevertheless, input enhancement was only one aspect of the study, which means that further research is evidently required.

A separate aspect that might inform further research concerns the order of L2 case acquisition. Naturally, the target constructions do not allow for insights into the overall L2 Lithuanian case acquisition order, but they at least include the nominative, genitive, and accusative. So far, most insights have concerned L2 Russian case, which has shown that the three target cases are acquired in the order 'nominative > accusative > genitive' (Cherepovskaia et al., 2022; Rubinstein, 1995a). While the L1 acquisition order shows the concurrent emergence of both the genitive and the accusative, it nevertheless echoes the frequency distributions of case (nom. 32.6%, acc. 25.3%, gen. 22%, Zemskaia, 1979). However, according to this study's results on L2 Lithuanian, performance in marking the genitive appears to be much stronger than in marking the accusative, suggesting that, in L2 Lithuanian, the order of acquisition for these cases might be reversed, i.e., 'nominative > genitive > accusative'. Interestingly, this seems to be mostly echoed both in L1 acquisition (with acquisition of the genitive (1;7) being slightly earlier than the accusative (1;8) (Savickienė, 2003) and in frequency distributions such as

the one by Savickienė (2005; nom. 32%, gen. 28%, acc. 18%), as has been suggested in Section 2.2.3.

While L1 studies and frequency distributions seem to allow for predictions about the order of L2 case acquisition, L2 Russian data show notable disparities. For example, L2 learners acquire the locative/prepositional case quite early, even concurrently with or before the accusative, despite it emerging far later in L1 development and its significantly lower frequency than the accusative and genitive in L1 Russian. Such disparities may also exist in L2 Lithuanian and warrant investigation to improve L2 teaching materials.

The study might also inform test design, especially the EOIT. Applying such tests appears to entail additional challenges stemming from the nature of morphologically complex languages. It is important to consider that target nouns presented not in their base form but in another case might be less readily recognized. Additionally, syllable counts might be less relevant since extensive morphological markers lengthen the sentence; however, they are not as salient as separate (content) words. Counts in terms of content words, therefore, might be more appropriate. Lastly, an alternative to providing ungrammatical sentences amongst the items could be proposed. Participants could respond to belief statements not only in terms of ‘true’ or ‘false’ but also by changing the contents of said statements accordingly. In the current study, these changes are merely dichotomous and could easily be included in mixed-effects models as a random intercept. Moreover, truth values can be partially manipulated by providing statements about a participant’s surroundings and by presenting additional stimuli in those surroundings that pertain to those statements. Simply placing a bottle on the table while playing *Ant stalo nėra butelio* ‘There is no bottle on the table’ already provides more control.

CONCLUSIONS

This dissertation investigated whether explicit grammar instruction distributed throughout the task cycle (TSLT) offers advantages over a more limited focus on form, reserved primarily for the post-task stage (TBLT), in the acquisition of a complex grammatical category in a morphologically complex L2, namely Lithuanian case. It is still often thought that teaching morphologically complex L2s requires abundant explicit grammar instruction. Under this assumption, TBLT, traditionally aimed primarily at fostering implicit L2 knowledge by maintaining a focus on meaning and limiting explicit grammar instruction, is supposedly ill-suited for this purpose. By contrast, TSLT, which allows for preemptive explicit focus on form, could be a viable alternative for teaching morphologically complex L2s. Given the ongoing debates about the role of grammar in TBLT, the present study sought to determine whether these assumptions hold empirically.

Overall, the findings provide no support for the claim that extensive explicit grammar instruction throughout the task cycle is necessary for acquiring complex morphology. Both instructional approaches led to clear gains in explicit knowledge of case. These findings suggest that additional grammar instruction during the pre-task stage does not confer additional acquisitional benefits in this regard, a conclusion that may strongly inform task design. The data also indicate an advantage for the TBLT treatment, although potential baseline differences favoring the TBLT group warrant some caution in interpreting these results. Furthermore, the results indicate an advantage for TBLT in acquiring implicit knowledge of case. Only the TBLT group showed evidence of gains on the EOIT relative to the control group and strongly outperformed the TSLT group at the delayed posttest. Notwithstanding certain methodological limitations of the EOIT, these findings lend weight to both TBLT as a method for fostering implicit L2 knowledge and the EOIT as a measure of such knowledge.

Importantly, the results also indicate that even low-proficiency adult learners can achieve durable learning of complex grammatical structures. Apparently, a narrower focus on target structures that informs task and test design can lead to lasting improvements in performance with complex grammatical structures. Importantly, however, only TBLT seems conducive to fostering durable implicit grammar knowledge.

Taken together, these findings suggest that directing learners' attention to meaning before and during task performance, while restricting explicit focus on form(s) to the post-task stage, is both sufficient for acquiring explicit grammar knowledge and particularly conducive to developing

implicit grammar knowledge. Importantly, this supports the cognitive-interactionist framework that primarily informs TBLT and challenges the assumption that teaching morphologically complex L2s, especially at beginner levels, requires continuous explicit grammar instruction.

Regarding treatment differences in developmental patterns, some limited evidence was found for increased case-assignment skills accompanied by erroneous paradigm selections in the TBLT group, but the number of instances was too small to draw firm conclusions. Similarly, although the data showed that the TSLT treatment led to gradual case development and the TBLT treatment led directly to fully accurate performance, this pattern was only partially supported and thus requires further research.

At the same time, the study highlights several other factors that may have shaped the learning outcomes. First, participants' typologically similar L1s showed only limited effects. Although explicit grammar knowledge showed an immediate influence that diminished over time, no clear L1 effects were found for implicit case gains, leaving the overall findings inconclusive. Second, performance varied substantially by target syntactic role, with object marking proving more challenging than subject marking. This difference was most likely driven by participants' difficulties with the accusative, as strong learning gains were mainly observed for the genitive. The substantial gains in marking the genitive might have resulted from its high frequency in the input, its strong contingency following markers of negation, and the adjacency between these markers and subsequent genitive-marked nouns. Lastly, the paradigm analyses revealed that the overall dictionary frequency of paradigms in the L1 might be less relevant than their practical frequency in L2 input. This may explain the superior performance on lower-frequency paradigms, although potential spelling effects should not be disregarded.

The EOIT warrants additional consideration because certain methodological issues may have affected the results. The statements may have been relatively easy to process, potentially shortening the time between the audio input, truth evaluation, and sentence repetition. Additionally, participants who realized that the test contained grammatically incorrect sentences might have tapped into their explicit L2 knowledge rather than the intended implicit variety. Repeated test-taking not only increased this risk but also introduced training effects. Suggestions for improving the EOIT for L2 case acquisition research may inform future studies. Despite this, the evidence of TBLT-group learning gains provided a sufficient basis for interpreting the EOIT results.

In sum, this dissertation provides evidence that TBLT-aligned instruction can facilitate the acquisition of complex grammatical structures, even in morphologically complex L2s. This applies to both explicit and implicit case knowledge. The finding that TBLT can be equally conducive to TSLT approaches for acquiring explicit knowledge of a complex grammatical construction is particularly important and may inform further research. In contrast, extensive explicit grammar instruction throughout the task cycle does not necessarily enhance learning outcomes, especially for implicit L2 acquisition of grammatical constructions. Therefore, rather than expanding explicit instruction throughout the task cycle, the results suggest that maintaining a primary focus on meaning, complemented by explicit post-task grammar instruction, may offer a more effective approach to grammar instruction in TBLT, even when teaching morphologically complex L2s such as Lithuanian.

REFERENCES

- Akhutina, T. V. (2003). Is agrammatism an anomaly? *Journal of Russian & East European Psychology*, 41(3–4), 75–95.
<https://doi.org/10.2753/rpo1061-040541030475>
- Artoni, D. (2019). *Case within the phrasal procedure stage*.
<https://doi.org/10.1075/palart.7.05art>
- Balčiūnienė, A., Drukteinis, A., Kazlauskaitė, R., & Vaskelienė, J. (2024). *Praktinė bendrinė lietuvių kalbos gramatika [Practical general grammar of the Lithuanian language]*. Vilnius University Press.
- Baten, K. (2013). *The acquisition of the German case system by foreign language learners*. John Benjamins.
- Berry, D. C., & Broadbent, D. E. (1984). On the relationship between task performance and associated verbalizable knowledge. *The Quarterly Journal of Experimental Psychology Section A*, 36(2), 209–231.
<https://doi.org/10.1080/14640748408402156>
- Bley-Vroman, R., & Chaudron, C. (1994). Elicited imitation as a measure of second-language competence. In E. E. Tarone, S. M. Gass, & A. D. Cohen (Eds.), *Research methodology in second-language acquisition* (pp. 245–262). Routledge.
- Bresnan, J. (2000). *Lexical-functional syntax*. John Wiley & Sons.
- Brezina, V., & Pallotti, G. (2019). Morphological complexity in written L2 texts. *Second Language Research*, 35(1), 99–119.
<https://doi.org/10.1177/0267658316643125>
- Brinkutė, R. (2018). *Gramatinių kategorijų pasiskirstymas morfologiškai anotuotame lietuvių kalbos tekstyne [The distribution of grammatical categories in a morphologically annotated Lithuanian corpus]* [Master's thesis, Vytautas Magnus University]. <https://www.vdu.lt/cris/entities/etd/274f59e1-61b2-46c0-843d-d295d784da66>
- Bružaitė-Liseckienė, J. (2021). *Objekto raiška lietuvių kalboje: gimnazijų rusų ugdomąja kalba moksleivių atvejis [Object marking in the Lithuanian language: The case of students in Russian-language gymnasiums]* [Doctoral dissertation, Vilnius University].
<https://doi.org/10.15388/VU.THESIS.157>
- Buivydienė, V., & Žukienė, R. (2007). Komunikacinės metodikos taikymas mokant (-is) lietuvių kalbos kaip svetimosios [The application of communicative approaches in teaching Lithuanian as a foreign language]. *Filologija*, 15(4), 23–31.

- Bürkner, P. C. (2017). brms: An R package for Bayesian multilevel models using Stan. *Journal of Statistical Software*, 80, 1–28. <https://doi.org/10.18637/jss.v080.i01>
- Bybee, J. (2008). Usage-based grammar and second language acquisition. In P. Robinson & N. C. Ellis (Eds.), *Handbook of cognitive linguistics and second language acquisition* (pp. 226–246). Routledge.
- Chaudron, C. (1985). Intake: On models and methods for discovering learners' processing of input. *Studies in Second Language Acquisition*, 7(1), 1–14. <https://doi.org/10.1017/S027226310000512X>
- Cherepovskaia, N., Slioussar, N., & Denissenko, A. (2022). Acquisition of the nominal case system in Russian as a second language. *Second Language Research*, 38(3), 555–580. <https://doi.org/10.1177/0267658320988058>
- Čižik-Prokaševa, V. (2003). Linksnių atrakcija lyginamosiose konstrukcijose [The attraction of case in comparative constructions]. *Acta Linguistica Lithuanica*, 48, 1–17.
- Comer, W. J. (2012). Communicative language teaching and Russian: The current state of the field. In V. Makarova (Ed.), *Russian Language Studies in North America* (pp. 133–159). Anthem Press. <https://doi.org/10.7135/upo9780857286505.007>
- Corder, S. P. (1967). The significance of learner's errors. *IRAL - International Review of Applied Linguistics in Language Teaching*, 5(1–4), 161–170. <https://doi.org/10.1515/IRAL.1967.5.1-4.161>
- Council of Europe. (2020). *Common European Framework of Reference for Languages: Learning, teaching, assessment*. Cambridge University Press.
- Dabašinskienė, I., & Čubajevaitė, L. (2009). Acquisition of case in Lithuanian as L2: Error analysis. *Eesti Rakenduslingvistika Uhingu Aastaraamat*, 5, 47–65.
- Dadurkevičius, V. (2020). *Wordlist of lemmas from the joint corpus of Lithuanian*. Vilnius University Press. <https://clarin.vdu.lt/xmlui/handle/20.500.11821/41>
- Damos, D. (1991). *Multiple task performance*. CRC Press.
- Day, E., & Shapson, S. (1994). *Studies in Immersion Education*. Multilingual Matters. <https://doi.org/10.2307/329215>
- DeKeyser, R. (2014). Skill acquisition theory. In B. VanPatten, G. D. Keating, & S. Wulff (Eds.), *Theories in Second Language Acquisition: An Introduction* (pp. 106–124). Routledge. <https://doi.org/10.4324/9780203628942-11>

- Denissenko, A. (2016). *L'adquisició del rus com a llengua estrangera en un context d'instrucció formal a l'aula: Complexitat, Correcció i Fluïdesa i ús del cas en la producció escrita [The acquisition of Russian as a foreign language in a context of formal classroom instruction: Complexity, correctness and fluency and use of case in written production]* [Doctoral dissertation, Pompeu Fabra University].
- Doughty, C. J. (2003). Instructed SLA: constraints, compensation, and enhancement. In C. J. Doughty & M. H. Long (Eds.), *The handbook of second language acquisition* (pp. 256–310). Wiley.
- Dubasava, A. (2020). Acquisition of noun inflection in Lithuanian as a foreign language: A qualitative study. *Respectus Philologicus*, 37(42), 62–77. <https://doi.org/10.15388/RESPECTUS.2020.37.42.39>
- Ellis, N. C. (1996). Sequencing in SLA: Phonological memory, chunking, and points of order. *Studies in Second Language Acquisition*, 18(1), 91–126.
- Ellis, N. C. (1998). Emergentism, connectionism and language learning. *Language Learning*, 48(4), 631–664. <https://doi.org/10.1111/0023-8333.00063>
- Ellis, N. C. (2005). At the interface: Dynamic interactions of explicit and implicit language knowledge. *Studies in Second Language Acquisition*, 27(2). <https://doi.org/10.1017/S027226310505014X>
- Ellis, N. C. (2006). Cognitive perspectives on SLA: The associative-cognitive CREED. *AILA Review*, 19(1), 100–121. <https://doi.org/10.1075/aila.19.08ell>
- Ellis, N. C. (2008a). The psycholinguistics of the interaction approach. In A. Mackey & C. Polio (Eds.), *Multiple Perspectives on Interaction: Second Language Research in Honor of Susan M. Gass* (pp. 11–40). Routledge. <https://doi.org/10.4324/9780203880852>
- Ellis, N. C. (2008b). Usage-based and form-focused language acquisition: The associative learning of constructions, learned attention, and the limited L2 endstate. In P. Robinson & N. C. Ellis (Eds.), *Handbook of cognitive linguistics and second language acquisition* (pp. 382–415). Routledge. <https://doi.org/10.4324/9780203938560-24>
- Ellis, N. C. (2009). Optimizing the input: Frequency and sampling in usage-based and form-focused learning. In M. H. Long & C. J. Doughty (Eds.), *The Handbook of Language Teaching* (pp. 139–158). Blackwell. <https://doi.org/10.1002/9781444315783.CH9>
- Ellis, N. C. (2022). Second language learning of morphology. *Journal of the European Second Language Association*, 6(1), 34–59.

- Ellis, N. C., & Sagarra, N. (2010). The bounds of adult language acquisition: Blocking and learned attention. *Studies in Second Language Acquisition*, 32(4), 553–580. <https://doi.org/10.1017/S0272263110000264>
- Ellis, R. (1997). *SLA research and language teaching*. Oxford University Press.
- Ellis, R. (2003). *Task-based Language Learning and Teaching*. Oxford University Press.
- Ellis, R. (2005). *Planning and task performance in a second language*. John Benjamins.
- Ellis, R. (2009a). Measuring implicit and explicit knowledge of a second language. In C. Elder, R. Ellis, R. Erlam, S. Loewen, J. Philp, & H. Reinders (Eds.), *Implicit and explicit knowledge in second language learning, testing and teaching* (pp. 31–64). Multilingual Matters. <https://doi.org/10.21832/9781847691767-004>
- Ellis, R. (2009b). Task-based language teaching: Sorting out the misunderstandings. *International Journal of Applied Linguistics*, 19(3), 221–246. <https://doi.org/10.1111/j.1473-4192.2009.00231.x>
- Ellis, R. (2024). Task-based and Task-supported Language Teaching. *International Journal of TESOL Studies*, 6(4), 1-13. <https://doi.org/10.58304/ijts.20240401>
- Ellis, R., Basturkmen, H., & Loewen, S. (2002). Doing focus-on-form. *System*, 30(4), 419-432. [https://doi.org/10.1016/S0346-251X\(02\)00047-7](https://doi.org/10.1016/S0346-251X(02)00047-7)
- Ellis, R., Li, S., & Zhu, Y. (2019). The effects of pre-task explicit instruction on the performance of a focused task. *System*, 80, 38-47. <https://doi.org/10.1016/j.system.2018.10.004>
- Ellis, R., Loewen, S., & Erlam, R. (2006). Implicit and explicit corrective feedback and the acquisition of L2 grammar. *Studies in Second Language Acquisition*, 28(2), 339–368.
- Ellis, R., Skehan, P., Li, S., Shintani, N., & Lambert, C. (2019). *Task-based language teaching: Theory and practice*. Cambridge University Press. <https://doi.org/10.1017/9781108643689>
- Erlam, R. (2009). The elicited oral imitation test as a measure of implicit knowledge. In C. Elder, R. Ellis, R. Erlam, S. Loewen, J. Philp, & H. Reinders (Eds.), *Implicit and Explicit Knowledge in Second Language Learning, Testing and Teaching* (pp. 65–93). Multilingual Matters. <https://doi.org/10.21832/9781847691767-005>

- Erlam, R., & Ellis, R. (2018). Task-based language teaching for beginner-level learners of L2 French: An exploratory study. *Canadian Modern Language Review*, 74(1), 1–26.
- Frigo, L., & McDonald, J. L. (1998). Properties of phonological markers that affect the acquisition of gender-like subclasses. *Journal of Memory and Language*, 39(2), 218–245. <https://doi.org/10.1006/jmla.1998.2569>
- Gagarina, N., & Voeikova, M. D. (2009). Acquisition of case and number in Russian. In U. Stephany & M. D. Voeikova (Eds.), *Development of nominal inflection in first language acquisition: A cross-linguistic perspective* (pp. 179–216). De Gruyter Brill. <https://doi.org/10.1515/9783110217117.179/PDF>
- Gallimore, R., & Tharp, R. G. (1981). The interpretation of elicited sentence imitation in a standardized context. *Language Learning*, 31(2), 369–392. <https://doi.org/10.1111/j.1467-1770.1981.tb01390.x>
- Gass, S. M. (1997). *Input, interaction, and the second language learner*. Routledge. <https://doi.org/10.4324/9780203053560>
- Gass, S. M., & Mackey, A. (2014). Input, interaction, and output in second language acquisition. In B. VanPatten, G. D. Keating, & S. Wulff (Eds.), *Theories in Second Language Acquisition: An Introduction* (pp. 194–220). Routledge. <https://doi.org/10.4324/9780203628942-15>
- Gilabert, R., & Castellví, J. (2019). Task and syllabus design for morphologically complex languages. In J. Schwieter & A. Benatti (Eds.), *The Cambridge Handbook of Language Learning* (pp. 527–549). Cambridge University Press.
- Goldberg, A. E., & Casenhiser, D. (2008). Construction learning and second language acquisition. In P. Robinson & N. C. Ellis (Eds.), *Handbook of Cognitive Linguistics and Second Language Acquisition* (pp. 207–225). Routledge. <https://doi.org/10.4324/9780203938560-17>
- Goldberg, A. E., Casenhiser, D. M., & Sethuraman, N. (2004). Learning argument structure generalizations. *Cognitive Linguistics*, 15(3), 289–316. <https://doi.org/10.1515/cogl.2004.011>
- Goo, J., Granena, G., Yilmaz, Y., & Novella, M. (2015). Implicit and explicit instruction in L2 learning: Norris & Ortega (2000) revisited and updated. In P. Rebuschat (Ed.), *Implicit and Explicit Learning of Languages* (pp. 443–482). John Benjamins.
- Gurzynski-Weiss, L. (2023). Let's talk tasks. *TASK. Journal on Task-Based Language Teaching and Learning*, 3(1), 191–208. <https://doi.org/10.1075/task.00019.gur>

- Gvozdev, A. N. (1961). *Voprosy izučeniya detskoj reči [Issues in the study of children's speech]*. Izdatel'stvo Akademii pedagogičeskix nauk RSFSR.
- Hamayan, E., Saegert, J., & Larudee, P. (1977). Elicited imitation in second language learners. *Language and Speech*, 20(1), 86-97. <https://doi.org/10.1177/002383097702000109>
- Han, Y., & Ellis, R. (1998). Implicit knowledge, explicit knowledge and general language proficiency. *Language Teaching Research*, 2(1), 1-23. <https://doi.org/10.1177/136216889800200102>
- Harley, B. (1998). The role of focus-on-form tasks in promoting child L2 acquisition. In C. J. Doughty & J. Williams (Eds.), *Focus on form in classroom second language acquisition* (pp. 156–174). Cambridge University Press.
- Henry, K. (1996). Early L2 writing development: A study of autobiographical essays by university-level students of Russian. *The Modern Language Journal*, 80(3), 309–326.
- Herron, D. T., & Bates, E. A. (1997). Sentential and acoustic factors in the recognition of open- and closed-class words. *Journal of Memory and Language*, 37(2), 217–239. <https://doi.org/10.1006/jmla.1997.2514>
- Hinz, J., Krause, C., Rast, R., Shoemaker, E. M., & Watorek, M. (2013). Initial processing of morphological marking in nonnative language acquisition. *EUROSLA Yearbook*, 13(1), 139–175. <https://doi.org/10.1075/eurosla.13.08hin>
- Holvoet, A., & Semėnienė, L. (2004). Linksnio teorijos pagrindai [Fundamentals of case theory]. In A. Holvoet & L. Semėnienė (Eds.), *Gramatinių kategorijų tyrimai* (pp. 11–33). Lietuvių kalbos institutas.
- Howatt, A. P. R. (1984). *A history of English language teaching*. Oxford University Press.
- Ishikawa, T. (2008). *Task complexity, reasoning demands and second language speech production* [Doctoral dissertation, Aoyama Gakuin University].
- Izumi, S. (2002). Output, input enhancement, and the noticing hypothesis. *Studies in Second Language Acquisition*, 24(4), 541–577. <https://doi.org/10.1017/S0272263102004023>
- Janowska, I. (2020). Zadaniowa koncepcja kształcenia językowego w materiałach do nauczania języka polskiego jako obcego [The task-based approach to language education in teaching materials for Polish as a foreign language]. *Acta Universitatis Lodziensis. Kształcenie Polonistyczne Cudzoziemców*, (27), 505–526.
- Johnson, K. (1982). *Communicative syllabus design and methodology*. Pergamon.

- Keck, C. M., Iberri-Shea, G., Tracy-Ventura, N., & Wa-Mbaleka, S. (2006). Investigating the empirical link between task-based interaction and acquisition. In L. Ortega & J. Norris (Eds.), *Synthesizing research on language learning and teaching* (pp. 91–132). John Benjamins. <https://doi.org/10.1075/llt.13.08kec>
- Kempe, V., & Brooks, P. J. (2008). Second language learning of complex inflectional systems. *Language Learning*, 58(4), 703–746. <https://doi.org/10.1111/j.1467-9922.2008.00477.x>
- Kempe, V., & MacWhinney, B. (1998). The acquisition of case marking by adult learners of Russian and German. *Studies in Second Language Acquisition*, 20(4), 543–587. <https://doi.org/10.1017/S0272263198004045>
- Kenanidis, P., Dąbrowska, E., Llompart, M., & Pili-Moss, D. (2023). Can adults learn L2 grammar after prolonged exposure under incidental conditions? *PLoS ONE*, 18(7 July), 1–40. <https://doi.org/10.1371/journal.pone.0288989>
- Kogan, V. V., & Bondarenko, M. (2022). Russian and Russia through tasks for beginners. In S. V. Nuss & W. W. Martelle (Eds.), *Task-Based Instruction for Teaching Russian as a Foreign Language* (pp. 77–97). Routledge. <https://doi.org/10.4324/9781003146346-6>
- Kostromitina, M., & Plonsky, L. (2022). Elicited imitation tasks as a measure of L2 proficiency. *Studies in Second Language Acquisition*, 44(3), 886–911. <https://doi.org/10.1017/S0272263121000395>
- Krashen, S., & Scarcella, R. (1978). On routines and patterns in second language acquisition and performance. *Language Learning*, 28(2), 283–300. <https://doi.org/10.1111/j.1467-1770.1978.tb00135.x>
- Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, 33(1), 159–174. <https://doi.org/10.2307/2529310>
- Larsen-Freeman, D. (2011). A complexity theory approach to second language development/acquisition. In D. Atkinson (Ed.), *Alternative Approaches to Second Language Acquisition* (pp. 48–72). Routledge.
- Laskowski, R. (1989). Markedness and the category of case in Polish. In O. M. Tomić (Ed.), *Markedness in synchrony and diachrony* (pp. 207–226). Mouton de Gruyter.
- Laugalienė, A. (2020). Partitivity and object marking in Finnish and Lithuanian. *Philologia Estonica Tallinnensis*, 5(1), 235–267.

- Lee, S. K., & Huang, H. T. (2008). Visual input enhancement and grammar learning: A meta-analytic review. *Studies in Second Language Acquisition*, 30(3), 307–331. <https://doi.org/10.1017/S0272263108080479>
- Leung, J. H. C., & Williams, J. N. (2011). The implicit learning of mappings between forms and contextually derived meanings. *Studies in Second Language Acquisition*, 33(1), 33–55. <https://doi.org/10.1017/S0272263110000525>
- Levelt, W. J. M. (1989). *Speaking: from intention to articulation*. MIT Press.
- Levkina, M. (2014). *The role of task sequencing in L2 development as mediated by working memory capacity* [Doctoral dissertation, University of Barcelona].
- Littlewood, W. (2007). Communicative and task-based language teaching in East Asian classrooms. *Language Teaching*, 40(3), 243–249. <https://doi.org/10.1017/S0261444807004363>
- Littlewood, W. (2014). Communication-oriented language teaching: Where are we now? Where do we go from here? *Language Teaching*, 47(3), 349–362. <https://doi.org/10.1017/S0261444812000134>
- Long, M. H. (1981). Input, interaction, and second language acquisition. *Annals of the New York Academy of Sciences*, 379(1), 259–278. <https://doi.org/10.1111/j.1749-6632.1981.tb42014.x>
- Long, M. H. (1985). A role for instruction in second language acquisition: Task-based language teaching. In K. Hyldenstam & M. Pienemann (Eds.), *Modelling and assessing second language acquisition* (pp. 77–100). Multilingual Matters.
- Long, M. H. (1991). Focus on form: A design feature in language teaching methodology. In K. De Bot, R. B. Ginsberg, & C. Kramsch (Eds.), *Foreign Language Research in Cross-cultural Perspective* (pp. 39–52). John Benjamins.
- Long, M. H. (1996). The role of the linguistic environment in second language acquisition. In W. C. Ritchie & T. K. Bhatia (Eds.), *Handbook of Second Language Acquisition* (pp. 413–468). Academic Press. <https://doi.org/10.1016/b978-012589042-7/50015-3>
- Long, M. H. (2000). Focus on Form in Task-Based Language Teaching. In R. D. Lambert & E. Shohamy (Eds.), *Language Policy and Pedagogy* (pp. 179–192). John Benjamins. <https://doi.org/10.1075/z.96.11lon>
- Long, M. H. (2014). *Second language acquisition and task-based language teaching*. John Wiley & Sons.
- Long, M. H. (2016). In Defense of Tasks and TBLT: Nonissues and Real Issues. *Annual Review of Applied Linguistics*, 36, 5–33. <https://doi.org/10.1017/S0267190515000057>

- Lyster, R. (1994). The effect of functional-analytic teaching on aspects of French immersion students' sociolinguistic competence. *Applied Linguistics*, 15(3), 263-287. <https://doi.org/10.1093/applin/15.3.263>
- Lyster, R., & Saito, K. (2010). Oral feedback in classroom SLA: A meta-analysis. *Studies in Second Language Acquisition*, 32(2), 265-302. <https://doi.org/10.1017/S0272263109990520>
- Mačiukaite, S. (2008). Differences in adult L2 grammars of Lithuanian: The case of gender agreement. *Eesti Rakenduslingvistika Uingu Aastaraamat*, (4), 87-103. <https://doi.org/10.5128/ERYa4.06>
- Mackey, A. (1999). Input, interaction, and second language development: An empirical study of question formation in ESL. *Studies in Second Language Acquisition*, 21(4), 557-587.
- Magnani, M., & Artoni, D. (2015). Teaching learnable grammar in Russian as a second language: A syllabus proposal for case. In E. F. Quero Gervilla, B. Barros García, & T. R. Kopylova (Eds.), *Trends in Slavic studies* (pp. 57-70). URSS.
- Malicka, A. (2020). The role of task sequencing in fluency, accuracy, and complexity: Investigating the SSARC model of pedagogic task sequencing. *Language Teaching Research*, 24(5), 642-665. <https://doi.org/10.1177/1362168818813668>
- Markina, E. (2018). *Comparing focus on forms and task-based language teaching in the acquisition of Russian as a foreign language* [Doctoral dissertation, University of Barcelona].
- Milton, J. (2010). The development of vocabulary breadth across the CEFR levels. A common basis for the elaboration of language syllabuses, curriculum guidelines, examinations, and textbooks across Europe. In I. Bartning, M. Martin, & I. Vedder (Eds.), *Communicative proficiency and linguistic development: intersections between SLA and language testing research* (pp. 211-232). Eurosla.
- Morey, R. D., & Rouder, J. N. (2026). BayesFactor: Computation of Bayes factors for common designs. R package version 0.9.12-4.7. In *R package version 0.9*.
- Muranoi, H. (2000). Focus on form through interaction enhancement: Integrating formal instruction into a communicative task in EFL classrooms. *Language Learning*, 50(4), 617-673. <https://doi.org/10.1111/0023-8333.00142>
- Newport, E. L., & Aslin, R. N. (2004). Learning at a distance I. Statistical learning of non-adjacent dependencies. *Cognitive Psychology*, 48(2), 127-162.

- Norris, J. M., & Ortega, L. (2000). Effectiveness of L2 instruction: A research synthesis and quantitative meta-analysis. *Language Learning*, 50(3), 417-528. <https://doi.org/10.1111/0023-8333.00136>
- Nunan, D. (1989). *Designing tasks for the communicative classroom*. Cambridge University Press.
- Nuss, S. V. (2022). Morphology acquisition research meets instruction of L2 Russian. In S. V. Nuss & W. W. Martelle (Eds.), *Task-Based Instruction for Teaching Russian as a Foreign Language* (pp. 15–35). Routledge. <https://doi.org/10.4324/9781003146346-2>
- Oh, S.-Y. (2001). Two Types of Input Modification and EFL Reading Comprehension: Simplification versus Elaboration. *TESOL Quarterly*, 35(1), 69-96. <https://doi.org/10.2307/3587860>
- Paradis, M. (2009). *Declarative and Procedural Determinants of Second Languages*. John Benjamins.
- Pienemann, M. (1984). Psychological Constraints on the Teachability of Languages. *Studies in Second Language Acquisition*, 6(2), 186-214. <https://doi.org/10.1017/S0272263100005015>
- Pienemann, M. (1998). *Language processing and second language development*. John Benjamins.
- Pienemann, M., & Lenzing, A. (2020). Processability Theory. In B. VanPatten & J. Williams (Eds.), *Theories in second language acquisition* (pp. 162–191). Routledge.
- Prabhu, N. S. (1987). *Second language pedagogy*. Oxford University Press.
- Pribušauskaitė, J., Ramonienė, M., Stumbrienė, V., & Vilkienė, L. (2009). *Lietuvių kalbos kaip svetimosios mokymas: Keletas metodikos klausimų [Teaching Lithuanian as a foreign language: Some methodological issues]*. Vilnius University Press.
- R Core Team. (2023). R: A language and environment for statistical computing. R foundation for statistical computing. <https://www.R-project.org/>. *R Foundation for Statistical Computing*.
- Ramonaitė, J. T. (2015). Kaip lietuviškai šneka užsieniečiai? Lietuvių kaip antrosios kalbos įsisavinimas [How do foreigners speak Lithuanian? Learning Lithuanian as a second language]. *Baltistica*, 50(2), 295–330.
- Ramonaitė, J. T. (2017). What does say these forms as VALGU or ŽINĖJAU, created by foreign speakers. *Lietuvių Kalba*, (11), 1–25.
- Ramonaitė, J. T. (2020). Can we repeat what we do not say in L2? *Taikomoji Kalbotyra*, (14), 164–182.

- Rebuschat, P., & Williams, J. (2009). Implicit Learning of Word Order. In N. Taatgen & H. van Rijn (Eds.), *Proceedings of the Annual Meeting of the Cognitive Science Society* (pp. 425–430). Cognitive Science Society.
- Reeder, P. A., Newport, E. L., & Aslin, R. N. (2009). The role of distributional information in linguistic category formation. *Proceedings of the Annual Meeting of the Cognitive Science Society*, 31(31), 2564–2569.
- Reeder, P. A., Newport, E. L., & Aslin, R. N. (2010). Novel words in novel contexts: The role of distributional information in form class category learning. *The 32nd Annual Conference of the Cognitive Science Society*, 32(32), 2063–2068.
- Richards, J. (2015, August 27). *Bridging the gap between receptive and productive competence*. Cambridge. <https://www.cambridge.org/elt/blog/2015/08/27/bridging-gap-receptive-productive-competence/>
- Ringbom, H., & Jarvis, S. (2009). The Importance of Cross-Linguistic Similarity in Foreign Language Learning. In M. H. Long & C. J. Doughty (Eds.), *The Handbook of Language Teaching* (pp. 106–118). Blackwell. <https://doi.org/10.1002/9781444315783.ch7>
- Robinson, P. (2007a). Criteria for classifying and sequencing pedagogic tasks. In M. d. P. García Mayo (Ed.), *Investigating tasks in formal language learning* (pp. 7–26). Multilingual Matters.
- Robinson, P. (2007b). Task complexity, theory of mind, and intentional reasoning: Effects on L2 speech production, interaction, uptake and perceptions of task difficulty. *IRAL - International Review of Applied Linguistics in Language Teaching*, 45(3), 193–213. <https://doi.org/10.1515/iral.2007.009>
- Robinson, P. (2012). Abilities to Learn: Cognitive Abilities. In N.M. Seel (Ed.), *Encyclopedia of the Sciences of Learning* (pp. 17–20). Springer Nature. https://doi.org/10.1007/978-1-4419-1428-6_620
- Robinson, P. (2015). The Cognition Hypothesis, second language task demands, and the SSARC model of pedagogic task sequencing. In M. Bygate (Ed.), *Domains and directions in the development of TBLT: A decade of plenaries from the international conference* (pp. 87–121). John Benjamins. <https://doi.org/10.1075/tblt.8.04rob>
- Robinson, P. (2022). The cognition hypothesis, the triadic componential framework and the SSARC model: An instructional design theory of pedagogic task sequencing. In M. J. Ahmadian & M. H. Long (Eds.), *The Cambridge handbook of task-based language teaching* (pp. 205–225). Cambridge University Press.

- Robinson, P., & Ellis, N. C. (2008). Conclusion: Cognitive linguistics, second language acquisition and L2 instruction—Issues for research. In P. Robinson & N. C. Ellis (Eds.), *Handbook of Cognitive Linguistics and Second Language Acquisition* (pp. 499–556). Routledge. <https://doi.org/10.4324/9780203938560-28/CONCLUSION-COGNITIVE-LINGUISTICS-SECOND-LANGUAGE-ACQUISITION-L2-INSTRUCTION>
- Rubinstein, G. (1995a). On acquisition of Russian cases by American classroom learners. *IRAL - International Review of Applied Linguistics in Language Teaching*, 33(1), 9–34. <https://doi.org/10.1515/iral.1995.33.1.9>
- Rubinstein, G. (1995b). On case errors made in oral speech by American learners of Russian. *The Slavic and East European Journal*, 39(3), 408–429. <https://doi.org/10.2307/308241>
- Ruigendijk, E. (2002). *Case assignment in agrammatism: a cross-linguistic study* [Doctoral dissertation, University of Groningen].
- Sanders, A. F. (1998). *Elements of human performance: Reaction processes and attention in human skill*. Psychology Press.
- Sarandi, H. (2015). Reexamining elicited imitation as a measure of implicit grammatical knowledge and beyond...? *Language Testing*, 32(4), 485–501. <https://doi.org/10.1177/0265532214564504>
- Sato, M., & Loewen, S. (2018). Metacognitive instruction enhances the effectiveness of corrective feedback: Variable effects of feedback types and linguistic targets. *Language Learning*, 68(2), 507–545. <https://doi.org/10.1111/lang.12283>
- Saturno, J. (2023). Transfer (or lack thereof) and the accusative case in L2 Polish. In C. Granget, I. Repiso, & F. S. Guillaume (Eds.), *Language, creoles, varieties: From emergence to transmission* (pp. 303–342). Language Science Press. <https://doi.org/10.5281/zenodo.10280612>
- Savickienė, I. (2003). *The acquisition of Lithuanian noun morphology*. Österreichischen Akademie der Wissenschaften.
- Savickienė, I. (2005). Linksnių vartojimo dažnumas ir daiktavardžio reikšmė [Frequency of case usage and the meaning of the noun]. *Acta Linguistica Lithuanica*, 52, 59–65.
- Savickienė, I. (2006). Linksnio kategorijos įsisavinimas: Lietuvių kalba kaip gimtoji ir svetimoji [Acquiring the category of case: Lithuanian as a native and foreign language]. *Kalbotyra*, 56(3), 122–129.

- Schad, D. J., Vasishth, S., Hohenstein, S., & Kliegl, R. (2020). How to capitalize on a priori contrasts in linear (mixed) models: A tutorial. *Journal of Memory and Language*, 110, 1–40. <https://doi.org/10.1016/j.jml.2019.104038>
- Schmidt, R. W. (1990). The role of consciousness in second language learning. *Applied Linguistics*, 11(2), 129–158. <https://doi.org/10.1093/applin/11.2.129>
- Schmidt, R. W. (2001). Attention. In P. Robinson (Ed.), *Cognition and Second Language Instruction* (pp. 3–32). Cambridge University Press. <https://doi.org/10.1017/CBO9781139524780.003>
- Schmidt, R. W., & Frota, S. (1986). Developing basic conversational ability in a second language: A case study of an adult learner of Portuguese. In R. R. Day (Ed.), *Talking to learn: Conversation in second language acquisition* (pp. 237–326). Newbury House.
- Sheen, R. (2003). Focus on form—a myth in the making? *ELT Journal*, 57(3), 225–233. <https://doi.org/10.1093/elt/57.3.225>
- Shehadeh, A. (2012). Broadening the perspective of task-based language teaching scholarship: The contribution of research in foreign language contexts. In A. Shehadeh & C. A. Coombe (Eds.), *Task-Based Language Teaching in Foreign Language Contexts: Research and implementation* (pp. 1–20). John Benjamins.
- Shintani, N. (2015). The incidental grammar acquisition in focus on form and focus on forms instruction for young beginner learners. *Tesol Quarterly*, 49(1), 115–140.
- Shintani, N., & Ellis, R. (2010). The incidental acquisition of English plurals by Japanese children in comprehension-based and production-based lessons: A process-product study. *Studies in Second Language Acquisition*, 32(4), 607–637.
- Skehan, P. (1998). *A cognitive approach to language learning*. Oxford University Press.
- Skehan, P. (2009). Modelling second language performance: Integrating complexity, accuracy, fluency, and lexis. *Applied Linguistics*, 30(4), 510–532. <https://doi.org/10.1093/applin/amp047>
- Skehan, P. (2014). *Processing perspectives on task performance*. John Benjamins.
- Skehan, P. (2015). Limited attention capacity and cognition: Two hypotheses regarding second language performance on tasks. In M. Bygate (Ed.), *Domains and directions in the development of TBLT: A decade of plenaries from the international conference* (pp. 123–55). John Benjamins.

- Skehan, P., & Foster, P. (1997). The influence of planning and post-task activities on accuracy and complexity in task based learning. *Language Teaching Research*, 1(3), 185–211.
- Skehan, P., & Foster, P. (1999). The influence of task structure and processing conditions on narrative retellings. *Language Learning*, 49(1), 93–120. <https://doi.org/10.1111/1467-9922.00071>
- Slabakova, R. (2019). The Bottleneck Hypothesis updated. In T. Ionin & M. Rispoli (Eds.), *Three Streams of Generative Language Acquisition Research: Selected Papers from the 7th Meeting of Generative Approaches to Language Acquisition - North* (pp. 319–346). John Benjamins. <https://doi.org/10.1075/lald.63.16sla>
- Slobin, D. I. (1966). The acquisition of Russian as a native language. In K. Smith & G. A. Miller (Eds.), *The genesis of language: A psycholinguistic approach* (pp. 129–135). MIT Press.
- Smoczyńska, M. (1985). The acquisition of Polish. In D.I. Slobin (ed.), *The Crosslinguistic Study of Language Acquisition. Vol. 1: The Data* (pp. 595–681). Lawrence Erlbaum Associates.
- Spada, N., & Lightbown, P. M. (1999). Instruction, first language influence, and developmental readiness in second language acquisition. *The Modern Language Journal*, 83(1), 1–22.
- Spada, N., Shiu, J. L. J., & Tomita, Y. (2015). Validating an elicited imitation task as a measure of implicit knowledge: Comparisons with other validation studies. *Language Learning*, 65(3), 723–751. <https://doi.org/10.1111/lang.12129>
- Stanley, W. B., Mathews, R. C., Buss, R. R., & Kotler-Cope, S. (1989). Insight Without Awareness: On the Interaction of Verbalization, Instruction and Practice in a Simulated Process Control Task. *The Quarterly Journal of Experimental Psychology Section A*, 41(3), 553–577. <https://doi.org/10.1080/14640748908402382>
- Stepšys, J., & Kamandulytė-Merfeldienė, L. (2024). Būdvardiškųjų žodžių derinimo sunkumai: eksperimentinis dvikalbių vaikų ir svetimkalbių suaugusiųjų kalbos tyrimas [Difficulties in adjective agreement: An experimental study of the language of bilingual children and foreign-language-speaking adults]. *Deeds and Days*, 81, 117–136. <https://doi.org/10.7220/2335-8769.81.7>
- Stumbrienė, V., Daraškienė, I., & Vaškevičienė, L. (2022). *Sėkmės! [Good luck!] (books 1 and 2)*. Vilnius University Press.

- Stumbrienė, V., & Kaškelevičienė, A. (2018). *Nė dienos be lietuvių kalbos* [Not a day goes by without the Lithuanian language]. Vilnius University Press.
- Suga, K., & Loewen, S. (2023). Potential test-learning effects of an oral elicited imitation test: Methodological considerations for form-focused instruction studies. *Research Methods in Applied Linguistics*, 2(1), 1-13. <https://doi.org/10.1016/j.rmal.2022.100035>
- Swan, M. (2005). Legislation by hypothesis: The case of task-based instruction. *Applied Linguistics*, 26(3), 376–401. <https://doi.org/10.1093/applin/ami013>
- Tavakoli, P., & Skehan, P. (2005). Strategic planning, task structure and performance testing. In R. Ellis (Ed.), *Planning and Task Performance in a Second Language* (pp. 239–273). John Benjamins. <https://doi.org/10.1075/llt.11.15tav>
- Tomasello, M. (2003). *Constructing a language: a usage-based theory of language acquisition*. Harvard University Press.
- Tomasello, M. (2008). *Origins of human communication*. The MIT Press. <https://doi.org/10.7551/mitpress/7551.001.0001>
- Udes, R. (2025a). The role of grammar instruction in Task-Based Language Teaching: acquisition of case in L2 Lithuanian with different focus-on-form treatments. *Taikomoji Kalbotyra*, 22, 50–75.
- Udes, R. (2025b). Užduotimis grindžiamo kalbos mokymo (UGKM) metodikos potencialas mokant(is) sudėtingos morfologijos svetimųjų kalbų [Exploring the potential of task-based language teaching (TBLT) for teaching second languages with complex morphological systems]. *Verbum*, 16, 1–10. <https://doi.org/10.15388/verb.16.09>
- Van den Branden, K. (2006). *Task-based language education: From theory to practice*. Cambridge University Press.
- VanPatten, B. (2014). Input processing in adult second language acquisition. In B. VanPatten & J. Williams (Eds.), *Theories in Second Language Acquisition* (pp. 125–146). Routledge.
- VanPatten, B., & Keating, G. (2007, April). Getting tense: Lexical preference, L1 transfer and native and non-native processing of temporal reference. *Paper Delivered at the Annual Meeting of the American Association for Applied Linguistics*.
- Wagenmakers, E.-J., Wetzels, R., Borsboom, D., & Van Der Maas, H. L. J. (2011). Why psychologists must change the way they analyze their data: The case of psi: Comment on Bem. *Journal of Personality and Social Psychology*, 100(3), 426–432.

- White, J. (1998). Getting the learners' attention: a typographical input enhancement study. In C. J. Doughty & J. Williams (Eds.), *Focus on form in classroom second language acquisition* (pp. 85–113). Cambridge University Press.
- Wickens, C. D. (2007). Attention to the second language. *IRAL - International Review of Applied Linguistics in Language Teaching*, 45(3), 177–191. <https://doi.org/10.1515/iral.2007.008>
- Williams, J. N., & Kuribara, C. (2008). Comparing a nativist and emergentist approach to the initial stage of SLA: An investigation of Japanese scrambling. *Lingua*, 118(4), 522–553. <https://doi.org/10.1016/j.lingua.2007.03.003>
- Willis, D., & Willis, J. (2007). *Doing task-based teaching*. Oxford University Press.
- Willis, J. (1996). *A framework for task-based learning*. Longman.
- Wilson, B., Spierings, M., Ravignani, A., Mueller, J. L., Mintz, T. H., Wijnen, F., van Der Kant, A., Smith, K., & Rey, A. (2020). Non-adjacent dependency learning in humans and other animals. *Topics in Cognitive Science*, 12(3), 843–858.
- Wong, M. H. I. (2025). A Bayesian approach to small samples: Mixed-effects modeling in L2 interventional research. *Research Methods in Applied Linguistics*, 4(3), 1–11. <https://doi.org/10.1016/j.rmal.2025.100231>
- Wulff, S., & Ellis, N. C. (2018). Usage-based approaches to second language acquisition. In D. T. Miller, F. Bayram, J. Rothman, & L. Serratrice (Eds.), *Bilingual cognition and language: the state of the science across its subfields* (pp. 37–56). John Benjamins. <https://doi.org/10.1075/sibil.54.03wul>
- Wurzel, W. U., Panagl, O., Mayerthaler, W., & Dressler, W. U. (1987). *Leitmotifs in natural morphology*. John Benjamins.
- Yang, Y., & Lyster, R. (2010). Effects of form-focused practice and feedback on Chinese EFL learners' acquisition of regular and irregular past tense forms. *Studies in Second Language Acquisition*, 32(2), 235–263. <https://doi.org/10.1017/S0272263109990519>
- Yilmaz, Y. (2013). The relative effectiveness of mixed, explicit and implicit feedback in the acquisition of English articles. *System*, 41(3), 691–705.
- Zemskaja, E. A. (1979). *Russkaja razgovornaja reč: lingvističeskij analiz i problemy obučenija [Russian colloquial speech: Linguistic analysis and teaching problems]*. Russkij jazyk.
- Zygmantas, J. (2011) *Adult newcomers' difficulties in learning Lithuanian: an ethnographic case study* [Doctoral dissertation, Vilnius University].

APPENDICES

Appendix A – Information sheets and consent forms for participants

Information sheets received by the participants as well as the consent forms signed by them are stored in this dissertation's project on the OSF (Open Science Framework) website. Access to these files is granted through the following link:

https://osf.io/wfpe8/overview?view_only=7a3501c9c43341abbb54b5363741c3c4

The documents can be found in the folder titled 'Information sheets and consent forms'.

Appendix B – Task structures

Task 1		
Expressing your opinion on different residential areas		
Pre-task		
Part	Activities	
<i>Introduction</i>	Together with the class, the teacher discusses what a <i>rajonas</i> 'city district/neighborhood' is. They discuss in what district they live, whether they know the names of more districts, and what they know about them.	
<i>Reading</i>	Students read a text about four different city districts in Vilnius and some facts of interest about them. Based on the text, they list advantages and disadvantages of each district. Afterwards, they discuss in pairs which district they prefer.	
	TBLT	TSLT
	Finally, they write an interesting fact about their own hometown for homework.	Finally, they complete a table involving singular and plural locative, plural genitive endings for six paradigms.
<i>Listening</i>	Students listen to an interview between a journalist and a resident of the <i>Antakalnis</i> district. Students complete some exercises to check for comprehension and fill in gaps in a transcript of the listening text.	
	TBLT	TSLT
	Finally, they fill in the missing phrases in a transcript of the listening text.	Finally, they fill in the missing grammatical endings in a transcript of the listening text.

<i>Subtasks</i>	1. Student are given a selection of well-known places in Vilnius. They are then tasked to find out in what city districts these places are located.
	2. Students are given a simplified map of the district <i>Žvėrynas</i> and a list of questions about the available facilities. After answering these questions, they then express their opinion on the residential area.
Main task	
	In pairs, each student looks at a simplified map of a different residential area; pairs ask and answer questions about it.
	Pairs list the advantages and disadvantages of each residential area and express their preference.
Post-task	
<i>Results</i>	In class, several of each pair's choices and reasons are discussed.
	Students share in class what they learned about residential areas in Vilnius.
	Students revisit the reading and listening texts to identify phrases they consider useful when communicating in Lithuanian.
<i>Grammar</i>	Students complete an exercise about the locative case.
	Students complete an exercise about subject marking in affirmative and negated contexts.
	Students complete an exercise on the gender of numbers.
	Students complete two exercises on verbal paradigms.

Task 2		
Comparing rental apartments and selecting the most suitable one		
Pre-task		
Part	Activities	
<i>Introduction</i>	Together with the class, the teacher discusses whether students are satisfied with their current place of residence and what they like or dislike about it. They also discuss what they pay attention to when looking for a new place to live.	
<i>Reading</i>	Students read listings of two rental apartments and mark what rooms each apartment has. Afterwards, based on the texts, they list advantages and disadvantages of each apartment. Lastly, they discuss which apartment they prefer and write down three reasons in pairs.	
	TBLT	TSLT
	Students provide three advantages and disadvantages.	Students provide only two advantages and disadvantages.

		They complete a table involving the singular genitive, and singular and plural instrumental endings for six paradigms.
Listening	Students listen to a conversation of two friends who are looking for a new place to live. Students complete some exercises to check for comprehension.	
	TBLT	TSLT
	Finally, they fill in the missing phrases in a transcript of the listening text.	Finally, they fill in the missing grammatical endings in a transcript of the listening text.
Subtasks	1. Students are given the layout and a short list of details of two apartments. In pairs, they compare the apartments and list what is present in both and absent in one of them. Afterwards, they discuss which one they prefer.	
	2. The teacher provides two possible features of an apartment three times. Each time, in pairs, students are required to discuss which one is more important to them. They write down their choice and provide simple reasoning for it.	
Main task		
	In pairs, each student looks at a simplified listing of a different apartment; pairs ask and answer questions about them.	
	Pairs decide which apartment they prefer and provide several reasons.	
Post-task		
Results	In class, several of each pair’s choices and reasons are discussed.	
	Students are required to write a listing based on their current home.	
	Students revisit the reading and listening texts to identify phrases they consider useful when communicating in Lithuanian.	
Grammar	Students complete an exercise about subject marking in affirmative and negated contexts.	
	Students complete an exercise on nominative singular and plural and genitive plural marking following numbers.	
	Students complete an exercise on the instrumental case after the preposition <i>su</i> ‘with’.	
	Students complete an exercise on adjective–noun agreement based on grammatical gender.	
	Students fill in adjectival case endings based on the nouns. All sentences were taken from the reading and listening texts.	
	Students complete an exercise on relative pronouns.	

Task 3		
Deciding what furniture to buy		
Pre-task		
Part	Activities	
Introduction	Together with the class, the teacher discusses what types of furniture or objects are generally found in living rooms and bedrooms. They also discuss what they find important to have in these rooms. Lastly, they discuss a picture of a living room from the reading text that follows and what they feel it is lacking.	
Reading	Students read a forum discussion on how to furnish the living room pictured in the first post. Students mark which forum user mentions what furniture or object. Then, students mark whose advice they like and discuss their choices in pairs.	
	TBLT	TSLT
	Finally, students write their own response for on the forum.	Finally, they complete a table involving singular and plural genitive and accusative endings for five paradigms.
Listening	Students listen to a conversation of two friends who are looking for new furniture. Students complete some exercises to check for comprehension.	
	TBLT	TSLT
	Finally, they fill in the missing phrases in a transcript of the listening text.	Finally, they fill in the missing grammatical endings in a transcript of the listening text.
Subtasks	1. Students work in pairs. Each student is given a different picture of a living room. Together, they describe the living rooms and list what is present in both and absent in one of them.	
	2. The teacher provides pictures of two objects three times. Each time, in pairs, students are required to discuss which one they find more useful. They write down their choices and provide simple reasoning for it.	
Main task		
	Each pair looks at the picture of the living room and discusses what pieces of furniture are needed.	
	Each student has the inventory list of a different store, which they cannot show to each other. Together, the students orally exchange what items are available in each store.	
	Each pair can only afford four pieces of furniture. Students discuss and decide which they would like to buy and write down the agreed-upon list of items.	

Post-task	
Results	In class, several of each pair's choices and reasons are discussed.
	Students revisit the reading and listening texts to identify phrases they consider useful when communicating in Lithuanian.
Grammar	Students complete an exercise about subject marking in affirmative and negated contexts.
	Students complete an exercise about direct-object marking in affirmative and negated contexts.
	Students complete an exercise on genitive marking on objects after the verbs <i>trūkti</i> 'lack', <i>ieškoti</i> 'look for', <i>reikėti</i> 'need'.
	Students complete an exercise about direct-object marking in affirmative and negated contexts as well as marking on objects after the verbs <i>trūkti</i> 'lack', <i>ieškoti</i> 'look for', <i>reikėti</i> 'need'.
	Students complete an exercise on the comparative form of adverbs and adjectives.
	Students complete an exercise on the use of the subjunctive when giving advice.

Task 4		
Planning a menu for a dinner party		
Pre-task		
Part	Activities	
Introduction	Together with the class, the teacher discusses what an <i>įkurtuvės</i> 'housewarming party' is. They discuss whether they have ever hosted a housewarming party and in what ways you can host one. The teacher then mentions a dinner party as an option. The teacher and students then discuss what food and drinks can be offered during a dinner party and what other details are important when hosting such a party.	
Listening	Students listen to a radio interview with a professional party planner. She and the host discuss what to pay attention to when organizing a dinner party. Students complete some exercises to check for comprehension. In pairs, they then discuss the party planner's advice and think of their own menu.	
	TBLT	TSLT
	Finally, they fill in the missing phrases in a transcript of the listening text.	Finally, they fill in the missing grammatical endings in a transcript of the listening text.
Reading	Students read a chat between a couple who are hosting a dinner party. Students fill in some of the missing messages. They then write down which dishes and ingredients are not suitable for any of the guests and think of alternatives.	

	TBLT	TSLT
	Finally, in pairs, students discuss and write down what other dishes or ingredients, mentioned in the dialog, could be changed or replaced.	Finally, they complete a table involving singular and plural dative endings for six paradigms.
<i>Subtasks</i>	1. Students think of more examples of foods and drinks for each category. Then, in pairs, they discuss their likes and dislikes.	
	2. In pairs, students consider the information about two guests to their dinner party and several suggestions for dishes. They then determine whether these dishes suit their guests and, if not, what they need to substitute or change.	
Main task		
	In pairs, each student looks at the information on the food preferences and allergies of the guests they are planning to invite. They provide this information together with their own food preferences and allergies.	
	Pairs decide on an appropriate appetizer for their dinner party. They can come up with their own dish or choose one from a list of suggestions. Ingredients can be swapped out if necessary.	
	Pairs decide on an appropriate entrée for their dinner party. They can come up with their own dish or choose one from a list of suggestions. Ingredients can be swapped out if necessary.	
	Pairs decide on an appropriate dessert for their dinner party. They can come up with their own dish or choose one from a list of suggestions. Ingredients can be swapped out if necessary.	
Post-task		
<i>Results</i>	In class, each pair's menu is explained and discussed.	
	Students revisit the reading and listening texts to identify phrases they consider useful when communicating in Lithuanian.	
<i>Grammar</i>	Students complete an exercise about direct-object marking in affirmative and negated contexts.	
	Students complete an exercise about the dative case in the context of discussing allergies.	
	Students complete an exercise on the accusative and instrumental cases in the construction <i>pakeisti ką kuo</i> 'substitute smth.ACC with smth.INST'.	
	Students complete two exercises about prepositional marking involving the genitive and accusative cases.	

Task 5		
Shopping for groceries for a party		
Pre-task		
Part	Activities	
Introduction	Together with the class, the teacher discusses how often they usually go for groceries and what they often buy. They also discuss whether they have a preference for a specific grocery store or supermarket chain. Lastly, they discuss deals and discounts, and whether they are important to the students.	
Reading	In pairs, students look at the weekly advertisement from a local supermarket chain. Together, they discuss and select deals for items in several categories. Afterward, they discuss what they would buy if given 30 euros to spend.	
	TBLT	TSLT
	Finally, students discuss what products they would like to see deals for and what discount that deal might provide.	Finally, they complete a table involving singular and plural dative endings for six paradigms.
Listening	Students listen to a conversation of a couple who are discussing what to buy at the grocery store. Students complete some exercises to check for comprehension.	
	TBLT	TSLT
	Finally, they fill in the missing phrases in a transcript of the listening text.	Finally, they fill in the missing grammatical endings in a transcript of the listening text.
Subtasks	1. In groups, each student is given a set of bills and coins. Together, they calculate the total amount without showing their money.	
	2. Three times, the teacher provides an occasion and an unfinished grocery list. In pairs, students then have to think of what to add to each list.	
	3. Students work in pairs and are each given a selection of nine deals. They then have to discuss and decide on what to buy with a budget of 10 euros.	
Main task		
	No sequence of steps specified.	
	- Pairs have information on different food preferences and allergies of some of their guests.	
	- Each student receives the weekly advertisement from a different supermarket.	
	- Pairs have a budget of 90 euros.	

	- Pairs have to decide on the following purchases: ○ snacks ○ drinks ○ candy/sweets ○ other products
Post-task	
<i>Results</i>	In class, each pair's choices and reasons are discussed.
	Students revisit the reading and listening texts to identify phrases they consider useful when communicating in Lithuanian.
<i>Grammar</i>	Students complete an exercise about subject marking in affirmative and negated contexts.
	Students complete an exercise about direct-object marking involving (partitive) genitive marking and accusative marking with definite amounts.
	Students complete an exercise on the use of the subjunctive.
	Students complete an exercise about direct-object marking in affirmative and negated contexts as well as marking on objects after the verbs <i>trūkti</i> 'lack', <i>ieškoti</i> 'look for', <i>reikėti</i> 'need'.
	Students complete an exercise on the dative case in the context of deals and discounts.

Appendix C – Task materials

Task materials are stored in this dissertation's project on the OSF (Open Science Framework) website. Access to these files is granted through the following link:

https://osf.io/wfpe8/overview?view_only=7a3501c9c43341abbb54b5363741c3c4

The task materials can be found in the folder titled 'Task materials'.

Pictures were generated by using the AI image generator provided by Muse AI (<https://miramuseai.net/>). Apartment floorplans were created by using the tool provided by Floor Plan Creator (floorplancreator.net).

Appendix D – Target nouns, accompanying verbs, and target items

Gap-fill test

Target noun	Place in frequency list	Present in textbook	Test time	Target syntactic role
<i>aikštė</i> 'square'	996	yes	Delayed posttest	Object
<i>automobilis</i> 'car'	175	yes	Delayed posttest	Subject
<i>balkonas</i> 'balcony'	4736	yes	Posttest	Object
<i>bendrabutis</i> 'student dormitory'	3852	yes	Pretest	Subject
<i>butas</i> 'apartment'	454	yes	Pretest	Object
<i>darbas</i> 'work'	37	yes	Pretest	Object
<i>darželis</i> 'kindergarten'	1898	yes	Posttest	Subject
<i>eilė</i> 'line'	852	yes	Posttest	Subject
<i>informacija</i> 'information'	173	yes	Delayed posttest	Object
<i>jogurtas</i> 'yoghurt'	10729	yes	Posttest	Object
<i>kambarys</i> 'room'	1311	yes	Posttest	Subject
<i>katė</i> 'cat'	2462	yes	Delayed posttest	Object
<i>kavinė</i> 'café'	1932	yes	Posttest	Subject
<i>kelionė</i> 'trip'	628	yes	Pretest	Object
<i>klausimas</i> 'question'	137	yes	Pretest	Object
<i>knyga</i> 'book'	472	yes	Posttest	Subject
<i>kokybė</i> 'quality'	706	no	Pretest	Object

Target noun	Place in frequency list	Present in textbook	Test time	Target syntactic role
<i>kompiuteris</i> ‘computer’	1047	yes	Delayed posttest	Subject
<i>laikas</i> ‘time’	103	yes	Pretest	Subject
<i>laivas</i> ‘boat’	464	<i>no</i>	Delayed posttest	Subject
<i>langas</i> ‘window’	1354	yes	Posttest	Subject
<i>maistas</i> ‘food’	384	yes	Delayed posttest	Object
<i>miškas</i> ‘forest’	471	yes	Pretest	Subject
<i>mokinys</i> ‘student (masc.)’	826	yes	Pretest	Object
<i>mokykla</i> ‘school’	185	yes	Pretest	Object
<i>mokytoja</i> ‘teacher (fem.)’	2715	yes	Delayed posttest	Object
<i>namas</i> ‘house’	534	yes	Posttest	Object
<i>nuomonė</i> ‘opinion’	286	yes	Posttest	Object
<i>Panevėžys</i> ‘a city in Lithuania’	555	yes	Posttest	Object
<i>parduotuvė</i> ‘shop’	750	yes	Pretest	Subject
<i>pavardė</i> ‘last name’	823	yes	Delayed posttest	Object
<i>pavasaris</i> ‘spring’	1157	yes	Delayed posttest	Object
<i>pokalbis</i> ‘conversation’	1297	yes	Pretest	Object
<i>prezidentas</i> ‘president (masc.)’	193	yes	Pretest	Object
<i>renginys</i> ‘event’	505	yes	Delayed posttest	Object
<i>rūsys</i> ‘basement’	3486	yes	Post-test	Object
<i>salė</i> ‘hall’	721	yes	Delayed posttest	Subject
<i>sąskaita</i> ‘check/bill’	672	yes	Pretest	Subject
<i>saulė</i> ‘sun’	1123	yes	Posttest	Object
<i>savaitgalis</i> ‘weekend’	1235	yes	Pretest	Object
<i>šeima</i> ‘family’	255	yes	Posttest	Object
<i>sostinė</i> ‘capital’	639	yes	Delayed posttest	Subject
<i>spektaklis</i> ‘play/performance’	1284	yes	Delayed posttest	Object
<i>stalas</i> ‘table’	1417	yes	Posttest	Subject

Target noun	Place in frequency list	Present in textbook	Test time	Target syntactic role
<i>straipsnis</i> ‘article’	20	no	Posttest	Object
<i>šventė</i> ‘holiday/celebration’	578	yes	Pretest	Object
<i>teatras</i> ‘theater’	566	yes	Delayed posttest	Subject
<i>telefonas</i> ‘telephone’	617	yes	Delayed posttest	Subject
<i>traukinys</i> ‘train’	1882	yes	Pretest	Subject
<i>universitetas</i> ‘university’	449	yes	Pretest	Subject
<i>uostas</i> ‘port’	620	no	Posttest	Subject
<i>vaikas</i> ‘child’	88	yes	Posttest	Object
<i>vakaras</i> ‘evening’	518	yes	Delayed posttest	Object
<i>varškė</i> ‘curd’	6905	yes	Posttest	Object
<i>veidas</i> ‘face’	1089	yes	Delayed posttest	Object
<i>vieta</i> ‘place’	75	yes	Delayed posttest	Subject
<i>virtuvė</i> ‘kitchen’	2352	yes	Pretest	Subject

Accompanying verb	Place in frequency list	Present in textbook
<i>auginti</i> ‘grow/keep’	1346	no
<i>būti</i> ‘be’	2	yes
<i>gauti</i> ‘get/receive’	96	yes
<i>leisti</i> ‘spend’	249	no
<i>matyti</i> ‘see’	330	yes
<i>mėgti</i> ‘like’	1382	yes
<i>organizuoti</i> ‘organize’	525	no
<i>pirkti</i> ‘buy’	564	yes
<i>planuoti</i> ‘plan’	544	no
<i>pradėti</i> ‘begin’	144	yes
<i>rasti</i> ‘find’	359	no
<i>skaityti</i> ‘read’	800	yes
<i>turėti</i> ‘have’	19	yes
<i>užduoti</i> ‘give/set’	4519	no
<i>valgyti</i> ‘eat’	1437	yes
<i>žiūrėti</i> ‘watch’	905	yes

Elicited oral imitation test

<i>Target noun</i>	Place in frequency list	Present in textbook	Test time	Target syntactic role
<i>bankas</i> ‘bank’	209	yes	Pretest	Subject
<i>aikštė</i> ‘square’	996	yes	Posttest	Subject
<i>alkoholis</i> ‘alcohol’	744	no	Pretest	Object
<i>automobilis</i> ‘car’	175	yes	Pretest	Object
<i>balkonas</i> ‘balcony’	4736	yes	Pretest	Object
<i>bananas</i> ‘banana’	7052	yes	Posttest	Subject
<i>baras</i> ‘bar’	2023	yes	Posttest	Subject
<i>biblioteka</i> ‘library’	1529	yes	Pretest	Subject
<i>brolis</i> ‘brother’	1253	yes	Delayed posttest	Object
<i>butelis</i> ‘bottle’	2819	no	Posttest	Subject
<i>darbas</i> ‘work’	37	yes	Delayed posttest	Object
<i>filmas</i> ‘movie’	496	yes	Posttest	Object
<i>fotelis</i> ‘armchair’	15213	yes	Pretest	Subject
<i>futbolas</i> ‘soccer’	685	yes	Pretest	Object
<i>gatvė</i> ‘street’	316	yes	Delayed posttest	Object
<i>jogurtas</i> ‘yoghurt’	10729	yes	Delayed posttest	Object
<i>jūra</i> ‘sea’	532	yes	Delayed posttest	Object
<i>katė</i> ‘cat’	2462	yes	Posttest	Subject
<i>kava</i> ‘coffee’	2408	yes	Pretest	Object
<i>kelionė</i> ‘trip’	628	yes	Delayed posttest	Object
<i>kiaušinis</i> ‘egg’	2509	yes	Delayed posttest	Subject
<i>kompiuteris</i> ‘computer’	1047	yes	Posttest	Object
<i>krepsinis</i> ‘basketball’	538	yes	Posttest	Object
<i>langas</i> ‘window’	1354	yes	Pretest	Subject
<i>mėsa</i> ‘meat’	1698	yes	Posttest	Object
<i>miškas</i> ‘forest’	471	yes	Delayed posttest	Subject

Target noun	Place in frequency list	Present in textbook	Test time	Target syntactic role
<i>mokykla</i> ‘school’	185	yes	Delayed posttest	Subject
<i>numeris</i> ‘number’	708	yes	Delayed posttest	Object
<i>obuolys</i> ‘apple’	3692	yes	Delayed posttest	Subject
<i>parduotuvė</i> ‘shop’	750	yes	Posttest	Object
<i>pienas</i> ‘milk’	991	yes	Posttest	Object
<i>prezidentė</i> ‘president (fem.)’	884	yes	Posttest	Object
<i>restoranas</i> ‘restaurant’	1860	yes	Posttest	Object
<i>saulė</i> ‘sun’	1123	yes	Pretest	Subject
<i>savaitė</i> ‘week’	287	yes	Pretest	Object
<i>sofa</i> ‘couch’	8551	yes	Posttest	Subject
<i>sportas</i> ‘sports’	368	yes	Delayed posttest	Object
<i>stalas</i> ‘table’	1417	yes	Pretest	Object
<i>stotelė</i> ‘bus stop’	4246	yes	Delayed posttest	Subject
<i>sūris</i> ‘cheese’	3037	yes	Pretest	Subject
<i>teatras</i> ‘theater’	566	yes	Posttest	Subject
<i>telefonas</i> ‘telephone’	617	yes	Pretest	Subject
<i>traukinys</i> ‘train’	1882	yes	Posttest	Subject
<i>universitetas</i> ‘university’	449	yes	Delayed posttest	Subject
<i>upė</i> ‘river’	1665	yes	Pretest	Subject
<i>varškė</i> ‘curd’	6905	yes	Pretest	Object
<i>virtuvė</i> ‘kitchen’	2352	yes	Delayed posttest	Subject
<i>žurnalas</i> ‘magazine’	1457	no	Delayed posttest	Subject

Accompanying verb	Place in frequency list	Present in textbook
<i>būti</i> ‘be’	2	yes
<i>gerti</i> ‘drink’	1395	yes
<i>matyti</i> ‘see’	330	yes

Accompanying verb	Place in frequency list	Present in textbook
<i>mėgti</i> ‘like’	1382	yes
<i>planuoti</i> ‘plan’	544	<i>no</i>
<i>turėti</i> ‘have’	19	yes
<i>valgyti</i> ‘eat’	1437	yes
<i>žaisti</i> ‘play’	548	yes
<i>žiūrėti</i> ‘watch’	905	yes

Target items

The complete lists of target and filler items are stored in this dissertation’s project on the OSF (Open Science Framework) website. Access to these files is granted through the following link:

https://osf.io/wfpe8/overview?view_only=7a3501c9c43341abbb54b5363741c3c4

The lists can be found in the folder titled ‘Items.’

Appendix E – R code and prior data

The code used for computing the Bayesian models fitted to this study’s data as well as the prior distributions used for testing are stored in this dissertation’s project on the OSF (Open Science Framework) website. Access to these files is granted through the following link:

https://osf.io/wfpe8/overview?view_only=7a3501c9c43341abbb54b5363741c3c4

The task materials can be found in the folder titled ‘R code for models.’

Appendix F – Alternative models fitted for exploring the data

Bayesian posterior distribution summary for participants' responses on the gap-fill test of the two treatment groups (binary scoring) including the three-way interaction between the parameters Test Time, Treatment Group, and Syntactic Role.

Parameter	Median	HDI (89%) Low	HDI (89%) High	<i>pd</i>	% in ROPE (± 0.18)
Intercept	-0.68	-1.26	-0.10	96.96	6.43
Test Time: Posttest	1.36	0.85	1.89	99.99	0.00
Test Time: Delayed posttest	1.37	0.86	1.90	99.99	0.00
Treatment: TBLT	0.80	-0.08	1.68	92.69	9.69
L1: Balto-Slavic	-0.93	-1.61	-0.25	98.39	1.55
Syntactic Role: Object	0.33	-0.56	1.22	72.68	22.80
Test Time: Posttest x TBLT	0.35	-0.07	0.78	91.12	24.59
Test Time: Delayed posttest x Treatment: TBLT	0.41	-0.01	0.85	94.04	17.55
Test Time: Posttest x Syntactic Role: Object	-0.31	-1.26	0.66	69.72	21.39
Delayed posttest x Syntactic Role: Object	-0.52	-1.48	0.43	80.91	17.53
Treatment: TBLT x Syntactic role: Object	0.47	-0.12	1.04	89.45	18.70
Test Time: Posttest x Treatment: TBLT x Syntactic Role: Object	0.14	-0.67	0.93	60.69	27.91
Test Time: Delayed Posttest x Treatment: TBLT x Syntactic Role: Object	-0.58	-1.39	0.23	87.09	15.74

Note. The reference levels for each variable were as follows: Test Time: Pretest; Treatment: TSLT; L1: non-Balto-Slavic; Syntactic Role: Subject.

HDI = Highest Density Interval; *pd* = Probability of Direction; ROPE = Region of Practical Equivalence.

Bayesian posterior distribution summary for participants' responses on the gap-fill test of the two treatment groups (binary scoring) including the three-way interaction between the parameters Test Time, Treatment Group, and L1.

Parameter	Median	HDI (89%) Low	HDI (89%) High	<i>pd</i>	% in ROPE (± 0.18)
Intercept	-0.71	-1.28	-0.14	97.61	4.51
Test Time: Posttest	1.46	0.95	1.97	100.00	0.00
Test Time: Delayed posttest	1.37	0.85	1.87	99.99	0.00
Treatment: TBLT	0.99	0.10	1.90	96.12	5.01
L1: Balto-Slavic	0.02	-0.90	0.92	51.27	25.70
Syntactic Role: Object	-1.20	-1.61	-0.79	100.00	0.00
Test Time: Posttest x TBLT	0.36	-0.08	0.80	90.14	24.72
Test Time: Delayed posttest x Treatment: TBLT	0.45	0.00	0.91	94.27	15.59
Test Time: Posttest x L1: Balto-Slavic	0.77	0.33	1.21	99.79	0.00
Test Time: Delayed posttest x L1: Balto-Slavic	0.18	-0.27	0.64	73.97	41.88
Treatment: TBLT x L1: Balto-Slavic	1.20	-0.41	2.81	88.42	7.31
Test Time: Posttest x Treatment: TBLT x L1: Balto- Slavic	-0.45	-1.29	0.40	79.69	19.97
Test Time: Delayed Posttest x Treatment: TBLT x L1: Balto- Slavic	0.27	-0.61	1.15	69.51	24.28

Note. The reference levels for each variable were as follows: Test Time: Pretest; Treatment: TSLT; L1: non-Balto-Slavic; Syntactic Role: Subject.
HDI = Highest Density Interval; *pd* = Probability of Direction; ROPE = Region of Practical Equivalence.

Bayesian posterior distribution summary for participants' responses on the gap-fill test of the two treatment groups (binary scoring) including the three-way interactions between the parameters Test Time, Treatment Group, and Case, as well as Test Time, Treatment Group, and Paradigm.

Parameter	Median	HDI (89%) Low	HDI (89%) High	<i>pd</i>	% in ROPE (± 0.18)
Intercept	-0.55	-1.08	-0.02	94.97%	11.41%
Test Time: Posttest	0.92	0.54	1.30	99.98%	0%

Parameter	Median	HDI (89%) Low	HDI (89%) High	<i>pd</i>	% in ROPE (±0.18)
Test Time:	1.40	1.01	1.79	100%	0%
Delayed posttest					
Treatment: TBLT	0.58	-0.31	1.49	84.86%	15.23%
Case: Genitive	-0.59	-1.19	0.02	93.92%	11.58%
Case: Accusative	-1.02	-1.76	-0.29	98.62%	0.91%
Paradigm: -is/-ys	-0.37	-1.11	0.34	79.77%	24.43%
Paradigm: -è	-1.63	-2.40	-0.88	99.94%	0%
Paradigm: -a	1.08	0.12	2.00	96.54%	4.34%
Syntactic marking: Object	-0.41	-0.77	-0.06	97.02%	12.42%
L1: Balto-Slavic	0.33	-0.58	1.25	71.40%	21.80%
Test Time: Posttest x	2.20	1.36	3.06	99.98%	0%
Case: Genitive					
Test Time: Delayed posttest x Case: Genitive	0.80	-0.06	1.64	93.06%	9.34%
Test Time: Posttest x	-1.44	-2.38	-0.47	98.98%	0%
Case: Accusative					
Test Time: Delayed posttest x Case: Accusative	-1.27	-2.19	-0.32	98.39%	0.93%
Case: Genitive x	0.85	0.08	1.63	96.06%	6.00%
Treatment: TBLT					
Case: Accusative x	0.84	0.00	1.69	94.58%	8.32%
Treatment: TBLT					
Test Time: Posttest x	0.03	-0.98	1.07	51.51%	23.23%
Paradigm: -is/-ys					
Test Time: Delayed posttest x Paradigm: -is/-ys	0.28	-0.72	1.28	67.69%	21.50%
Test Time: Posttest x Paradigm: -è	0.46	-0.55	1.48	76.74%	18.05%
Test Time: Delayed posttest x Paradigm: -è	-0.09	-1.13	0.95	55.56%	23.43%
Test Time: Posttest x Paradigm: -a	-0.78	-2.09	0.56	82.67%	11.40%
Test Time: Delayed posttest x Paradigm: -a	0.38	-0.94	1.72	67.83%	16.72%
Paradigm: -is/-ys x Treatment: TBLT	-0.44	-1.36	0.48	77.86%	19.35%
Paradigm: -è x	0.29	-0.69	1.28	68.17%	21.73%

Parameter	Median	HDI (89%) Low	HDI (89%) High	<i>pd</i>	% in ROPE (±0.18)
Treatment: TBLT Paradigm: -a x Treatment: TBLT	-0.42	-1.54	0.72	72.55%	17.38%
Test Time: Posttest x Treatment: TBLT x Case: Genitive	-0.01	-1.09	1.08	50.34%	22.18%
Test Time: Delayed posttest x Treatment: TBLT x Case: Genitive	-0.12	-1.21	0.99	56.67%	21.61%
Test Time: Posttest x Treatment: TBLT x Case: Accusative	0.10	-1.14	1.33	55.12%	19.55%
Test Time: Delayed posttest x Treatment: TBLT x Case: Accusative	-0.68	-1.87	0.52	81.46%	13.22%
Test Time: Posttest x Treatment: TBLT x Paradigm: -is/-ys	0.76	-0.52	2.03	83.00%	11.93%
Test Time: Delayed posttest x Treatment: TBLT x Paradigm: -is/-ys	0.08	-1.21	1.33	53.97%	18.82%
Test Time: Posttest x Treatment: TBLT x Paradigm: -è	0.38	-0.91	1.67	68.15%	16.70%
Test Time: Delayed posttest x Treatment: TBLT x Paradigm: -è	0.59	-0.74	1.88	76.72%	13.91%
Test Time: Posttest x Treatment: TBLT x Paradigm: -a	-1.23	-2.82	0.32	89.66%	7.12%
Test Time: Delayed posttest x Treatment: TBLT x Paradigm: -a	-0.07	-1.67	1.56	52.95%	14.78%

Note. The reference levels for each variable were as follows: Test Time: Pretest; Treatment: TSLT; Case: Nominative; Paradigm: -as, Correctness: Correct (+); L1: non-Balto-Slavic; Syntactic Role: Subject.

HDI = Highest Density Interval; *pd* = Probability of Direction; ROPE = Region of Practical Equivalence.

Bayesian posterior distribution summary for participants' responses on the elicited oral imitation test, including the control group, as well as a three-way interaction between the parameters Test Time, Treatment Group, and Grammatical Correctness

Parameter	Median	HDI (89%) Low	HDI (89%) High	<i>pd</i>	% in ROPE (± 0.18)
Intercept	-0.16	-0.79	0.47	65.28	34.80
Test Time: Posttest	0.57	-0.16	1.34	89.34	15.14
Treatment: TSLT	-0.23	-0.87	0.41	71.17	31.55
Treatment: TBLT	-0.14	-0.77	0.49	64.05	35.24
Correctness: Incorrect (-)	-4.82	-5.82	-3.78	100.00	0.00
Syntactic Role: Object	-0.08	-0.73	0.58	58.08	36.68
L1: Balto-Slavic	0.29	-0.13	0.71	86.48	31.79
Test Time: Posttest x Treatment: TSLT	0.25	-0.45	0.95	71.24	28.23
Test Time: Test Time x Treatment: TBLT	0.93	0.23	1.64	98.11	2.08
Test Time: Posttest x Correctness: Incorrect (-)	0.93	-0.49	2.24	85.79	9.16
Treatment: TSLT x Correctness: Incorrect (-)	-1.03	-2.05	-0.02	94.99	6.52
Treatment: TBLT x Correctness: Incorrect (-)	-1.25	-2.27	-0.26	97.85	1.74
Test Time: Posttest x Treatment: TSLT x Correctness: Incorrect (-)	-0.07	-1.33	1.19	53.25	19.07
Test Time: Posttest x Treatment: TBLT x Correctness: Incorrect (-)	-0.33	-1.58	0.94	66.09	17.07

Note. The reference levels for each variable were as follows: Test Time: Pretest; Treatment: TSLT; Correctness: Correct (+); L1: non-Balto-Slavic; Syntactic Role: Subject.

HDI = Highest Density Interval; *pd* = Probability of Direction; ROPE = Region of Practical Equivalence.

Bayesian posterior distribution summary for participants' responses on the elicited oral imitation test of the two treatment groups, including a three-way interaction between the parameters Test Time, Treatment Group, and Grammatical Correctness

Parameter	Median	HDI (89%) Low	HDI (89%) High	<i>pd</i>	% in ROPE (± 0.18)
Intercept	-0.40	-0.91	0.12	89.23	22.20
Test Time: Posttest	0.95	0.33	1.57	99.29	0.04
Test Time: Delayed posttest	1.10	0.50	1.72	99.76	0.00
Treatment: TBLT	0.06	-0.66	0.79	55.59	32.07
Correctness: Incorrect (-)	-6.16	-7.04	-5.32	100.00	0.00
Syntactic Role: Object	-0.64	-1.13	-0.14	98.03	4.67
L1: Balto-Slavic	0.28	-0.28	0.85	78.16	31.26
Test Time: Posttest x TBLT	0.63	-0.10	1.35	91.96	12.82
Test Time: Delayed posttest x Treatment: TBLT	0.84	0.11	1.57	96.77	5.16
Test Time: Posttest x Correctness: Incorrect (-)	1.02	-0.17	2.12	91.58	7.86
Delayed posttest x Correctness: Incorrect (-)	0.91	-0.28	2.05	89.46	9.26
Treatment: TBLT x Correctness: Incorrect (-)	-0.40	-1.40	0.65	73.66	19.26
Test Time: Posttest x Treatment: TBLT x Correctness: Incorrect (-)	-0.20	-1.47	1.06	59.66	18.18
Test Time: Delayed Posttest x Treatment: TBLT x Correctness: Incorrect (-)	0.36	-0.93	1.64	67.24	17.18

Note. The reference levels for each variable were as follows: Test Time: Pretest; Treatment: TSLT; L1: non-Balto-Slavic; Syntactic Role: Subject.

HDI = Highest Density Interval; *pd* = Probability of Direction; ROPE = Region of Practical Equivalence.

Bayesian posterior distribution summary for participants' responses on the elicited oral imitation test of the two treatment groups, including a three-way interaction between the parameters Test Time, Treatment Group, and Syntactic Role

Parameter	Median	HDI (89%) Low	HDI (89%) High	<i>pd</i>	% in ROPE (± 0.18)
Intercept	-0.29	-0.81	0.22	82.27	30.69
Test Time: Posttest	0.85	0.21	1.48	98.14	2.46
Test Time: Delayed posttest	1.03	0.40	1.68	99.43	0.00
Treatment: TBLT	0.16	-0.52	0.84	64.76	32.62
Correctness: Incorrect (-)	-0.71	-1.55	0.10	92.01	11.34
L1: Balto-Slavic	-5.54	-6.13	-4.99	100.00	0.00
Syntactic Role: Object	0.27	-0.29	0.84	78.44	31.53
Test Time: Posttest x TBLT	0.50	-0.17	1.16	88.54	17.89
Test Time: Delayed posttest x Treatment: TBLT	0.75	0.09	1.42	96.55	6.07
Test Time: Posttest x Syntactic Role: Object	0.45	-0.70	1.63	73.42	17.03
Delayed posttest x Syntactic Role: Object	-0.29	-1.42	0.87	65.92	19.55
Treatment: TBLT x Syntactic role: Object	-0.62	-1.51	0.29	86.11	14.93
Test Time: Posttest x Treatment: TBLT x Syntactic Role: Object	1.09	-0.12	2.27	92.51	7.30
Test Time: Delayed Posttest x Treatment: TBLT x Syntactic Role: Object	0.90	-0.31	2.10	88.31	9.84

Note. The reference levels for each variable were as follows: Test Time: Pretest; Treatment: TSLT; L1: non-Balto-Slavic; Syntactic Role: Subject.

HDI = Highest Density Interval; *pd* = Probability of Direction; ROPE = Region of Practical Equivalence.

Appendix G – Caterpillar plots

The caterpillar plots used for checking potential negative item outliers of the gap-fill test and elicited oral imitation test are stored in this dissertation's project on the OSF (Open Science Framework) website. Access to these files is granted through the following link:

https://osf.io/wfpe8/overview?view_only=7a3501c9c43341abbb54b5363741c3c4

The task materials can be found in the folder titled 'Caterpillar plots.'

The caterpillar plots show probability scales for each item divided over Treatment and Test Time. Each dot represents the estimated probability of a correct response for one item, with horizontal intervals indicating uncertainty. Dots further to the left indicate more difficult items, whereas dots further to the right indicate easier items. The items are ordered from most challenging to least challenging for each test. Numbers indicate item IDs. Missing numbers represent filler items.

SANTRAUKA

Šioje disertacijoje siekiama ištirti užduotimis grindžiamo kalbos mokymo (UGKM) veiksmingumą mokant linksnio gramatinės kategorijos lietuvių kaip svetimšios kalbos (K2) mokymo(si) kontekste. Nors UGKM pripažįstamas veiksminga K2 mokymo metodika (R. Ellis ir kt., 2019), neretai laikomasi nuostatos, kad ši metodika nėra tinkama sudėtingos morfologijos K2 mokymui (Gilabert ir Castellví, 2019; Markina, 2018). Manoma, kad tokių kalbų sudėtingos gramatinės kategorijos, tarp jų – ir linksnių sistema, reikalauja intensyvaus į formą orientuoto mokymo. Todėl tyrėjai linkę manyti, kad komunikacinio pobūdžio UGKM metodika, kurioje gramatikos mokymui numatyta pagalbinė funkcija, negali užtikrinti pakankamai gerų mokymosi rezultatų. Vis dėlto, kadangi UGKM taikymo sudėtingos morfologijos K2 mokymui galimybės iki šiol tirtos ribotai, šia disertacija siekiama kritiškai įvertinti ir patikrinti minėtas nuostatas.

Tyrimai K2 mokymo(si) srityje vis dažniau rodo, kad UGKM gali pasirodyti veiksmingesnis nei tradiciniai mokymo metodai (R. Ellis et al., 2019). Šis metodas pabrėžia K2 vartojimą autentiškose, realaus gyvenimo komunikacinius kontekstus imituojančiose užduotyse – nuo kavos užsakymo kavinėje iki mokesčių deklaracijos pateikimo. Tokios realaus gyvenimo užduotys pritaikomos ugdymo reikmėms ir transformuojamos į vadinamąsias pedagogines užduotis (angl. *pedagogical tasks*; Long, 2014). Nors mokslinėje literatūroje pateikiama įvairių užduoties apibrėžčių, viena iš plačiausiai pripažįstamų yra R. Ellis ir kt. (2019, p. 10) pateikta apibrėžtis. Joje pabrėžiama, kad UGKM užduotys turėtų būti orientuotos į raišką, numatyti tam tikrą informacijos spragą ar trūkumą, kurie skatina besimokančiuosius sąveikauti tarpusavyje arba prasmingai apdoroti mokomąją medžiagą; užduotys taip pat turi sudaryti sąlygas besimokantiejiems remtis savo lingvistiniais ir nelingvistiniais komunikaciniais ištekliais bei siekti aiškaus nelingvistinio komunikacinio rezultato. UGKM užduotys paprastai turi trijų etapų struktūrą (R. Ellis, 2003; J. Willis, 1996). Pirmasis etapas – *pasirengimo* etapas (angl. *pre-task*), kurio metu pristatoma užduotis ir (arba) jos tema, pateikiama aktuali K2 įvestis ir atliekamos su ja susijusios veiklos. Vėliau vykdomas pagrindinis užduoties *atlikimo* etapas (angl. *main task*), prieš kurį gali būti numatytas planavimo etapas. Galiausiai pereinama prie *apibendrinimo* etapo (angl. *post-task*), kuriame aptariami užduoties rezultatai ir, o tai šiam tyrimui ypač aktualu, analizuojami su užduotimi susiję gramatiniai aspektai.

Šie keturi užduočių kriterijai ir trijų etapų užduoties struktūra buvo sukurti siekiant pagrindinio UGKM tikslo – skatinti implicitinį K2 įsisavinimą

(R. Ellis ir kt., 2019). Implicitinės žinios suprantamos kaip žinios, kuriomis remiamasi intuityviai ir kurias dažnai sunku paaiškinti, pavyzdžiui, tokios yra gimtosios kalbos (K1) žinios, kuriomis operuoja gimtakalbiai. Kitame žinių įsisąmoninimo skalės gale yra eksplikitinės žinios – sąmoningai išmoktos žinios, kurias kalbos vartotojas gali paaiškinti, tačiau ne visada geba spontaniškai pritaikyti kalbinėje veikloje, pvz., K2 gramatikos taisyklės. Tradiciniais kalbų mokymo metodais dažniausiai siekiama formuoti eksplikitines K2 žinias (R. Ellis ir kt., 2019; Long, 2014). Tokie sintetiniai mokymo metodai tikslinę K2 suskaido į atskirus vienetus, pavyzdžiui, tikslinius leksinius vienetus (pavienius žodžius ar kolokacijas) ar gramatinės taisyklės, juos išdėsto tam tikra seka ir pateikia besimokantiejiems nuosekliai, po vieną. Tada besimokančiojo užduotis yra integruoti šiuos pavienius kalbos elementus siekiant veiksmingai komunikuoti tikslinę K2. Long (2014) šį mokymo metodą apibūdina kaip dėmesį formoms (angl. *focus on forms*). Vis dėlto tokie skirti formų mokyti metodai, pirmiausia skatinantys eksplikitinių K2 žinių įgijimą, sulaukė reikšmingos kritikos. Reaguojant į šią kritiką, ilgainiui susiformavo forma į antrą planą nustūmusi komunikacinę K2 mokymo kryptis ir UGKM metodika, orientuota ne tiek į eksplikitinių, kiek į implicitinių K2 žinių ugdymą (R. Ellis ir kt., 2019).

Long (2014) teigimu, būtent implicitinis, o ne eksplikitinis mokymasis yra pagrindinis suaugusiųjų K2 išmokimo mechanizmas. Todėl K2 mokymui(si) tinkamesni yra analitiniai mokymo metodai, skatinantys besimokančiuosius aktyviai įsitraukti į K2 pateiktą užduoties įvestį ir sąveikauti su kitais besimokančiais. Šie metodai vadinami analitiniais, nes neskaido tikslinės K2 į iš anksto nustatytus mokymo vienetus, o sudaro sąlygas besimokantiejiems analizuoti K2 įvestį ir savo kalbinę raišką bei indukcinio būdu generuoti kalbos struktūras ir dėsningumus. Vis dėlto išskirtinis dėmesio sutelkimas į raišką taip pat kelia tam tikrų sunkumų.

Anot Long (2014), nors implicitinis mokymasis yra pagrindinis suaugusiųjų mokymo(si) būdas, suaugusiųjų gebėjimas implicitiniu būdu įsisavinti K2 yra ribotas. Atsižvelgdamas į tai, Long (2014) siūlo kitą analitinį mokymo būdą – dėmesį formai (angl. *focus on form*). Dėmesys formai apibrėžiamas kaip trumpi į formą orientuoti epizodai, skirti komunikaciniams sunkumams, su kuriais besimokantieji susiduria prasminės sąveikos metu, įveikti. Priešingai nei dėmesys formoms, kai gramatinės struktūros pateikiamos dar prieš prasidedant besimokančiųjų sąveikai pagal numatytą mokymo seką, dėmesys formai aktualizuoja konkrečias gramatinės konstrukcijas kaip reakciją į komunikacijos metu iškilusias problemas. Vis dėlto šį metodinį principą galima papildyti didesniu dėmesiu formoms

užduoties *apibendrinimo* etape, t. y. aiškiai struktūruotomis į formą orientuotomis veiklomis. Toks sprendimas leidžia besimokantiesiems *pasirengimo* ir pagrindinės užduoties *atlikimo* etapuose išlaikyti dėmesį komunikacijos tikslą turinčiai sąveikai (Skehan, 1998; D. Willis ir Willis, 2007), o vėlesnės, sutelkiančios dėmesį į formą veiklos padeda spręsti užduoties atlikimo metu išryškėjusias ar užfiksuotas kalbos raiškos ir formos problemas (Long, 2016). Šis į turinį dėmesį kreipiančios sąveikos ir ją papildančio eksplisitinio dėmesio kalbos formai derinys sudaro kognityvinio-interakcionistinio modelio (angl. *cognitive-interactionist model*), kuriuo grindžiamas UGKM, pagrindą (R. Ellis ir kt., 2019; Long, 2014). K2 mokymo prieigų tyrimai vis dažniau patvirtina, kad UGKM gali būti veiksmingesnis už tradicinius kalbų mokymo metodus (R. Ellis ir kt., 2019).

Vis dėlto sudėtingos morfologijos kalbos, tokios kaip rusų, suomių ir lietuvių, UGKM šalininkams tebėra rimtas iššūkis. Morfosintaksinių konstrukcijų įsisavinimas dažnai laikomas vienu iš sudėtingiausių K2 mokymosi aspektų (Slabakova, 2019). Dėl šių priežasčių daugelis sudėtingos morfologijos K2 mokytojų pirmenybę teikia tradiciškesniems, eksplisitiniu mokymu grindžiamiems kalbos formų mokymo metodams (Comer, 2012; Gilabert ir Castellví, 2019; Markina, 2018). Be to, vyrauja nuostata, kad UGKM gali būti netinkamas šių kalbų mokymui, ypač mokant žemesniųjų mokėjimo lygių pradedančiuosius besimokančiuosius (Kogan ir Bondarenko, 2022).

Daugiausia sunkumų kitakalbiam mokiniams kelia gramatinė linksnio kategorija. Dėl savo sudėtingumo ir morfologinio turtingumo ji vienas iš didžiausių iššūkių daugelio kalbų, tarp jų – lietuvių kalbos, besimokantiesiems (Savickienė, 2006). Sunkumus lemia ne tik didelė morfologinių formų įvairovė, bet ir ribotas jų ryškumas (angl. *salience*), žemas sąlygiškumo laipsnis (angl. *contingency*), t. y. nenuosekli ir silpna formos ir reikšmės sąsaja, bei galimas pertekliškumas (angl. *redundancy*) iš besimokančiųjų K1 perspektyvos (N. C. Ellis, 2022). Dėl šių priežasčių daugelis sudėtingos morfologijos K2 mokytojų abejoja komunikacinių metodų, tokių kaip UGKM, veiksmingumu ir jų pagrindiniu tikslu – implicitiniu būdu ugdyti komunikacinę kompetenciją.

Tokios abejonės bent iš dalies atrodo pagrįstos. Tyrėjai (N. C. Ellis, 2006a, 2009; Frigo ir McDonald, 1998) sutaria, kad implicitinis mokymasis yra veiksmingiausias perprantant gretimąsias priklausomybes (angl. *adjacent dependencies*), dažnai pasikartojančius žodžių junginius (angl. *frequent chunks*) ir vadinamąsias neapibrėžtąsias taisykles (angl. *fuzzy rules*), bet tik jeigu mokymas grindžiamas gausia K2 kalbinės medžiagos įvestimi. Dėl šios

priežasties sudėtingos morfologijos K2 mokytojai linkę teikti pirmenybę tradiciniams, eksplcitiniams kalbos formų mokymo metodams (Comer, 2012; Gilabert ir Castellví, 2019; Markina, 2018), nepaisydami šių metodų ribotumų (Long, 2014). Atsižvelgiant į tai, kad besimokantieji, apdorodami K2 įvestį, linkę daugiau dėmesio skirti turinio požiūriu svarbiems, o ne funkciniais kalbos elementams (VanPatten, 2014), toks pasirinkimas gali būti laikomas pagrįstu.

Vis dėlto kai kurie tyrimai kelia abejonių dėl eksplcitinių mokymo metodų veiksmingumo. Remiantis Pienemanno kalbos apdorojimo teorija (angl. Processability Theory; 1998), intensyvus eksplcitinis gramatikos mokymas negali svariai nulemti sudėtingos morfologijos K2 mokymo(si) rezultatų, nes visi K2 besimokantieji pereina tuos pačius raidos etapus. Manoma, kad besimokantieji negali pereiti į aukštesnį kalbos įsisavinimo os etapą, kol nėra tam pasirengę, nepriklausomai nuo gramatikos mokymo(si) intensyvumo. Teigiama, kad tas pats principas taikytinas ir tipologiškai panašios K1 poveikiui (Saturno, 2023). Be to, ankstesni tyrimai (Kempe ir Brooks, 2008; Kempe ir MacWhinney, 1998) patvirtina, kad besimokantieji, perprasdami linksnių sistemą, dažniau remiasi ne taisyklėmis grįstomis, o kalbiniais vienetais grįstomis strategijomis (angl. *item-based strategies*). Tai leidžia daryti prielaidą, kad gausia K2 įvestimi grindžiami metodai, tokie kaip UGKM, gali būti palankesni linksnių kategorijai įsisavinti. Todėl išlieka neaišku, ar eksplcitiniai mokymo metodai iš tiesų yra veiksmingesni mokant morfologiškai sudėtingų K2.

Mokslo literatūroje pateikiama keletas pasiūlymų, kaip pritaikyti UGKM metodiką sudėtingos morfologijos K2 mokymui (Gilabert ir Castellví, 2019; Kogan ir Bondarenko, 2022; Udes, 2025). Pavyzdžiui, Gilabert ir Castellví (2019) teigia, kad mokant sudėtingos morfologijos K2 dėmesys formai galėtų būti integruojamas į visus užduoties etapus, o ne paliekamas vien tik *apibendrinimo* etapui. Vis dėlto didesnis dėmesys gramatikai visos užduočių sekos metu iš dalies prieštarauja pagrindiniam UGKM tikslui – skatinti implicitinį mokymąsi. Šiame kontekste Ellis (2009b) aprašo vadinamąsias fokusuotas užduotis (angl. *focused tasks*). Tai pedagoginės užduotys, atitinkančios visus anksčiau aptartus UGKM kriterijus, tačiau kryptingai orientuotos į konkrečių gramatinių konstrukcijų perpratimą ir apimančios išankstinį dėmesį gramatikai *pasirengimo* etape. Mokymo programos, grindžiamos tokiomis fokusuotomis užduotimis, dažnai priskiriamos ne UGKM, o užduotimis paremtam kalbos mokymui (UPKM; angl. *Task-Supported Language Teaching*).

Tokie alternatyvūs UGKM metodikos taikymo būdai išlieka iki galo neaiškūs dėl to, kad nežinoma, kokio pobūdžio K2 žinias įgyti jie skatina. Pagrindinis klausimas – ar didesnis eksplikitinis dėmesys gramatikai ankstesniuose užduoties etapuose, būdingesnis UPKM metodui, neslopina implicitinių žinių įsisavinimo. UPKM metodas gana glaudžiai siejasi su įgūdžių įgijimo teorija (angl. *Skill Acquisition Theory*; DeKeyser, 2014), pagal kurią deklaratyvosios žinios, t. y. eksplikitinės žinios apie tam tikrą kalbos konstrukciją ar taisyklę, gali būti procedūrizuojamos ir iš dalies automatizuojamos, taip įgydamos implicitinėms žinioms būdingų savybių. Kita vertus, kyla klausimas, ar tradiciškesnis UGKM metodas pirmiausia skatina implicitinių žinių įgijimą, ar kartu stiprina ir eksplikitines K2 žinias. Kadangi UGKM tradiciškai apima dėmesio formai technikas (angl. *focus-on-form techniques*) ir dėmesį formoms užduoties *apibendrinimo* etape, tai leidžia daryti prielaidą, kad tradicinis UGKM metodas gali būti pakankamai veiksmingas perprasti net ir santykinai sudėtingas gramatinės kategorijas.

Nepaisant šių abejonių, empirinių tyrimų, nagrinėjančių UGKM taikymą sudėtingos morfologijos K2 mokymui, vis dar publikuota nedaug. Šia disertacija siekiama užpildyti šią spragą, tiriant lietuvių K2 linksnių sistemos mokymą(si) taikant su UGKM ir UPKM metodikomis suderintas užduotis. Šiame kontekste siekiama atsakyti į du pagrindinius **klausimus**: ar UGKM gali palengvinti sudėtingos gramatinės kategorijos, tokios kaip linksnio kategorija, įsisavinimą ir ar mokymosi rezultatai skiriasi priklausomai nuo taikomos metodikos. Be to, siekiama įvertinti ne tik trumpalaikius, bet ir ilgalaikius mokymosi rezultatus. Markina (2018) nustatė, kad rusų kaip K2 besimokantieji, kuriems buvo taikytas UGKM metodas, demonstravo reikšmingą trumpalaikę pažangą įsisavinant linksnių sistemą. Vis dėlto atidėtojo baigiamojo testo metu dalyvių rezultatai grįžo į pradinį lygį. Denissenko (2016) tyrimas parodė, kad 150 akademinių valandų trukusi PPP intervencija nesukėlė reikšmingos pažangos perprantant linksnių sistemą keturiuose kalbos mokėjimo lygiuose (A1, A2, B1.1 ir B1.2). Šie rezultatai leidžia daryti prielaidą, kad vien formalaus gramatikos mokymo gali būti nepakankama sėkmingai įsisavinti linksnio kategoriją. Autoriaus žiniomis, intervencinių tyrimų, skirtų būtent linksnių sistemai išmokti, yra nedaug. Lietuvių kalbos kaip K2 kontekste tokių tyrimų iki šiol neatlikta.

Be bendrų mokymosi rezultatų, disertacijoje nagrinėjami kokybiniai K2 linksnių sistemos įsisavinimo skirtumai. Denissenko (2016) ir Cherepovskaia ir kt. (2022) parodė, kad linksnių sistemos klaidų tipų analizė gali atskleisti skirtingų besimokančiųjų grupių įsisavinimo tendencijų skirtumus net ir tais atvejais, kai bendras klaidų skaičius išlieka nepakitęs.

Dabašinskienė ir Čubajevaitė (2009) nustatė, kad per dažnas vardininko vartojimas, tam tikrų netiesioginių linksnų painiojimas ir daiktavardžio bei linksniaavimo paradigmos neatitiktys yra būdingos lietuvių K2 besimokančiųjų klaidos. Todėl daroma prielaida, kad klaidų tipų analizė gali padėti paaiškinti lietuvių K2 linksnų sistemos įsisavinimo raidos skirtumus, todėl ji įtraukta į šį tyrimą.

Taigi pagrindinis šios disertacijos **tikslas** – įvertinti lietuvių K2 kurso dalyvių mokymosi pažangą, kuri matuojama dviem testais, skirtais linksnų sistemos žinioms tikrinti, ir testų rezultatais po dviejų skirtingų UGKM principais grindžiamų mokymo intervencijų. Šios intervencijos skiriasi eksplcitinio gramatikos mokymo integravimu: UGKM atveju eksplcitinis gramatikos mokymas įtraukiamas tik užduoties *apibendrinimo* etape, o UPKM atveju dėmesys gramatikai nuosekliai integruojamas į visą užduoties ciklą.

Siekiant šio tikslo, suformuluoti šie **uždaviniai**:

1. Apžvelgti mokslinę literatūrą ir aprašyti pagrindinius UGKM metodikos principus, ypatingą dėmesį skiriant eksplcitiniam gramatikos mokymui, taip pat užduočių kūrimo principams ir priemonėms, susijusioms su K2 morfologinės sistemos mokymu.
2. Apžvelgti naujausius linksnio kategorijos sistemos išmokymo K2 kontekste tyrimus, siekiant nustatyti esamą žinių lygį šioje srityje.
3. Parengti UGKM principais pagrįstą lietuvių K2 mokymo intervenciją – dviejų versijų kursą, kuris skirtųsi eksplcitinio gramatikos mokymo vieta ir apimtimi.
4. Sukurti testavimo priemonės, skirtas tiek eksplcitinėms, tiek implicitinėms lietuvių K2 linksnio kategorijos žinioms vertinti.
5. Surengti lietuvių K2 kursą dviem dalyvių grupėms ir organizuoti kursų dalyvių žinių tikrinimą (testavimą) tris kartus iš anksto numatytais laiko momentais, siekiant įvertinti dalyvių mokymosi pažangą.
6. Surinkti ir koduoti duomenis bei atlikti jų statistinę analizę.
7. Aprašyti tyrimo rezultatus, prireikus patikslinti hipotezes ir suformuluoti tyrimo išvadas, aptarti jų reikšmę nagrinėjamai mokslo sričiai.

Šioje disertacijoje taikomi daugiausia kiekybiniai tyrimo **metodai**. Tyrimas grindžiamas empiriniais eksperimento duomenimis, surinktais taikant mišrų eksperimentinį planą, apimantį tiek tarp subjektinį (angl. *between-groups*), tiek vidus subjektinį (angl. *within-groups*) dizainą.

Pagrindinis tyrimo etapas, apėmęs testų rengimą ir vykdymą bei kursų vedimą eksperimentinėse grupėse, vyko nuo 2024 m. lapkričio 4 d. iki 2025 m. vasario 4 dienos. Papildomas kontrolinės grupės testavimas buvo vykdomas nuo 2025 m. gegužės iki rugpjūčio mėn. Tyrimo dalyviai buvo su

Vilniaus universitetu nesusiję suaugę lietuvių K2 besimokantieji, užsiregistravę į kursą užpildę paraiškos formą, paskelbtą Vilniaus universiteto Lituanistinių studijų katedros interneto svetainėje.

Nors dalyvių lietuvių K2 mokėjimo lygis nebuvo formaliai patikrintas prieš tyrimą, paraiškos formoje buvo nurodyta, kad kursas skirtas A1–A2 kalbos mokėjimo lygio besimokantiems pagal „Bendruosius Europos kalbų metmenis“ (Europos Taryba, 2020). Dalyvių pasiekimai kurso metu patvirtino, kad iki kursų pradžios jie jau buvo įgiję lietuvių kalbos žinių, jų žinios ir kompetencijos atitiko numatytąjį kurso lygį. A1–A2 lygio pasirinkimą lėmė kelios priežastys:

- Kadangi šio lygio besimokantieji jau yra įgiję pagrindinių lietuvių kalbos žinių, tai suteikė galimybes lanksčiau kurti užduotis, skirtas žodinei sąveikai.
- Manoma, kad linksnių skirtys (angl. *case opposition*) K2 įsisavinimo procese formuojasi palaipsniui ir pereina kelias raidos stadijas, kol bazinė vardininko forma besimokančiojo mentaliniame leksikone pradeda sistemingai kontrastuoti su kitais, netiesioginiais linksniais (Baten, 2013; Cherepovskaia ir kt., 2022; Magnani ir Artoni, 2015). Todėl į tyrimą buvo įtraukti dalyviai, jau turintys lietuvių K2 žinių, tačiau dar nepasiekę tokio mokėjimo lygio, kuriame linksnių skirtys būtų visiškai susiformavusios, ypač implicitinių žinių lygmeniu.
- Šioje disertacijoje aiškiai skiriamos eksplicitinės ir implicitinės K2 žinios. Kadangi besimokantieji, pradedantys mokytis nuo pradinio lygio, neturi jokių implicitinių lietuvių K2 kalbos žinių, į tyrimą buvo siekiama įtraukti jau šiek tiek kalbos mokančius dalyvius. A1–A2 lygiai geriausiai atitiko tyrimui reikalingą kalbos žinių apimtį.

Dalyviai atsitiktinės atrankos būdu buvo paskirstyti į UGKM (Užduotimis grindžiamas kalbos mokymas) intervencijos grupę, kurioje eksplicitinis gramatikos mokymas buvo taikomas tik užduoties *apibendrinimo* etape, ir į UPKM (Užduotimis paremtas kalbos mokymas) intervencijos grupę, kurioje dėmesys gramatikai buvo integruojamas į visą užduoties ciklą. Sklandžiam mokymo proceso organizavimui užtikrinti abi intervencinės grupės papildomai suskirstytos į keturias mažesnes grupes. Kiekvienos grupės mokymo intervencijos trukmė, sutampanti su kurso trukme, buvo 18 valandų. Visose grupėse mokytojo pareigas atliko šios disertacijos autorius.

Tyrimė naudota lietuvių K2 mokymo medžiaga buvo parengta specialiai šiam tyrimui. Užduotys pateikiamos šios disertacijos prieduose (*Appendix C*). Ją sudarė penkios pedagoginės užduotys, kurios buvo parinktos atsižvelgiant į jų komunikacinę aktualumą asmenims, neseniai atvykusiems

gyventi į Lietuvą. Jos taip pat buvo parinktos atsižvelgiant į jų potencialą skatinti šiame tyrime tiriamų tikslinių gramatinių konstrukcijų vartojimą. Tikslinės konstrukcijos apima veiksnio žymėjimą teigiamuosiuose (vardininko) ir neigiamuosiuose (kilmininko) kontekstuose bei tiesioginio objekto žymėjimą teigiamuosiuose (galininko) ir neigiamuosiuose (kilmininko) kontekstuose.

Parengtos šios užduotys:

1. Išreikšti savo nuomonę apie miesto rajonus.
2. Palyginti butus ir išsirinkti tinkamiausią.
3. Aptarti, kokių baldų trūksta, ir susitarti, ką reikia nusipirkti.
4. Suplanuoti meniu vakarėliui.
5. Apsipirkti vakarėliui.

Visos pagrindinės užduotys buvo atliekamos poromis. Pirmosios užduoties metu kiekvienas dalyvis gavo supaprastintą vieno Vilniaus rajono žemėlapi, apie kurį turėjo pateikti informaciją savo pašnekovui (t. y. kitam dalyviui) ir paskui paklausinėti jį apie kitą rajoną, kurio žemėlapi turėjo pašnekovas. Užduoties pabaigoje dalyviai išreiškė savo nuomonę apie abu rajonus. Antrosios užduoties metu kiekvienas poros dalyvis gavo informaciją apie skirtingus nuomojamus butus. Dalyviai keitėsi informacija ir kartu rinkosi tinkamiausią variantą. Trečiojoje užduotyje dalyviai turėjo aptarti, kaip įrengti svetainę, tačiau kiekvienas lankėsi skirtingoje baldų parduotuvėje. Jie keitėsi savo turima informacija apie parduotuvėse matytus baldus ir turėjo susitarti dėl keturių baldų, kuriuos pasirinktų svetainės interjerui. Ketvirtojoje užduotyje dalyviai keitėsi informacija apie planuojamų pakviesti svečių maisto pasirinkimus ir kartu sprendė, koks turėtų būti vakarėlio meniu. Penktojoje užduotyje kiekvienas dalyvis gavo skirtingo prekybos tinklo kainų leidinį ir, keisdamiesi informacija, turėjo sudaryti vakarėliui reikalingų maisto produktų pirkinių sąrašą.

Kiekvienos užduoties pasirengimo etapą sudarė:

- temos pristatymas;
- skaitymo tekstas su skirtingomis veiklomis;
- klausymo tekstas su skirtingomis veiklomis;
- parengiamosios užduotys (angl. *subtasks*), t. y. smulkesnės ir lengviau įgyvendinamos veiklos, skirtos pasirengti pagrindinei užduočiai.

Užduoties *apibendrinimo* etapas visada apėmė naudingos leksikos įtvirtinimo veiklas ir, svarbiausia, su tikslu užduotimi susijusių gramatikos pratimų rinkinį.

Kiekvienai užduočiai buvo parengtos dvi versijos, kurios skyrėsi eksPLICITINIO gramatikos mokymo vieta ir apimtimi. Pirmiausia gramatiniai

elementai mokomojoje įvestyje buvo tipografiškai paryškinti, daugiausia linksnių galūnės. Linksnių galūnės buvo pabraukiamos, o skirtingiems linksniams žymėti naudotos skirtingos spalvos. Kitas svarbus skirtumas tarp UGKM ir UPKM užduočių versijų buvo tam tikrų komunikacinių veiklų pakeitimas eksPLICITINėmis į formą orientuotomis veiklomis. Klausymo dalyje, užuot pildę trūkstamus žodžius ar potencialiai naudingas frazes, dalyviai turėjo įrašyti trūkstamas gramatines formas, dažniausiai linksnių galūnes. Panašiai skaitymo dalyje komunikacinės veiklos buvo pakeistos linksnių galūnių lentelėmis. Dalyviai turėjo užpildyti šias lenteles, remdamiesi skaitymo tekste pateiktais žodžiais ir jų galūnėmis.

Tyrimo priemonės buvo sukurtos siekiant įvertinti dalyvių eksPLICITINES ir IMPLICITINES lietuvių K2 linksnių sistemos žinias. Šių testų tikslinės gramatinės konstrukcijos buvo veiksnio žymėjimas teigiamuose ir neigiamuose kontekstuose bei tiesioginio objekto žymėjimas teigiamuose ir neigiamuose kontekstuose. Siekiant įvertinti dalyvių eksPLICITINES lietuvių K2 linksnių sistemos žinias, buvo parengtas žodžių įrašymo testas (angl. *gap-fill test*). Dalyviams buvo pateikti sakiniai su tarpais, kuriuose jie turėjo įrašyti skliaustuose pateiktą žodį parinkdami jam tinkamą linksnio galūnę. Prie kiekvienos praleistos vietos skliaustuose buvo pateikiamas tikslinis žodis jo bazinė forma, t. y. vardininko vienaskaitos forma, pavyzdžiui:

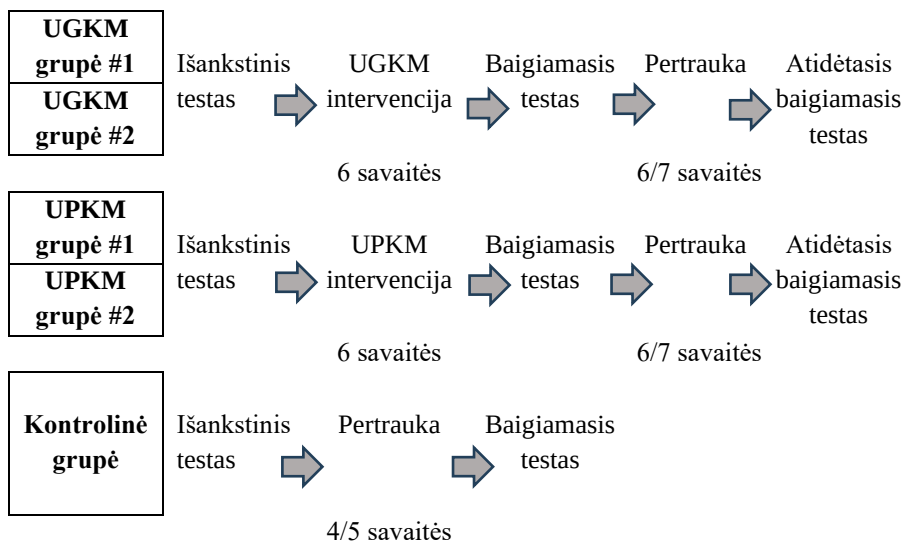
(36) *Nežinau, kiek mokėti, nes nėra _____ (sąskaita).*

Siekiant įvertinti dalyvių IMPLICITINES lietuvių K2 linksnių sistemos žinias, buvo pasirinktas paskatinto žodinio pakartojimo testas (angl. *elicited oral imitation test*). Dalyviai klausėsi įrašytų teiginių, įvertino jų turinį pasirinkdami vieną iš atsakymų variantų („teisingai“, „neteisingai“, „nesuprantu“, „nežinau“) ir paskui turėjo garsiai pakartoti išgirstą sakinį. Šio testo taikymas grindžiamas prielaida, kad dalyviai, sutelkdami dėmesį į teiginio turinį, ypač esant laiko apribojimams, sąmoningai nekreipia dėmesio į jo gramatinę raišką. Todėl kartodami teiginius jie remiasi savo turimomis IMPLICITINėmis K2 žiniomis. Pažymėtina, kad dalis teiginių buvo sąmoningai suformuluoti su gramatinėmis klaidomis. Daroma prielaida, kad tuo atveju, jei dalyvių IMPLICITINės konkrečios gramatinės kategorijos žinios yra pakankamai tvirtos, jie spontaniškai ištaisys išgirstą klaidą ir pakartodami teiginį jį pasakys teisingai. Toliau pateikiami gramatiškai taisyklingo ir netaisyklingo teiginio pavyzdžiai:

(37) *Čia ant stalo yra mano telefonas.*

(38) **Čia ant stalo nėra sūris.*

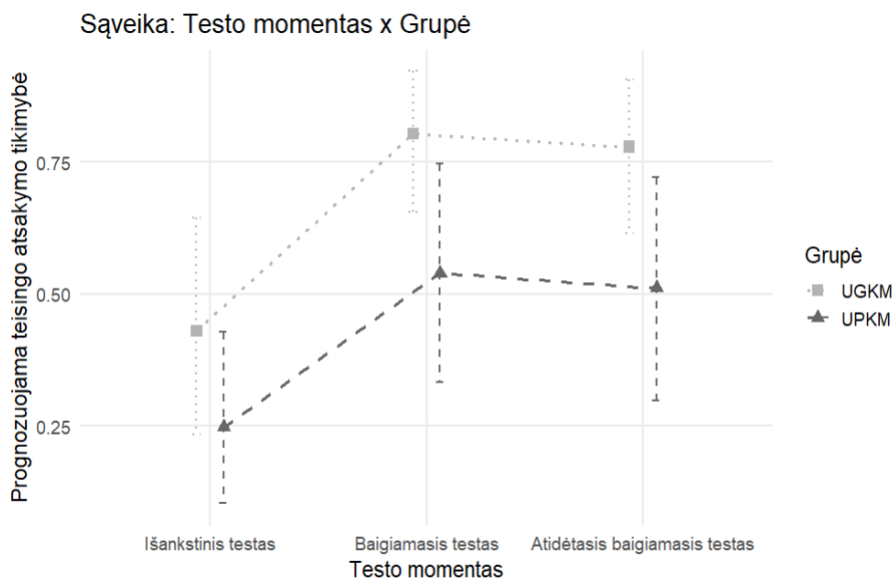
Dalyviai abu testus atliko tris kartus: prieš intervenciją (išankstinis testas, angl. *pretest*), siekiant nustatyti pradinį žinių lygį; iškart po intervencijos (baigiamasis testas, angl. *immediate posttest*), trumpalaikiams (angl. *immediate*) mokymosi rezultatams įvertinti; ir praėjus 1,5 mėnesio po intervencijos (atidėtas baigiamasis testas, angl. *delayed posttest*), siekiant įvertinti ilgalaikius mokymosi rezultatus. Didžiąją dalį testavimo organizavo ir vykdė disertacijos autorius, tačiau implicitines žinias vertinantys testai buvo administruojami padedant Vilniaus universiteto kolegoms. Tais atvejais, kai dalyviai negalėjo dalyvauti grupinėse testavimo sesijose, testai buvo atliekami individualiai – kontaktiniu arba nuotoliniu būdu. Kontrolinės grupės testavimas taip pat buvo vykdomas individualiai. Dalyviai testus atliko du kartus, darant 4–5 savaitių pertrauką, kontaktiniu arba nuotoliniu būdu, dalyvaujant disertacijos autoriui. 15 paveiksle pateikiama bendra tyrimo dizaino schema.



15 paveikslas. Šios disertacijos tyrimo dizaino schema

Dalyvių atsakymai į testų klausimus buvo koduoti ir suvesti į „Microsoft Excel“ programą. Statistinė duomenų analizė atlikta naudojant programinę aplinką „R“ (R Core Team, 2013). Duomenys pirmaisia analizuoti taikant Bajeso mišriųjų efektų modelius (angl. *Bayesian mixed-effects models*), sudarytus naudojant paketą *brms* (Bürkner, 2017). Siekiant palengvinti modelių interpretavimą, poriniai palyginimai (angl. *pairwise comparisons*) buvo atlikti naudojant *emmeans()* ir *contrast()* funkcijas iš *emmeans* paketo. Be to, įrašymo testo atsakymai papildomai analizuoti pagal klaidų tipus.

Tyrimo **rezultatai** rodo, kad UGKM gali būti veiksmingesnė (lietuvių) K2 mokymo metodika nei UPKM. Vertinant eksplicitinių linksnių sistemos žinių pokyčius pagal įrašymo testo rezultatus, abiejų grupių pasiekimai buvo panašūs. Tiek UGKM, tiek UPKM grupėse nustatyta reikšminga ir didelė mokymosi pažanga baigiamojo ir atidėtojo baigiamojo testų metu. Vis dėlto, priešingai nei tikėtasi, UPKM grupė neparodė didesnės mokymosi pažangos nei UGKM grupė nei baigiamojo, nei atidėtojo baigiamojo testo metu, nepaisant to, kad šiems dalyviams naudota intensyvesnio eksplicitinio gramatikos mokymo medžiaga. Be to, Bajeso mišriųjų efektų modelis be kontrolinės grupės duomenų pateikė stiprius įrodymus, kad UGKM grupės dalyvių pažanga tarp išankstinio ir baigiamojo testų buvo didesnė nei UPKM grupės dalyvių. Rezultatai taip pat leidžia daryti prielaidą apie galimą UGKM pranašumą ir atidėtojo baigiamojo testo metu. Aprašytos tendencijos vaizduojamos 16 paveiksle.



16 paveikslas. Modelio prognozuotos teisingo atsakymo tikimybės pagal testo momentą ir grupę (įrašymo testo duomenys, neįskaitant kontrolinės grupės). Paklaidos juostos žymi 89 % didžiausio tankio intervalus (angl. *HDI*)

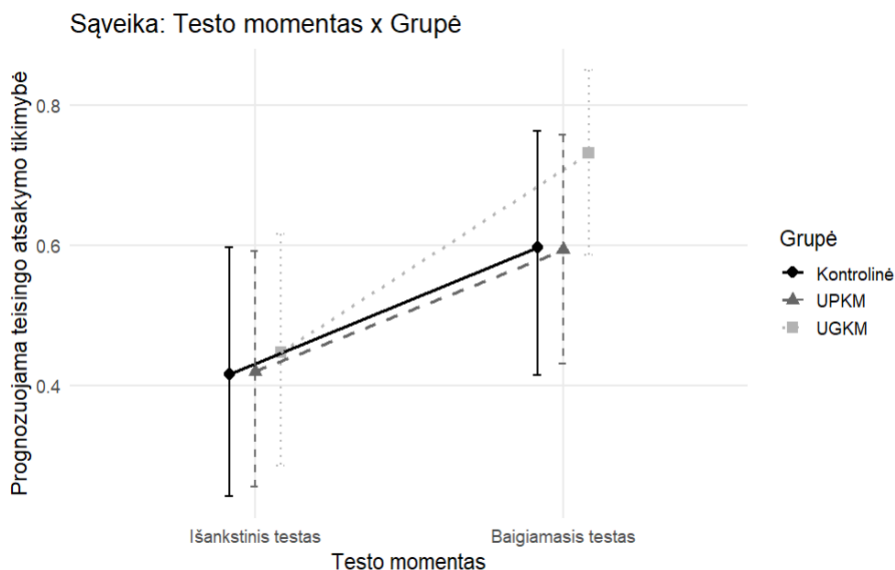
Gauti rezultatai gali būti aiškinami keliais būdais. Pirmasis galimas paaiškinimas susijęs su testų ypatumų ir mokymo intervencijų sąveika. Nors įrašymo testas buvo skirtas eksplicitinėms linksnių sistemos žinioms vertinti, tokios tyrimo priemonės retai atspindi griežtą eksplicitinių ir implicitinių žinių skirtį. Kai dalyviai nebuvo tikri, kokią gramatinę taisyklę taikyti konkrečiam

užduoties elementui, jie greičiausiai rėmėsi intuicija, o tokie atsakymai labiau atspindi implicitines žinias. Tikėtina, kad UPKM grupės dalyvių įgytos žinios daugiausia išliko eksplikitinės, o UGKM grupėje taikytas ribotas eksplikitinis gramatikos mokymas skatino intensyvesnę kalbos formų pastebėjimą (angl. *noticing*) ir didesnę jautrumą kalbos pavyzdžiams (angl. *exemplars*). Todėl UGKM grupės dalyviai sėkmingiau spėjo remdamiesi implicitinėmis žiniomis. Rezultatai taip pat atskleidė tendenciją, kad UGKM grupės dalyviai dažniau nepaisė rašybos ar linksniavimo paradigmos klaidų ir iš karto pateikdavo taisyklingas linksnių galūnes. Svarbu pažymėti, kad toks rezultatų aiškinimas gali būti laikomas papildomu kognityvinio-interakcionistinio modelio, sudarančio UGKM teorinį pagrindą, patvirtinimu.

Kitas galimas rezultatų aiškinimas susijęs su grupių skirtumais. Bajeso mišriųjų efektų modelis, į kurį nebuvo įtraukti kontrolinės grupės duomenys, pateikė vidutinio stiprumo įrodymus, kad eksperimentinės grupės skyrėsi jau išankstinio testavimo metu. Vis dėlto tam tikri įrodymai apie galimus grupių skirtumus išlieka gana prieštaringi, ypač tai pasakytina apie dalyvių pasirengimą mokytis (angl. *developmental readiness*) pagal jų turėtas eksplikitines linksnių sistemos žinias. Klaidų tipų analizė atskleidė panašias abiejų grupių linksnių sistemos raidos tendencijas. Nors modelis be kontrolinės grupės duomenų rodė geresnius UGKM grupės rezultatus žymint kilmininką ir galininką, trijų veiksnių – testavimo laiko, intervencijos tipo ir linksnio – sąveikos efekto nebuvo, o tai leidžia teigti, kad mokymosi pažanga šių linksnių žymėjimo srityje tarp eksperimentinių grupių reikšmingai nesiskyrė. Šie rezultatai leidžia daryti prielaidą, kad eksperimentinės grupės buvo panašios pasirengimo mokytis požiūriu, vertinant jų imlumą eksplikitinėms linksnių sistemos žinioms. Galiausiai pažymėtina, kad poriniai palyginimai pateikė tvirtus įrodymus apie abiejų grupių mokymosi pažangą.

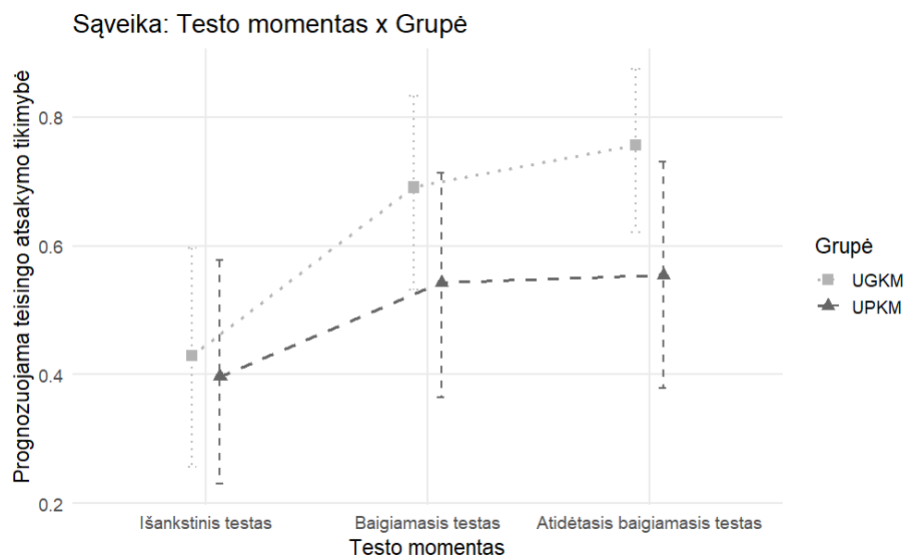
Paskatinto žodinio pakartojimo testo rezultatai atspindi implicitinių linksnių sistemos žinių įgijimą. Vis dėlto šios tyrimo dalies rezultatai yra mažiau akivaizdūs. Pirma, pradinis Bajeso mišriųjų efektų modelis, į kurį buvo įtraukta kontrolinė grupė, pateikė vidutinio stiprumo įrodymus, kad kontrolinės grupės rezultatai pagerėjo nuo išankstinio iki baigiamojo testo. Tuo tarpu UGKM grupės pažangos skirtumas, palyginti su kontroline grupe, buvo pagrįstas tik silpnais įrodymais. UPKM grupės rezultatai iš esmės nesiskyrė nuo kontrolinės grupės rezultatų. Poriniai palyginimai atskleidė indikacinius įrodymus, kad kontrolinės ir UPKM grupių rezultatai pagerėjo nuo išankstinio iki baigiamojo testo, ir tvirtus įrodymus, kad toks pagerėjimas nustatytas UGKM grupėje. Šią tendenciją papildomai patvirtino porinių palyginimų rezultatai, rodantys tvirtai pagrįstus skirtumus tarp UGKM grupės

ir tiek UPKM, tiek kontrolinės grupių baigiamojo testo metu. Tokių skirtumų išankstinio testavimo metu nenustatyta. Ši tendencija vaizduojama 17 paveiksle. Nors šie rezultatai sudarė prielaidas tolesnei duomenų analizei, įskaitant atidėtojo baigiamojo testo rezultatų vertinimą, interpretuojant gautus duomenis būtina atsižvelgti į galimą kontrolinės grupės pažangą.



17 paveikslas. Modelio prognozuotos teisingo atsakymo tikimybės pagal testo momentą ir grupę (paskatinto žodinio pakartojimo testo duomenys, įskaitant kontrolinę grupę). Paklaidos juostos žymi 89 % didžiausio tankio intervalus (angl. *HDI*)

Poriniai palyginimai, apskaičiuoti pagal tolesnį Bajeso mišriųjų efektų modelį, apimančią ir atidėtojo baigiamojo testo duomenis, pateikė tvirtus įrodymus, patvirtinančius mokymosi pažangą nuo išankstinio iki baigiamojo testo ir nuo išankstinio testo iki atidėtojo baigiamojo testo tik UGKM grupėje, UPKM grupėje tokie įrodymai buvo silpni. Be to, testavimo laiko ir intervencijos tipo sąveikos analizė modelyje atskleidė tvirtus įrodymus, kad mokymosi pažanga tarp grupių reikšmingai skyrėsi atidėtojo baigiamojo testo metu – UGKM grupės naudai. Šios tendencijos vaizduojamos 18 paveiksle. Gauti rezultatai leidžia daryti prielaidą, kad UGKM grupėje pasireiškė aiški implicitinių linksnių sistemos žinių pažanga, o UPKM grupėje reikšmingų pokyčių nenustatyta.



18 paveikslas. Modelio prognozuotos teisingo atsakymo tikimybės pagal testo momentą ir grupę (paskatinto žodinio pakartojimo testo duomenys, neįskaitant kontrolinės grupės). Paklaidos juostos žymi 89 % didžiausio tankio intervalus (angl. *HDI*)

Šių rezultatų galima pateikti keletą paaiškinimų. Pirma, didesnė UGKM grupės mokymosi pažanga atitiko išankstines prielaidas. Remiantis kognityviniu-interakcionistiniu modeliu (R. Ellis ir kt., 2019), intensyvesnė įraišką orientuota sąveika su kitais dalyviais ir gausesnė K2 įvestis turėjo sudaryti palankesnes sąlygas implicitinių linksnių sistemos žinioms formuotis. Vis dėlto stebina tai, kad UPKM grupėje nebuvo nustatyta reikšmingos mokymosi pažangos. Vadovaujantis tuo pačiu teoriniu požiūriu, galima daryti prielaidą, kad didesnė eksPLICITINIO gramatikos mokymo apimtis pirmiausia sustiprino eksPLICITINES dalyvių žinias, tačiau kartu galėjo riboti jų gebėjimą pastebėti kalbos formas. Kitas galimas paaiškinimas yra susijęs su išankstiniu dėmesiu formai. Gali būti, kad dėl to užduočių atlikimas dalyviams tapo ne sąveika, kurioje svarbiausia – raiška, o veikla, kurioje reikia taisyklingai vartoti gramatinės formas. Būtent apie tokį poslinkį yra įspėję UGKM tyrėjai Long (2014) bei Willis ir Willis (2007). Tikėtina, kad tai galėjo apsunkinti implicitinių linksnių sistemos žinių įgijimą. Dar aptartinas ir rezultatas, kad, palyginti su kontroline ir UPKM grupėmis, atidėtojo baigiamojo testo rezultatai atskleidė, kad UGKM grupė ne tik išlaikė pasiektą pažangą, bet ir ją dar labiau sustiprino. Tai leidžia daryti prielaidą apie tvaresnį UGKM metodikos poveikį. Mažesnė eksPLICITINIO gramatikos mokymo apimtis galėjo paskatinti dalyvius kitaip apdoroti K2 įvestį ir savo kalbinę

raišką, daugiau dėmesio skiriant reikšmei ir dėl to veiksmingiau pastebint, kokios K2 konstrukcijos yra reikalingos komunikacinei užduočiai atlikti. Pažymėtina, kad išankstinio testavimo metu nebuvo nustatyta jokių reikšmingų skirtumų tarp eksperimentinių grupių.

Vis dėlto negalima ignoruoti tam tikrų su gautais rezultatais susijusių apribojimų. Vidutinio stiprumo įrodymai, rodantys kontrolinės grupės pažangą nuo išankstinio iki baigiamojo testo, kelia abejonių dėl pradinės mokymosi pažangos abiejose eksperimentinėse grupėse. Todėl negalima atmesti galimybės, kad geresnius rezultatus lėmė ne intervencijos poveikis, o kiti eksperimento metu nekontroliuoti veiksniai. Tai kelia svarbių metodologinių klausimų. Nors ankstesni tyrimai pateikė patikimų įrodymų, pagrindžiančių paskatinto žodinio pakartojimo testo tinkamumą implicitinėms gramatikos žinioms vertinti (Erlam, 2009; Spada ir kt., 2015), tam tikrų sunkumų kyla tada, kai dalyviai suvokia, kad dalis teiginių yra gramatiškai netaisyklingi. Tokiais atvejais dalyviai gali pradėti taisyti sakinius arba pakartoti juos tiksliai taip, kaip išgirdo garso įrašę. Pirmuoju atveju kyla pagrįstų abejonių, ar taisydami sakinius dalyviai iš tiesų remiasi implicitinėmis žiniomis, nes jų dėmesys nukrypsta nuo reikšmės į gramatinę formą. Be to, paskatinto žodinio pakartojimo testo taikymas sudėtingos morfologijos kalbų kontekste kelia papildomų iššūkių. Tiksliniai daiktavardžiai, pateikti ne bazine, o linksniuota forma, gali būti sunkiau atpažįstami.

Toliau aptariami ilgalaikiai mokymosi rezultatai, vertinti naudojant atidėtąjį baigiamąjį testą. Kaip minėta, tyrimo duomenys rodo, kad tam tikra ilgalaikė mokymosi pažanga iš tiesų buvo pasiekta. Įrašymo testo rezultatai atskleidė reikšmingą abiejų grupių eksplicitinių linksnių sistemos žinių pagerėjimą. Tuo tarpu paskatinto žodinio pakartojimo testas parodė reikšmingą ir ilgalaikę implicitinių žinių pažangą tik UGKM grupėje. Ypač svarbu pabrėžti, kad ilgalaikiai mokymosi rezultatai buvo pasiekti. Kaip minėta, Markinos (2018) tyrime atidėtojo baigiamojo testo rezultatai parodė, kad dalyvių linksnių sistemos žinios grįžo į pradinį lygį, o Denissenko (2016) tyrime nebuvo nustatyta reikšmingos pažangos perprantant linksnių sistemą, net ir po intensyvios 150 valandų trukmės mokymo intervencijos.

Tada kyla klausimas, kodėl šio tyrimo dalyviai pasiekė ne tik trumpalaikių, bet ir ilgalaikių linksnių sistemos mokymosi rezultatų. Prielaida, kad šio tyrimo eksperimentinės grupės turėjo palankesnes ilgalaikio linksnių sistemos mokymosi sąlygas, grindžiama atidesniu dėmesiu konkrečioms tikslinėms konstrukcijoms. Denissenko (2016) tyrime linksnių sistema buvo nagrinėjama visapusiškai, apimant platų linksnių funkcijų ir paradigų

spektrą. Be to, mokymo intervenciją sudarė įprasti formalūs rusų K2 užsiėmimai, kurie nebuvo pritaikyti konkrečioms tyrimo tikslams. Markinos (2018) tyrime pasirinktas siauresnis tikslinių konstrukcijų laukas – daugiausia dėmesio skirta prielinksniniam linksnių valdymui, o užduotys buvo orientuotos būtent į šį reiškinį. Vis dėlto prielinksniai apima platų semantinių reikšmių spektrą ir yra susiję su penkiais linksniais. Palyginti su minėtais tyrimais, šio tyrimo tikslinės konstrukcijos yra konceptualiai nuoseklesnės – veiksnio žymėjimas, tiesioginio objekto žymėjimas ir neiginių konstrukcijos – bei apėmė tik tris linksnius.

K2 mokymo(si) literatūroje pateikiama keletas galimų paaiškinimų, kodėl toks tikslinių konstrukcijų pasirinkimas galėjo turėti teigiamą įtaką ilgalaikiam linksnio kategorijos įsisavinimui. Pirma, remiantis Magnani ir Artoni (2015) hipotetiniu rusų K2 linksnių mokymosi modeliu, siekiant išmokyti taisyklingai vartoti tiriamąsias konstrukcijas reikia pereiti iš antrojo, *kategorijų*, etapo į trečiąjį, *frazinį*, etapą. Viena iš trijų būtinų šio perėjimo sąlygų yra veiksmazodžio ir objekto linksnių derinimas. Tačiau šiame tyrime tikslinės konstrukcijos buvo dar labiau susiaurintos, daugiausia dėmesio skiriant tik tiesioginio objekto žymėjimui ir neįtraukiant veiksmazodžių, valdančių tik kilmininką, naudininką ar įnagininką. Tikėtina, kad toks dėmesys konkrečiai šio raidos etapo pakategorei palengvino dalyvių perėjimą į *frazinį* etapą. Be to, nedidelė tikslinių konstrukcijų apimtis lėmė didesnę jų dažnį mokomojoje medžiagoje. Taip pat tikėtina, kad dėmesys neiginių konstrukcijoms padidino kilmininko sąlygiškumą (N. Ellis, 2022), nes neigimas nuosekliai siejamas su kilmininko žymėjimu. Padidėjęs sąlygiškumas galėjo lemti geresnius dalyvių rezultatus žymint kilmininką, palyginti su vardininku ir ypač galininku.

Siekiant ištirti galimus eksperimentinių grupių mokymosi raidos skirtumus, buvo atlikta įrašymo testo atsakymų klaidų analizė. Buvo keliamas hipotezė, kad UGKM grupėje kiekvienu testavimo etapu mažės linksnio parinkimo klaidų (angl. *case assignment errors*), o su paradigmos pasirinkimu susijusių klaidų skaičius išliks stabilus arba padidės. Tuo tarpu UPKM grupėje buvo tikimasi priešingos tendencijos – didesnio linksnio parinkimo klaidų skaičiaus ir mažesnio paradigminių klaidų skaičiaus. Veiksnio ir tiesioginio objekto žymėjimo rezultatai rodo, kad tik UGKM grupės mokymosi tendencijos iš esmės atitinka iškeltas hipotezes. Galima daryti prielaidą, kad 1,5 mėnesio pertrauka labiau paveikė žinias apie linksniavimo paradigmas nei gebėjimą parinkti tinkamą linksnį, kaip ir buvo numatyta hipotezėje. Šią prielaidą papildė teiginys, kad implicitiniu mokymusi grindžiamų metodų poveikis yra tvaresnis ilgalaikėje perspektyvoje. Tai galėtų paaiškinti UGKM

grupės pranašumą ir didesnę jos naudą linksnio parinkimo žinioms (Long, 2016, p. 15). Vis dėlto paradigminių klaidų nustatyta palyginti nedaug, todėl jų procentiniai pokyčiai nėra itin informatyvūs. Be to, klaidų tipų pasiskirstymas pagal atskirus linksnius atskleidė įvairias ir nevienodas tendencijas, kurios tik iš dalies atitinka iškeltas hipotezes. Todėl remiantis šiais duomenimis sudėtinga nustatyti aiškias linksnio kategorijos įsisavinimo raidos tendencijas.

Svarbu pažymėti, kad vienas iš modelių, į kurį nebuvo įtraukti kontrolinės grupės duomenys, vis dėlto atskleidė galimus eksperimentinių grupių raidos tendencijų skirtumus. Rezultatai rodo, kad UPKM grupės dalyvių linksnių sistemos žinios gerėjo palaipsniui: visiškai neteisingus atsakymus pamažu keitė iš dalies teisingi atsakymai. Tuo tarpu UGKM grupės dalyviai iš esmės išvengė tarpinio etapo, kuriam būdingos rašybos ir (arba) paradigminės klaidos, ir iškart perėjo nuo visiškai neteisingų prie visiškai teisingų atsakymų. Vis dėlto modelis nepateikė tvirtų įrodymų, patvirtinančių UGKM grupės pranašumą visiškai teisingų atsakymų kategorijoje baigiamojo testo metu. Todėl, nors gauti rezultatai leidžia kelti prielaidą, kad UGKM ir UPKM metodai gali būti susiję su skirtingomis linksnio kategorijos įsisavinimo trajektorijomis, šiuo metu turimi įrodymai išlieka neaiškūs.

Atsižvelgiant į šiuos disertacijos rezultatus, formuluojami šie **ginamieji teiginiai**:

1. Ribotas dėmesys formai, apimantis trumpus į formą orientuotus epizodus, kylančius reaguojant į besimokančiųjų komunikacinius sunkumus, ir eksplisitinis dėmesys formoms tik užduoties *apibendrinimo* etape (t. y. UGKM metodą), yra pakankamas sėkmingai įgyti eksplisitinių linksnių sistemos žinių. Nuolatinis dėmesys gramatikai viso užduoties ciklo metu (t. y. UPKM metodas) nebūtinai lemia geresnį eksplisitinių linksnių sistemos žinių įgijimą.
2. UGKM metodui būdingi tik epizodiniai dėmesio formai elementai užduoties *pasirengimo* ir *atlikimo* etapuose bei eksplisitinis dėmesys formoms užduoties *apibendrinimo* etape leidžia išlaikyti į raišką orientuotą sąveiką ir sudaro prielaidas implicitinių linksnių sistemos žinių formavimuisi. Nuolatinis dėmesys formai viso užduoties ciklo metu (t. y. UPKM metodas) gali trukdyti įgyti implicitinių linksnių sistemos žinių.
3. Lietuvių K2 linksnių sistemos mokymas, taikant UGKM principais grindžiamas užduotis, gali lemti ne tik trumpalaikę, bet ir ilgalaikę mokymosi pažangą, jei mokymo trukmė yra pakankama, o tikslinės konstrukcijos – aiškiai apibrėžtos.

4. Eksplicitinio gramatikos mokymo vieta ir intensyvumas užduočių cikle turi ribotą poveikį linksnių sistemos įsisavinimo raidos trajektorijoms, vertinamoms pagal klaidų tipus.
5. Perprantant K2 linksnių sistemą gimtosios K1 tipologinis panašumas į K2 gali turėti trumpalaikį poveikį, o įtaka ilgalaikiam K2 linksnio kategorijos įsisavinimui yra ribota.
6. Sudėtingos morfologijos K2 linksnių sistemos mokymas, taikant UGKM principais grindžiamas užduotis, gali lemti implicitinių žinių formavimąsi net ir suaugusiems besimokantiejiems, ir tai patvirtina kognityvinio-interakcionistinio modelio prielaidas. Siekiant skatinti implicitinių žinių įgijimą, eksplicitinis gramatikos mokymas turėtų būti sutelkiamas į užduoties *apibendrinimo* etapą.
7. Sudėtingos morfologijos K2 linksnių sistemos galima sėkmingai mokyti taikant UGKM principais grindžiamas užduotis net ir žemesnio kalbos mokėjimo lygio besimokantiejiems.

Šios disertacijos rezultatai svariai papildė esamus K2 mokymo(si) tyrimus tiek teoriniu, tiek praktiniu požiūriu. Svarbiausia, gauti duomenys rodo, kad UGKM gali būti veiksmingas mokant sudėtingos morfologijos K2, ypač įsisavinant sudėtingas gramatines kategorijas, tokias kaip linksnių sistema. Tyrimo rezultatai leidžia daryti prielaidą, kad UGKM principais grindžiami kursai gali skatinti ne tik trumpalaikę, bet ir ilgalaikę mokymosi pažangą, įskaitant implicitinių žinių formavimąsi. Tikėtina, kad šios išvados prisidės prie tolesnių UGKM taikymo sudėtingos morfologijos K2 mokymo kontekste tyrimų, taip pat praplės žinias apie K2 mokymą(si) ir praturtins linksnio kategorijos įsisavinimo tyrimus. Teoriniu požiūriu gauti rezultatai papildė empirinių įrodymų bazę, pagrindžiančią kognityvinį-interakcionistinį modelį, kuriuo grindžiama UGKM metodika, ir kartu suteikia naujų įžvalgų kitoms K2 mokymo(si) teorijoms. Praktiniu požiūriu šioje disertacijoje parengta mokomoji medžiaga gali būti naudojama kaip pavyzdys kuriant UGKM principais grindžiamas mokymo programas, skirtas sudėtingos morfologijos K2 mokymui, tiek būsimuose tyrimuose, tiek praktiniuose K2 kursuose.

ABOUT THE AUTHOR

Richard Udes

richard.udes@flf.vu.lt

Richard Udes graduated from the University of Amsterdam (the Netherlands) in 2011, where he obtained a BA in History and a BA in Slavic Languages & Cultures: Russian. As part of the latter program, he participated in an exchange program at Saint Petersburg State University (Russia). After spending a year in Ukraine to learn Ukrainian, he returned to the University of Amsterdam, where he completed an MA in Linguistics of European languages: Slavonic in 2013. His MA thesis examined definiteness in Russian within the Functional Discourse Grammar framework and involved experimental research for which he briefly returned to Saint Petersburg. Following graduation, he moved to Vilnius, Lithuania, in 2014. He had already begun learning Lithuanian in 2013 with the aim of achieving fluency. Between 2013 and 2021, he worked as a freelance teacher of Dutch, English, and, later, Lithuanian as second languages. During this period, he became interested in Task-Based Language Teaching (TBLT) and increasingly incorporated its principles into his teaching. This experience ultimately inspired him to return to academia and pursue a PhD position at Vilnius University (Lithuania), focusing on teaching Lithuanian as a second language through TBLT. Richard Udes is currently a second-language instructor in Dutch and Lithuanian, as well as a researcher in the Psycholinguistics Research Group at Vilnius University.

Conference presentations

- 2022 - Poster presentation “Adapting TBLT to morphologically complex languages: approaches to the teaching of cases in Lithuanian” (*Widening the Horizon of Task-Based Language Teaching*, University of Innsbruck, Austria).
- 2024 – “Exploring the compatibility of TBLT and morphologically complex languages: an analysis of learner performance on an L2 Lithuanian task” (*EuroSLA 33*, Paul Valéry University, Montpellier, France).
- 2025 – “Exploring the effectiveness of TBLT for teaching morphologically complex languages: Case acquisition in L2 Lithuanian” (*Task-Based Leveraging of Technology*, University of Groningen, the Netherlands).

- 2025 – “Task-Based Language Teaching and morphologically complex languages: A study on case acquisition in L2 Lithuanian” (*Languages and people*, Vilnius University, Lithuania).

Publications

- Udes, R. (2025a). Užduotimis grindžiamo kalbos mokymo (UGKM) metodikos potencialas mokant(is) sudėtingos morfologijos svetimųjų kalbų [Exploring the potential of task-based language teaching (TBLT) for teaching second languages with complex morphological systems]. *Verbum*, 16, 1–10. <https://doi.org/10.15388/verb.16.09>
- Udes, R. (2025b). The role of grammar instruction in Task-Based Language Teaching: acquisition of case in L2 Lithuanian with different focus-on-form treatments. *Taikomoji kalbotyra*, 23, 50–75.

Other PhD activities

- 2021 - Autumn Linguistics School organized by Vilnius University, online.
- 2023 - Study visit to Poland: the PROM exchange program at Jagiellonian University in Kraków, Poland. Host professor: Dr. I. Janowska.
- 2023 - Internship at the University of Barcelona, Faculty of Philology and Communication, Spain. Host professors: Prof. R. Gilabert, Dr. J. Castellví.
- 2024 - PhD summer school on statistics at the Catholic University of Louvain, Belgium.
- 2025 - The 22nd Salos summer school in Antalietpė, Lithuania.

NOTES

Vilnius University Press
9 Saulėtekio Ave., Building III, LT-10222 Vilnius
Email: info@leidykla.vu.lt, www.leidykla.vu.lt
bookshop.vu.lt, journals.vu.lt
Print run 20