

# The role of grammar instruction in Task-Based Language Teaching: acquisition of case in L2 Lithuanian with different focus-on-form treatments

**Richard Udes**

Vilnius University

[richard.udes@ff.vu.lt](mailto:richard.udes@ff.vu.lt)

ORCID: 0009-0002-0266-4735

ROR: <https://ror.org/03nadee84>

**Abstract.** Task-Based Language Teaching (TBLT) is a second language (L2) teaching methodology that emphasizes the completion of authentic, real-world tasks. The role of grammar instruction within the task cycle remains a point of discussion. TBLT is often argued to primarily foster the acquisition of implicit, rather than explicit, L2 knowledge. This is considered less favorable for the teaching of morphologically complex languages such as Lithuanian, which features complex grammatical categories like case. The present paper seeks to examine these claims. Eighteen adult L2 Lithuanian learners participated in an experimental study. One group ( $n = 8$ ) completed a task version with a strong focus on grammar, the other ( $n = 10$ ) without. Although the results did not reveal significant treatment differences in gains in explicit case knowledge, the findings suggest differences that could divert developmental patterns related to case and paradigm selection, which may inform future task and test design.

**Key words:** *task-based language teaching, second language acquisition, Lithuanian as a second language, acquisition of morphology, case*

## Gramatikos mokymas užduotimis grindžiamame kalbų mokyme (UGKM): lietuvių K2 kalbos linksnio įsisavinimas lyginant skirtingas gramatinės formos mokymo priemogis

**Santrauka.** Užduotimis grindžiamas kalbų mokymas (UGKM) yra svetimšios kalbos (K2) mokymo metodika, kurioje akcentuojamas autentiškų, realaus pasaulio komunikaciją imituojančių užduočių atlikimas. Nors UGKM jau seniai pripažintas kaip pažangus kalbų mokymo metodas, mokslininkai iki šiol nesutaria, kiek jame turėtų likti vietos gramatikos mokymui. Dažnai teigiama, kad UGKM pirmiausia skatina implicitinį, o ne eksplicitinį K2 žinių įsisavinimą. Kita vertus, pripažįstama, kad eksplicitinis gramatikos mokymas yra ypač svarbus mokant sudėtingos morfologijos kalbų, tokių kaip lietuvių, turinčių itin sunkiai įsisavinamų gramatinių kategorijų, pavyzdžiui, linksnio kategorija. Šio straipsnio tikslas – aprašyti gramatikos komponento vaidmenį UGKM užduočių cikle, skirtame lietuvių kalbos linksnio kategorijos mokymui. Žvalgomajame tyrime, kurį sudarė išankstinis testas, intervencija ir baigiamasis testas, dalyvavo 18 suaugusiųjų lietuvių K2 mokinių. Viena grupė ( $n = 10$ ) atliko UGKM kriterijus atitinkančią užduotį su ribotu eksplicitiniu gramatikos mokymu, o kita ( $n = 8$ ) – tą pačią užduotį su išsamesniu gramatikos mokymu.

Eksperimento tikslinės struktūros buvo subjekto ir tiesioginio objekto raiška su neiginiu ir be jo, o tai apima daiktavardžio vardininko, kilmininko ir galininko formas.

Nors abiejų grupių dalyviai sėkmingai atliko užduotį, gramatikos žinias, kaip rodo baigiamasis testas, reikšmingiau pagerino tik ta grupė, kurioje taikytas ribotas gramatikos mokymas. Vis dėlto statistinė analizė nerodė ryškaus dviejų grupių skirtumo tarp išankstinio ir baigiamojo testo. Be to, kai kurie esminiai skirtumai tarp tiriamųjų grupių galėjo turėti įtakos rezultatams. Grupė, dirbusi su išsamesniu gramatikos pristatymu užduotyje, jau išankstiniame teste demonstravo geresnes linksnių žinias, todėl galimybė pasiekti akivaizdžią pažangą buvo mažesnė. Be to, kiekvienos grupės dalyviai lankė skirtingus lietuvių kalbos kursus. Tikėtina, kad įtaką rezultatams turėjo ir tam tikrų tikslinių žodžių ir konstrukcijų dažnumas ir ryškumas. Nepaisant to, klaidų analizė atskleidė tam tikrus skirtumus tarp grupių, kurie iš dalies patvirtina skirtingą linksnio įsisavinimo raidą, priklausomai nuo taikytos intervencijos. Tyrimo įžvalgos gali būti naudingos kuriant užduotis ir testus tolesniems tyrimams.

**Raktažodžiai:** *užduotimis grindžiamas kalbos mokymas, svetimšios kalbos įsisavinimas, lietuvių kaip svetimoji kalba, morfologija, linksnio kategorija*

## 1. Introduction

Task-Based Language Teaching (TBLT) is a second language (L2) instructional approach grounded in the principles of communicative language teaching but distinguished by its emphasis on completing authentic, real-world tasks. Although its popularity continues to grow (Ellis et al. 2019) and the body of research on TBLT is expanding, there are still areas of disagreement. One of the more contested topics in the TBLT literature is the role of grammar within the task cycle. This debate encompasses several issues, including where grammar fits in the cycle, the application of focus-on-form techniques, and the incorporation of so-called focused tasks (Ellis 2003) into TBLT syllabi. These ongoing discussions may partly explain the persistent misconception that grammar plays a minimal role in TBLT (Ellis et al. 2019). As a result, it is still commonly believed that TBLT, with its emphasis on incidental learning, primarily supports the development of implicit L2 knowledge and communicative competence, while giving less attention to explicit learning. This, in turn, has led to the perception that TBLT is less effective for teaching grammatical forms (Sheen 2003; Swan 2005).

The role of explicit grammar instruction within the task cycle is particularly relevant when considering morphologically complex languages such as Finnish, Russian, or Lithuanian. Traditionally, the teaching of these languages has placed strong emphasis on grammar acquisition from the outset (Comer 2012). Despite some developments, many instructors remain skeptical of communicative L2 teaching methodologies like TBLT, especially at the beginner level (Kogan and Bondarenko 2022: 78–79). Although some scholars have proposed adaptations of TBLT for the instruction of morphologically rich languages (Gilabert and Castellví 2019; Kogan and Bondarenko 2022), empirical research in this area remains limited (Kogan, Révész, and Cheung 2024; Pomarolli et al. 2023). This is especially true for studies focusing on the acquisition of L2 morphology (Markina 2018), which is often regarded as one of the most challenging aspects of learning such languages (Savickienė 2006). The perceived difficulty of teaching L2 morphology is a key factor driving teachers to favor structurally organized syllabi and extensive use of explicit instruction (Comer 2012: 152; Gilabert and Castellví 2019: 530).

This exploratory study seeks to challenge several prevailing assumptions surrounding grammar instruction in TBLT. Specifically, it investigates whether the amount and placement of explicit instruction within the task cycle influences learners' development of explicit L2 knowledge of a complex grammatical category after completing a single task. The target language is Lithuanian, widely regarded as morphologically complex, with grammatical case being the category of interest. The next section

of this article provides a brief overview of the central tenets of the TBLT methodology and revisits the debate on grammar within the TBLT literature, highlighting how the teaching of morphologically rich languages may challenge some of these views. Section 3 offers a brief introduction to the target constructions of the study and the challenges faced by learners of L2 Lithuanian in acquiring them. Section 4 presents the aims and questions of the present study, while Section 5 outlines its design and methodology. Finally, Sections 6 and 7 report the results and discussion.

## 2. TBLT and the role of grammar instruction

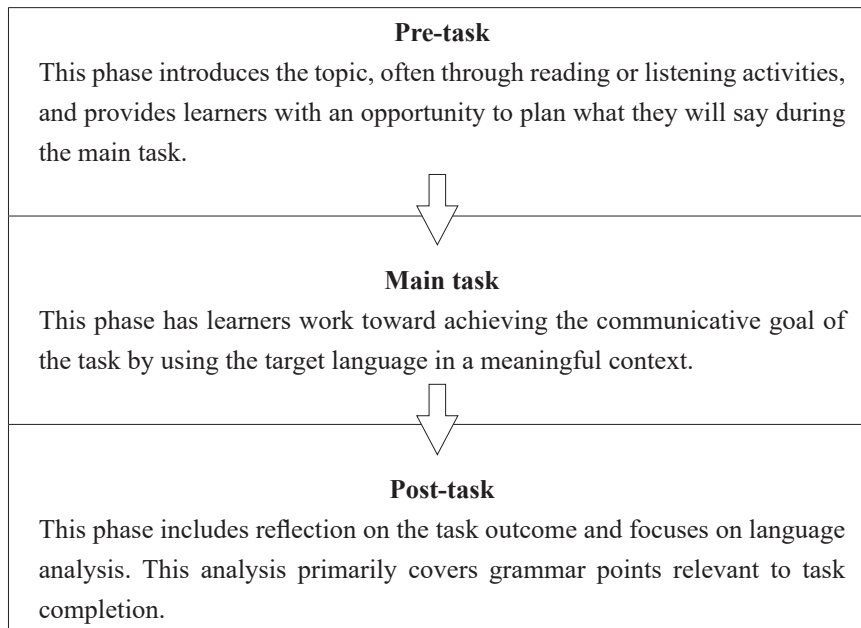
Task-Based Language Teaching is a second language teaching approach that promotes realistic language use in authentic communicative contexts through purposefully designed tasks. Rooted in communicative language teaching, TBLT emphasizes the language functions learners need and aims to develop fluency in the target language. The most important component of TBLT curricula is the tasks which are supposed to reflect real-world language demands. These tasks can in context and complexity ranging from ordering a cup of coffee in a café to conducting a routine vehicle inspection as a border patrol agent (R. Ellis et al., 2019: 23, 163).

Although numerous definitions of a task have been proposed in the TBLT literature, one of the most widely cited is based on the following four criteria (Ellis et al. 2019: 10):

1. **Meaning-focused communication:** The task should be structured in such a way as to ensure that learners concentrate on conveying and understanding meaning.
2. **Presence of a gap:** A task must include some form of ‘gap’ that encourages learners to communicate in order to bridge it. Such gaps can be informational, reasoning, or opinion-based,
3. **Learner-driven language use:** Students are expected to draw on their own linguistic and non-linguistic resources, which means the task should not explicitly encourage the use of certain L2 constructions or grammar rules.
4. **Clear communicative outcome:** Task success is not based on linguistic accuracy, but the successful achievement of the intended communicative goal of the task.

Lastly, tasks are typically organized into three phases (Ellis 2003: 244; Willis 1996: 52). A typical outline of a task is illustrated in Figure 1. As this structure suggests, grammar instruction is typically addressed only *after* the task has been carried out.

A central tenet of the TBLT methodology is its emphasis on incidental learning through interaction, i.e., learning without conscious intention (Ellis et al. 2019; Long 2014). This process is believed to lead to the development of implicit knowledge, which is unconscious and automatized, similar to that of one’s knowledge of a first language (L1). This emphasis on the acquisition of implicit L2 knowledge stands in stark contrast to traditional structural approaches such as Presentation–Practice–Production (PPP). Such approaches typically break the target language into discrete grammatical units, which are then taught and practiced sequentially and which, supposedly, primarily lead to the development of explicit knowledge with only limited potential for implicit L2 acquisition. While Long (2014) argues that implicit learning is the default mechanism not only for L1 acquisition in children but also for L2 acquisition in adults, he acknowledges that adults have a reduced capacity to acquire a second language in this way. This is where focus on form becomes essential: it refers to techniques that draw learners’ attention to linguistic forms in a way that supports the development of explicit knowledge without undermining their focus on meaning during task performance. In Long’s (2014) cognitive-interactionist framework, the explicit knowledge gained through focus on form taps into the way L2 learners attend to the input through noticing, which can indirectly facilitate the acquisition of implicit knowledge.



**Figure 1.** Outline of a typical task according to the TBLT methodology. Based on: R. Ellis, 2003, p. 244; J. Willis, 1996, p. 52

The extent and place of focus on form within the task cycle have been a longstanding subject of debate in the literature. According to Long's (2014) original definition, focus on form should be reactive (emerging in real-time as a response to learner errors), incidental (occurring during meaning-focused interaction), and brief (as not to overshadow the communicative goal of the task). Conversely, Willis and Willis (2007) argue that even brief attention to form during task performance may cause learners to shift their focus away from meaning and toward linguistic structures. Ellis (2009) takes the concept further, allowing for preemptive focus on form during the pre-task stage. He also allows for so-called 'focused tasks' which are tasks designed to elicit specific linguistic features while still maintaining a meaning-oriented interaction. This interpretation of focus on form is sometimes referred to as Task-Supported Language Teaching (TSLT), often regarded as a weaker version of TBLT (Ellis 2009). Notably, the post-task phase remains the least controversial stage in the task cycle for explicit instruction.

The teaching of morphologically complex L2s may require a distinct approach to focus on form. It is still widely believed that strong communicative methodologies, such as TBLT, are poorly suited for the instruction of languages with rich morphological systems (Comer 2012:152; Gilabert and Castellví 2019:530; Kogan and Bondarenko 2022:79). Long (2014:43) describes inflectional morphology as a "fragile" linguistic feature, noting that adult learners often lose the ability to notice subtle forms, such as verbal endings or nominal case markers, implicitly. As a result, these forms are typically acquired late or sometimes not at all. Studies on the Basic Variety, i.e., naturally acquired L2s, seem to confirm this, as they tend to be largely devoid of grammatical morphology, including case markers (Ellis 2022; Klein 1998; Perdue 1993). Functional morphology (e.g., verbal tense markers, nominal case endings) possesses properties that make it difficult to detect (e.g., low perceptual salience) and, consequently, to acquire (N. Ellis 2022). Furthermore, L2 learners naturally prioritize meaning over semantically less salient elements during input processing, which leads them to overlook inflectional morphology (VanPatten 2015). While some scholars argue that grammar gains importance only at more advanced L2 stages when learners aim to express more nuanced meanings (Ellis et al. 2019), this view may not hold for morphologically complex languages. Slabakova's (2019) Bottleneck Hypothesis claims that

the acquisition of functional morphology represents the most challenging aspect of L2 learning and its slow acquisition impedes overall L2 development (Nuss 2022). Furthermore, it seems the more complex an L2's morphological system is, the longer it takes for L2 learners to develop inflectional competence (Brezina and Pallotti 2019).

Two notable proposals have been put forward to adapt TBLT for the teaching of morphologically complex L2s. The first, by Gilabert and Castellví (2019), argues that most research on task design has largely overlooked the linguistic complexity of target L2s, focusing primarily on the conceptual complexity of tasks, i.e., the contextual demands and cognitive effort required to achieve a task's communicative goal. The key challenge for task design for the teaching of morphologically complex L2s lies in balancing the conceptual and linguistic demands. To address this, Gilabert and Castellví propose breaking the main task into smaller, preparatory sub-tasks, which L2 learners carry out beforehand, and are aimed to reduce the conceptual load imposed by the main task. The second proposal by Kogan and Bondarenko (2022) focuses specifically on task design for beginner learners of morphologically rich L2s. Its main claim is that it is essential to offer cognitively engaging tasks, even at early stages of language learning. One approach the authors propose is the use of solo tasks, which are essentially input-based tasks and designed to stimulate learners cognitively while minimizing output demands.

One study that investigated the effectiveness of TBLT for the acquisition of complex L2 morphology is by Markina (2018) on L2 Russian. The study compared a TBLT group and a TSLT group, each receiving 28 hours of instruction: the first 14 hours focused on case marking following prepositions, and the second 14 hours on verbs of motion. While both groups demonstrated lasting improvement in verbs of motion, gains in case marking after prepositions were evident in the immediate posttest but declined in the delayed posttest. Markina attributes this decline to VanPatten's (2015) Input Processing Theory: unlike verbs of motion, which often contain semantically rich morphemes such as tense, gender, and path, prepositional case marking is largely grammatical and thus more easily overlooked during input processing. Interestingly, Markina does not view the relatively short 14-hour instructional period as a limitation. She cites a study by Denissenko (2016), which involved 150 hours of PPP-based L2 Russian instruction and resulted in no case improvement at all levels (0, A1, A2, B1). This suggests that formal instruction alone may be insufficient for facilitating the acquisition of case marking in L2 learners.

### 3. Challenges for learners of L2 Lithuanian

The studies by Markina (2018) and Denissenko (2016) clearly demonstrate that grammatical case in L2 Russian poses a significant challenge for learners. Evidence suggests that L2 Lithuanian is no exception in this regard (Dabašinskienė and Čubajevaitė 2009; Savickienė 2006). Dabašinskienė and Čubajevaitė (2009) note that the majority of errors produced by L2 learners across different proficiency levels involve the grammatical cases, namely the nominative, genitive, and accusative. Nevertheless, these cases are generally introduced at an early stage, typically at the A1 level, in the context of subject and direct object marking. In Lithuanian, subjects in affirmative sentences are marked with the nominative case, but when negated, they shift to the genitive.

- (1) *But-e*                      *yra*                      *balkon-as*.  
apartment-LOC            is                      balcony-NOM  
'There is a balcony in the apartment.'
- (2) *But-e*                      *nėra*                      *balkon-o*.  
apartment-LOC            NEG.is                balcony-GEN  
'There is no balcony in the apartment.'

An exception concerns sentences containing the verb *patikti* ('to like'). This verb requires an indirect object in the dative case (the person who likes) and a nominative subject (the item that is liked). Unlike standard subject constructions, however, the subject of *patikti* does not change to the genitive case when the verb is negated but remains in the nominative:

- (3) *Man patinka jūsų balkon-as.*  
I.DAT like your.PL balcony-NOM  
I like your balcony.'
- (4) *Man ne-patinka jūsų balkon-as.*  
I.DAT NEG-like your.PL balcony-NOM  
I do not like your balcony.'

Direct objects, however, follow a similar pattern like regular subjects: they are typically marked in the accusative case in affirmative contexts, but switch to the genitive under negation, a pattern that is often challenging for L2 learners (Dabašinskienė and Čubajevaitė 2009). Learners must first acquire accusative case marking for direct objects, then modify this knowledge to accommodate the genitive under negation.

- (5) *But-as turi balkon-q.*  
apartment-NOM has balcony-ACC  
'The apartment has a balcony.'
- (6) *But-as ne-turi balkon-o.*  
apartment-NOM NEG-have balcony-GEN  
'The apartment does not have a balcony.'

Notably, the nominative singular corresponds to the base form of lexical items, which is the unmarked form that L2 learners typically encounter in dictionaries and word lists in L2 textbooks (Savickienė 2005). As such, the base form often serves as a foundation for learning to mark other cases, both for the L1 and L2 (Gagarina N. and Voeikova M. D. 2009). Consequently, L2 Lithuanian learners often overgeneralize the nominative case when other oblique cases are appropriate (Dabašinskienė and Čubajevaitė 2009).

An additional challenge for L2 learners of Lithuanian is not just assigning the correct case but also selecting the appropriate ending based on the declension pattern to which a noun belongs. Lithuanian features five main declension patterns, however, due to variations in these main declensions, this typically means L2 learners need to master at least four additional deviating paradigms, especially since these different paradigms are usually presented separately in L2 textbooks. For example, the noun *balkonas* ('balcony'), which takes the nominative ending *-as*, has the genitive ending *-o* (*balkono*) and the accusative ending *-q* (*balkonq*). In contrast, the noun *dieną* ('day') takes the nominative ending *-a*, with the genitive ending *-os* (*dienos*) and the, in this case, identical, accusative ending *-q* (*dienq*).<sup>1</sup> This can lead to mistakes such as in (7), reflecting accurate case assignment, but the wrong ending:

- (7) \**But-as ne-turi balkon-os.*  
apartment-NOM NEG-have balcony-GEN  
'The apartment does not have a balcony.'

<sup>1</sup> The nouns *balkonas* ('balcony') and *diena* ('day') differ not only in declension pattern/paradigm but also in grammatical gender, with *balkonas* being a masculine and *diena* a feminine noun. However, in the absence of adjective or participle agreement and pronominal deixis, grammatical gender does not function as an independent factor in the case-marking process. Under these conditions, the different endings can be considered primarily a matter of declension pattern.

Difficulties in selecting the correct endings from the appropriate paradigms have been observed by Dabašinskienė and Čubajevaitė (2009), Bružaitė-Liseckienė (2021), and Stepšys and Kamandulytė-Merfeldienė (2024) in the context of adjective-noun agreement.

## 4. Aims and research questions

Research into the teaching of morphologically complex L2s through the TBLT framework, particularly with regard to the acquisition of L2 morphology, is scarce. Given the limited number of studies and challenging findings, the present study adopts an exploratory approach. It aims to contribute to the field by investigating the effectiveness of TBLT for the acquisition of grammatical case in L2 Lithuanian under two different approaches to explicit grammar instruction. Specifically, two groups of participants will be presented with the same task differing in the place and extent of focus on form within the task cycle. Learning gains in both groups will be compared by examining potential improvement from pre- to post-test. Additionally, since it is often argued that TBLT fosters mainly implicit knowledge, focus will be put on potential explicit learning gains.

The study addresses two primary research questions: (1) Do the timing and extent of explicit grammar instruction within the task cycle influence learning outcomes related to the Lithuanian case? (2) Can measurable gains in case acquisition be observed at the A1 level following a brief instructional intervention? The target structures to be investigated will be the aforementioned subject (involving the nominative and genitive cases) and direct object marking (involving the accusative and genitive) in affirmative and negative sentences. Beyond overall learning gains, potential treatment-related differences in developmental patterns are also of interest, leading to a third research question: (3) Do the two treatments give rise to different developmental patterns in the acquisition of grammatical case?

## 5. Methods

### 5.1. Participants

The participants in the study were 18 adult learners of L2 Lithuanian. Although small, this number was deemed sufficient for an exploratory study. The participants were divided into two groups: a TBLT group ( $n = 10$ ; 6 males, 4 females; age range: 18–56) and a TSLT group ( $n = 8$ ; all female; age range: 35–58). The TBLT group comprised a linguistically diverse range of L1s, including English, Japanese, Persian, Russian, Spanish, Polish (with 2 participants), and Ukrainian (with 3 participants). In contrast, all participants in the TSLT group were native speakers of Russian and/or Ukrainian. Clearly, the groups were quite different, which will be addressed further in the Discussion section. The study was conducted at a university in Vilnius, Lithuania, where participants were already enrolled in an L2 Lithuanian language course. Their level was estimated to be within the A1–A2 range according to the CEFR framework.

### 5.2. Research design

The study employed a pretest–treatment–posttest design and was conducted over four sessions. The first and fourth sessions lasted around 30 minutes during which participants completed the pre- and post-test. The treatment was administered over the second and third sessions (a total of four hours) with each group receiving a different instructional approach. The TBLT group performed a TBLT-aligned task, featuring explicit grammar instruction only in the post-task phase. In contrast, the TSLT group completed a task-supported version of the same task, which incorporated focus on form throughout

the entire task cycle. These additional focus-on-form activities replaced certain more communicative exercises used with the TBLT group, resulting in roughly equivalent treatment lengths across groups. This way, there was no potential instructional time advantage for any of the groups. The two groups participated at different times, and the author of this study served as the teacher for both treatments.

### 5.3. The two versions of the task

The task designed for this study was formulated as follows: “Discuss which pieces of furniture are necessary for a living room and agree on what to purchase.” It required participants to identify what furniture they need and collaboratively decide which items to buy. This scenario was selected based on the assumption that furnishing a home is a common communicative need among many L2 learners living abroad. Moreover, the input and output generated by this task were likely to elicit frequent use of subject and direct object marking, both in affirmative and negative contexts, which were the target grammatical structures of the study. Considering this, the task can be classified as a semi-focused task, as it was designed to include and promote frequent use of subject and direct-object case marking, while these features were not the sole linguistic elements necessary for task completion.

The task structure follows the framework illustrated in Figure 1. It was operationalized as follows:

#### 1. Pre-task Phase

1. Introduction: The topic of furniture and room furnishing is introduced. Participants brainstorm words for furniture and reflect on important items.
2. Reading: Learners read a simulated forum discussion in which a user seeks advice on how to furnish a room. The activity reinforces furniture-related vocabulary and provides input flooding, with numerous instances of subject and direct-object case marking.
3. Listening: Participants listen to a dialogue between two roommates as they shop for furniture. This audio input provides further exposure to case marking as well as useful vocabulary and phrases for expressing preferences and agreement or disagreement.
4. Interaction (based on Gilabert & Castellví’s (2019) concept of sub-tasks): The main task was divided into two parts:
  - o Expressing what is present and absent: in pairs, each participant has a different picture of a living room that they cannot show to their partner, prompting them to describe their respective rooms and identify similarities and differences in terms of the furniture present.
  - o Expressing preferences: in three rounds of paired comparisons between different pieces of furniture or devices round, participants discuss their preferences in pairs and try to reach a joint decision.

#### 2. Main Task: “Discuss which pieces of furniture are necessary for a living room and agree on what to purchase.”

- Procedure: Working in pairs, participants need to furnish a living room featured in a visual prompt. Each participant then visits a different furniture store, which means they see different store inventories. The participants are required to inform each other about the items available at their respective stores and together decide on a maximum of four pieces of furniture or appliances to purchase.

#### 3. Post-task Phase

- Comparison: Pairs share and compare the furniture they selected with the rest of the class.

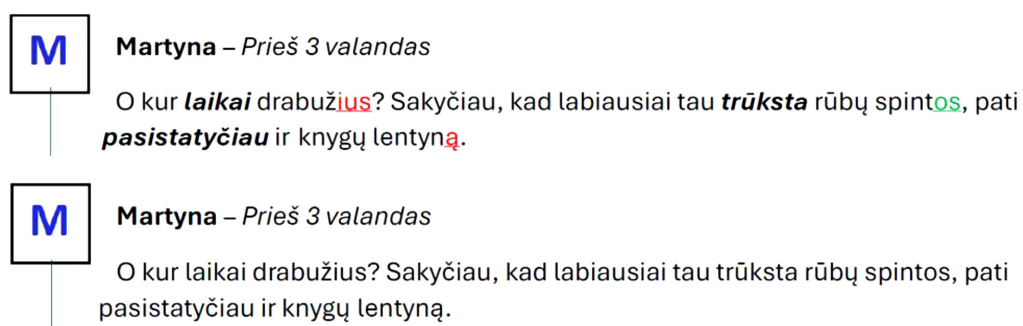
- Language Review: Participants review the reading and listening materials to look for useful vocabulary and phrases.
- Grammar Focus: Learners complete a set of targeted grammar exercises focusing on subject marking, direct object marking, and the use of comparative adjectives and adverbs. These linguistic features were deemed relevant to the completion of this particular task.

As can be seen above, in the TBLT version, explicit focus on form was limited to the post-task phase. Exercises on the target constructions took around 20-30 minutes for participants to complete. No direct references to grammatical forms were made during the pre-task or main task stages, although brief, incidental form-focused feedback was permitted in response to student questions. Such feedback remained minimal and did not divert attention from the communicative purpose of the task. The only implicit focus on form in this version was provided through input flooding in the reading and listening texts.

The TSLT version of the task was almost identical but included two key differences:

- Explicit grammar exercises in the pre-task stage: In the reading activity, learners completed a table, which outlined declension patterns for nouns in the genitive and accusative cases, with most of the relevant endings appearing in the reading text. This replaced an activity for the TBLT group in which participants had to write their own forum post. In the listening task, participants were given a script with omitted grammatical endings, which they were required to fill in while listening. In contrast, the TBLT group filled in missing phrases from the same script. These activities were designed to explicitly draw learners' attention to case morphology before engaging with the task.
- Highlighted case forms in all input and materials: The reading and listening texts, example sentences for sub-tasks, and the store inventories used in the main task featured visually highlighted genitive and accusative forms in subject and direct-object contexts (see Figure 2). Although this method is often considered less intrusive and more implicitly form-focused, it was expected to significantly shift learner attention toward grammar. As a result, the sub-tasks in this version may have functioned more like grammar practice than opportunities to practice meaning-driven communication.

Importantly, while the task at all stages provided ample opportunities for noticing and/or employing the target structures of this study, participants in both groups were not explicitly encouraged to use them during task performance. The instructions for the interaction and main task stages made no explicit reference to the case and successful task completion was determined solely by attainment of the communicative goal.



**Figure 2.** Excerpts from the reading text from the TSLT version (above) and from the TBLT version of the task (below). The translation of the sentence is as follows: ‘Where do you keep your clothes? I would say that what you need most is a wardrobe, I would put a bookshelf as well’

## 5.4. Instruments

In both the pretest and posttest, participants completed a gap-fill exercise designed to assess their explicit knowledge of grammatical case. The target items tested the correct use of subject marking (nominative or genitive) and direct object marking (accusative or genitive) in both affirmative and negative statements. Most of the test sentences were adapted from authentic online sources to ensure natural language use. Besides the relevance of authentic input in communicative teaching approaches such as TBLT, the use of more naturalistic test items was considered appropriate because the target constructions are generally introduced at an early stage of L2 instruction. Learners at roughly the A1 level are therefore likely to have encountered subject and direct object marking in mechanical exercises commonly found in A1-level textbooks, making test items containing authentic language a suitable challenge.

Each sentence included a target noun provided in the nominative singular base form in brackets. Participants were instructed to fill in the correct case form in the blank. No translations were provided. See the example below:

- (8) *Jeį butas neturi \_\_\_\_\_ (balkonas), tai yra didelį minusą.*  
'When an apartment does not have a (balcony), that is a big disadvantage.'

Each test consisted of 26 items: four fillers, 11 items requiring subject marking, of which four included constructions with the verb *patikti* ('like'), and 11 targeting direct object marking. Out of the subject marking items, six targeted the nominative and five the genitive. For the direct object marking items, the pretest included five targeting the genitive and six the accusative, while the posttest included six items targeting the genitive and five the accusative. All target words were singular nouns, as plural forms are typically introduced at a later stage in case instruction. This also reduced the risk of subjects and direct objects being marked with the genitive according to its partitive function, i.e., denoting indefinite quantities. Furthermore, target nouns were always presented without accompanying adjectives.

Target nouns belonged to four different declension paradigms (*-as*, *-is/-ys*, *-a*, *-ė*), which are considered the most frequent in Lithuanian (Bružaitė-Liseckienė, 2021, p. 58). Besides the four noun paradigms mentioned above, there are eight further paradigms, which, according to calculations presented in Bružaitė-Liseckienė (2021 p. 58) account for approximately 4% of all nouns in Lithuanian.<sup>2</sup> Earlier studies (Bružaitė-Liseckienė 2021; Stepšys and Kamandulytė-Merfeldienė 2024) have shown that low-frequency paradigms are often replaced by high-frequency ones by L2 learners. Given the very low frequency of these remaining paradigms, the decision was made to exclude them in this experiment to enhance the focus on case assignment and the most frequent instances of case marking. Nevertheless, the frequency amongst the four most common paradigms is not uniform, which was reflected in the study design. The number of items for each target paradigm further depended on their relative frequency: eight items contained a target word ending with *-as*, six – with *-is/-ys*, six – with *-ė*, and two – with *-a*.

Target nouns were supposed to belong to the 2500 most frequent lemmas according to the Wordlist of lemmas from the joint corpus of Lithuanian (Dadurkevičius 2020).<sup>3</sup> If they did not, they had to be present in either one of the most frequently used textbooks of L2 Lithuanian (*Nė dienos be lietuvių kalbos* (2018) or *Sėkmės!* (2022)). The same criteria applied to the verbs accompanying these target

<sup>2</sup> These numbers, however, disregard the frequency of use of each separate noun. Hence, high frequency nouns can belong to relatively unproductive paradigms, such as *vanduo* 'water' and the paradigm *-uo*.

<sup>3</sup> The 2500 lemma range is taken from Milton (2010:224).

items, as valency is considered important for the acquisition of case marking. The only exception was the verb *užduoti* ‘give/set’, which did not meet both criteria, but was nevertheless included due to use in the collocational context of *užduoti klausimą* ‘ask a question’, which L2 students were thought to encounter regularly. Some target nouns were used twice in the tests, although never with the same type of syntactic marking. The full list of target nouns and verbs is provided in Appendix A. The full list of items is provided in Appendix B.

## 5.5. Data analysis

Participant responses were evaluated for accuracy, with a few exceptions. Although plural case endings were not required, they were accepted as correct if they reflected the appropriate case.

- (9) *Stoties perone jau nebuvo **traukinių**, mes pavėlavome.*  
‘There were no more **trains** on the station platform, we arrived too late.’  
Potential contextual inconsistencies were disregarded.
- (10) *Sidnėjus – šalies ekonominis centras, uostas. Mieste yra ir **universitetai**.* [should be genitive, *universitetų* ‘some universities’]  
‘Sydney is the country’s economic center and port. It is also home to universities.’
- (11) *Jei nepatinka **mokyklos** – keiskite **ją**.*  
‘If you do not like the **schools**, change **it**.’

Additionally, spelling errors related to case endings were disregarded as irrelevant for the aims of the study. These could be omissions of diacritics or the erroneous absence/presence of the soft sign *i*.

- (12) *Mūsų kaime nėra **parduotuvės**.* [should be *parduotuvės* ‘shop’]  
‘There is no shop in our village.’
- (13) *Vis **lekiamo**, dirbame, nėra **laikio** susitikti su draugais...* [should be *laiko* ‘time’]  
‘We are always on the go, working, no time to see our friends...’

Errors were further analyzed according to type: a different oblique case than appropriate for the given context (oblique case assignment error); incorrect use of the base form, which coincides with the nominative singular (base form error); or the correct case, but with an ending from a different declensional paradigm (paradigm error). While all target nouns in this study belonged to the most frequent declensional paradigms, competition between the most frequent declensional paradigms can still be expected, which is why this aspect was included in the analysis.

Statistical analyses were performed by fitting a Bayesian mixed-effects model using the *brms* package (Bürkner 2017) in *R* (R Core Team 2023). Bayesian models and frequentist logistic mixed-effects models differ in the sources of information that are used for parameter estimations. While estimates in frequentist models are made by using only the data from a single study, Bayesian models include not only the study’s data, but also priors, data we know or assume beforehand. Combining these two sources of information, through Bayes’ theorem, leads to a posterior distribution, which is the updated range of plausible values for each parameter based on the priors and the data. Preference was given to Bayesian models over frequentist logistic mixed-effects models as Bayesian models are said to perform better with small sample sizes which are common in exploratory classroom-based intervention studies such as this one (Wong 2025).

Since the design of this study closely matched that of Wong (2025), most of the model settings were adopted from there. The response variable was binary, requiring an analysis employing the Bernoulli

family with a logit function. As no interventionist research has so far been conducted on the acquisition of case in L2 Lithuanian, weakly informative priors were employed: fixed effects were set for a normal(0, 2) distribution, and random effect standard deviations were applied a Cauchy(0, 1) distribution. Model estimation used four Markov Chain Monte Carlo (MCMC) chains, each with 6000 iterations and a 2000 warm-up. The control parameter `adapt_delta` was set to 0.99. Successful convergence was established by checking `Rhat` (=1.00). Regarding the arguments within the model, the binary outcome variable was accuracy (1 = correct, 0 = incorrect), with fixed effects for Test Time (PRE, POST), Treatment Group (TSLT, TBLT), Syntactic Marking (Subject, Object), and their interactions. L1 (Slavic, non-Slavic) was also added as a fixed effect, given typological similarities with Lithuanian such as the presence of a nominal case system. Items and Participants were included as random effects. Additionally, two alternative models were fitted to examine the influence of potentially confounding variables. One model included only participants with a Slavic L1 (TBLT:  $n = 6$ ), given the disproportionate distribution of Slavic L1s across the groups. The second alternative model treated paradigm errors as accurate responses, since, although paradigm errors are technically mistakes, they nonetheless indicate successful case assignment.

Instead of accepting or rejecting the alternative hypothesis based on a  $p$ -value, Bayesian models are guided by probabilities, which comes down to establishing the strength of the evidence for an effect or interaction. In order to report the findings, Highest Density Intervals (HDIs) were used to summarize posterior uncertainty and a Region of Practical Equivalence (ROPE) was defined to assess the practical significance of the findings. HDIs constitute a range of values from a Bayesian posterior distribution that contains the most credible values of a parameter. In this study, the HDI was set at 89%, which included all values that account for 89% of the most likely parameter estimates. In other words, one could argue, “there is an 89% probability that the true value lies within the reported interval”. A ROPE is a small range of values around zero and is usually employed to check whether an effect is negligible, not so much statistically, but practically. In this study, the ROPE was defined as  $\pm 0.18$  on the log-odds scale, and in cases where large portions of the posterior distribution are situated within the ROPE, this effect is said to be most likely negligible. Additionally, the Probability of Direction ( $pd$ ) was computed to display the degree of certainty that an effect is either positive or negative. Values close to 100% indicate strong evidence for the positive/negative direction.

Contrast coding (specifically, sum-to-zero contrasts) was applied to the categorical predictors as comparisons involved two levels (two treatment groups) without a baseline condition (e.g., a control group). Instead of testing direct group comparisons, sum contrasts assess statistical differences relative to the overall mean across both groups (Schad et al. 2020). This means that, under sum coding, the coefficients for main effects represent the average effect across all levels of the predictor, while interaction terms represent how individual groups deviate from that average. To evaluate group-specific improvement from pre- to posttest, Bayesian hypothesis tests were performed by using the `hypothesis()` function in the `brms` package). They use posterior samples to estimate the probability of whether a hypothesis is true given the data and model (Post.Prob.), the extent to which the data more strongly supports the tested hypothesis over its alternative (Evidence Ratio) as well as the range within which the true value is likely to lie with a given probability (Credible Intervals; CI). Post.Prob values of over 0.95 are considered indicators of strong evidence for an effect.

## 6. Results

The descriptive statistics (Table 1) indicate an improvement in case accuracy for the TBLT group, whereas no such improvement was observed in the TSLT group. Overall, the TSLT group performed better than the TBLT group on both the pretest and posttest.

**Table 1.** Results of the fill-in-the-gap test by treatment group

Treatment group		Mean (max = 22)	SD
TBLT (n = 10)	Pretest	14.6	5.89
	Posttest	16.4	5.24
TSLT (n = 8)	Pretest	17.8	2.33
	Posttest	17.6	2.96

Table 2 shows the total number of errors and identifiable error types, categorized by the type of syntactic role of the items, and divided by the treatment groups. In terms of subject marking, the TBLT showed a remarkable decrease in errors for all error types, while the TSLT group showed only a modest decrease. Furthermore, while the TBLT group demonstrated a similar number of errors for direct object marking, the TSLT group’s number of errors demonstrated a small increase in total errors. Interestingly, the TBLT group showed an increase in mainly wrong paradigms, which means they assigned the correct case but the wrong ending. The TSLT group, on the other hand, demonstrated a significant increase in oblique case errors.

**Table 2.** Error types by type of syntactic role, treatment group, and test time. Percentages reflect the share of error types out of the total number of identifiable errors within the same test

Syntactic role	Group	Test	Oblique case assignment errors		Base form errors		Paradigm errors		Unid. /NA	Total
Subject	TBLT	<i>Pre</i>	26	(74%)	7	(20%)	2	(6%)	5	<b>40</b>
		<i>Post</i>	18	(78%)	3	(13%)	2	(9%)	0	<b>23</b>
	TSLT	<i>Pre</i>	8	(80%)	1	(10%)	1	(10%)	0	<b>10</b>
		<i>Post</i>	6	(86%)	1	(14%)	0	(0%)	0	<b>7</b>
Object	TBLT	<i>Pre</i>	23	(72%)	7	(22%)	2	(6%)	2	<b>34</b>
		<i>Post</i>	14	(46%)	8	(27%)	8	(27%)	3	<b>33</b>
	TSLT	<i>Pre</i>	12	(55%)	8	(36%)	2	(9%)	2	<b>24</b>
		<i>Post</i>	21	(78%)	4	(15%)	2	(7%)	1	<b>28</b>

Table 3 presents the distribution of total errors and identifiable error types across grammatical cases by treatment group. The TBLT group showed an increase in performance for the nominative with a noticeable reduction of total errors. The number of nominative case-marking errors remained relatively stable for the TSLT group. Both groups showed stable performance in terms of genitive case errors, despite a higher number of genitive-targeting items at the posttest. However, the TBLT group performed better in assigning the correct case, yet more frequently chose the wrong paradigm. Regarding the accusative, the TBLT group showed a slight decrease in errors, possibly caused by the smaller number of items at the posttest, whereas the TSLT group showed an increase, mainly due to assigning the wrong case. This suggests that particularly the TSLT group had problems with the accusative case.

**Table 3** Error types by nominal case, treatment group, and test time. Percentages reflect the share of error types out of the total number of identifiable errors within the same test

Case	Group	Test	Oblique case assignment errors		Base form errors		Paradigm errors		Unid. /NA	Total
Nominative	TBLT	Pre	21	(95%)	//		1	(5%)	1	<b>23</b>
		Post	10	(100%)	//		0	(0%)	0	<b>10</b>
	TSLT	Pre	6	(100%)	//		0	(0%)	0	<b>6</b>
		Post	4	(100%)	//		0	(0%)	0	<b>4</b>
Genitive	TBLT	Pre (n=5)	14	(56%)	9	(36%)	2	(8%)	5	<b>30</b>
		Post (n=6)	10	(37%)	8	(30%)	9	(33%)	2	<b>29</b>
	TSLT	Pre (n=5)	8	(67%)	1	(8%)	3	(25%)	0	<b>12</b>
		Post (n=6)	5	(56%)	2	(22%)	2	(22%)	1	<b>10</b>
Accusative	TBLT	Pre (n=6)	14	(70%)	5	(25%)	1	(5%)	1	<b>21</b>
		Post (n=5)	12	(75%)	3	(19%)	1	(6%)	1	<b>17</b>
	TSLT	Pre (n=6)	6	(43%)	8	(71%)	0	(0%)	2	<b>16</b>
		Post (n=5)	18	(86%)	3	(14%)	0	(0%)	0	<b>21</b>

Table 4 shows the parameter estimates derived by the Bayesian model. The model showed good convergence ( $\hat{R} = 1.00$ ). The significance of the results can be determined by the width of the HDIs and whether it contains zero, the directional strength of the *pd*, and the ROPE overlap with the posterior distribution in percentages, with low overlap denoting more significant results. There appears to be a slight positive effect ( $\beta = 0.19$ ) for the posttest over the pretest on average for both groups; however, the 89% HDI [-0.07, 0.44] contains zero, and a substantial number of estimates lies within the ROPE (47.88%), making the evidence for improvement weak to moderate at best. As for individual group performance, Bayesian hypothesis tests indicated very strong evidence supporting pre- to posttest for the TBLT group (Estimate = 0.31, 89% CrI [0.02, 0.60], Evid.Ratio = 23.17, P(diff > 0) = 0.96). In contrast, evidence for improvement in the TSLT group was weak (Estimate = 0.07, 89% CrI [-0.26, 0.39], Evid.Ratio = 1.76, P(diff > 0) = 0.64).

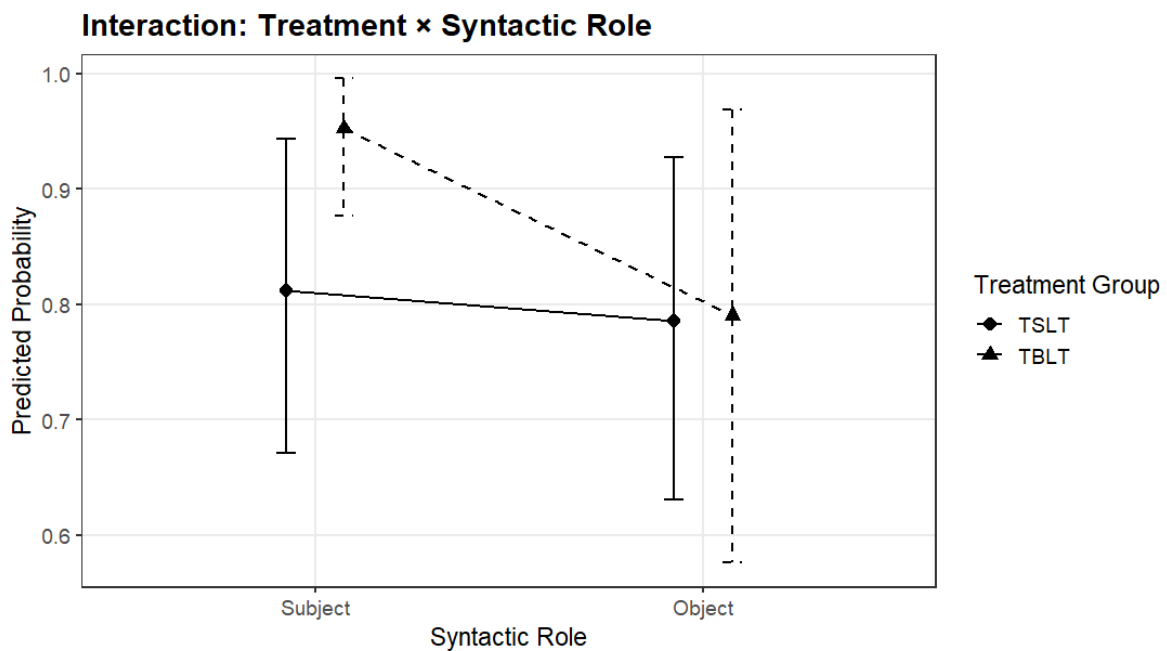
**Table 4.** Bayesian posterior summary for fill-in-the-gap test

Parameter	Median	HDI (89%)		<i>pd</i>	% in ROPE ( $\pm 0.18$ )
		Low	High		
Intercept	1.77	0.93	2.64	99.93%	0%
Test Time	0.19	-0.07	0.44	88.23%	47.88%
Treatment	-0.39	-1.11	0.34	81.59%	22.22%
Syntactic Role	-0.46	-0.72	-0.21	99.69%	1.42%
L1	-0.21	-1.05	0.65	66.30%	26.86%
Test time : Treatment	0.12	-0.05	0.29	87.62%	72.27%
Syntactic Role : Treatment	0.38	0.20	0.55	99.99%	0.86%
Test time : Syntactic Role	-0.24	-0.49	0.01	94.04%	33.36%

*Note.* The reference levels for each variable were as follows: Test Time: Pretest; Treatment: TSLT, Syntactic Role: Subject; L1: non-Slavic. These same levels also applied to the interactions. HDI = Highest Density Interval, *pd* = Probability of Direction, ROPE = Region of Practical Equivalence.

Continuing with the other variables, while the TSLT group seemingly outperformed the TBLT group over both tests ( $\beta = -0.39$ ), the 89% HDI [-1.11, 0.34] is wide, contains zero, the *pd* is relatively weak (81.59%), and a moderate chunk of estimates overlaps with the ROPE (22.22%), suggesting there is little evidence for substantial group differences. Importantly, the target syntactic role of items did seem to impact the results heavily. Object marking was clearly more difficult for the participants ( $\beta = -0.46$ ). The evidence for this is very strong and robust, as the 89% HDI [-0.72, -0.21] does not contain zero and overlap with the ROPE is minimal (1.42%). Finally, whether the participants' L1 belonged to the Slavic language group seemed not to have any bearing on the results, judging from the wide 89% HDI [-1.05, 0.65], which also contains zero, and the very weak *pd* (66.30%).

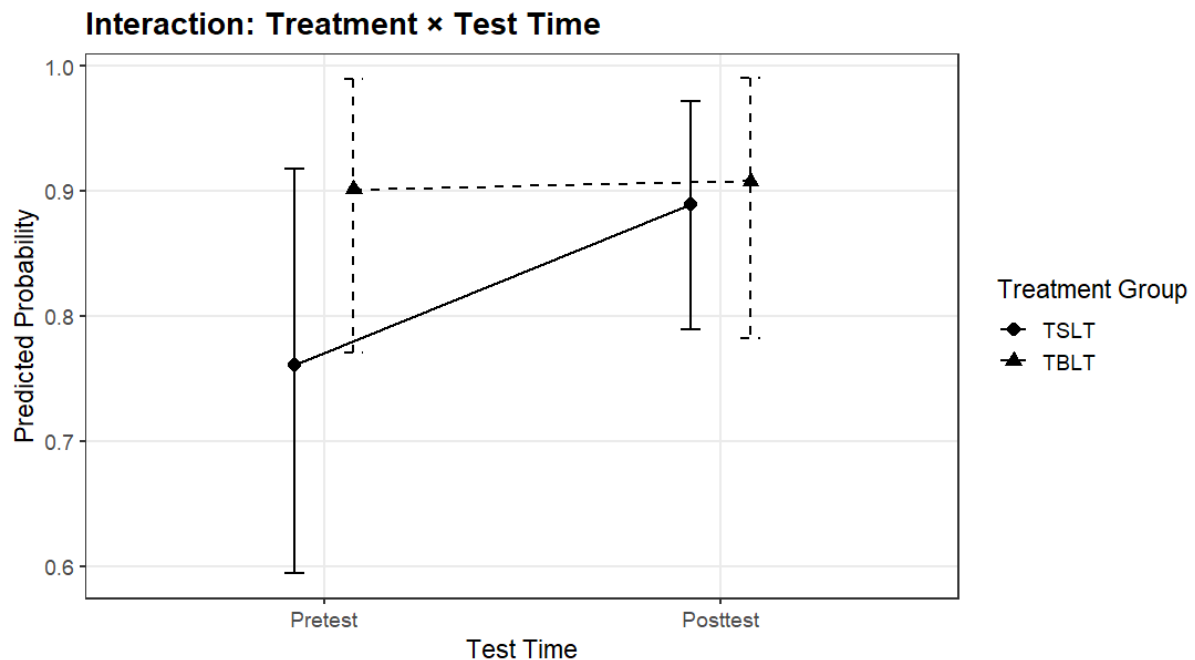
Examining interactions, there is little evidence of a difference in pre-post improvement between groups ( $\beta = 0.12$ , 89% HDI [-0.05, 0.29], *pd* = 87.62%, ROPE = 72.27%). In other words, although the TBLT group showed strong evidence of improvement on its own, there is little to no evidence that its gains exceeded those of the TSLT group. The interaction between Test and Syntactic role seems slightly more substantial, which might indicate worse performance or impeded improvement for object marking than subject marking in the posttest ( $\beta = -0.24$ , 89% HDI [-0.49, 0.01], *pd* = 94.04%, ROPE = 33.36%). Nevertheless, the evidence is not very strong. However, a strong and robust interaction can be found between the treatment group and their performance on either subject or direct object marking items ( $\beta = 0.38$ , 89% HDI [0.20, 0.55], *pd* = 99.99%, ROPE = 0.86%). As can be seen in Figure 3, the TBLT group's performance on both subject and object marking items was stable, whereas the TSLT group showed far superior performance on Subject marking than on Object marking items. As shown in Table 2, not only was the difference in the number of errors of the TSLT group between subject and object items quite large at the pretest, but the number of errors for object items increased in the posttest, which most likely contributed to this robust interaction between Treatment and Syntactic Role.



**Figure 3.** Predicted probabilities of correct responses by Treatment Group and Syntactic Role. Error bars represent the 89% HDIs. The figure illustrates the interaction between Treatment and Syntactic Role based on Bayesian logistic regression estimates

Pairwise comparisons further corroborated these findings. Participants in the TSLT group had markedly lower odds of answering correctly on object marking items compared to subject marking ones (OR = 0.189, 89% HDI [0.08, 0.33]), while, in contrast, almost no difference was found in the TBLT group (OR = 0.849, 89% HDI [0.41, 1.35]). Conversely, the TSLT demonstrated a much higher likelihood of providing a correct response than the TBLT group for subject-marking items (OR = 0.215, 89% HDI [0.01, 0.68]), while object-marking items showed no difference between the groups (OR = 0.962, 89% HDI [0.05, 2.98]). In summary, these results suggest that linguistic features of test items, rather than treatment group alone, played a key role in explaining learners' performance.

Finally, the results from the two additional models need to be considered (see Appendix C for their Bayesian posterior summaries). The first alternative model included only participants with a Slavic L1. The results for all main effects and interactions were very similar to those of the original model and will therefore not be discussed further here. The other alternative model included responses in which paradigm errors were accepted. While the results were largely consistent with those reported in Table 4, the interaction between Treatment Group and Test Time showed moderately strong evidence for superior pre- to posttest improvement of the TBLT group compared with the TSLT group ( $\beta = 0.22$ , 89% HDI [0.04, 0.39],  $pd = 97.75\%$ ,  $ROPE = 36.07\%$ ). A strong increase in performance for the TBLT group can be seen in Figure 4, in contrast to the very stable performance of the TSLT group. The sudden increase in strength of this interaction when accepting paradigm errors suggests particular difficulty for the TBLT group in selecting endings from the appropriate paradigms.



**Figure 4.** Predicted probabilities of correct responses by Treatment Group and Test Time. Error bars represent the 89% HDIs. The figure illustrates the interaction between Treatment and Test Time based on Bayesian logistic regression estimates

## 7. Discussion

Treatment consisting of one task yielded inconclusive results. While both groups successfully completed the task, the Bayesian hypothesis tests revealed strong evidence of progress for the TBLT group, but not for the TSLT group. Moreover, there was little evidence for superior pre-post improvement of the TBLT group compared with the TSLT group, unless responses with paradigm errors were considered acceptable. These findings could be interpreted as potentially supporting the effectiveness of the TBLT approach, with its limited inclusion of explicit grammar instruction, relative to the TSLT approach. Nevertheless, it is surprising that only the TBLT group showed an increase in performance for case marking, while the TSLT group did not, not even according to the mean results. Given that gap-fill exercises are typically designed to assess explicit knowledge, improved performance would be more likely for the TSLT group, not the TBLT group. In Markina's (2018) study, both groups showed gains in mean performance for the gap-fill exercises on the immediate posttest.<sup>4</sup> This further underscores the surprising lack of improvement of the TSLT group in the present study.

Our data also allow for a comparison of group performance in terms of the error amounts and error types by type of syntactic marking and case. Firstly, the performance of both treatment groups, as presented in Table 2 and corroborated by the Bayesian model, varies noticeably depending on whether items involve subject or object marking. The TSLT group demonstrated a high baseline performance in subject marking, whereas the TBLT group performed significantly worse. This strong initial performance by the TSLT group likely limited the potential for observable improvement in the posttest. In contrast, the TBLT group demonstrated learning gains, with a reduction in errors across almost all types. Object marking, however, displayed a markedly different pattern. While the TSLT group outperformed the TBLT group in both the pre- and posttests, their performance nevertheless slightly declined in the posttest. In contrast, the TBLT group's performance remained stable. Notably, although the overall number of errors in the TBLT group did not change significantly, there was a marked reduction in oblique case errors and a corresponding increase in paradigm errors. This suggests that participants improved in correctly assigning grammatical case but were still refining their understanding of the complex declension paradigms. The decline in the TSLT group's performance appears to be due specifically to increased oblique case errors, rather than paradigm-related issues. Finally, the model identified a significant performance difference between subject- and object-marking items. This discrepancy may be attributed to the relative ease of guessing correct answers for subject-marking items: target words were consistently presented in their base form, i.e., the nominative singular, which was the correct form for all non-negated subject items and for those involving the verb *patikti* ('to like').

An analysis of errors and error types by grammatical case, as presented in Table 3, reveals several other noteworthy patterns. For the genitive case, the TBLT group demonstrated almost no change in base form errors, but this coincided with a decrease in oblique case and unidentifiable errors, as well as an increase in paradigm errors. This might suggest that participants became better at identifying the correct grammatical case but continued to struggle with selecting the appropriate morphological forms. The TSLT group, however, maintained a strong baseline performance for both the nominative and genitive cases, which resulted in stable posttest outcomes. Nevertheless, a different trend can be discerned for the accusative case. While the TBLT group showed a decrease in errors, the TSLT group demonstrated an increase, specifically in errors related to case assignment. This indicates that the TSLT group's difficulties with direct object marking were mainly related to challenges in assigning the accu-

<sup>4</sup> Importantly, besides the two different target L2s (L2 Lithuanian and L2 Russian), Markina's study also examined a different target construction, namely prepositional case government. Nevertheless, this linguistic operation is quite comparable to subject and direct object marking as both are similarly grammatical and carry minimal semantic content.

sative case, rather than the genitive. From this follows that TSLT participants, by the posttest, had become less proficient in accurately assigning the accusative case to direct objects of non-negated verbs.<sup>5</sup>

Nevertheless, an analysis of the items that the majority of participants responded to erroneously reveals some potential limitations related to item design. Starting with the TSLT group, the problematic items all involved accusative marking, which aligns with the patterns observed in the error analyses. One important reason for the TSLT group's difficulties with the accusative case might be rooted in the target words themselves. Although the variable *item* was included in the statistical model as a random effect, closer analysis of the outlier items revealed the influence of possible frequency and salience effects. This can be seen in some of the following example answers by the TSLT group in the posttest:

- (14) \**Kalėdų eglutė ir vėl puošia miest-e.*  
Christmas tree again decorates city-LOC  
The Christmas tree once again decorates **in the city.**'
- (15) \**1946 m. persikėlėme į Panevėžį, ten tėvai nusipirko nam-o.*  
in 1946 moved.1p.PL to Panevėžys.ACC there parents bought **house-GEN**  
'In 1946 we moved to Panevėžys, there my parents bought **home.**'
- (16) \**Ši įmonė atidarė viešbuč-io gyvūnams.*  
this company opened hotel-GEN animals.DAT  
'This company opened a **hotel** for pets.'

When the target word referred to a location, learners of both groups frequently used the locative case in inappropriate contexts like in example 14. This is understandable, as such words are more commonly encountered and used in the locative case in everyday language. Another clear example is the target word *namas* 'house', which was often marked in the genitive as *nām-o* 'house-GEN'. More precisely, learners likely intended to use the adverb *namo* 'home', which is identical in form to the genitive but differs in stress and might be more salient within participants' interlanguage system. Finally, the TSLT group predominantly favored the genitive form of the noun *viešbutis* 'hotel', which may reflect a salience effect resulting from the palatalized ending (i.e., *-čio* rather than *-tio*). On the other hand, the overall frequency of the genitive case in the Lithuanian language should not be ignored. In L1 Lithuanian texts of various genres the genitive is the most frequent case, even more so than the nominative base form (Brinkutė 2018: 53). This might also have led to overuse of genitive forms in these contexts.

An alternative explanation for these and other frequent errors may lie in the selection process of target nouns and their accompanying verbs. Example 14, for instance, included the verb *puošti* 'decorate', a relatively low-frequency verb that may have been unfamiliar to learners, despite its inclusion in one of the textbooks. Two other items that proved challenging for the majority of the TSLT group were *įjungti virdulį* 'turn on the kettle' and *užduoti klausimą* 'ask a question'. *Įjungti virdulį*, especially, contains two relatively low-frequency words. Furthermore, the rationale that the presence of *užduoti* in a collocation might facilitate processing proved false, at least for the TSLT group. As beginner L2 learners seem to primarily rely on item-based learning for case when encountering new nouns Kempe and Brooks (2008), a stricter selection process for both target nouns and their accompanying verbs seems warranted. Both items also included an imperative and past-tense form, possibly further complicating comprehension of the items and underscores the need for more uniformity in verbal forms.

<sup>5</sup> These findings can also be compared to Denissenko's (2016) study, where learner performance was similarly broken down by case. In the posttest, Denissenko's A1-A2 group exhibited a slight decrease in nominative errors, a substantial reduction in genitive errors, and a slight increase in accusative errors. Participants in our study showed somewhat comparable patterns, with decreases in errors for the nominative for the TBLT group and an increase in accusative errors for the TSLT group.

In summary, these observations suggest that the salience of lexical items and/or grammatical constructions, case frequency, and lemma frequency should be carefully considered when designing experimental instruments, a concern that can be at least partially addressed through piloting.

Another challenge for this study concerning the items was the inclusion of target sentences containing the verb *patikti* ‘to like’. As mentioned previously, unlike standard subject and direct object constructions, the subject of *patikti* does not change to the genitive case when the verb is negated but remains in the nominative. Although this is an important grammatical feature, such constructions were hardly represented in the task input and, in hindsight, do not align well with the study’s primary focus on subject and direct object case marking. This can be illustrated by the performance of the TBLT group, as the only two items with which the majority of this group struggled across both tests were two negated *patikti* items at the pretest. Furthermore, items involving negated *patikti* posed an additional challenge for accurate evaluation. The *-a* and *-ė* paradigms share identical forms for the nominative plural and genitive singular (e.g., *mokyklos* ‘school.GEN/schools’), making it difficult to determine whether participants had erroneously selected the genitive case or had selected an alternative nominative plural form. Although most of these plural forms were contextually inappropriate, the decision was made to disregard contextual discrepancies when evaluating alternative plural responses. To assess the impact of these items, the results were reanalyzed by fitting another Bayesian mixed-effects logistic regression model with the exclusion of the relevant sentences (four target items per test). While this model revealed only minimal changes, the Bayesian hypothesis test for the TBLT group now yielded only weak to moderate evidence for pre- to posttest improvement (Estimate = 0.15, 89% CrI [-0.15, 0.46], Evid.Ratio = 4.08, P(diff > 0) = 0.80). Hence, it would be advisable to repeat the experiment without these items in order to better align input and target structures and to maximize the number of analyzable target items.

Despite potential limitations of the target items, data analysis by error types might shed light on differing developmental pathways in this study. The Bayesian model showed a significant interaction between Treatment and Syntactic Role, which indicated that the TBLT group’s performance was less affected by these roles than the TSLT group’s. While the significant interaction between treatment group and syntactic role was mainly caused by the superior subject marking knowledge of the TSLT group, their inferior performance on object marking at the posttest likely only widened the gap. The error analyses described above have shown that (a) the TBLT group showed a decrease in case assignment errors for object marking that was accompanied by a rise in paradigm errors, which show accurate case assignment, and (b) the decrease in object marking performance at the posttest for the TSLT group was mainly due to a rise in case assignment errors, specifically the accusative. Observation (a) is additionally corroborated by the alternative model treating responses containing paradigm errors as correct, which showed strong evidence for superior pre-post improvement for the TBLT group, suggesting strong influence of inaccurate paradigm selections for this group. Importantly, both observations might indicate a subtle difference in developmental patterns. It could be argued that the true TBLT task, of which the aim was to mainly foster communicative skills, was more effective in instilling the logic of case assignment, however, less so in teaching the appropriate endings. The TSLT task version, on the other hand, included a much heavier focus on case endings, which may have been to the detriment of the acquisition of appropriate case assignment. This might also explain why primarily the TSLT group used inappropriate yet frequent and/or salient case endings.<sup>6</sup> However, in order to illuminate this, a bigger sample size and a longer intervention are surely needed.

<sup>6</sup> As one anonymous reviewer notes, the importance of lexical variety in the input should not be underestimated in the process of L2 acquisition of declensional paradigms. However, while both task versions had identical lexical variety in the input, the TSLT version included input enhancement of case forms, which is believed to enhance noticing (Ellis et al. 2019). As a result, while input enhancement might have drawn participants’ attention to endings, it might also have distracted from accurate case assignment.

Lastly, the differences between the two groups highlight a potential limitation of the study. While the Bayesian model did not reveal strong evidence for group differences over all tests, it should be remembered that the groups differed markedly in their baseline performance levels. While the TBLT group began with a mean accuracy of approximately 66% on the pretest and improved to 75% on the posttest, the TSLT group demonstrated substantially higher accuracy at the outset (81%) and remained at the same level following the intervention (80%). This suggests that the TSLT group had less room for improvement due to their already more advanced grammatical knowledge. The TSLT group in Markina's (2018) study demonstrated superior case knowledge on the pretest compared to the TBLT group, achieving superior learning gains; however, this was after considerably longer treatment.

One possible reason for these differences might be the L1 composition of both groups. Many studies suggest that a learner's L1 significantly impacts the development of case marking (Bružaitė-Liseckienė 2021; Kenanidis et al. 2023; Stepšys and Kamandulytė-Merfeldienė 2024). On the other hand, several studies present contrasting findings. Dubasava (2020) asserts that having a morphologically complex L1 offers little clear advantage for the successful acquisition of L2 case, having observed very similar L2 Lithuanian errors among learners with non-Slavic L1s. Similarly, Saturno (2023) argues that the potential of positive transfer in the L2 due to a typologically similar L1 should not be overstated, noting frequent mistakes by Russian L1 in L2 Polish object marking that were also common to speakers of other L1s.<sup>7</sup> These perspectives and findings are consistent with the results of the Bayesian models, which showed little to no evidence of L1 influence. Nevertheless, positive L1 transfer may occur in a more subtle manner: a complex nominal system in the L1 may support case assignment but not the selection of the relevant paradigm ending. Indeed, Saturno (2023) found that L1 Russian learners of L2 Polish frequently marked the accusative correctly but with erroneous Russian endings. Accommodating this phenomenon would require a more lenient approach to paradigm errors. Naturally, the sample size of this study was too small and homogeneous to allow for robust analyses. Future studies should therefore include groups with a wider variety of L1 backgrounds to explore these claims more thoroughly.

The instructional contexts of the two groups might have also impacted the results as they differed quite strongly: the TBLT group was enrolled in a course with a stronger emphasis on communicative competence, whereas the TSLT group's course was more focused on preparing students for a state exam, which included greater attention to grammar instruction. This might explain the superior baseline knowledge of grammatical case for the TSLT group, despite supposedly similar proficiency levels between groups. A more important limitation, however, may have been the small number of participants and/or the limited treatment length. Therefore, while only the TBLT group demonstrated strong evidence of case-related learning gains from pre- to posttest, the high initial baseline knowledge of the TSLT group, along with other limitations discussed above, precludes drawing firm conclusions regarding the effectiveness of a single task in promoting explicit case knowledge, particularly with regard to the timing and extent of grammar instruction within the task cycle.

## 8. Conclusion

In the present study, two groups of learners completed a TBLT-aligned task in L2 Lithuanian, a morphologically complex language. The task was semi-focused, meaning it was designed to provide input rich in subject and direct object marking and to (implicitly) elicit these constructions in learners' out-

<sup>7</sup> Granted, Saturno (2023) additionally mentions that despite comparable errors, typologically similar L1s can speed up the acquisition process. This seems to align with the claim made by Lightbown and Spada (2021:104) that the influence of the L1 increases with the proficiency level of the L2. As the participants in this study were still beginner learners, it may be argued that positive L1 transfer must have been quite limited, though this requires more research.

put. These structures constituted the primary focus of the study. The two task versions differed in the placement and extent of focus on form: the TBLT group received explicit focus on form only during the post-task phase, whereas the TSLT group carried out a version that included focus on form throughout the entire task cycle. The study aimed to examine the efficacy of TBLT in facilitating the acquisition of complex grammatical categories, such as case, and in fostering explicit grammatical knowledge of such categories, by exploring the optimal timing and extent of focus on form within the task cycle.

Although participants in both groups successfully completed the task, statistical analyses provided strong evidence for learning gains in explicit knowledge of grammatical case only for the TBLT group. There was no strong evidence for superior improvement in comparison to the TSLT group, however, unless paradigm errors were considered acceptable. It is striking that the TBLT group, rather than the TSLT group, demonstrated learning progress, despite the tests being designed to assess explicit knowledge. Analysis of learner errors per case and type showed further differences between the groups, which could indicate differing developmental patterns depending on the type of treatment received. The results suggest that the TBLT group struggled mainly with selecting appropriate nominal paradigms, whereas the TSLT group was mostly challenged by the accurate assignment of the accusative case.

However, certain item- and group-related limitations may have influenced the results. Frequency and/or salience effects of certain target words and constructions, case frequency, as well as limitations due to item design, may have affected participants' performance. Differences between the two groups may also have played a role in shaping the results. The TSLT group demonstrated stronger case knowledge from the pretest, leaving less room for measurable improvement. Additionally, participants in each group were enrolled in very different L2 Lithuanian courses, potentially shaping the nature of their baseline case knowledge. And while the model showed no significant effect of participants' L1, the groups were disproportionate in terms of the distribution of typologically close L1s, an imbalance that should be addressed in future follow-up studies.

The results of this study indicate that the TBLT methodology has the potential to foster explicit knowledge of complex grammatical categories, such as case, even after the completion of just one task. Furthermore, the results indicate certain potentially differing developmental patterns that need to be explored further. Nevertheless, given the high baseline knowledge exhibited by the TSLT group and the other limitations provided above, no definitive conclusions can currently be drawn regarding the specific role of the timing and extent of explicit focus on form within the task cycle may play in this process. Further research, addressing these limitations with a larger sample size and/or extended treatment, is needed to explore these findings more thoroughly.

## Acknowledgment

This research was conducted in cooperation with the VDU Education Academy in Vilnius, Lithuania. The author gratefully acknowledges Vilma Leonavičienė and the Academy's staff and teachers for their support and contributions to this study.

## References

- Brezina, V., G. Pallotti. 2019. Morphological complexity in written L2 texts. *Second Language Research*, 35(1), 99–119. doi:10.1177/0267658316643125.
- Bružaitė-Liseckienė, J. 2021. *Objekto raiška lietuvių kalboje: gimnazijų rusų ugdomąja kalba moksleivių atvejis*. Vilnius University, Vilnius. Doctoral dissertation.

- Bürkner, P. C. 2017. Brms: An R Package for Bayesian multilevel models using Stan. *Journal of Statistical Software*, 80, 1–28. doi:10.18637/jss.v080.i01.
- Comer, W. J. 2012. Communicative language teaching and Russian: The current state of the field. *Russian Language Studies in North America*. V. Makarova (ed.). London: Anthem Press. 133–59.
- Dabašinskienė, I., L. Čubajevaitė. 2009. Acquisition of case in Lithuanian as L2: Error analysis. *Eesti Rakenduslingvistika Uingu Aastaraamat*, (5), 47–65.
- 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]*. Universitat Pompeu Fabra, Barcelona. Doctoral dissertation.
- Dubasava, A. 2020. Acquisition of noun inflection in Lithuanian as a foreign language: A qualitative study. *Respectus Philologicus*, 37(42), 62–77. doi:10.15388/RESPECTUS.2020.37.42.39.
- Ellis, N. C. 2022. Second language learning of morphology. *Journal of the European Second Language Association*, 6(1), 34–59.
- Ellis, R. 2003. *Task-Based Language Learning and Teaching*. Oxford University Press.
- Ellis, R., P. Skehan, S. Li, N. Shintani, C. Lambert. 2019. *Task-Based Language Teaching: Theory and Practice*. Cambridge University Press.
- Ellis, R. 2009. Task-based language teaching: Sorting out the misunderstandings. *International Journal of Applied Linguistics*, 19(3), 221–46. doi:10.1111/j.1473-4192.2009.00231.x.
- Gagarina N., M., D. Voeikova. 2009. Acquisition of case and number in Russian. *Development of Nominal Inflection in First Language Acquisition: A Cross-Linguistic Perspective*. U. Stephany, M. D. Voeikova (eds.). Berlin: De Gruyter Brill. 179–216.
- Gilabert, R., J. Castellví. 2019. Task and syllabus design for morphologically complex languages. *The Cambridge Handbook of Language Learning*, J. Schwieter, A. Benatti (eds.). Cambridge University Press. 527–49.
- Kempe, V., P.J. Brooks. 2008. Second language learning of complex inflectional systems. *Language Learning*, 58(4). doi:10.1111/j.1467-9922.2008.00477.x.
- Kenanidis, P., E. Dąbrowska, M. Llompart, D. Pili-Moss. 2023. Can Adults Learn L2 Grammar after Prolonged Exposure under Incidental Conditions? *PLoS ONE*, 18(7 July), 1–40. doi:10.1371/journal.pone.0288989.
- Klein, W. 1998. The contribution of second language acquisition research. *Language Learning*, 48(4), 527–49. doi:10.1111/0023-8333.00057.
- Kogan, V., A. Révész, S. Cheung. 2024. The effect of task authenticity on second language writing product and process: The case of a morphologically complex language–Russian. *Foreign Language Annals*, 481–500. doi:10.1111/FLAN.12789.
- Kogan, V., M. Bondarenko. 2022. Russian and Russia through tasks for beginners. *Task-Based Instruction for Teaching Russian as a Foreign Language*, S. V. Nuss, W. W. Martelle (eds.). London: Routledge. 77–97.
- Lightbown, P.M., N. Spada. 2021. *How Languages are Learned*. Oxford: Oxford University Press.
- Long, M. H. 2014. *Second Language Acquisition and Task-Based Language Teaching*. John Wiley & Sons.
- Markina, E. 2018. *Comparing Focus on Forms and Task-Based Language Teaching in the Acquisition of Russian as a Foreign Language*. University of Barcelona. Doctoral dissertation.
- 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. *Communicative Proficiency and Linguistic Development: Intersections Between Sla and Language Testing Research*. I. Bartning, M. Martin, I. Vedder (eds.). Eurosla. 211-232.
- Nuss, S. V. 2022. Morphology acquisition research meets instruction of L2 Russian. *Task-Based Instruction for Teaching Russian as a Foreign Language*, S. V. Nuss, W. W. Martelle (eds.). London: Routledge. 15–35.

- Perdue, C. 1993. *Adult Language Acquisition: Crosslinguistic Perspectives*. edited by C. Perdue. Cambridge: Cambridge University Press.
- Pomarolli, G., D. Artoni, M. Boschiero, S. Piccinin. 2023. Inclusive and accessible teaching of Russian L2 to absolute beginners: A comparison between TBLT and PPP. *Instructed Second Language Acquisition*, 7(2), 166-193. doi:10.1558/isla.26047.
- R Core Team. 2023. *R: A Language and Environment for Statistical Computing*. Internet access: <https://www.r-project.org>.
- Saturno, J. 2023. Transfer (or lack thereof) and the accusative case in L2 Polish. *Language, Creoles, Varieties: From Emergence to Transmission*. C. Granget, I. Repiso, F.S. Guillaume (eds.). Berlin: Language Science Press. 303-342.
- Savickienė, I. 2005. Linksnių vartojimo dažnumas ir daiktavardžio reikšmė. *Acta Linguistica Lithuanica*, 52, 59–65.
- Savickienė, I. 2006. Linksnių kategorijos įsisavinimas: lietuvių kalba kaip gimtoji ir svetimoji. *Kalbotyra*, 56(3), 122–29.
- Schad, D. J., S. Vasishth, S. Hohenstein, R. Kliegl. 2020. How to capitalize on a priori contrasts in linear (mixed) models: A tutorial. *Journal of Memory and Language*, 110, 1–40. doi:10.1016/j.jml.2019.104038.
- Sheen, R. 2003. Focus on Form—a Myth in the Making? *ELT Journal*, 57(3), 225–33. doi:10.1093/elt/57.3.225.
- Slabakova, R. 2019. The bottleneck hypothesis updated. *Three Streams of Generative Language Acquisition Research: Selected Papers from the 7th Meeting of Generative Approaches to Language Acquisition – North*. T. Ionin, M. Rispoli (eds.). Amsterdam: John Benjamins. 319–46.
- Stepšys, J., L. Kamandulytė-Merfeldienė. 2024. Būdvardiškųjų žodžių derinimo sunkumai: eksperimentinis dvikalbių vaikų ir svetimkalbių suaugusiųjų kalbos tyrimas. *Deeds and Days*, 81, 117-136. doi:10.7220/2335-8769.81.7.
- Swan, M. 2005. Legislation by hypothesis: The case of task-based instruction. *Applied Linguistics*, 26(3), 376–401. doi:10.1093/applin/ami013.
- Stumbrienė, V., I. Daraškienė, L. Vaškevičienė. 2022. *Sėkmės!* (Books 1 and 2). Vilniaus universiteto leidykla.
- Stumbrienė, V., A. Kaškelevičienė. 2018. *Nė dienos be lietuvių kalbos*. Vilniaus universiteto leidykla.
- VanPatten, B. 2015. Input processing in adult SLA. *Theories in Second Language Acquisition: An Introduction*. B. VanPatten, J. Williams (eds.). Routledge. 113–34.
- Willis, D., J. Willis. 2007. *Doing Task-Based Teaching*. Oxford University Press.
- Willis, J. 1996. *A Framework for Task-Based Learning*. Longman.
- 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. doi:<https://doi.org/10.1016/j.rmal.2025.100231>.

---

Richard Udes is a doctoral student at the Faculty of Philology, Vilnius University. His research focuses on second language acquisition, particularly Task-Based Language Teaching and the acquisition of morphologically complex languages, with a special emphasis on grammatical case.

---

*Submitted September 2025*  
*Accepted January 2026*

## Appendixes

### Appendix A – List of target nouns and accompanying verbs

Target noun	Place in frequency list	Present in textbook	Accompanying verb	Place in frequency list	Present in textbook
<i>baimė</i> ‘fear’	2194	<i>no</i>	<i>atidaryti</i> ‘open’	1472	yes
<i>balkonas</i> ‘balcony’	4736	yes	<i>baigti</i> ‘finish’	376	yes
<i>butas</i> ‘apartment’	454	yes	<i>būti</i> ‘be’	2	yes
<i>darbas</i> ‘work’	37	yes	<i>įjungti</i> ‘turn on’	4070	yes
<i>darželis</i> ‘kindergarten’	1898	yes	<i>jausti</i> ‘feel’	1345	<i>no</i>
<i>galimybė</i> ‘possibility’	232	<i>no</i>	<i>lankyti</i> ‘visit’	2052	yes
<i>grupė</i> ‘group/band’	167	yes	<i>matyti</i> ‘see’	330	yes
<i>kambarys</i> ‘room’	1311	yes	<i>mėgti</i> ‘like’	1382	yes
<i>kavinė</i> ‘cafe’	1932	yes	<i>nusipirkti</i> ‘buy’	1910	yes
<i>klausimas</i> ‘question’	137	yes	<i>patikti</i> ‘like’	1178	yes
<i>laikas</i> ‘time’	103	yes	<i>puošti</i> ‘decorate’	3597	yes
<i>langas</i> ‘window’	1354	yes	<i>skaityti</i> ‘read’	800	yes
<i>miestas</i> ‘city’	105	yes	<i>sutvarkyti</i> ‘tidy’	2086	yes
<i>mokykla</i> ‘school’	185	yes	<i>turėti</i> ‘have’	19	yes
<i>namas</i> ‘house’	534	yes	<i>užduoti</i> ‘give/set’	4519	<i>no</i>
<i>nuomonė</i> ‘opinion’	286	yes			
<i>paplūdimys</i> ‘beach’	2898	yes			
<i>parduotuvė</i> ‘shop’	750	yes			
<i>prieškambaris</i> ‘antechamber/hall’	21903	yes			
<i>rūsys</i> ‘basement’	3486	yes			
<i>sąskaita</i> ‘check’	672	yes			
<i>saulė</i> ‘sun’	1123	yes			
<i>šeima</i> ‘family’	255	yes			
<i>stalas</i> ‘table’	1417	yes			
<i>straipsnis</i> ‘article’	20	<i>no</i>			
<i>traukinys</i> ‘train’	1882	yes			
<i>universitetas</i> ‘university’	449	yes			
<i>vardas</i> ‘name’	357	yes			
<i>veidas</i> ‘face’	1089	yes			
<i>viešbutis</i> ‘hotel’	1495	yes			
<i>virdulys</i> ‘kettle’	21216	yes			
<i>virtuvė</i> ‘kitchen’	2352	yes			

## Appendix B – List of target items in the instruments

Item id	Test	Item
1	Pre	Man patinka tavo _____ ( <i>klausimas</i> ).
2	Pre	Vis lekiamo, dirbame, nėra _____ ( <i>laikas</i> ) susitikti su draugais...
3	Pre	Sunku patikėti, kad pasaulyje yra _____ ( <i>miestas</i> ), kuriame gyvena apie 38 milijonai žmonių.
4	Pre	Sidnėjus – šalies ekonominis centras, uostas. Mieste yra ir _____ ( <i>universitetas</i> ).
5	Pre	Ką daryti, jei mums nepatinka _____ ( <i>viešbutis</i> )?
6	Pre	Jei nepatinka _____ ( <i>paplūdimys</i> ) priešais vieną viešbutį, galite eiti į kitą.
7	Pre	Stoties perone jau nebuvo _____ (traukinys), mes pavėlavome.
8	Pre	Nežinau kiek mokėti, nes nėra _____ ( <i>sąskaita</i> ).
9	Pre	Labai keista, kad bute nėra _____ ( <i>virtuvė</i> ).
10	Pre	Mūsų kaime nėra _____ ( <i>parduotuvės</i> ).
11	Pre	Man nepatinka _____ ( <i>grupė</i> ) „Fojė“.
12	Pre	Kai turi _____ ( <i>butas</i> ) centre, kas nors užaina į svečius beveik kiekvieną vakarą.
13	Pre	Jei butas neturi _____ ( <i>balkonas</i> ), tai yra didelis minusas.
14	Pre	Ekskursijos pabaigoje gidė uždavė _____ ( <i>klausimas</i> ) apie Vilnių.
15	Pre	Aš nebaigiau _____ ( <i>universitetas</i> ), nes mokytis man buvo nuobodu.
16	Pre	Jonas dar nelanko vaikų _____ ( <i>darželis</i> ).
17	Pre	Jei vaikas sutvarko savo _____ ( <i>kambarys</i> ), svarbu jį pagirti.
18	Pre	Įpilkite vandens ir įjunkite _____ ( <i>virduklis</i> ).
19	Pre	Aš esu vienas iš tų, kurie mėgsta _____ ( <i>mokykla</i> ).
20	Pre	Gyvenu Londone ir, deja, neturiu _____ ( <i>galimybė</i> ) dažnai grįžti į Lietuvą.
21	Pre	Deja, bibliotekoje neturime _____ ( <i>kavinė</i> ).
22	Pre	Jeigu mėgsti _____ ( <i>saulė</i> ), atostogauk rugpjūtį!
27	Post	Kambaryje nėra _____ ( <i>langas</i> ) ir šviesa patenka pro lauko duris.
28	Post	O ką daryti, jei virtuvėje net nėra _____ ( <i>stalas</i> )?
29	Post	Man smagu, kad mano _____ ( <i>darbas</i> ) patinka kolegoms.
30	Post	Mano _____ ( <i>vardas</i> ) man nepatinka, bet dabar jau nieko nebepakeisiu.
31	Post	Dar viena problema – mūsų kaime nėra vaikų _____ ( <i>darželis</i> ).
32	Post	Pirmame aukšte yra _____ ( <i>prieškambaris</i> ).
33	Post	Ispanai nemėgsta skaičiaus 13, net viešbučiuose nėra _____ ( <i>kambarys</i> ) su šiuo numeriu.
34	Post	Jei nepatinka _____ ( <i>mokykla</i> ) – keiskite ją.
35	Post	Šalia yra _____ ( <i>kavinė</i> ), kurioje skanus maistas ir malonus aptarnavimas.
36	Post	Kai nėra _____ ( <i>galimybė</i> ) valgyti sveikiau, bandykite bent nevalgyti užkandžių ir saldainių.
37	Post	Kam nepatinka _____ ( <i>saulė</i> )?
38	Post	1946 m. persikėlėme į Panevėžį, ten tėvai nusipirko _____ ( <i>namas</i> ).
39	Post	Žinau, kad tai blogai, bet neturiu _____ ( <i>laikas</i> ) sportuoti.
40	Post	Kalėdų eglutė ir vėl puošia _____ ( <i>miestas</i> ).
41	Post	Kartais sapnuoju žmogų, bet nematau jo _____ ( <i>veidas</i> ).
42	Post	Ši įmonė atidarė _____ ( <i>viešbutis</i> ) gyvūnams.
43	Post	Butas taip pat turi _____ ( <i>rūsys</i> ) su metalinėmis durimis.
44	Post	Neskačiau _____ ( <i>straipsnis</i> ), iš nuotraukos viskas aišku.
45	Post	Na, ir kas, kad neturiu _____ ( <i>šeima</i> ). Mano prioritetas - karjera.
46	Post	Ar būna žmonių, kurie visai nejaučia _____ ( <i>baimė</i> )?
47	Post	Šiuo klausimu neturiu _____ ( <i>nuomonė</i> ).
48	Post	Mes atidarėme savo _____ ( <i>parduotuvė</i> ) internete.

## Appendix C – Alternative models fitted to the data

Bayesian posterior summary for fill-in-the-gap test of the participants whose L1 was Slavic

Parameter	Median	HDI (89%)		pd	% in ROPE (±0.18)
		Low	High		
Intercept	1.53	0.88	2.20	99.93%	0%
Test Time	0.15	-0.11	0.42	81.78%	57.84%
Treatment	-0.40	-1.02	0.23	85.32%	22.18%
Syntactic Role	-0.43	-0.70	-0.17	99.52%	3.59%
Test time : Treatment	0.10	-0.08	0.28	82.04%	78.07%
Syntactic Role: Treatment	0.38	0.20	0.57	99.96%	1.17%
Test time : Syntactic Role	-0.19	-0.46	0.07	87.64%	47.39%

*Note.* The reference levels for each variable were as follows: Test Time: Pretest; Treatment: TSLT, Syntactic Role: Subject. These same levels also applied to the interactions.

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

Bayesian posterior summary for fill-in-the-gap test in which paradigm errors were accepted

Parameter	Median	HDI (89%)		pd	% in ROPE (±0.18)
		Low	High		
Intercept	1.93	1.11	2.82	99.96%	0%
Test Time	0.25	-0.03	0.54	92.44%	34.42%
Treatment	-0.31	-1.02	0.38	76.84%	26.86%
Syntactic Role	-0.35	-0.64	-0.06	97.38%	15.31%
L1	-0.14	-1.00	0.69	61.37%	28.34%
Test time : Treatment	0.22	0.04	0.39	97.75%	36.07%
Syntactic Role: Treatment	0.44	0.27	0.61	100.00%	0.00%
Test time : Syntactic Role	-0.17	-0.45	0.12	82.62%	53.04%

*Note.* The reference levels for each variable were as follows: Test Time: Pretest; Treatment: TSLT, Syntactic Role: Subject; L1: Non-Slavic. These same levels also applied to the interactions.

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

Bayesian posterior summary for fill-in-the-gap test excluding items containing the verb *patikti* ‘to like’

Parameter	Median	HDI (89%)		pd	% in ROPE (±0.18)
		Low	High		
Intercept	1.78	0.90	2.71	99.85%	0%
Test Time	0.07	-0.18	0.33	67.16%	73.64%
Treatment	-0.33	-1.09	0.45	76.34%	24.47%
Syntactic Role	-0.44	-0.71	-0.18	99.64%	3.28%
L1	-0.18	-1.11	0.73	63.02%	25.68%
Test time : Treatment	0.08	-0.09	0.26	77.33%	83.26%
Syntactic Role: Treatment	0.34	0.15	0.54	99.81%	6.65%
Test time : Syntactic Role	-0.12	-0.38	0.13	78.42%	64.23%

*Note.* The reference levels for each variable were as follows: Test Time: Pretest; Treatment: TSLT, Syntactic Role: Subject; L1: Non-Slavic. These same levels also applied to the interactions.

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