Light Verb Constructions in Lithuanian: Identification and Classification

Leksinės analitinės konstrukcijos: nustatymas ir klasifikacija

LINGUISTICS / KALBOTYRA

Jolanta Kovalevskaitė
Vytautas Magnus University, Lithuania

Erika Rimkutė
Vytautas Magnus University, Lithuania

Laura Vilkaitė-Lozdienė
Vilnius University, Lithuania

Light verb constructions (LVCs) are verb-noun constructions in which the noun carries the semantic meaning and the verb is semantically reduced, when compared with its main meaning, for example, *atlikti analizę* ('to perform an analysis'). LVCs in Lithuanian have not been addressed much so far. The analysis of Lithuanian LVCs was carried out as a part of the PARSEME project on verbal identification of multiword expressions (MWE). This paper aims at presenting some initial findings on the identification of LVCs in Lithuanian, based on the 1st edition of the PARSEME shared-task results (2017). We describe the identification process according to the semantic and syntactic features of LVCs (PARSEME guidelines 1.0 2017) and discuss the grammatical features of the identified Lithuanian LVCs.

LVCs seem to be less frequent in Lithuanian than in other languages: they make up about 0.2% (215 instances) of the analysed 200,000 token corpus. Based on the number of different LVCs, there seem to be two groups of verbs functioning as light verbs: a relatively small group of common light verbs used in the most prototypical examples of Lithuanian LVCs (e.g., *vykdyti* 'to perform', *atlikti* 'to perform', *daryti* 'to do', and *turėti* 'to have') and a larger group of less common light verbs. Most of the nouns in analysed LVCs have suffixes *-imas* and *-ymas*, which are the most typical Lithuanian suffixes for deriving a noun from a verb. Almost 40% of all LVCs are used with 1–3 words intervening between a verb and a noun.

KEYWORDS: verbal multi-word expression, light verb construction, Lithuanian language, annotation, corpus, PARSEME COST Action IC1207.

Abstract

Introduction

A light verb construction (LVC) has been defined as 'a verb-complement pair in which the verb has little lexical meaning (is 'light') and much of the semantic content of the construction is obtained from the complement' (Tan et al., 2006, p. 31). Some examples of LVCs in English are *to make a decision, to take sth into consideration*. While the verb in an LVC is a syntactic head of the phrase, the noun is a semantic head that carries the semantic information (Nagy et al., 2013). As the verb in an LVC does not have an independent semantic meaning, the LVC can usually be paraphrased by a verbal form of the noun used in the construction without losing any meaning (e.g., *to give a lecture = to lecture*) (Hwang et al., 2010). However, 'total synonymy is rare and by no means systematic phenomenon in language' (Bergs, 2005, p. 214). Nagy et al. (2013, p. 329) noted that '[s]ince the syntactic and the semantic head of the construction are not the same, they require special treatment when parsing' and, therefore, their identification is important for NLP tasks.

LVCs form a subtype of multiword expressions as their elements are rather fixed and their meanings cannot be interpreted word by word. According to Sag et al. (2002, p.7), 'although such phrases [LVCs] are sometimes claimed to be idioms, this seems to be stretching the term too far: the noun is used in a normal sense, and the verb meaning appears to be bleached, rather than idiomatic.' In some cases, though, the distinction between idioms and LVCs is not straightforward, such as in *to take charge* where *to take* could indicate an LVC but *charge* is also used figuratively (Hwang et al., 2010). However, apart from the borderline cases, rather than being idioms, most of the LVCs seem to be a special type of collocations, in the sense that the noun and the verb are frequently used together, only in this specific case the verb carries very little semantic meaning. Butt (2010, p. 21) notes that verbs do not entirely lose their meanings in LVCs, rather 'they seem to modulate or structure a given event predication, but not supply their own event'.

Therefore, as a type of MWEs, LVCs share some properties with idioms and some with collocations. However, the picture is more complex than that. LVCs in themselves seem to be rather heterogeneous as a class and various proposed classifications of LVCs exist. One example could be Kearns' (2002) classification of light verbs into true light verbs (such as *have* in the phrase *to have a read*) and vague action verbs (such as *make* as in *to make an inspection*). This distinction is made mostly by looking at the nature of the noun in the phrase: for the verb to be a true light verb, the stem of the noun has to be identical to a stem of a verb, while in case of vague action verb, the noun is derived from a verb. In this specific case, *to read* and *a read* have the same stem, while *an inspection* is derived from *to inspect*. As explained in Samardžić (2008), for the true light verbs, the complement is more verbal, i.e., the noun is identical in its form with the corresponding verb (*to take a look > to look*). For the vague action verbs, the complement is more nominal, a noun is derived with a suffix, can have different modifiers and can move more freely: *to make a decision > to make a difficult decision*. However, as Samardžić and Merlo (2010) noted, the difference between these two types of LVC is sometimes not clear-cut[1]. While Kearns (2002) made a compelling argument for this distinction, for the sake of analysing the Lithuanian language, it seems

[1] However, LVCs can form simple predicates or complex predicates, and, according to Samardžić (2008), LVCs with 'true light verbs' have a more stable structure, are more fixed, allow for fewer insertions or modification: these LVCs could be 'considered as simple predicates, with all the arguments belonging to the complement', 'while the constructions with vague action verbs could be seen as complex predicates, where some arguments belong to the verb and some to the complement' (Samardžić, 2008, p.19). Storrer (2007) in her study on German LVCs found the correlation between LVC's structure and syntactic behaviour: a) more fixed LVCs are those with noun introduced by a preposition (*tritt in Verbindung* 'contacts'): for these constructions the syntactic modifications are impossible; b) LVCs where the verb is followed by a direct object-noun (as in *trifft eine Entscheidung* 'decides') are less fixed: for these constructions more syntactic modifications (modifications with adjectives, number and determiner variation, negation with *kein*) are possible.

hardly relevant, as in Lithuanian, the verb and the noun cannot have identical forms which makes the true light verbs impossible by definition.

Another classification was put forwards by Bergs (2005) who also claims that 'LVCs do not form a unified and easily identifiable group' (Bergs, 2005, p. 210). In Bergs (2005, p. 210–215), four (sub)-types of LVCs are described: 1) *have a walk, take a shower*, etc.: this type is based on a light verb and a deverbal, eventive, "action" nominal which is formally identical to the corresponding verb (total conversion); 2) LVCs of type 2 (*have an agreement, take action*) have the same structure as those of type 1; however, the eventive noun is derived from a verb through other derivational processes (e.g., suffixation); and 3) LVCs of type 3 show essentially the same syntax as those of type 1 and 2, but their nominal part may be compounded (e.g., *have a heart-attack, do somebody's homework*). LVCs of type 4 (*have in keeping, have in command*) deviate from other types in terms of syntax and morphology, and thus, are most distant from the LVC core – type 1. Comparing with the core, type 2 lacks the prototypical morphology, and type 3 lacks the corresponding simple verbs (Bergs, 2005, p. 214).

While these various possible classifications of LVCs exist, this paper does not aim to classify them but rather to identify them in the Lithuanian part of the multilingual PARSEME corpus[2] as one of the classes of multiword constructions, alongside with idioms, following the guidelines of PARSEME shared task. A list of Lithuanian LVCs could help identify the verbs that are common in these constructions. As Butt (2010) argues, light verbs should be defined as a separate syntactic class, as they are different from auxiliaries and from verbs. Tan et al. (2006) note that in many languages there seems to be a finite set of light verbs. Therefore, it may be possible to list light verbs for each language and to use this predefined list for the NLP purposes, in order to identify the LVCs easier. A similar list is already compiled of grammatical multi-word units in Lithuanian and it is used for automatic morphological annotation (Rimkutė, 2009).

In Lithuanian, no special attention has been paid to LVCs so far: if addressed, they were taken as an example of collocations (e.g., Marcinkevičienė, 2010). This might be due to the fact that they would be considered a small group of constructions, not typical for Lithuanian (c.f. such subtypes of multi-word verbs as particle or phrasal verbs that are not relevant for Lithuanian). This paper aims at presenting some initial findings on the identification of LVCs in Lithuanian, based on the 1st edition of the PARSEME shared-task[3] results on verbal MWE identification version 1.0 and in the PASTOVU[4] project. It aims at presenting more data about the Lithuanian LVCs and their main structural and grammatical features. A detailed investigation of the nature and behaviour of verbal MWEs as LVCs, could enhance the output and results of various NLP applications and syntactic analysis.

Lithuanian LVCs were identified and annotated manually by two linguists in a 200,000 token subcorpus of articles from a popular Lithuanian news portal DELFI.lt.[5] The texts for the corpus were taken from the portal between August and September 2016. Texts on various topics (such as business, cars, sport, news, celebrities, science, etc.) were analysed. LVCs were annotated using the *brat rapid annotation tool*[6] and applying PARSEME shared-task annotation guidelines (PARSEME guidelines 1.0 2017).

LVC Identification Methodology

---

[2] *PARSEME multilingual corpus of verbal multiword expressions, 1.0.* https://lindat.mff.cuni.cz/repository/xmlui/handle/11372/LRT-2282

[3] The PARSEME shared task on automatic identification of verbal multiword expressions is a collective effort undertaken by the European PARSEME COST Action IC1207. The shared task on verbal MWE identification aims at identifying verbal MWEs in running texts in 18 languages from several language families since verbal MWEs frequently introduce issues, which are central to deep parsing: https://typo.uni-konstanz.de/parseme/index.php/2-general/142-parseme-shared-task-on-automatic-detection-of-verbal-mwes

[4] http://mwe.lt/en_US/

[5] htttp://www.delfi.lt

[6] http://brat.nlplab.org/index.html

In the PARSEME shared-task project, several types of the verbal MWEs were annotated (Savary et al., 2017). In the Lithuanian corpus, though, only two universal categories of verbal MWEs were annotated: idioms and LVCs (only the latter category is discussed in this paper). In the guidelines, LVCs are defined as verbal MWEs which function as (possibly unsaturated) verb phrases, that is, their syntactic heads are verbs in finite forms and their other lexicalised components are dependents of the verb (e.g., *made a decision*) (PARSEME guidelines 1.0 2017). While annotating verbal MWEs, first, a verbal phrase (or an infinitival/nominal/participial variant of a verbal phrase) was identified and then it was checked whether it followed the indicated criteria for an LVC.
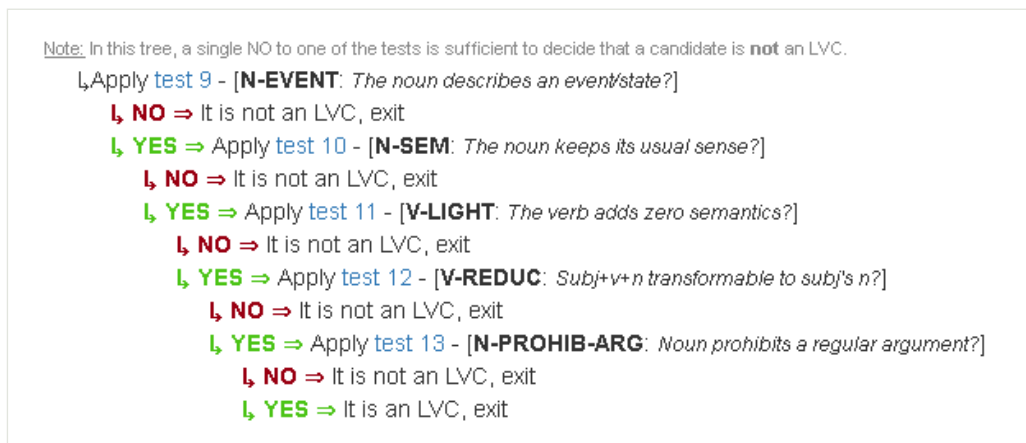
LVCs have the following general characteristics ((PARSEME guidelines 1.0 2017):

1   They are formed by a verb (V) and a noun (N), which either directly depends on the verb (*to give a lecture*), or is introduced by a preposition (*to come into bloom*).

2   **a)** A noun typically refers to an event (*decision, visit*) or a state (*fear, courage*). The noun has one of its regular meanings (which can be retrieved even in the absence of the verb).

   b) The verb is 'light', i.e. it contributes to the meaning of the whole only to a small degree. It only contributes morphological features (tense, mood, person, number, etc.).

When annotating LVCs in the Lithuanian corpus, the above mentioned definition was applied. Every candidate for LVCs was evaluated according to the LVC-specific decision tree (PARSEME guidelines 1.0 2017). In this tree (see Fig. 1), a single negative answer to one of the tests was sufficient to decide that a candidate phrase was not an LVC.

Fig. 1

LVC-specific decision tree (PARSEME guidelines 1.0 2017)

Note: In this tree, a single NO to one of the tests is sufficient to decide that a candidate is **not** an LVC.
ↄApply test 9 - [**N-EVENT**: *The noun describes an event/state?*]
  ↄ **NO** ⇒ It is not an LVC, exit
  ↄ **YES** ⇒ Apply test 10 - [**N-SEM**: *The noun keeps its usual sense?*]
    ↄ **NO** ⇒ It is not an LVC, exit
    ↄ **YES** ⇒ Apply test 11 - [**V-LIGHT**: *The verb adds zero semantics?*]
      ↄ **NO** ⇒ It is not an LVC, exit
      ↄ **YES** ⇒ Apply test 12 - [**V-REDUC**: *Subj+v+n transformable to subj's n?*]
        ↄ **NO** ⇒ It is not an LVC, exit
        ↄ **YES** ⇒ Apply test 13 - [**N-PROHIB-ARG**: *Noun prohibits a regular argument?*]
          ↄ **NO** ⇒ It is not an LVC, exit
          ↄ **YES** ⇒ It is an LVC, exit

This decision tree was followed step by step. For instance, *priimti sprendimą* 'to make a decision' was identified as an LVC, as it passed all five tests:

1   the noun *sprendimas* refers to an event 'a decision' and is derived from a verb *spręsti* 'to decide';

2   the noun *sprendimas* has a literal meaning and is used in its original sense;

3   the verb *priimti* 'to take' adds no meaning to *sprendimas* besides that of performing an activity;

4   the NP such as *Seimas priėmė sprendimą* 'the Parliament has made a decision' is transformable to a phrase *Seimo sprendimas* 'Parliament's decision' and both phrases refer to the same event;

5   in the phrase, *Seimas priėmė sprendimą,* the noun *sprendimas* 'decision' cannot be modified, e.g.*Seimas priėmė vyriausybės sprendimą* *'The Parliament has made the government's decision'.

There were quite a few LVC candidates that failed to meet the fifth criterion: 'noun prohibits a regular argument' (PARSEME guidelines 1.0 2017), although they fully satisfied the other four criteria. For example, in *jis pateikė pasiūlymą* 'he put forward a proposal', it is possible to modify the noun *pasiūlymas* 'proposal': *jis pateikė kolegos/mano pasiūlymą* 'he put forward colleague's/my proposal'. Following the guidelines, such verbal MWEs were excluded from the list of LVCs. The interpretations of these cases, though, raised many doubts. On the one hand, as shown by this example, the noun does not prohibit another argument (*Jis pateikė **kolegos/mano** pasiūlymą*), so it clearly fails the test number 5. On the other hand, despite the fact that this argument is possible theoretically and practically, it remains questionable if the case should not be regarded as an LVC. Issues like this one showed that test 5 is problematic and needs further clarification. Drawing on annotators' feedback, in the new version of the guidelines (PARSEME guidelines 1.1. 2018), tests number 4 and 5 were joined. Accordingly, a new wording was chosen: 'Is the subject of the verb a semantic argument of the noun?' (Savary et al., 2018, p. 103). For example: a) *John made a presentation to his boss* → John is a semantic argument (the presenter) of the noun (*John's presentation*), thus, this case would be an acceptable LVC; b) *John made his boss' presentation*, where John is the subject of the verb, but is not a semantic argument of the noun, would be not acceptable (it is not *John's presentation*, but rather *his boss' presentation*).

In view of the corrections, the verbal MWE in the previously discussed example *Seimas priėmė sprendimą* becomes an acceptable LVC, because the subject of the verb is a semantic argument of the noun, i.e., *Seimo sprendimas* 'the Parliament's decision'. The verbal MWE *jis pateikė pasiūlymą* 'he put forward a proposal' could also be counted as LVC (e.g., *jo pasiūlymas* 'his proposal'). On the contrary, if there is an indication in the context that *jis* is the subject of the verb, but not a semantic argument of the noun, the case cannot be accepted as an LVC. Therefore, following this decision tree from the PARSEME guidelines 1.0, a part of verbal MWEs excluded from the research could be included in the next stages of the shared task.

Frequency of LVCs

In the Lithuanian data, 215 LVCs were identified (including repetitions of the same construction or different grammatical forms of the same construction). They made up about 0.2% of the 200,000 token corpus analysed. It is a rather small percentage as, for example, idioms were also annotated in the same corpus and 292 idioms were found (i.e. more idioms than LVCs). Also, when looking at Lithuanian language, in the Corpus of Contemporary Lithuanian Language of 100 million tokens, automatically detected MWEs covered 68.1% of the corpus (Marcinkevičienė et al., 2005, p.32), suggesting that the overall percentage of MWEs in Lithuanian is high, while the percentage of LVCs seems to be low.

For a comparison, we can look at the PARSEME data from other languages (Savary et al. 2018). For example, Bulgarian and Polish corpora were from the same Baltic and Slavic language group[7] and of a similar size but the number of LVCs in these three corpora differed considerably (see Table 1).

Characteristics of Lithuanian LVCs

Table 1

Number of LVCs in different languages

Language	Tokens	Total number of verbal MWEs	Number of LVCs
Bulgarian	200,128	2,406	511
Lithuanian	256,235	502	215
Polish	220,934	3,649	1,453

7   Lithuanian, Bulgarian, Polish, Czech, and Slovene were grouped together in the PARSEME project (partially for convenience reasons) as they were the only languages from Baltic and Slavic families, analysed during the project.

These differences are rather surprising, considering that the languages have similar structures. There are some potential reasons for these differences. For example, as Savary et al. (2018, p.108) note, during the annotation 'a language specific interpretation of the guidelines could not be avoided and this was mainly due to different linguistic sensitivities and traditions, language-specific challenges and incompleteness or imprecision of the guidelines.' Therefore, one potential reason for the differences can be a still scarce understanding of the LVC in Lithuanian (for annotators with a Lithuanian linguistic background, the LVCs are a rather foreign phenomenon). Also, as already mentioned above, difficulties and inaccuracies when applying some tests might have played a role. Another potential reason might be the fact that standard written Lithuanian prefers verbal, rather than nominal, constructions (Leonavičienė et al., 2013; Pažūsis, 2014). This might be the reason, why LVCs are not very frequent in Lithuanian: they are simply not typical for Lithuanian. However, a more detailed study of the Lithuanian LVCs is needed to test this hypothesis.

On the other hand, the percentage of the LVCs in Lithuanian (0.2%) seems to be not so low when compared with English data. For instance, Ronan and Schneider (2015) came to a conclusion that LVCs make up about 1,600 tokens in a million token corpus (i.e., about 0.16%). Hence, the frequencies of LVCs in English and Lithuanian seem to be surprisingly similar.

Grammatical Features of LVCs

Verbs and Nouns in LVCs

Two groups of verbs were identified in the Lithuanian LVCs: common light verbs (4 verbs in 55 different LVC-types in total) and less common light verbs (17 verbs in 38 different LVC-types in total) (see Table 2).

According to the dictionary of word frequencies (Utka, 2016), the verbs from the 1st group are some of the most frequent verbs in Lithuanian overall (e.g., *turėti* 'to have': 18th in the frequency list, *atlikti* 'to perform': 86th position, *daryti* 'to do' – 150th position, *padaryti* 'to make' – 207th position, *vykdyti* 'to perform' – 422nd position (cf. Nagy et al., 2013 for English). Lithuanian grammar studies classify these verbs as partially independent, auxiliary (Labutis, 1998; Valeckienė, 1998).

In Table 2, verbs that combine with a larger number of different nouns (ten or more) to form LVCs are labelled as common light verbs, while others are labelled as less common light verbs. When analysing the common light verbs in the Lithuanian LVCs (see Table 1), it turned out that these verbs (*vykdyti* 'to perform', *atlikti* 'to perform', *daryti* 'to do/to make', and *turėti* 'to have') at least partially correspond to the most frequent light verbs in English: *do, give, have, make, take* (Baldwin et al., 2010). Also, some English verbs have more than one potential equivalent in Lithuanian: cf. *vykdyti, atlikti* are semantically similar to *daryti* (*do* or *make*). *To give* and *to take* are used less frequent as light verbs in Lithuanian LVCs (see Table 2 for the group of less common light verbs).

As Storrer (2007) notes, 'the meaning of support verb [the author uses the term *support verb constructions* instead of LVCs] is not identical to the meaning of the homonymous main verb', e.g. *I **took** a cup / I **took** a look* (cf. Storrer 2007, p. 2). A similar observation can be made concerning the light verbs from both Lithuanian groups. Verbs of the second group *teikti, leisti, patirti, skirti, duoti, kelti* are light verbs in LVCs, but they could have a clear lexical meaning in other phrases (e.g., *skirti dėmesio* 'to pay attention' would be a light verb construction, while in *skirti premiją* 'to award a prize', the verb *skirti* has an independent semantic meaning as a verbal part of collocation). For example, in English, the verb *to deliver* in the LVC *to deliver a*

Common light verbs	Number of different LVC-types	Examples
(į)vykdyti 'to perform'	19	*vykdyti patikrinimus* 'to carry out inspections'; *įvykdyti nusikaltimą* 'to commit a crime'
atlikti 'to accomplish/to perform'	14	*atlikti analizę* 'to perform the analysis'
turėti 'to have'	12	*turi galimybę* 'has an opportunity'
(pa)daryti 'to do/to make'	10	*padarė poveikį* 'had an influence'; *daro spaudimą* 'puts pressure (on)'
Less common light verbs		
su-/pa-teikti 'to give'	6	*suteikė galimybę* 'gave a chance'; *pateikė protestą* 'filed a protest'
(su)rengti 'to arrange/to organize'	4	*rengti išpuolius* 'to organize attacks'; *surengė derybas* 'organized negotiations'
imtis 'to undertake'	3	*imtis veiksmų* 'to take actions'
priimti 'to accept'	3	*priimti sprendimą* 'to make a decision'
patirti 'to experience'	3	*patirti nuostolius* 'to suffer losses'
skirti 'to dedicate'	3	*skirti pastangų* 'to put effort'
duoti 'to give'	3	*duoti parodymus* 'to give testimony'
pri-/su-daryti 'to make/to create'	3	*pridarė nuostolių* 'caused losses'; *sudaryti sąlygas* 'to create conditions'
(į)vesti 'to introduce/to conduct'	2	*įvesti apribojimus* 'to impose restrictions'
kelti 'to raise'	2	*kelti grėsmę* 'to cause threat'
leisti 'to spend'	2	*leisti laiką* 'to spend time'
dėti 'to put'	1	*deda pastangas* 'puts effort'
laikytis 'to hold'	1	*laikosi požiūrio* 'takes an approach'
(pa)siekti 'to achieve'	1	*pasiekti susitarimą* 'to achieve agreement'
(su)laukti 'to wait'	1	*sulaukė pasisekimo* 'gained success'

Note: *the translations are provided seeking an equivalent meaning in English; hence, some English phrases might not seem to be clear LVCs in translated examples.*

speech does not have his common meaning (as 'to bring goods to specific places'), but rather is used as syntactic operator (which is a common feature for the most frequency light verbs such as *make, take*) (PARSEME guidelines 1.0 2017). In our data, such verbs could be *priimti* (*priimti sprendimą* 'to make a decision'), *leisti* (*leisti laiką* 'to spend time'), *kelti* (*kelti grėsmę* 'to cause threat'), *laikytis* (*laikytis požiūrio* 'to take an approach'), *pasiekti* (*pasiekti susitarimą* 'to reach an agreement'), and *sulaukti* (*sulaukė pasisekimo* 'gained success').

The LVCs with the common light verbs seem to be the most prototypical examples of the LVCs in Lithuanian: e.g., *atlikti analizę* 'to perform the analysis', *daryti spaudimą* 'to put pressure on', etc. 55 different LVC-types were formed with the common light

---

[8]   In Table 2, where the prefix only adds the meaning of perfective aspect (e.g. *vykdė > įvykdė* 'was performing > performed'), it is provided in the parentheses *(į)vykdyti* and both forms are counted as one verb. Prefixes that contribute to the meaning of the verb are provided with a slash (*su-/pa-teikti*) and are counted as separate verbs. Reflexive forms are listed as separate verbs (*laikytis, imtis*).

verbs[9]. The less common light verbs formed 38 different LVC-types. If we compare the number of LVCs in this group with the former group, the numbers are pretty similar, but the second group consists of 4 times more verbs (4 versus 17).

It was noted during the annotation stage that some of the verbs tend to be used in several different derivational forms with various prefixes. For example, one of the most frequent light verb *daryti* 'to do' was used with the most different prefixes: **pa***daryti* 'to make', **su***daryti* 'to create', **su***sidaryti* 'to form', **pri***daryti* 'to cause'. The use of many derivative forms could be attributed to the fact that a lot of forms of LVCs were used in the past tense (in written language, around 38% of declined verbs are used in the past tense (Rimkutė, 2006)) and past forms tend to be the ones with prefixes. The prefixes essentially mark an event aspect, when a writer signals that an action is already accomplished (*daryti* 'to do' and **pa***daryti* 'to make'). In Table 2, the verb **pa***daryti* is grouped with DARYTI, as in the analysed cases, the prefix only added the meaning of perfective aspect. However, the cases where the prefix adds additional meaning are given in the second group of verbs as separate verbs (*pridaryti, sudaryti*).

There were extremely many verbs with prefixes in the less common light verbs group: these prefixes modify their meanings. For example, a general meaning of the verb *teikti* could be described as 'to provide/to give' (*teikia konsultacijas* 'consults'), *suteikti* means 'to give/ to grant' (*suteikti galimybę* 'to give an opportunity', *suteikia galią* 'gives power', *suteikia įžvalgų* 'gives insights'), while its form with another prefix *pateikti* means 'to give/to submit' (*pateikė protestą* 'filed a protest', *pateikė paaiškinimą* 'gave an explanation'). Verbs with prefixes usually express more independent lexical meanings; therefore, they are not as clearly light verbs as those from the common light verbs group. As Butt (2010) notes, light verbs tend to modulate the meaning of LVCs in terms of providing some additional information. The use of these verbs with different prefixes or their reflexive forms seems to do so: add some extra meaning to the construction.

Other LVC studies (e.g., Storrer, 2007) mention that some light verbs contribute specific semantic or grammatical features such as aspect or causality to the meaning of the LVC. In the second stage of the PARSEME shared-task, such LVCs are already analysed and a distinction between full LVCs (e.g., *to have the right*) and causative LVCs (*to grant the right*) is made (PARSEME guidelines 1.1, 2018). However, at the time of this annotation study, such a distinction was not yet made.

The extracted LVCs contained nouns derived from verbs. Most of these derivatives are derived from verbs with suffixes that show abstract meanings *–imas, –ymas* that are the most typical Lithuanian suffixes for deriving a noun from a verb, e.g., *pasirinkimas* 'choice' from *pasirinkti* 'to choose'; *stebėjimas* 'observation' from *stebėti* 'to observe'; *patikrinimas* 'examination' from *patikrinti* 'to examine'. A smaller part of the nouns was derived from verbs adding nominal inflections or prefixes, e.g., with inflections: *skrydis* 'flight' from *skristi* 'to fly', *poveikis* 'influence' from *paveikti* 'to influence'. Only a small part of nouns was not derived from verbs (*reputacija* 'reputation', *laikas* 'time', *dėmesys* 'attention').

Nouns in the LVCs had meanings of an event or a state, and retained them in the LVC (this was one of the identification criteria). Some of the nouns were used only in plural forms either because only a plural from of a specific noun exists in Lithuanian (such as in *derybos* 'negotiation') or because the plural form is chosen to indicate the named object (e.g., *nuostoliai* 'loss', *nurodymai* 'guidelines', *priekaištai* 'reproaches').

---

[9] LVCs with different word order, e.g. *daryti įtaką, įtaką daro* 'have an influence' were counted as the same LVC – *daryti įtaką*. LVC verbs used in their neutral and negative forms, such as *turi įtaką* 'has an influence', *neturėjo įtakos* 'did not have influence', were also seen as the same LVC.

Syntactic Features of LVCs

In most cases, the verb directly governs the noun. All the verbs in the constructions were transitive and required a direct object. In Lithuanian, it is usually marked by an accusative or a genitive. Therefore, an accusative is usually the case of the nouns in LVCs, for example, *turės poreikį* 'will have an influence' or *vykdė priežiūrą* 'supervised'. A further object is often used in a dative: *kelti grėsmę* **kam** 'to cause threat **to whom**', *skirti laiką/dėmesį* **kam** 'to dedicate time/attention **for what**'.

Most of the LVCs analysed followed the word order of verb + noun. As Lithuanian has a rather flexible word order, 20 token instances (i.e., about 9% of all the LVCs) had an opposite word order, for example, *įtaką daro* and *daro įtaką* 'has an influence'. The noun-verb word order is a marked word order and it depends on the sentence structure.

In PARSEME annotation guidelines, nominal, participial and other syntactic variants of prototypical verbal MWEs were included, e.g., *decisions which we made, decision making, the decision that the director has to make*. When identifying the LVCs in Lithuanian, there were a number of cases when a light verb was used as a participle form that modifies a noun (e.g., *daromas pranešimas* *'a being made notification'). These cases were not counted as LVCs as they were not predicates according to the Lithuanian grammar; however, they could also be interpreted as cases of passivation or as syntactic variability of an LVC (cf. *a demo was given* (Sag et al., 2002)). Potentially not applying this criterion (i.e., the LVC is used as a predicate) would make the identification of LVCs more consistent for Lithuanian (cf. Nagy et al., 2013, for English).

LVCs in Lithuanian were usually made of two words (only one construction with a preposition was identified). However, other words (1–3 words) could be used in between of the verb and the noun. For example, *patyrė pralaimėjimų* 'experienced failures' used as *patyrė didžiausių iki šiol užfiksuotų pralaimėjimų* 'experienced the largest so far registered failures'. Sag et al. (2002) defines LVCs with insertions as a kind of syntactic variability and calls these insertions internal modifications (e.g., *to give a revealing demo*).

Of all 215 LVCs, 81 LVCs were with insertions (37.7%). 50 LVCs (the majority of the cases) had one intervening word. When one word intervenes, usually it is a modifier of the noun (e.g., *padarė* **šiurkščių** *pažeidimų* 'made **serious** violations', *vykdyti* **savo** *įsipareigojimą* 'to fulfil **one's** obligations', *priėmė* **įstatymo** *pataisą* 'adopted **law's** amendment'). Less often, in case of an inversion in the sentence, the inserted word can be a part of the complex predicate (e.g., *išvadas* **turi** *daryti* '**must** make conclusions', *vertinimą* **galės** *atlikti* '**will be able** to perform the translation'). Another half of LVCs with insertions were LVCs with two (19 cases) or three (12 cases) inserted words. In these LVCs, the insertions were also mostly modifiers but there are some cases of adverbials as well (*įdėti* **išties daug** *pastangų* 'to put **really much** effort', *priėmė* **daug neteisingų** *sprendimų* 'made **many wrong** decisions').

Conclusion

As a class of multiword expressions, i.e., multi-word verbs, LVCs seem to be relevant for Lithuanian and, therefore, should receive more attention. The structure of the LVCs in Lithuanian seems to be similar to the structure of LVCs in other languages, such as English or Polish. In most cases, the verb directly governs the noun; prepositional LVCs are very rare in Lithuanian.

Nouns in the LVCs have a meaning of an event or a state. The majority (90%) of nouns in the analysed LVCs have suffixes. The most often used suffixes are *–imas* and *–ymas*, which are the most typical Lithuanian suffixes for deriving a noun from a verb. The common light verbs in Lithuanian LVCs are *vykdyti* 'to fulfil', *atlikti* 'to perform', *daryti* 'to do', and *turėti* 'to have'. These verbs correspond to the most frequent light verbs in English. Following the idea of Tan et al. (2006) that light verbs tend to form a finite list, this small-scale study is a first step towards developing such a list for Lithuanian.

The LVCs with the common LVC verbs seem to be the most prototypical examples of the LVCs in Lithuanian. In the group of the less common light verbs, there is significantly higher diversity of verbs. In this group, there are quite a few verbs with prefixes: part of these verbs with prefixes mark an event aspect, but in some of these verbs, prefixes also add some semantic meaning. Although functioning as light verbs, they tend to contribute to the meaning of the whole LVC more than the most prototypical light verbs such as *atlikti* 'to perform' or *daryti* 'to do' (e.g. *pridaryti nuostolių* 'to cause losses', *sudaryti sąlygas* 'to create conditions').

Almost 40% of LVCs can have other words inserted between their components. In the majority of the cases (50 LVCs identified), the insertion consists of one word: most often it is the modifier preceding the noun, less often (in case of the opposite word order) it is a part of a complex predicate. The data from other languages show that potential syntactic transformations might also be important for classification and identification of LVCs. Thus, for Lithuanian, a more detailed study of the usage of LVCs is needed.

According to the PARSEME shared task edition 1.0, in Lithuanian, the density of LVCs is rather low. LVCs seem to be less frequent in Lithuanian than they are in other typologically similar languages such as Polish: they make up about 0.2% (215 instances) of the 200,000 token corpus analysed. The following editions of the PARSEME shared-task project could be a possibility to collect more data to further investigate and to revise the initial findings on Lithuanian LVCs.

Overall, LVCs seem to be used in Lithuanian as they are in other languages and the same identification criteria seem to be mostly applicable, although a language specific interpretation of the guidelines could not be avoided and some language-specific issues should be reconsidered. For example, it might be worth counting MWEs where the verb is in a participle form (*atliktas tyrimas* 'conducted research') as LVCs, despite the fact that this phrase is an attributive rather than a predicative phrase. In this research, we treated phrases with non-finite verb forms as LVCs, if these verb forms were used as predicates, e.g., *įvykdęs nusikaltimą* 'having committed a crime', *atlikusi analizę* 'having conducted a research', *pasiekusi susitarimą* 'having reached consensus'.

References

1 Baldwin, T. and Kim, S. N., 2010. Multiword Expressions. In: Handbook of Natural Language Processing, (eds.) Indurkhya, N. and Damerau, F. J, pp. 267- 292. Boca Raton: CRC Press.

2 Bergs, A. 2005. Social Networks and Historical Socialinguistics: Studies in Morphosyntactic Variation in the Paston Letters. In: Topics in English Linguistics, 51, (eds.) Traugott, C. E. and Kortmann, B. pp. 210-215. Berlin and New York: Mouton de Gruyter. https://doi.org/10.1515/9783110923223

3 Butt, M. 2010. The Light Verb Jungle: Still Hacking Away. In: Complex Predicates in Cross-Linguistic Perspective, (eds.) Baker, B., Amberber, M. and Harvey, M., pp. 48-78. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511712234.004

4 Hwang, J. D., Bhatia, A., Bonial, C., Mansouri, A., Vaidya, A., Xue, N. and Palmer M. 2010. Propbank Annotation of Multilingual Light Verb Constructions. In: Proceedings of the Fourth Linguistic Annotation Workshop, pp. 82-90.

5 Kearns, K., 2002. Light Verbs in English. Available at: https://pdfs.semanticscholar.org/f96d/9feb01f336f3669f65304a244ee0234f1413.pdf [Accessed March 2017].

6 Labutis, V. 1998. Lietuvių kalbos sintaksė. Vilnius: Vilniaus universiteto leidykla.

7 Leonavičienė, A. and Liepuoniūtė, D. 2013. Europos Sąjungos Teisingumo Teismo generalinio advokato išvadų leksinių analitinių konstrukcijų vertimas iš prancūzų kalbos į lietuvių kalbą. In: Kalbos kultūra, 86, pp. 95-109.

8 Marcinkevičienė, R. and Grigonytė, G. 2005. Lexicogrammatical Patterns of Lithuanian

Phrases. In: Proceedings of the Second Baltic Conference on Human Language Technologies, (eds.) Langemets, M. and Penjam, P., pp.299-305.

9   Marcinkevičienė, R. 2010. Lietuvių kalbos kolokacijos. Kaunas: Vytautas Magnus University.

10 Nagy, I., Vincze, V. and Farkas, R. 2013. Full-coverage Identification of English Light Verb Constructions. In: Proceedings of the International Joint Conference on NLP, pp. 329-337.

11 PARSEME Guidelines 1.0. 2017. PARSEME shared-task annotation guidelines version 1.0: Available at: http://parsemefr.lif.univ-mrs.fr/parseme-st-guidelines/1.0/ [Accessed January 2017].

12 PARSEME Guidelines 1.1. 2018. PARSEME shared-task annotation guidelines version 1.1. Available at: http://parsemefr.lif.univ-mrs.fr/parseme-st-guidelines/1.1/ [Accessed October 2018].

13 Pažūsis, L. 2014. Kalba ir vertimas. Vilnius: Vilnius University. Available at: http://www.esparama.lt/es_parama_pletra/failai/ESFproduktai/2014_Monografija_Vertimas_ir_kalba.pdf [Accessed March 2017].

14 Rimkutė, E. 2006. Morfologinio daugiareikšmiškumo ribojimas kompiuteriniame tekstyne. PhD dissertation. Kaunas: Vytautas Magnus University.

15 Rimkutė, E. 2009. Gramatinė morfologinių samplaikų klasifikacija. In: Kalbų studijos, 14, pp. 32-38.

16 Ronan, P. and Schneider G. 2015. Determining Light Verb Constructions in Contemporary British and Irish English. In: International Journal of Corpus Linguistics, 20(3), pp. 326-354. https://doi.org/10.1075/ijcl.20.3.03ron

17 Sag, I. A., Baldwin, T., Bond, F., Copestake and A. Flickinger, D. 2002. Multiword Expressions: A Pain in the Neck for NLP. In: Proceedings of the Third International Conference on Computational Linguistics and Intelligent Text Processing, CICLing'02, pp. 1-15. London: Springer-Verlag. https://doi.org/10.1007/3-540-45715-1_1

18 Samardžić, T. 2008. Light Verbs and the Lexical Category Bias of Their Complements. Available at: http://www.unige.ch/lettres/linguistique/samardzic/dea2.pdf [Accessed October 2017].

19 Samardžić, T. and Merlo, P. 2010. Cross-lingual Variation of Light Verb Constructions: Using Parallel Corpora and Automatic Alignment for Linguistic Research. In: Proceedings of the 2010 Workshop on NLP and Linguistics: Finding the Common Ground, pp. 52-60.

20 Savary, A., Ramisch, C., Cordeiro, S., R., Sangati, F., Vincze, V., Zadeh, B. Q, Candito, M., Cap, F., Giouli, V., Stoyanova, I. and Doucet, A. 2017. The PARSEME Shared Task on Automatic Identification of Verbal Multiword Expressions. In: Proceedings of the 13th Workshop on Multiword Expressions (MWE 2017), pp. 31-47. Available at: http://www.aclweb.org/anthology/W17-1704 [Accessed October 2017]. https://doi.org/10.18653/v1/W17-1704

21 Savary A., Candito M., Barbu Mititelu V., Bejček E., Cap F., Čéplö S., Cordeiro S. R., Eryiğit G., Giouli V., Gompel M. van, HaCohen-Kerner Y., Kovalevskaitė J., Krek S., Liebeskind Ch., Monti J., Escartín C. P., Plas L. van der, and QasemiZadeh B. 2018. PARSEME Multilingual Corpus of Verbal Multiword Expressions. In: Multiword Expressions at Length and in Depth: Extended Papers from the MWE 2017 Workshop, (eds.) Markantonatou, S., Ramisch, C., Savary, A. and Vincze, V. p. 87-147. Berlin: Language Science Press.

22 Storrer, A., 2007. Corpus-based Investigations on German Support Verb Constructions. In: Collocation and Idioms: Linguistic, Lexicographic, and Computational Aspects, (ed.) Fellbaum, C. London: Continuum Press.

23 Tan, Y. F., Kan, M. Y. and Cuir, H. 2006. Extending Corpus-based Identification of Light Verb Constructions Using a Supervised Learning Framework. In: Proceedings of the EACL Workshop on MultiWord Expressions in a Multilingual Contexts, pp. 49-56.

24 Utka, A. 2016. Lemmatised Wordlist of 1 m. Corpus of Contemporary Lithuanian, CLARIN-LT Digital Library in the Republic of Lithuania. Available at: http://hdl.handle.net/20.500.11821/12 [Accessed August 2019]

25 Valeckienė, A. 1998. Funkcinė lietuvių kalbos gramatika. Vilnius: Mokslo ir enciklopedijų leidybos institutas.

Santrauka

Jolanta Kovalevskaitė, Erika Rimkutė, Laura Vilkaitė-Lozdienė. Leksinės analitinės konstrukcijos: nustatymas ir klasifikacija

Leksinės analitinės konstrukcijos (LAK) – tai pastovieji žodžių junginiai, kuriuose, nors sintaksiškai daiktavardį valdo veiksmažodis, bet didžiausią reikšminį krūvį perteikia daiktavardis, o veiksmažodis yra išblukusios reikšmės (desemantizuotas), pvz., *priimti sprendimą, atlikti analizę, vykdyti patikrinimus, daryti spaudimą*. Lietuvių kalbos LAK – iki šiol mažai analizuotas pastoviųjų žodžių junginių tipas. Šios konstrukcijos tirtos vykdant PARSEME COST Action IC1207 projekto tęstinę veiklą, skirtą veiksmažodinių pastoviųjų žodžių junginių anotavimui daugiakalbiame tekstyne. Šiuo straipsniu norima pristatyti lietuvių kalbos LAK tyrimą, paremtą duomenimis iš 1-osios PARSEME daugiakalbio tekstyno versijos (2017). Straipsnyje aprašytas LAK nustatymo procesas remiantis semantiniais ir sintaksiniais LAK požymiais, analizuojamos LAK gramatinės ypatybės.

Atlikus tyrimą, paaiškėjo, kad lietuvių kalbos LAK yra rečiau nei kitose kalbose vartojami pastovieji žodžių junginiai: lietuviškoje daugiakalbio tekstyno (jį sudaro internetinės periodikos tekstai) dalyje (200 tūkst. žodžių tekstyne) tokios konstrukcijos sudaro apie 0,2 proc. (rasta 215 LAK).

LAK sudarančius veiksmažodžius galima suklasifikuoti į dvi grupes: 1) veiksmažodžius, dažniausiai pasitaikančius lietuviškuose LAK, dėl dažnumo galima laikyti tipiškais desemantizuotais veiksmažodžiais (pvz., *vykdyti, atlikti, daryti, turėti*), ir 2) ne tokius tipiškus desemantizuotus veiksmažodžius, kurių grupėje užfiksuota didesnė leksemų įvairovė. Dauguma LAK sudarančių daiktavardžių turi priesagas *–imas* ir *–ymas*. Apie 40 proc. LAK vartosenos atvejų sudaro kolokacijos su 1–3 žodžių įsiterpimais tarp veiksmažodžio ir daiktavardžio.

About the Authors

Jolanta Kovalevskaitė

Dr., Vytautas Magnus University, Lithuania

Research interests
Corpus linguistics, multi-word expressions, terminology, lexicography

Address
V. Putvinskio g. 23-216, Kaunas, Lithuania

E-mail
jolanta.kovalevskaite@vdu.lt

Erika Rimkutė

Dr., Associate professor at Vytautas Magnus University, Lithuania

Research interests
Corpus linguistics, computational linguistics, automatic grammatical analysis, multi-word expressions

Address
V. Putvinskio g. 23-216, Kaunas, Lithuania

E-mail
erika.rimkute@vdu.lt

Laura Vilkaitė–Lozdienė

Dr., Vilnius University, Lithuania

Research interests
Second language acquisition, multi-word expressions, psycholinguistics

Address
Universiteto g. 5, LT-01513 Vilnius, Lithuania

E-mail
laura.vilkaite@flf.vu.lt