Advanced search tips

Usage patterns

➢ Usage patterns are associated with specific meanings of the headword (displayed by clicking on the tab *Meanings* in the tool *Headword Search*). After selecting a specific pattern, a three-colour table is displayed, in which the individual columns represent the grammatical, semantic, and lexical levels of the pattern. The model is read from top to bottom: the first level is grammatical (marked in blue), the second level is semantic (pink), and the third level is lexical (purple).

The example could be the 3rd meaning of the verb *bėgti* ('to run') presented as pattern 3.1 in the database. In this pattern, the information at the grammatical level shows that this verb in the present tense form (marked as BĖGTI_prs) is typically used with the subject (denoting an agent) and the adverbial of manner (marked as Adv). At the semantic and lexical level, we can additionally learn that the agent in this model is abstract (*laikas* 'time'), and the adverbial of manner is expressed by the adverb *greitai* ('quickly'). This information helps to distinguish the 3rd meaning of the verb *bėgti* ('to run') from the 1st one (where the subject is a human being or an animal) and the 2nd one (where the subject is a liquid, such as water or blood). The types of the adverbials in the latter patterns are also different.

➢ Sometimes the differences in meaning are already apparent at the grammatical level of the pattern, but most often they emerge at the semantic and lexical level. The meanings distinguished vary in frequency. In addition, some word meanings are expressed through more diverse forms (hence, more patterns exist), whereas others can be represented by a single pattern. For example, for the verb *bėgti* ('to run'), three meanings are distinguished: the first meaning is associated with three usage patterns, the second meaning appears in two usage patterns, and the third meaning is linked to only one pattern.

➢ At the grammatical level, the **syntactic functions** of the headwords (e.g. subject, object, etc.) and the **grammatical features** of the categories that convey these functions (e.g. case, tense, etc.) are indicated by using specific abbreviations (see the **List of morphological specifications** and the **List of abbreviations for syntactic functions**).

➢ At the semantic level, **semantic types (groups)** are specified. Adjectives are divided into three groups: those expressing physical qualities (e.g. *aukštas* ('tall, high')*, mažas* ('small'), *stiprus* ('strong')), evaluative adjectives (e.g. *geras* ('good')*, gražus* ('beautiful')*, malonus* ('pleasant')), and qualitative adjectives (e.g. *medinis* ('wooden')*, fizinis* ('physical')). Verbs are classified as denoting an active action (e.g. *bėgti* ('to run')*, skaityti* ('to read')*, daryti* ('to do')), a spontaneous process (e.g. *lyti* ('to rain')*, snigti* ('to snow')*, temti* ('to darken')), or a state (e.g. *mylėti* ('to love')*, gulėti* ('to lie'), *tikėtis* ('to expect')). For nouns, the list of groups was non-finite: nouns are semantically classified into groups denoting people, body parts, identity, names, professions, species, country names, plants, objects, places, dimensions, actions, etc.

➢ At the lexical level, **collocates** (words that commonly cooccur with the headword) are provided (e.g. the collocates of the verb *bėgti* ('to run') are *laikas* ('time') and *greitai* ('fast')). A word is considered to be a collocate of a headword if it is used 3 or more times with the headword in the Pedagogic Corpus.

➢ In all usage patterns, the vertical slash (|) is used to indicate alternative variants in the pattern. For example, in pattern 1.2 of the verb *bėgti* ('to run'), the variability is noted at the grammatical level: the slash is used to indicate that three types of prepositional constructions can be used for the adverbial.

Pronunciation

➢ To search for phonetic information and pronunciation, click on **Phonetic information**. The information here includes the accented word form, its transcription, and an audio recording. The sounds are transcribed using the symbols of the International Phonetic Alphabet (IPA): https://kalbu.vdu.lt/mokymosi-priemones/tartis/transkripcija/.

➢ The accentual class is indicated next to each noun and adjective. If a word has two variants of accent, both variants are provided (for example, the adjective *asmeninis* ('personal') can be accented according to the first or second accentual class):

➢ In the tool **Headword Search**, the tab **Word forms** lists all the word forms used in the corpus. If a word has two variants of accent, the inflectional forms are accentuated according to both variants.

Grammar

➢ After each headword, its part of speech is indicated. The part of speech is also indicated for idiomatic expressions (e.g. *akis paganyti* – verb ('to feast one's eyes')) and phrasal units that perform the function of an interjection (e.g. *labai malonu* – interjection ('very nice')). For compound nouns, morphological information is provided for each word in the compound (e.g. *oro uostas* – noun_sg.gen noun ('airport')).

➢ Additional labels are added to the part of speech if (1) the noun is used only in the singular or only in the plural form, e.g. *dėmesys* **noun_sg** ('attention'); *akiniai* **noun_pl** ('glasses'); or (2) the noun or verb is reflexive, e.g. *apsilankymas* **noun_refl** ('visit'); *klausytis* **verb_refl** ('to listen').

➢ The tab **Word forms** provides the inflected forms in which the headword is used in the corpus; the notation of morphological categories follows the conventions of the Leipzig Glossing Rules (see the **List of morphological specifications**).

➢ Information on the **grammatical collocability** of headwords is provided at the grammatical level of their **usage patterns** and includes information about their **syntactic features**. These are displayed by indicating their **syntactic functions**, such as Sub – subject, Obj – object, Pred – predicate, Adv – adverbials, etc. (for more detailed information, see the **List of abbreviations for syntactic functions**).

Information at the grammatical level also includes relevant **morphological features**, e.g. Obj_acc, which stands for the object (abbreviated as *Obj*) used in the accusative case (abbreviated as *acc*). The most common **morphological categories** indicated in the lexicon include case, number, and verb forms (e.g. imperative, third person, participial form) (for more detailed information, see the **List of morphological specifications**).

Here is an example of how grammatical collocability is represented at the grammatical level for the noun *sveikata*: pattern 1.5. shows that when used with a verb, this noun performs the function of an object and, depending on the verb, it is used either in the accusative case (SVEIKATA_acc), as in *saugoti sveikatą* ('to protect one's health'), or in the instrumental case, as in *pasiteirauti apie sveikatą* ('to inquire about someone's health').

➢ Auxiliary and semi-auxiliary verbs (e.g. *būti* ('to be'), *galėti* ('to be able'), *turėti* ('to have')) are frequent in the Pedagogic Corpus, but in the lexicon, they are not described in separate entries because their usage is clearly seen in the usage patterns of independent words. Some verbs (*imti* 'to take', *laikyti* 'to hold') can be used both as independent and auxiliary verbs; such verbs are included in the headword list and are represented in separate entries.

➢ In the headword list, there are very few **grammatical words** (e.g. particles, conjunctions, and prepositions, such as *tik* ('only'), *dar* ('more'), *turbūt* ('probably'), and *beje* ('by the way')). These parts of speech are described if they overlap with a certain inflected form of a particular noun or a verb. For example, since one of the inflected forms of the verb *vadintis* ('to be called') also performs the function of a particle, the lexical database contains two entries: one for the verb *vadintis* ('to be called') and another one for the particle *vadinasi* ('it means').

Collocations

➢ Information on **lexical collocability** is provided at the third – lexical – level of usage patterns. To represent this information, the entry provides **collocates**, which are defined as words commonly used with the headword (for example, the collocates of the verb *bėgti* ('to run') are *laikas* ('time') and *greitai* ('fast')). A word is considered to be a collocate of the headword if it is used 3 or more times with it in the Pedagogic Corpus.

➢ The tool **Collocate Search** allows you to find words for which the search word is a common collocate. For example, the search results for the verb *sveikinti* ('to greet') will display the noun *proga* ('occasion').
If you search for the noun *proga* ('occasion') by using the tool **Headword Search**, 8 patterns will be displayed. However, the **Collocate Search** will result in a single pattern, which is pattern 1.1, where the collocate *sveikinti* ('to congratulate') is used. The expanded information shows that when the noun *proga* ('occasion') is used with an attribute that does not agree with the noun, it refers to a reason (*gimtadienio proga* 'on one's birthday'); when used with the verb *sveikinti* ('to congratulate'), it forms the collocation *sveikinti gimtadienio proga* ('to congratulate someone on their birthday').

➢ In the **Collocate Search**, you can find information about some of the words that are not included in the headword list. For example, the noun *įspūdis* ('impression') is not on the list, so we will not find it with the tool **Headword Search**, but we can find it in the collocations of some other words. For example, if you enter *įspūdis* ('impression') in the **Collocate Search**, the search results will show that it collocates with some adjectives (*blogas įspūdis* ('bad impression')*, geras įspūdis* ('good impression'), *malonus įspūdis* ('pleasant impression')) and verbs (*padaryti įspūdį, palikti įspūdį* ('to make an impression')).

Word formation

➢ When searching for a part of a word using the tool **Headword Search**, the search results will display related derivatives, e.g. vadin -> *vadinti, vadintis, pavadinti, išvadinti*.

➢ Derivatives are indicated next to the relevant meaning of the headword and include items that occur in the Pedagogic Corpus. Derivatives have active shortcuts: clicking on the selected derivative opens its entry in a separate window.

➢ Words that are derivationally similar but semantically not related to the headword (e.g. *sakyti* and *užsakyti* ('to say' and 'to order')) are marked with a hashtag (#). If the interrelation is set only diachronically, it is not marked in the lexicon, e.g. *tikti, tikslus, tikslas* ('to fit', 'exact', 'purpose'); *gyti* and *gyvas* ('to heal' and 'alive'); *skiesti* and *skaidrus* ('to dilute' and 'clear').

➢ There may be some chains of derivatives with the gaps since not all derivatives from the chain are used in the Pedagogic Corpus. For example, the word *neįvykti* ('not to occur') does appear in it, but the base word *įvykti* ('to happen') is missing. In such cases, the gap word (here: the base word) is marked with the @ symbol.