Issues in Translation of Linguistic Collocations
Lingvistinių kolokacijų vertimo ypatumai

Translation / Vertimas

Miglė Sulpinaitė
Bachelor of Translation of Technical Texts, Kaunas University of Technology, Lithuania

Jolita Horbačauskienė
Dr., associate professor, Kaunas University of Technology, Lithuania

Ramunė Kasperavičienė
Dr., associate professor, Kaunas University of Technology, Lithuania.

Abstract

Collocation is a term used to describe two-word combinations when there is a limited number of words that may precede or follow another word. There are many discussions about the term itself; therefore, the characteristics and taxonomies of collocations are under debate. One of the accepted linguistic classifications of collocations is based on properties of collocations such as compositionality, modifiability, substitutability, and internal association (Lu & Xu, 2005, p.4). The aim of the paper was to analyse the characteristics of linguistic collocations in English and their possible translation transformations in Lithuanian. The research targets four types of linguistic collocations: strong, loose, fixed and fully fixed. In the analysis, two main types of translation shifts are considered: level shift and category shift, further subdivided into several types. However, the majority of the analysed collocations undergo a category shift in the translation process. The results of the research indicate that the most common type of linguistic collocations in popular science texts is strong collocations. Loose collocations account for one-third of the analysed instances. Fixed collocations account for the third largest group and the least common group is fully fixed collocations.

KEYWORDS: Linguistic collocation, translation transformation, category shift, rank shift, intra-system shift, structure shift.

Introduction

Collocation is defined as a set of words that commonly occur together (McCarthy, O’Dell, 2005). This linguistic phenomenon is significant in language learning as it is an element of formulaic language. The term collocation is relatively new to the branch of linguistics; however, the concept has been both intensively and extensively analysed. There are many discussions about the term itself; therefore, the characteristics and taxonomies of collocations are under debate. On the other hand, linguistic collocations are not as widely and comprehensively investigated. They are most often approached from the lexical perspective. Therefore, the aim of the paper is to analyse the characteristics of linguistic collocations in English and their possible translation transformations.

The research methods employed in the paper are quantitative and qualitative analysis. The research targets four types of linguistic collocations: strong, loose, fixed and fully fixed. For the research purposes, 250 examples of linguistic collocations and their translated equivalents were selected from popular science magazines.
The verb *collocate* originates from the Latin verb *collocare* and the translation given by the Online Etymology Dictionary is *place, arrange*. The term *collocation* is relatively new in linguistics, first introduced by Firth, who argued that collocation was ‘lexical meaning at the syntagmatic level’ (Firth, 1957, p.96). Despite his attempts to describe meaning at the syntagmatic level, many researchers and scientists still define collocations in terms of a tendency of words to occur with other words. The concept of collocation was later refined by Halliday and then by Sinclair and became central to their study of lexis (Palumbo, 2009).

The most commonly used definition states that collocation is a pair or set of words that commonly occur together (McCarthy, O’Dell, 2005; Handl, 2008). Parrot (2010) claims that collocation is a term used to describe two-word combinations when there is a limited number of words that may precede or follow another word. Some authors define collocations as a group of words that have ‘mutual expectancy’ (Jackson, 2007, p.106). It means that words have the ability to predict the likelihood of the occurrence of another word. There is a limited number of words that can be used with other words and the direction of words is important, as demonstrated in the collocation ‘green tea’: *green* may occur with many different words (*car, house, fence*, etc.), while the word *tea* is expected to occur only with a limited number of words (*green, black*). This example demonstrates that mutual expectancy of words can be stronger or weaker and it depends on the direction of words. Many word combinations are ‘so frequently recurrent that the choice of one of their constituents appears to automatically trigger their selection of one or more other constituents in their immediate contexts’ (Bartsch, 2004, p.11).

Sinclair (1991) defines collocation as the occurrence of two or more words with a small space between them in a text, dividing collocations into two groups: 1) upward – when the second word is the collocate and the first one is the node (usually composed of prepositions, adverbs, conjunctions, pronouns), and 2) downward – when the first word is the node and the second is the collocate (mostly composed of verbs, prepositions, adverbs, adjectives and nouns (pp.115–116). Hausmann (1985) states that collocations are composed of only content words (nouns, verbs, adjectives) and fails to have regard to function words (prepositions, pronouns, articles). However, this statement is unjustified since many collocations include function words, e.g. the indefinite article *a in to make a call or the preposition of in a sense of pride*.

Collocations are sometimes confused with idioms. McCarthy and O’Dell (2005) define idioms as groups of words in a fixed order that have a special meaning; therefore, the meaning of separate words cannot be guessed. The main difference between these linguistic phenomena is that the meaning of idioms is pre-constructed and not semantically literal; therefore, the words have not retained their conventional literal meanings and cannot be analysed as such, whereas collocations are pre-constructed and semantically literal (non-metaphorical), which means that the words have retained their conventional literal meanings (Benson et al., 1986).

Learning collocations is an effective way of learning the language. According to Hyland, collocations are ‘an important component of fluent linguistic production and a key factor in successful language learning’ (2008, p.4), as they make language sound in the most natural way. The ability to use collocations gives the opportunity to be more expressive and colourful as well as to be more precise and not to repeat what has already been said. Finally, learning collocations improves one’s style of writing. According to Cowie, ‘collocations present one of the biggest problems for the learners of English – the more information a student can find in a dictionary on this topic the better’ (1983). This point of view suggests solving the issue of learning the English language by learning collocations.

There is quite a significant number of classifications of collocations. This happens because of the disagreement between linguists about the definition and even typology of collocations. In
addition to this, there are no general criteria for a collocational classification since linguists
tend to differently approach them (from the statistical, semantic, functional, or syntactic per-
spectives). Even though there are many views as to the structure of collocations, there is
no consensus on a unified structure of collocations. This problem derives from the lack of
agreement between language experts about the definition of collocations. As the attitudes
towards it are divergent, the typology and the structure of this linguistic phenomenon remain
questionable.

One of the accepted linguistic classifications of collocations is based on properties such as
compositionality, modifiability, substitutability, and internal association (Li et al., 2005, p.4)
(see Table 1). **Idiomatic collocations (or fully fixed collocations)** have a fixed form and their
constituents cannot be replaced with any other words. It is a rigid combination; therefore,
the meaning of the collocation cannot be determined from the meaning of its parts (Dale,
Moisl, Somers, 2000), e.g., to catch a cold refers to becoming infected with a cold. **Fixed col-
locations** also have a fixed form and their constituents cannot be replaced with other words –
they cannot be modified or substituted. The constituents of these collocations repeatedly
c-o-occur with each other, are bound specifically to each other and do not co-occur with oth-
er lexical units freely (Loos et al., 2003), e.g., utter disaster is a fixed collocation since there
are not many words that could be used instead of word disaster. Moreover, the words may
not be replaced by other words to maintain the same meaning. **Strong collocations** can be
modified. However, the modification is very restricted. In cases when modification is possible,
the order of constituents cannot be changed. Accordingly, there are only a few synonyms
that could replace the constituent (Li et al., 2005). For example, in the collocation form alli-
ance, the word alliance tends not to be used with many other words. **Loose collocations (or
normal collocations)** have very loose restrictions. They allow more substitutions of compo-
nents; the change of the order of the sentence is also allowed. Furthermore, the components
may be substituted by synonyms and combinations that have a similar meaning (ibid.), e.g.,
lawful income and legitimate income. However, a limitation is still required to restrict the sub-
stitution. Synonym information and substitutability allows defining whether the collocation is
idiomatic, fixed, strong with very limited substitution, or loose with only limited substitution.

Four basic linguistic properties that demonstrate the difference between collocational ex-
pressions and non-collocational expressions are distinguished (Wermter, 2008).

<table>
<thead>
<tr>
<th>Type 0</th>
<th>Type 1</th>
<th>Type 2</th>
<th>Type 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Idiomatic collocation</td>
<td>Fixed collocation</td>
<td>Strong collocation</td>
<td>Loose collocation</td>
</tr>
<tr>
<td>Compositional</td>
<td>No</td>
<td>Limited to yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Synonym substitutable</td>
<td>No</td>
<td>No</td>
<td>Very limited</td>
</tr>
<tr>
<td>Order alter</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Modifiable</td>
<td>No</td>
<td>No</td>
<td>Very limited</td>
</tr>
<tr>
<td>Statistical significance</td>
<td>Not required</td>
<td>Not required</td>
<td>Required</td>
</tr>
<tr>
<td>Examples</td>
<td>Bluetooth</td>
<td>Diplomat immunity</td>
<td>Form alliances</td>
</tr>
</tbody>
</table>
co-occurrence involves frequency, regularity and textual proximity. Frequency is the key factor in distinguishing collocations. In most cases, words that conclude a collocation are close to each other. Non- or limited compositionality, in simplified terms a principle of compositionality, sometimes called ‘Frege’s Principle’, is defined as the meaning of every expression in a language having a function of the meaning of its parts (Goldberg, 1995). However, a collocational expression does not always represent the precise meaning of its parts. Idioms are sometimes considered as a subtype of collocations. Therefore, the meaning is different from its meaning as a free word combination, and this property marks the difference between collocations and free word combinations. Non- or limited substitutability represents a particular meaning in a collocation: if a word of a collocation is replaced with another word, then the whole expression loses its meaning. This property demonstrates that the components of collocation can be substituted neither syntactically nor semantically. This property is applied even in the case when a substitute word is the same part of speech or has a closely similar meaning – a synonym, e.g., strong tea vs. powerful tea. Even though words strong and powerful are synonymous, they cannot replace each other in this context. Another example is white wine vs. yellow wine. Although yellow seems to be a good description of the colour of wine, we cannot say yellow wine instead of white wine. Non- or limited modifiability describes the syntagmatic effect that there is only a limited number of collocations that can be modified because other lexical items cannot be inserted within a phrase. According to Benson (1989), this is mostly the case for idiomatic expressions, as in the example: kick the bucket vs. kick the large bucket. This example illustrates that this idiomatic expression cannot be modified by adding an additional lexical item since the phrase itself loses its collocational meaning and is unaccountable for natural speech.

Collocations may also be considered in terms of formulaic language. According to Schmitt, they are ‘recurrent multi-word lexical items that have a single meaning or function’ (2010, p.136). The importance of collocations as an element of formulaic language cannot be denied. Schmitt adds that the use of formulaic language ‘eases the cognitive burden’ for both the speaker and the listener and it has been seen to enhance the fluency of speech (ibid.). Non-native speakers who use formulaic language are regarded as more proficient in both written and spoken discourses. Wray & Perkins (2000) argue that the roles of formulaic language coincide with the roles of collocations in benefitting a non-native speaker. Formulaic language not only allows the speaker to use ordinary expressions but also eases the whole process of thinking. Kecskes states that ‘formulaic expressions ease the processing overload because they are already made and do not require of the speaker/ hearer any putting together’ (2014, p.112). The listener is able to process the information in a shorter period of time as expressions used for providing the information can be predictable.

The key factor of formulaic language is that expressions are like clichés. They are repeated by speakers all the time and they cannot be transformed. Therefore, this leads to the conclusion that every formulaic expression is like a multiword collocation (Lewis, 2000). No addition or removal of any lexical item is allowed since this affects the whole structure and the meaning of an expression. Collocations tend to convey holistic meanings that are more than the sum of the individual parts. As Gairns and Redman (1986) note, they diverge significantly from the literal meaning and operate as a single semantic unit.

Collocations cannot be translated on a word-by-word basis. In translation theory, changes are determined as shifts since they take place during the translation process. According to Munday (2001, p.55), translation shifts are linguistic changes that occur in translation from a source text (ST) to a target text (TT). Catford describes translation transformations as “departures from formal correspondence when translating from source language into target language” (Catford in Cyrus, 2009, p.90). This definition relies on formal correspondence and textual equivalence. The textual equivalence is obtained when source and target items are
interchangeable. This is achieved when the target language text is equivalent to the source language text. Meanwhile, formal correspondence is a relationship between two linguistic categories that occupy relatively the same place in their respective languages (ibid.).

Figure 1 illustrates the translation shifts according to Catford who claims that there are two main types of translation shifts: level shift and category shift, which is further subdivided into several types. The translation shift model is adopted in a number of studies as it helps to discover more information about the relations existing between diverse languages at the syntactic level.

The data for analysis was collected from popular science magazines. A total number of 250 English collocations were randomly collected from original English articles. Their Lithuanian versions taken from translated versions of the same magazines were analysed in terms of translation transformations.

The results on the distribution of linguistic collocations are provided in Figure 2. The most frequently used type of linguistic collocations in the investigated texts is strong. The total number of 132 strong collocations (53 %) was found. Natural surroundings is a collocation where modification is very restricted. The word surroundings is not frequently used with many other words. It mostly collocates with adjectives and there are quite many of them, e.g., attractive, pleasant, unfamiliar, etc. However, there is no synonym that could replace the constituent natural. That is why this collocation is considered to be strong. Political propaganda is another example of strong collocations. The most significant reason for the collocation being a strong one is that it does not have many words that could go with the word propaganda. The order of the constituents cannot be changed in this collocation. Cancer patients is a collocation where modification is also restricted; however, the order of the constituents can be changed though addition of a lexical unit is necessary, e.g., patients with cancer. Cancer does not collocate with many other words, but its constituent patients has a few synonyms that could replace it, e.g., sufferer, victim.
These qualities make the collocation strong. As the examples show, strong collocations can be modified in a very restricted way. Furthermore, the node occurs with only a limited number of words. Accordingly, there are only a few synonymous words that could replace the collocate (Li et al., 2005, p.126).

The second biggest group of linguistic collocations is loose collocations. Figure 3 demonstrates that loose collocations account for 32% (n=80) of all the linguistic collocations found. The examples provided below show the characteristics of this type of linguistic collocations and explain the reasons why they are attributed to this type. In the collocation taking notes, the word note collocates with a variety of parts of speech, such as adjectives, verbs and nouns. The collocate taking can be replaced with a number of words and, therefore, there are quite a lot synonyms possible, e.g., keep, make, etc. Additionally, the order of the components of the collocation can be changed and the modification is allowed. Main goal is a collocation in which the node goal can collocate with a relatively large number of words, mainly adjectives. Moreover, the collocate main can be replaced by many synonymous words, e.g., major, primary, prime. In the collocation large amounts, the node amount collocates with a number of words, mostly adjectives, many of which are exact synonyms: enormous, huge, massive, substantial, etc. Loose collocations allow more substitutions and modifications of components than any other type of linguistic collocations. However, as in other cases of collocations, limitation is still present.

Fixed collocations account for 8% (n=20) of the total number of linguistic collocations found. Even though the number is not as large as in previously discussed types, they are quite common in English. Black crows is a fixed collocation with a very limited number of words that can be substituted for the word black. Only a few verbs and adjectives that collocate with the noun crow are possible. Moreover, there are no synonyms that could replace the collocate since synonymy is not allowed in this type of collocations. Another significant feature of fixed collocations illustrated by this example is that none of the constituents can be replaced by other words to maintain the same meaning. In the collocation water vapour, the node vapour collocates with only a few words, such as cloud, form, emit. There is no possibility to replace the word water in this collocation. In great plague, the node plaque collocates with great. In fact, plague collocates with several other verbs and adjectives, e.g., bubonic, suffer, spread. However, there are no synonymous adjectives to great. In other contexts, this word could have a considerable number of synonyms, whereas in this case synonyms could not replace the word itself. Lexical units of this collocation are bound to each other and they cannot occur with other lexical units freely (Loos et al., 2003). Even though the number of found and analysed fixed collocations is not substantial, they are quite common. Not only the constituents of this type of collocations can be replaced to maintain the same meaning since words are bound to each other, but also the modification and substitution is not allowed.

Fully fixed, i.e. idiomatic, collocations account for 7% of the analysed occurrences (n=20), for example a fully fixed collocation greenhouse gasses is idiomatic, as its meaning cannot be determined from the meaning of its parts. In this case, green is not understood as a colour and the whole word does not mean that the house is green, resulting in an idiomatic collocation. Neither the modification is allowed nor the order of the constituents can be changed.

Translation transformations are commonly known as translation shifts and are determined as linguistic changes that occur in translation of the source text to the target text (Munday, 2001, p.55). In order to analyse the translation transformations of the linguistic collocations, translation shifts according to Catford were applied.

Figure 3 demonstrates the frequency of the translation transformations applied in the target text. The most frequently used shift in the investigated texts was category accounting for 88%
of the total number. Level shifts were discovered in only 14 occurrences in the source texts, which makes 12 % of the total number. Since the number is not significant, the research will focus on category shifts only. According to Catford, category shift is divided into four groups.

Figure 4 demonstrates the frequency of each category shift.

The most common form of category shift in the research data is rank – 37 %. The second most frequently applied shift is intra-system with 34 % of the total number. The number of class and structure shifts is relatively low, 16 % and 13 %, respectively. Strong collocations comprise the largest group of linguistic collocations, 53 % of the total number.

The distribution of category shifts of strong collocations is represented by 44 % of intra-system shifts, e.g. brain in the collocation brain activity in the ST is in singular, but its translation smegenų veikla in the TT is in plural; divorce in the collocation divorce rate in the ST was transformed from the singular into the plural skyrybų skaičius in the TT (see Figure 5). However, the opposite process of an intra-system shift was also noticed, e.g. surroundings in the collocation natural surroundings in the ST is in the plural, while natūrali aplinka is in the singular in the TT; data in the collocation raw data in the ST was transformed from the plural into the singular in neapdorota informacija in the TT.

Rank shift is the second largest group of category shifts in strong collocations accounting for 31 % of all the instances. The translation equivalent is at one rank in the ST and at the other rank in the TT, e.g. religious devotion of the ST became a word religingumas in the TT; became friends of the ST was transformed into
draugystė in the TT; the collocation *walking distance* of the ST was transformed into *pėsčiomis* in the TT.

Class shifts make 18% of all the category shifts, i.e. shifts from one part of speech to another. In the majority of cases, a noun was transformed into a verb, but other parts of speech were also observed, e.g. the noun *maturity* in the collocation *sexual maturity* of the ST was changed into the verb *subręsta* in the TT; the noun *attention* in the collocation *constant attention* of the ST was transferred as the verb *saugotis* in the TT; the noun *office* in the phrase *office block* became an adjective *administracinis* in the TT.

Structure shifts account for 7% of all the category shifts. According to Munday (2001, p.93), this is the most commonly occurring shift. However, as the data analysis shows, the results are opposite in the case of strong collocations found in popular science texts, since the minority of the translation transformation occurrences were structure, e.g. the verb and adverb collocation *rust away* of the ST became adverb and verb *kitur surūdytų* in the TT; the adverb and verb collocation *freely confess* of the ST was changed into the verb (object) and adverb combination *pripažįstu (tai) atvirai* in the TT; the noun phrase *oil revenue* of the ST changed the order in *pajamos iš naftos* in the TT into noun, preposition and noun.

Even though the transformations are essential for fluent translation, many of the collocations are not translated using shifts, which indicates that the most appropriate way of information rendering depends on the translator’s proficiency.

Distribution of category shifts of loose collocations is presented in Figure 6.

Rank shifts comprise the largest group with 12 instances found making 44% of all the category shifts in the group of loose collocations. The majority of the collocations were transformed from a phrase level to a word level. However, there were instances where collocations were transformed into the whole clause, e.g. the phrase *rarely used* of the ST was transformed into a single word *retas* in the TT; the collocation *major threat* of the ST became a clause *pavojingas tik vienas dalykas* in the TT; the collocation *main factors* of the ST was transformed as *veiksniai* in the TT. The examples above demonstrate that in the case of rank shift the majority of collocations fail to retain their collocational characteristics and become single lexical units.

The second largest group of category shifts applied to loose collocations was *intra-system shift*, as illustrated by the following examples: evidence in the collocation *evidence suggests* of the ST is in the singular, while the translation *radiniai atskleidžia* in the TT is in the plural; the singular *gate* in the collocation *front gate* of the ST was transformed into the plural *paradainiai vartai* in the TT; the plural *decisions* in the verb phrase *make the decisions* of the ST was transformed into the singular *priimti sprendimą* in the TT;

Class shifts account for 19% of all the category shifts, e.g., the noun *explanation* in the collocation *plausible explanation* of the ST was transformed into the verb *paaiškina* in the TT. Moreover, an adjective *plausible* in the same collocation in the ST became the adverb *patikimiau* in the TT.

![Figure 6: Category shifts of loose collocations](image-url)
Structure shifts comprise the minority of category shifts among loose collocations making 7% of the total number, e.g., the verb and adverb collocation declined sharply of the ST became the adverb and verb collocation labai sumažėjo in the TT; the noun (pronoun) and verb collocation sparks (that) ignited of the ST was changed into verb and noun įsižiebusi kibirkštis in the TT.

The third largest group of linguistic collocations found was fixed collocations. The category shifts of fixed collocations are presented in Figure 7.

The following examples illustrate rank shifts comprising 38% of all the transformations in this category, e.g., the collocation cooking pot of the ST was changed into the phrase indas gaminti valgiui in the TT; the collocation blood pressure of the ST was transformed into kraujosjąpos in the TT; bowling ball of the ST became a single word rutulys in the TT. Intra-system shifts accounted for 12% of all the category transformations in the group of fixed collocations. In the data analysed, every single instance was transformed from the singular to the plural, e.g. dig in the collocation archaeological dig was singular in the ST, while the translation archeologiniai kasinėjimai in the TT was in the plural; singular detector in the collocation metal detector of the ST was changed into plural metalo detektoriai in the TT.

Structure shifts occurred in 38% of the instances in the group of fixed collocations, e.g., word order was changed in the ST collocation job interview to pokalbis dėl darbo in the TT.

Class shifts were also less common translation transformations of fixed collocations, e.g., the adjective electric in the ST collocation electric cables was rendered as the noun elektros in the collocation elektros kabeliai in the TT.

Rank and structure shifts were the most common category shifts in the groups of fully fixed collocations. The following examples illustrate the grammatical transformation, e.g. the order of words in the ST collocation heart of the city was changed into miesto širdis in the TT; source of light of the ST was rendered as šviesos šaltinis in the TT; time of day of the ST was transformed into paros metas in the TT.

In the process of translation, fully fixed collocations were transformed from a phrase-level collocation to a word, a clause, or a three-word phrase, e.g., the ST collocation beneath the surface was transformed into a single word gilyn in the TT; pace of change of the ST was rendered as a single word pokyčiai in the TT; and the ST collocation cast of characters was transformed into the three-word collocation teatro personažų sąrašas in the TT. Class and structure shifts were rare. Besides, in the process of translation, some collocations lost their collocational characteristics.

Collocations are defined as a set of words that regularly occur together. There are four types of linguistic collocations: strong, loose, fixed and fully fixed. Strong collocations can be modified; however, the modification is very restricted and the order of constituents cannot be changed. Synonymous substitution is very restricted as well. Loose collocations are defined as collocations where the constituents can be modified and substituted by synonyms or words with a similar meaning. Although these are the least restricted collocations, the limi-
tation is still required. Fixed collocations have a fixed form and the constituents are bound to each other; therefore, they cannot be substituted or modified. Fully fixed collocations (idiomatic collocations) are rigid combinations and are the most restricted linguistic collocations. The results of the research indicate that the most common type of linguistic collocations in popular science texts is strong collocations with more than half of all the instances. Loose collocations account for one third of the analysed examples. Fixed collocations account for the third largest group and the least common group is fully fixed collocations.

Category shifts were identified in more than one-third of the analysed collocations. The results of the analysed data show that the most common form of translation transformation is rank shift, in which case a number of translated collocations lost their collocational characteristics and became single lexical units. The second largest group of category shifts applied in the analysed texts is intra-system shift with a morphological category change from singular to plural or vice versa. The occurrences of class shifts are quite frequent when collocations of the ST change their grammatical class in the TT via translation transformation. Structure shifts form the smallest group of translation transformation occurrences. Translation transformations are inevitable, especially in cases of fixed and fully fixed collocations where full equivalence is hardly possible.

References

Miglė Stulpinaitė, Jolita Horbačauskienė, Ramunė Kasperavičienė. Lingvistinių kolokacijų vertimo ypatumai