In the Company of Words in Translation: Conventionalisation of Collocations in Croatian Translations from English

Posedi, Patrik

Master's thesis / Diplomski rad

2024

Degree Grantor / Ustanova koja je dodijelila akademski / stručni stupanj: **University of Zagreb, Faculty of Humanities and Social Sciences / Sveučilište u Zagrebu, Filozofski fakultet**

Permanent link / Trajna poveznica: https://urn.nsk.hr/urn:nbn:hr:131:043051

Rights / Prava: In copyright/Zaštićeno autorskim pravom.

Download date / Datum preuzimanja: **2024-10-12**

University of Zagreb

Faculty of Humanities and Social Sciences

Department of English

Patrik Posedi

In the Company of Words in Translation:
Conventionalisation of Collocations in Croatian Translations from English

Master's Thesis

Supervisor: Prof. Nataša Pavlović, PhD

Zagreb, 2024

Sveučilište u Zagrebu

Filozofski fakultet

Odsjek za anglistiku

Patrik Posedi

U društvu prevedenih riječi:

konvencionalizacija kolokacija u hrvatskim prijevodima s engleskog

Diplomski rad

Mentorica: dr. sc. Nataša Pavlović, red. prof.

Zagreb, 2024.

Table of Contents

Abstract

The controversial notion of translation universals, described initially as absolute generalities of translation, is nowadays understood more along the lines of probabilistic claims about different subsets of translations. One such universal – conventionalisation – is in the focus of this study. Conventionalisation is described as a feature of translations according to which translations conform to the typical patterns of the target language, even to the degree of exaggerating them, meaning that unconventional source elements become conventional in the target. This universal is studied here by researching three conventional and three unconventional English collocations and their Croatian renderings. The aim is to determine if collocations become conventionalised through translation, with respect to genre (literary vs. popular scientific) and with respect to participants' perception of a source collocation's conventionality. The study combines an experiment, where participants are asked to translate two texts and answer questions about the collocations' conventionality, and a corpus-based analysis, aimed at assessing conventionality. The results of this small-scale study show evidence of conventionalisation, more prominently in the translations of the popular scientific text, while the participants' perception of a particular collocation also seems to have played a role. This study offers new insights into Croatian translations from English with respect to collocations and conventionalisation, as well as showing that translation universals can still be a fruitful topic of research.

Key words: translation universals, conventionalisation, collocations, genre, English-Croatian language pair

Sažetak

Kontroverzni pojam prijevodnih univerzalija, koje su prvobitno opisane kao apsolutne zakonitosti u prijevodima, danas je shvaćen više kao koncept koji opisuje probabilističke tvrdnje o različitim podskupovima prijevoda. Jedna takva univerzalija – konvencionalizacija – u fokusu je ovog istraživanja. Konvencionalizacija se opisuje kao karakteristika prijevoda prema kojoj se oni prilagođavaju tipičnim obrascima ciljnog jezika, čak do razine pretjerivanja, pa elementi koji su u izvorniku nekonvencionalni postaju konvencionalnima u prijevodu. Ta se prijevodna univerzalija ovdje istražuje s pomoću triju konvencionalnih i triju nekonvencionalnih engleskih kolokacija i njihovih prijevoda na hrvatski. Cilj je saznati postaju

li kolokacije zaista konvencionalnijima u prijevodima te postoji li razlika s obzirom na žanr (književni u odnosu na popularnoznanstveni tekst) i percepciju ispitanika/ispitanica o konvencionalnosti izvorne kolokacije. U ovom se istraživanju kombinira eksperiment, u kojem ispitanici/ispitanice prevode dva teksta i odgovaraju na pitanja o konvencionalnosti kolokacija, te analiza temeljena na korpusnom istraživanju čiji je cilj procijeniti konvencionalnost određene kolokacije. Rezultati ovog istraživanja manjeg opsega upućuju na postojanje konvencionalizacije, koja je istaknutija u prijevodima popularnoznanstvenog teksta, dok percepcije ispitanika/ispitanica također pokazuju značajnu ulogu u konvencionalizaciji. Ovim se istraživanje pružaju nova saznanja o hrvatskim prijevodima s engleskog jezika s obzirom na kolokacije i konvencionalizaciju te pokazuje da su prijevodne univerzalije još uvijek vrijedne istraživanja.

Ključne riječi: prijevodne univerzalije, konvencionalizacija, kolokacije, žanr, englesko-hrvatski jezični par

1. Introduction

The process of translation is not a disorganised activity. In every translation situation, translators have to juggle between different conventions, requirements or requests. Additionally, there are some regularities guiding the process, whether they are culture-specific *norms* or general *translation universals*. Identifying these regularities has been the goal of descriptive translation studies (DTS) since its emergence in the second half of the 20th century. DTS has developed as a separate paradigm within translation studies (TS) and it treats translations as "facts of target cultures" (Toury 2012: 23), trying to describe them as thoroughly as possible through its empirical, systemic and descriptive approach. However, the bulk of research produced so far has mainly focused on big languages, partly because of the volumes of texts translated from and into them and partly because such languages tend to have large languages resources such as digital corpora that make such research possible.

One such regularity on the lexical level that comes up in the relevant literature is the claim that translations show a tendency to conform to a target language (TL) to a degree that may not be seen at first glance. According to this, translators shape their target texts focusing more on the target circumstances than on the source ones to the point of exaggeration. However, this allegedly inherent feature of translation is not something that can be thought of as proven or as agreed on by the research community. New research is being produced that not only brings a better understanding of the topic, but also challenges it while offering valuable insights into the translation process itself. The goal of this study is to continue in the same vein, to describe further the translation process and Croatian translations from English by means of studying translated collocations. This is to be done with the DTS's theoretical background as the base and using large-scale languages resources that are nowadays more readily available for smaller languages, such as Croatian.

2. Theoretical background

2.1. Translation universals

The concept of translation universals has been a hotly debated issue since its introduction. This concept has its origin in the notion of *translation laws* (Toury 2012: 295). Toury (2012) described these laws as characteristics common to all translations regardless of the language combination, time period, text type or genre. He also proposed two such laws: the law of interference, i.e. transfer of source text phenomena into the target text (Toury 2012: 310), and the law of growing standardisation, i.e. the modifications of the source language to fit target

language rules and conventions (Toury 2012: 304), which motivated the work on translation universals. Baker (1993: 243) introduced these "universal features of translation" defining them as "features which typically occur in translated text rather than original utterances and which are not the result of interference from specific linguistic systems". In other words, translation universals are purported to be the regularities of translation that are true of all types of translations and that are typical of translations only, regardless of the language pair and the time period a text was translated in, i.e. regardless of the socio-cultural context[1] (Eskola 2004: 85).

With her seminal paper on the use of large electronic corpora in search of generalities of translation, in relation to either source texts or original texts written in a target language, Baker (1993) opened the door to the discussion on universals and their universality. Since then many such universals have been put forward by researchers, e.g. explicitation and conventionalisation (for an overview of the topic, see e.g. Chesterman 2004 or Kegalj 2023), and most of them, if not all of them, have been challenged in terms of their universality. That is why descriptive research on translation universals is necessarily probabilistic, as Toury (2004) reminds. Overgeneralisation (also in the form of translation ideals) is problematic in this sense since we cannot study all translations ever produced and there is always a risk that a culture plays a bigger role than perceived at first glance; however, we can make claims about particular types of translations and particular language pairs (Chesterman 2004: 41-43). This is again probabilistic but to a lesser degree than making claims about general features of all translations. In this sense, Chesterman (2004: 43-44) observes that some researchers prefer to use the term hypothesis instead of universal precisely due to the fact that they can never be proved to be *universal* of all translations. Consequently, the term translation universal is today established in the above-mentioned sense of a translation regularity (Kegalj 2023: 42).

To sum up, even though the goal of translation universals is to find regularities of translation that would hold across all translation types and across all languages, this is simply not feasible due to the sheer volume of translated texts and other extra-linguistic factors. Hence, a potential pitfall is making overgeneralisations. Instead, translation universals should be studied through a "bottom-up approach" (Chesterman 2004: 41), finding regularities first in different subsets of translations and then working our way up. After all, generalisations at the highest level are not

---

[1] A concept contrary to translation universals is the concept of translation norms. Also within DTS, Toury (2012: 61) described the act of translation as "a norm-governed activity", with these norms being highly target culture dependent, i.e. socio-culturally dependent, and as such not universal, i.e. true of all translations, as opposed to translation universals.

necessarily needed for further insights into the subject matter since, as Chesterman (2004: 43) points out, "[a]ny level of generalization can increase understanding"[2].

2.1.1. Conventionalisation

As mentioned above, more than a few translation universals have been put forward by researchers over the years. One such universal is *normalisation*, which Baker (1996: 45) defines as "the tendency to conform to patterns and practices which are typical of the target language, even to the point of exaggerating them". This concept manifests through "typical grammatical structures, punctuation and collocational patterns or clichés" (Baker 1996: 51). Baker (1996) also calls this universal *conservatism*, while Mauranen (2007: 12) uses the term *conventionalisation*, which seems to have become a more common name for this universal within TS, even though both terms can be found in the literature referring to the same phenomenon.

Apart from its various terms, conventionalisation has not been clearly delineated either. Mauranen (2007: 12-13) points out that conventionalisation overlaps with other translation universals: Toury's law of growing standardisation, described as the selection of typical constituents from a target language repertoire (also noticed by Kegalj 2023: 52), and simplification, commonly defined as frequent lexical choices. Another category that can also be found in the literature (e.g. Kegalj 2023: 51) is the category of a *textual fit*. This was briefly introduced by Chesterman (2004: 39) as a means to group universals that seek to explain translation phenomena in relation to how well they fit with non-translated texts in the target language, and as such, it can be understood as a category above conventionalisation, comprising other universals such as simplification. There is also one other universal that may be considered to overlap or perhaps to be a subtype of conventionalisation and that is *sanitisation*; for example, Bernardini (2007: 2) talks about "normalisation/sanitisation" and "conventionality". Sanitisation was first introduced by Kenny (1998: 1), who defines it as "the suspected adaptation of a source text reality to make it more palatable for target audiences". Even though she does not explain the universal in detail, she refers to it as conventionalisation on the lexical level and studies it on collocations (Kenny 1998). This, presumably, drives Chesterman (2004: 40) to describe sanitisation as concerning "more conventional collocations", while Kegalj (2023: 53) understands it more rigidly as not only conventional but rather "sanitised", taking

---

[2] Albeit debatable in more ways than mentioned above, Chesterman (2004: 46) observes the additional value of translation universals in the interdisciplinary methodological approaches that started to develop as part of their research and that has spread to other fields of research in translation studies.

into account both higher and lower lexical levels. Bernardini (2007: 2-3), on the other hand, describes sanitisation, which she equates with normalisation, as mentioned above, as "the tendency to select habitual target language expressions to render creative turns of phrases in the source text".

Another important point is that both Chesterman (2004) and Kegalj (2023) talk about the universal in its relation to the source text[3]. This is not in line with Baker's (1996: 52) note that conventionalisation is observed solely from the target language's point of view. Furthermore, Mauranen (2007: 13) criticises conventionalisation precisely on this point, in that it is not clear what is meant by the term: a translation's relation to its source text/language or to other non-translated target language texts/the target language[4]. More generally, this is also partly the cause of controversy regarding translation universals. Not only their universality, but also the fact that they are not clearly delineated and are differently understood among researchers contribute to their low status, and this is not only the case with conventionalisation but also with other translation universals, as Chesterman (2004: 44) remarks.

2.2. Collocations

The concept of collocation has been central to corpus linguistics since its beginnings. Even though the concept was already known, it was Firth who first used it as a means of explaining the meaning of words (Barnbrook, Manson & Krishnamurthy 2013: 35-38). Firth (1957: 11) utilised the concept of collocation by claiming that a certain word is recognised "by the company it keeps", i.e. by finding itself in the recurrent and frequent word combinations that one might expect. Such a word combination, a collocation, consists of a *node*, i.e. "the word whose lexical behaviour is under examination" (Sinclair 1991: 175), which is in the company of a *collocate*, i.e. "a word which occurs in close proximity to a word under investigation" (Sinclair 1991: 170). Hence, a collocation may be defined as "the occurrence of two or more words within a short space of each other in a text" (Sinclair 1991: 170).

According to Sinclair (1991), there are two ways in which a node can get its collocate: the open-choice principle and the idiom principle. The open-choice principle considers the selection of a

---

[3] The so-called S-universals, where S stands for "source", as opposed to T-universals, where T stands for "target", i.e. non-translated comparable texts (Chesterman 2004: 39).

[4] In TS, non-translated texts are comprised in corpora which Chesterman (2004: 44) simply calls corpora of "non-translated texts", noting that they are often referred to as comparable corpora. Furthermore, large monolingual corpora are referred to as "reference corpora" in this kind of research, while corpora containing source texts and their translations are called "parallel corpora" (see Bernardini 2007). In Croatian, Kegalj (2023) refers to a comparable corpus as "usporedivi korpus" and to a parallel corpus as "usporedni korpus".

collocate in terms of a choice among plenty of others on the part of a speaker, with the only restriction being grammaticalness (Sinclair 1991: 175.). On the other hand, according to the open-choice principle, a speaker has available a restricted number of collocates with a particular node because the node itself affects the selection of the collocate (Sinclair 1991: 173). In other words, it can be said that a collocation created according to the open-choice principle may be an infrequent collocation, i.e. *unconventional* – perhaps not yet observed in language. On the other hand, a speaker can choose a frequent collocation, i.e. *conventional* – already observed in language as a result of the idiom principle since the node selects the collocate that is in the "range of [its] idiomatic collocational preferences" (Hanks 2013: 15)[5]. To sum up, a collocation is a word combination consisting of at least two words, which do not have equal status as one word is subordinate to the other.

2.3. Genre and TS

Through the development of TS, genre has become an important concept in the study of translation (see e.g. Vermeer 1989). García Izquierdo and Montalt i Resurrecció (2002: 135) note that "[w]hen we translate, we do so for specific communicative situations and purposes; that is, we write translations that will fulfil the needs and conventions of specific textual genres in the target language and culture", and we do that both consciously and unconsciously. Schäffner (2000: 210) describes genres as "global linguistic patterns which have historically developed in a linguistic community for fulfilling specific communicative tasks in specific situations". In other words, speakers of a certain linguistic community perceive particular texts as belonging to a specific genre based on common conventional features those texts have. Such conventions comprise both formal and stylistic language features (Kegalj 2023).

However, the situation is not as clear-cut as this. Depending on different linguistic traditions, and the fact that genre has its origin in literary classifications, there are several different terms which are not clearly delineated and overlap significantly.[6] As Kegalj (2023: 23) points out, in Croatian linguistic tradition, genre typology has been reserved for literary works, while usage of language has primarily been classified into *functional styles* following the Russian and Czech functionalist approaches, as opposed to the Anglo-American pragmalinguistic approaches. The classification into functional styles rests on particular text functions and the linguistic features

---

[5] The notion of collocations is not a necessarily clear concept since different definitions have been proposed since Firth (1957) first introduced it; however, this is not relevant for the present study. For an overview of the issues, see Evert (2007).

[6] Apart from genre, these also include *register*, *text type* and *style*; however, this goes beyond the scope of the present study. For a detailed overview on the topic, especially with respect to TS, see Kegalj (2023).

employed on that basis in different contexts (Molnár 2019). In Croatian linguistics, five functional styles are commonly distinguished: journalistic, conversational, administrative, scientific/scholarly and literary (Silić 2006). For example, the scientific/scholarly functional style of Croatian is usually described as to-the-point, factually based, without much lexical variation; on the other hand, the literary style is described as the functional style of Croatian that allows most freedom, a style where everything is possible (Silić 2006).

3. Key concepts

For the purposes of the present study, the concepts presented and discussed above are defined in this section in the sense they are used in this study.

Based on the overview presented above, *translation universals* are understood here as necessarily probabilistic claims about regularities of translation that are observed in translated language (Toury 2004). As such, translation universals are considered to be features that can be hypothesised only of certain subsets of translation (Chesterman 2004: 41-43).

Regarding *conventionalisation* as a translation universal, this term is favoured in this study over the term *normalisation* and others that come up in the relevant literature as referring to the same phenomenon since it seems to be the most common, as discussed above. It is taken to signify a feature of translation evident in conforming to the typical patterns and practices of the target language insofar as an *unconventional* element of the source text is rendered *conventionally* in the target text (Baker 1996, Mauranen 2007). Such a perspective is necessarily investigated with respect to both source and target languages.

With respect to *collocations*, they are considered to be occurrences of at least two words in close proximity in a text, with one word, the *collocate*, being subordinate to the other, the *node* (Firth 1957, Sinclair 1991). Such a definition does not restrict the term to refer to only typical collocates of a node, but also atypical ones. Additionally, some restrictions are placed on the concept. First, the constituents of the collocation have to be in a syntactical relation with each other, which Evert (2007: 5) calls syntactic co-occurrence[7], giving the example of "a noun and its modifying adjective […] or a verb and its object noun". Second, idioms, semi-compositional expressions (e.g. *make a decision*) and merely lexically (e.g. *brush teeth*) or pragmatically restricted (e.g. *blue sea*) word combinations are excluded since they differ across languages due

---

[7] As opposed to surface co-occurrence (words just happen to appear close to each other in running text) and textual concurrence (words appearing together in a sentence or some other structure) (Evert 2007: 4-5).

to inherent differences in the language systems (Evert 2007), even though Sinclair (1991) considers them also to be collocations.

Finally, even though the notions of *genre* and *functional style* cannot be considered to refer to the same category, they do overlap insofar as they describe formal and stylistic linguistic features of particular texts. Hence, for the sake of simplicity, *genre* is used in the sense of a set of linguistic patterns that a certain group of texts share, which makes them perceived as a whole by a linguistic community (Schäffner 2000).

4. Previous research

Collocations in translated language have not been thoroughly researched, either with regard to translation universals or with respect to the English-Croatian language pair. However, since translation universals came into the limelight, some investigations into the topic have been conducted on different language pairs with different foci. One such investigation is Kenny's (1998) investigation into conventionalisation, i.e. sanitisation, to use her term, as discussed above. Her main aim is to provide a highly empirical way of researching conventionalisation on the lexical level. Hence, she proposes a contrastive monolingual and bilingual corpus approaches as the only way to evaluate if an item has been conventionalised. There are two phases: identifying a translation and its equivalent in the SL, focusing on the SL/TL and using reference corpora in the monolingual phase to identify what is common in the source and target languages. The author only informally puts her methodology to the test, and concludes that a more empirical and systematic approach, which is not intuition-dependent, is the way forward (Kenny 1998).

Later, Kenny (2001) conducts a study on lexical conventionalisation of creative, i.e. unconventional, collocations on the German-English language pair. Utilising a fairly similar methodology (both monolingual and bilingual phases) to the one described in Kenny (1998), the author discovers that creative collocations are not commonly conventionalised in their translations, with most of the translations being equally or similarly unconventional as their source, depending on the translator and the nature of the translated text (Kenny 2001: 210). However, as the author herself notes, this investigation is small-scale and on a deliberately selected node, and hence cannot be generalised (Kenny 2001: 210).

In addition to Kenny's (1998) investigation, Bernardini (2007) also presents methodological solutions for a corpus-based approach to the study of collocations in translated language.

However, she does not focus exclusively on conventionalisation[8], but rather to what happens to English collocations in Italian translations and vice-versa, and uses the notion only to explain translation shifts identified. Like Kenny (1998, 2001), the main parts of Bernardini's methodological approach are based on monolingual reference corpora. However, unlike the previously mentioned studies, Bernardini's investigation utilises automatic identification and extraction of particular types of possible collocations. In addition, she studies conventionalisation with regard to comparable TL texts, and not the source text. On the basis of higher *Mutual Information* values[9] in the translation corpora, Bernardini (2007: 14) comes to the conclusion that translated texts are more collocational than texts originally written in the respective TL. Even though Bernardini develops a sound methodological approach, some things remain ambiguous in her research, for example her criteria for conventionality. However, the methodological insights and the results should not be disregarded.

The studies presented so far have mainly focused on literary translation; however, one study that has looked at collocations in non-literary translation is Mauranen's (2000) study on word strings that carry text reflexive meaning. Mauranen (2000) focuses on the English-Finnish language pair and investigates textual connectors, selected verb-noun combinations, and different patterns in order to see if translations would show atypical combinations when compared to original Finnish texts. However, it is important to note that she focuses solely on translations as opposed to original writing without consulting the source texts. What she finds is that translations show different combinatorial preferences from the non-translated texts, i.e. different patterning, which is not ungrammatical or perceived as odd by the native speaker; the collocational behaviour is simply different (Mauranen 2000: 10). Mauranen's conclusion is, hence, that translators do what is possible but not necessarily typical.

Circling back to literary translation and some relatively recent research, Xia (2014) conducted a comprehensive diachronic study on conventionalisation on different language levels of Chinese literary translations from English from two different time periods. Utilising corpora of translated and non-translated texts, the author finds that collocations, as units that are in a

---

[8] As briefly mentioned above, Bernardini (2007: 2-3) uses the term *normalisation*, which she equates with *sanitisation*, and another, different phenomena she mentions is *conventionality*. Since she does not further elaborate on the differences between normalisation and conventionality, and since her definition of normalisation fits the definition of conventionalisation, conventionalisation is used in the descriptions of her study and findings.
[9] *Mutual Information* (MI) has long been used for collocation identification in corpus linguistics; however, the problem is that it exaggerates the significance of infrequent events, and should be, hence, avoided, especially since in recent years new and better statistical measures have been developed (Stefanowitsch 2020: 226-227).

grammatical relation with each other, idioms included (Xia 2014: 78), in translated texts from both time periods show a fairly similar degree of conventionalisation when compared to respective non-translated texts (2014: 94). However, Xia (2014: 94) also finds evidence of deconventionalisation in a corpus of translation from an earlier period in a contrastive bilingual phase and notes that some degree of conventionalisation is a result of inherently different language systems. Moreover, the author studies conventionalisation not as being universal, but rather socio-culturally dependent, i.e. as a translation norm (Xia 2014: 78). Hence, Xia attributes the deconventionalisation evidence from the earlier period to the prestige enjoyed by European languages in the Chinese culture in that period (2014: 94).

Another study pertinent to the subject at hand is a study of genre conventions of original Croatian texts and translations from English conducted by Kegalj (2023). She presents an investigation of original and translated maritime law texts with regard to their genre, and uses translation universals to explain some of the features noticed in the translated texts. Of interest here is her investigation into collocations present in translation. Kegalj (2023: 166-167) finds evidence of conventionalisation in translated collocations in that the verbs in translated verb-noun collocations exhibit less variation than those in source texts, as well as those in non-translated Croatian texts, meaning that translators conform to typical TL patterns and specific genre conventions. However, what she presumably fails to notice with regard to less variation than in source texts is that this perhaps stems from inherent differences between the two language systems, which constrict some collocations since they are, what Evert (2007: 4) calls, language-specific semi-compositional expressions. Furthermore, apart from the evidence in favour of conventionalisation of collocations, Kegalj (2023: 167) also finds some evidence of deconventionalisation in atypical collocations; however, the evidence of conventionalisation still prevails.

To summarise the findings of the studies mentioned above, evidence both in favour of and against conventionalisation of collocations in translated language have been identified in previous research. Overall, the tendencies in favour of collocation conventionalisation have been reported more frequently; however, the evidence of unconventional collocations cannot be disregarded. Furthermore, there is not enough evidence to make claims on the prevalence of one of these options in connection with genre or particular language pairs. One thing that is also evident from the studies presented are the methodological challenges of this type of research, some of which, though, have been resolved since the publications of the mentioned research articles with the development of more sophisticated corpus tools (see Gablasova et al. 2017). It

also seems important to note that research so far has mainly focused on typical collocations in literary texts and their translations or solely on collocational properties of translated language, and not contrastively on source language (SL) collocations and their TL counterparts, which is one avenue of research still worth exploring.

5. Aims and hypotheses

The main aim of this study is to examine the presence of conventionalisation as a purported translation universal, on the example of Croatian renderings of English collocations. The following hypotheses were tested:

H1 Both conventional and unconventional collocations from English source texts are rendered conventionally in Croatian target texts.

H2 The occurrence of conventionalisation depends on the genre of texts being translated, i.e.:

H2a Conventionalisation is more frequent in the translations of a popular scientific article.

H2b Conventionalisation is less frequent in the translations of a literary text.

H3 Translator's perception of an SL collocation as conventional or unconventional affects their renderings, i.e.:

H3a The more conventional a translator perceives an SL collocation, the more conventional its translation is.

H3b The less conventional a translator perceives an SL collocation, the less conventional its translation is.

The first hypothesis is based on the definition of conventionalisation itself, which states that unconventional SL elements, including collocations, become conventional TL elements (Baker 1996, Mauranen 2007), but also that conventional SL elements remain conventional TL elements by extension.

The second hypothesis and its subhypotheses stem from the fact that genre has been shown to play a major role in translation (e.g. Kegalj 2023). The difference between TL literary and popular scientific genre conventions are expected to influence how conventional a rendering of a collocation is. In other words, the literary genre is less rigid and more expressive, which means that unconventional collocations are not foreign to the genre, while the popular scientific genre

is described to be more strict and informative, which means that it does not have a high tolerance for unconventionality (Silić 2006).

Finally, the third hypothesis takes into account the translation process as a mental activity on part of the translator. Translators analyse the source text and consider different translation solutions that would match the source in more than one way, including the conventionality of collocations, while also thinking of the TL's possibilities and limitations, the different (genre) conventions, norms, expectations and all other factors involved in the translation process.

6. Methodology

Multiple data collection methods were used in this study. First, SL and TL reference corpora were used for assessing a collocation's conventionality. Second, an experiment (Appendix B) was conducted via *Lime Survey* on two occasions, where participants translated two texts of different genres that contained the same SL collocations, one text per occasion. Third, after the translation of each text, participants filled out a short questionnaire (Appendix B) concerning those collocations, also via *Lime Survey*, along with some basic demographic data. The translation of the second text and completion of its accompanying questionnaire took place after a minimum of three weeks had passed since the translation of the first text and completion of its accompanying questionnaire. This time interval was deemed sufficient for the collocations under examination to slip from the participants' memory in an attempt at making the collocations as inconspicuous as possible in the second text. All other details about data collection are explained in more detail below.

6.1. Text and collocation selection

Two texts of different genres were used in the experiment. The first one was a literary text and the second was a popular scientific one (see Appendix A). The texts were adapted to contain the same three conventional and three unconventional English collocations. The first text to be selected belongs to the literary genre since it was expected that such a text would be most likely to contain unconventional collocations, as well as conventional ones, and the goal was to find authentic unconventional collocations which would be used in the study. The novel *The Song of Achilles* (2011) by Madeline Miller was chosen for this purpose and an effort was made to go through the novel and try to find an excerpt that would contain three conventional collocations and three unconventional collocations, the criteria for which are discussed in more detail below. It was also important to bear in mind that the excerpt would not be too long lest participants of the experiment should be discouraged. Consequently, the length constricted the

number of collocations under examination; six was selected as an optimal number, three conventional and three unconventional collocations, so that they would not stand out from the rest of the text. An equal number of both was selected in order for them to be comparable.

Once the literary text and its collocations were selected, the next step was to find another text of another genre that would contain the selected collocations. Keeping in mind the fact that literary translation and specialised translation are more often than not contrasted as being on the other sides of the spectrum that the art of translation is considered to be (Newmark 2004), a result of different genre conventions, among others, a choice was made that the second text would be a popular scientific one. The genre of (popular) scientific texts was selected also because more and more such texts in English are written by non-native speakers, who might lack the linguistic ability, as well as knowledge of genre conventions in English (see e.g. Marina & Snuviškiene 2005), resulting in unconventional linguistic patterns in those texts.

It was clear from the start that it was not probable, or even possible, to find a text that would contain the collocations selected in the literary text. As a result, the second text was rewritten in order to incorporate the collocations from the novel; however, different relevant sources were consulted for the text to be factually correct. Furthermore, since the selected collocations were thematically linked, they could be connected into a meaningful whole. Special attention was also paid to the language itself – in other words, specialised terms were used with formal vocabulary clearly and precisely. However, as mentioned above as one of the prerequisites, in addition to conventional collocation, which one might expect to find in popular scientific writing, unconventional collocations were also included in the text. This type of collocations was blended in as much as possible. Similarly, it is important to note here that the literary text was also slightly modified for practical reasons, i.e. length and clarity, and in order to add one other collocation to be used in the study.

6.2. Determining the (un)conventionality of source collocations

As described above, three conventional and three unconventional collocations had to be selected for this study. Since conventionality may be viewed as a spectrum, it was important to select collocations that would represent the two ends of the spectrum. In order to determine this, a corpus-based approach was used.[10] Following the general criteria for collocations outlined in the section on *Key concepts*, a measure for determining SL (un)conventionality was established

---

[10] I would like to extend my gratitude to prof. Stanojević for his invaluable guidance and help with corpus tools, as well as statistical analyses, which helped me a great deal in understanding new concepts and using them in this thesis.

with the help of *enTenTen21*, a large monolingual English corpus containing English texts from the Internet (Sketch Engine, enTenTen – *English corpus from the web* n.d.). This corpus, which was used as an SL reference corpus, is especially useful since it is lemmatized and tagged and accessing it through *Sketch Engine[11]*, a corpus management tool, also provides additional metrics of researching language patterns, which are discussed in more detail with regard to this study below.

6.2.1. Conventional (C) English collocations

The metrics provided by *Sketch Engine* deemed most appropriate for determining conventional English collocations were:

a. the absolute frequency value
b. the logDice value.

Absolute frequency, i.e. the number of total occurrences of a collocation, was taken as a criterion according to which a potential conventional collocation occurs more than five times in a given corpus. As Evert (2007: 43) puts it, in order to identify conventional collocations, it is recommended "to apply a frequency threshold of $f \geq 5$ in order to weed out potentially spurious [unconventional] collocations". This was checked by simply looking up potential collocations in the *Concordance* option in *Sketch Engine*. However, the absolute frequency values alone are not enough to determine conventionality since they do not provide sufficient evidence for the attraction of two words, instead, a quantitative association measure is needed to further determine this (Evert 2007: 4-5).

There are several options when it comes to statistical association measures, such as t-score, MI score and logDice (for an overview on each, see Gablasova et al. 2017). In this research, the association score selected was logDice. Precisely this association score was chosen since it by default calculates the score for collocations for which a statistical significance exists. In other words, since the logDice score has a maximum value of 14, collocations that have a positive logDice value are statistically significant collocations, while collocations with a negative logDice value have no statistical significance (Rychlý 2008: 9). This means that the higher the score, the greater the statistical association between two words (Gablasova et al. 2017: 164). Another reason for choosing logDice as the association measure in this research is, as already stated, the fact that logDice has a maximum value and as such does not depend on corpus size,

---

[11] *Sketch Engine* is unfortunately not a free tool; however, students and faculty at the Faculty of Humanities and Social Sciences can access it through their institutional login.

which means that it is comparable across different corpora (Rychlý 2008: 9). Another important thing about logDice that has to be pointed out is essentially how it works. The score does not take into account only the (relative) frequency of a collocation but also the frequencies of each item on their own in a given corpus (Rychlý 2008: 9). That means that a collocation that speakers might consider quite common and conventional does not necessarily have a high logDice score if one or both of its components are items that appear frequently in a corpus. As a result, the absolute frequency of conventional collocations does not need to be extremely high.

Furthermore, logDice is an accessible measure through *Sketch Engine*. Once a corpus is selected, one only has to type in the required word in the *Word Sketch* search option. The word is automatically categorised in a word class and the absolute frequency and logDice are shown, when the right view options are enabled, next to a word on a list of words that frequently occur with the word that was looked up. However, the scores provided by *Sketch Engine* are not ideal. *Sketch Engine* only takes one word order into account at a time. In other words, this means that two logDice scores will be calculated for a collocation consisting of e.g. an adjective and a noun, one for when the adjective is in the predicative position and one for when the adjective is in the attributive position. This in turn means that logDice itself is insufficient without the absolute frequency.

To sum up, even though the absolute frequency and logDice values may differ depending on what is in focus, this does not have much bearing on the study at hand since a precise number for this measure for conventional collocations is not needed. Instead, a relatively good logDice value accompanied with an absolute frequency of no lower than five co-occurrences for the syntactical relation that words in a certain collocation have is a good enough measure for determining conventionality of an English collocation[12].

6.2.2. Unconventional (UC) English collocations

The process of determining unconventional English collocations was simpler compared to determining conventional collocations. What was needed here was simply a low absolute frequency. Hence, the bar was set at a maximum of five co-occurrences in the corpus since, as already mentioned, co-occurrences in a given corpus that occur five times or less can be

---

[12] It is worth noting here that even though the absolute frequency threshold was set at five co-occurrences, the absolute frequencies of the collocations under study were not so low. The absolute frequency serves as a means to eliminate collocations that may be unconventional and logDice then serves as a means to determine which collocations out of those potentially conventional, i.e. with an absolute frequency higher than five, are indeed conventional.

classified as unconventional collocations (Evert 2008: 43). To check this, *Sketch Engine's* option *Concordance* was used. However, it was not enough to just type in the potential collocation. Rather, in order to account for the fact that there is no set word order for some types of collocations (e.g. attributive and predicative positions for adjectives in adjective-noun collocations) and to account for intervening words, the Corpus Query Language (CQL) was utilised[13]. A specialised formula was created that took into account a maximum of four intervening words (according to Sinclair 1991: 170) between a collocate and a node within a sentence:

[lemma=""] []{0,4} [lemma=""] within <s/>

This formula was used for both cases where two word orders were possible – the collocate-node word order was simply reversed to the node-collocate word order. Subsequently, a search using the formula above yielded ambiguous results in some cases. These had to be then manually filtered according to the criteria selected. However, in most cases the formula identified true collocations, so not much effort was needed to further weed out false positives.

6.3. Determining the (un)conventionality of target collocations

Similarly to determining the (un)conventionality of source collocations, a corpus-based approach was also needed for target collocations' conventionality. In this case the reference corpus used was the Croatian *MaCoCu* corpus. *MaCoCu Croatian Web v2 (2021–2022)* is the largest currently available corpus of Croatian, which contains texts crawled from the Internet (Sketch Engine, *MaCoCu Corpora from the web* n.d.)[14]. However, a slightly different approach was needed for the analysis of target collocations. Since it cannot be expected that the translations of collocations are such that they can be packed neatly into the polar opposite categories of conventional and unconventional, as outlined above for the selected SL collocations, additional categories were needed for translations that found themselves somewhere between those two ends.

For the sake of consistency and comparison, the criteria for determining if something is conventional/unconventional in the TL remained mostly the same as it was for determining (un)conventionality in the SL. The main measure for the category of unconventional was

---

[13] The CQL is another option offered in Sketch Engine and it "is a special code or query language used […] to search for complex grammatical or lexical patterns or to use search criteria which cannot be set using the standard user interface" (Sketch Engine, *CQL – Corpus Query Language* n.d.).

[14] It should be noted that it is possible the corpus contains translations since it contains texts from the Internet; however, the majority of data is expected to reflect non-translated Croatian.

absolute frequency lower than five ($f \geq 5$), while for the category of conventional it was the logDice value, which may or may not be accompanied by a high absolute frequency. However, since logDice is calculated for word combinations consisting of two words only, additional help was needed in identifying the conventionality of multiword expressions. Lexicographic resources came in handy here since they enumerate typical patterns of language usage. In this sense, the *Croatian Collocation Database* (*Kolokacijska baza hrvatskog jezika*) and the *Croatian Language Portal* (*Hrvatski jezični portal*) were used in combination with corpus data.

Furthermore, two additional categories were needed to classify the examples that did not fit into the two mentioned categories, and those are the categories of *potentially unconventional* (p-UC) and *potentially conventional* (p-C). The renderings that were classified as potentially unconventional were those with the absolute frequency of above or equal to six on the one hand, and the absolute frequency of below or equal to 20 ($6 \leq f \leq 20$) on the other. The renderings that were classified as potentially conventional were those with the absolute frequency higher or equal to 16 ($f \geq 16$) and for which there were no further indications that they might be highly conventional, such as descriptions in lexicographic resources or a logDice higher or equal to five (logDice ≥ 5)[15]. The described criteria are summarised in Table 1.

Table 1: Criteria for determining TL unconventionality

	UC	p-UC	p-C	C
absolute frequency	≤ 5	$\geq 6, \leq 20$	≥ 21	–
logDice + lexicographic resources	–	–	< 5 when applicable	≥ 5 when applicable

Albeit briefly, some additional procedures that were used in this stage should be mentioned. First, complete or partial (only the node retained) omission of the collocation was regarded as a conventional rendering since the unconventional part is lost. Second, with multiword expressions, the structure was also partly assessed in the corpus, i.e. determining if it is common that a multiword expression is followed by a noun phrase or a clause. Thirdly, if a collocation was translated as a multiword expression comprised of two collocations, each collocation was separately researched against the corpus, as well as a whole, which served as the basis of their classification into one of the categories (see Appendix C for examples).

---

[15] Along with the absolute frequency boundaries for these two categories, this is also, of course, an arbitrary boundary since logDice is a value for *conventional* collocations. However, a clear boundary was needed to distinguish between more and less conventional collocations and setting it at five was felt to be a sound number taking into account the logDice values of the conventional SL collocations under study.

6.4. Participants

The number of participants that took part in the study was 15. Participants were aged 22 to 28. Three of them identified as male, while the other 12 identified as female. This sample is a part of a population of beginner translators from English into Croatian who are just starting their professional careers with a formal education in both English and translation, as well as some experience in translation. Specifically, this means that the participants were either graduate students of English and translation or translators who have recently completed their studies and do not have much experience. It is also important to mention that this is a small sample; however, due to the design of the study, no more participants could have been included.

6.5. Post-experiment questionnaire

After the experiment, the participants were asked to assess the conventionality of each SL collocation under investigation. The collocations were randomly listed, and the participants rated the conventionality of each on a 1-5 Likert-type scale, where one meant completely unconventional and five completely conventional.

6.6. Analysis

All renderings of the SL collocations under examination were first analysed together. As outlined above, each rendering was classified on a conventionality scale into one of the four categories and each category was assigned a number from 1 to 4 (UC – 1, p-UC – 2, p-C – 3, C – 4) on an ordinal scale in order for the statistical analysis to be possible. This data, along with the data from the post-experiment questionnaire concerning participants' perceptions of SL collocations' conventionality, underwent statistical analyses and correlation tests in JASP (version 0.18.3.0), which is an open-source program for statistical analysis. Finally, the data was interpreted with respect to the aim of the study and previous research.

7. Results

7.1. SL collocations under examination

Table 2 presents the six SL collocations that were selected for this study, along with their corpus values.

Table 2: SL collocations under study

Type	Collocation	Absolute frequency	logDice
C	rich land	7048	5.6
	notoriously poor	2037	6.8
	knobby knee	199	6.5
UC	shining cygnet	0	–
	offer patrimony	0	–
	severe beard	2	–

The absolute frequencies of conventional collocations varied from collocation to collocation; however, they were all well over the cut-off point of five. The logDice values were not extremely high but high enough to signify that a collocation was very conventional. On the other hand, for unconventional collocations, only *severe beard* was found in the corpus; the two other unconventional collocations had no occurrences in the corpus. It is also important to note that Table 2 presents the data for only one word order, the one indicated in the table, since this was sufficient proof of conventionality where two word orders were possible.

7.2. Collocations in translation

7.2.1. Descriptive results

The collected data from the participants' translation was first analysed together with regard to the four conventionality categories. As mentioned above, 15 participants took part in the study. This means that there were 30 renderings per collocation, 15 from the literary text and 15 from the popular scientific text, resulting in 180 renderings in total, 90 per SL collocation type. Table 3 presents the renderings classified into those categories.

Table 3: Renderings of SL collocations with respect to the categories of conventionality

SL collocations	UC lit.	UC pop. scient.	p-UC lit.	p-UC pop. scient.	p-C lit.	p-C pop. scient.	C lit.	C pop. scient.
rich land	0		0		11		19	
	0	0	0	0	5	6	10	9
notoriously poor	0		6		0		24	
	0	0	5	1	0	0	10	14
knobby knee	0		3		1		26	
	0	0	0	3	0	1	15	11
total for C	0		9		12		69	
shining cygnet	26		0		2		2	
	15	11	0	0	0	2	0	2
offer patrimony	9		2		3		16	
	8	1	2	0	1	2	4	12
severe beard	4		0		12		14	
	3	1	0	0	10	2	2	12
total for UC	39		2		17		32	
total	39		11		29		101	

(Rows *rich land*, *notoriously poor*, *knobby knee* grouped under **conventional**; rows *shining cygnet*, *offer patrimony*, *severe beard* grouped under **unconventional**.)

Out of all the renderings, 39 (22%) were unconventional, 11 (6%) possibly unconventional, 29 (16%) possibly conventional and 101 (56%) conventional, meaning that more than a half of the renderings were conventional regardless of the collocation type. However, if we look at the data separately for each collocation type, the picture changes slightly. The renderings of conventional SL collocations were as follows: none of them unconventional, nine (10%) were possibly unconventional, 12 (13%) possibly conventional and 69 (77%) conventional. This means that the conventional SL collocations were in most cases translated into the TL as equally conventional as they were in the SL with less than one fourth of example not matching the conventionality of the SL collocations. When it comes to unconventional SL collocations, 39 (43%) of them were unconventional, two (2%) possibly unconventional, 17 (19%) possibly conventional and 32 (36%) conventional. This tells us that more than a half of the renderings were more conventional than their sources and more than one third were translated as fully conventional in the TL.

7.2.2. Literary text

7.2.2.1. Renderings of conventional SL collocations

Concerning the literary text, the first variable that was analysed was the category a translation of a conventional SL collocation belonged to according to the classification outlined above (UC, p-UC, p-C, C), i.e. its frequency. The results are presented in Figure 1.
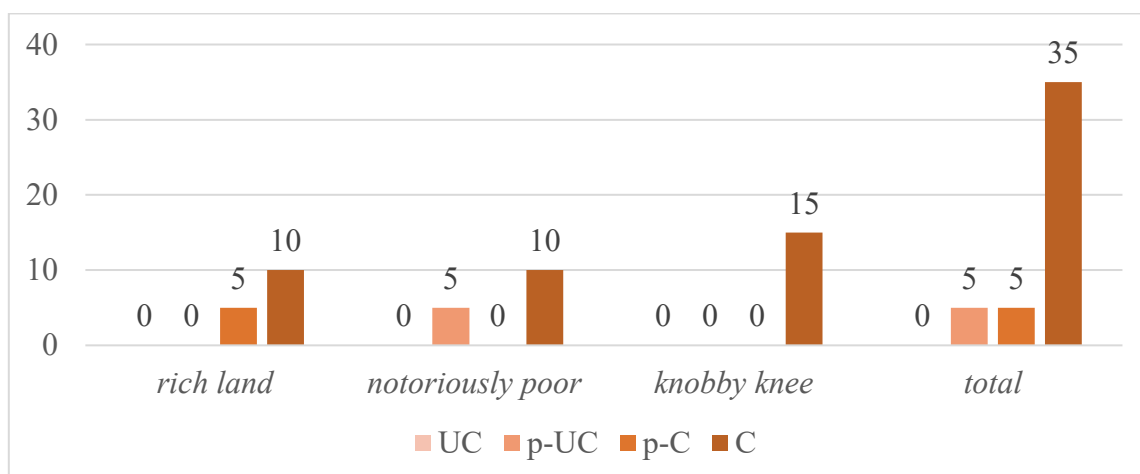
Figure 1: Absolute frequencies of renderings of conventional SL collocations in the literary text

As seen in the figure, conventional SL collocations were translated most commonly as equally conventional in the TL. More precisely, no conventional SL collocations were translated unconventionally, while two collocations (*rich land* and *notoriously poor*) were translated, in five instances each, less conventionally than their SL equivalents. One collocation (*knobby knee*) was translated in all instances as conventional as it was in the SL. This figure also shows that the frequency distribution was left-skewed, meaning that the frequency distribution was more in favour of the conventional end.

7.2.2.2. Renderings of unconventional SL collocations

When it comes to unconventional SL collocations, the absolute frequency values of a particular type of rendering are available in Figure 2.
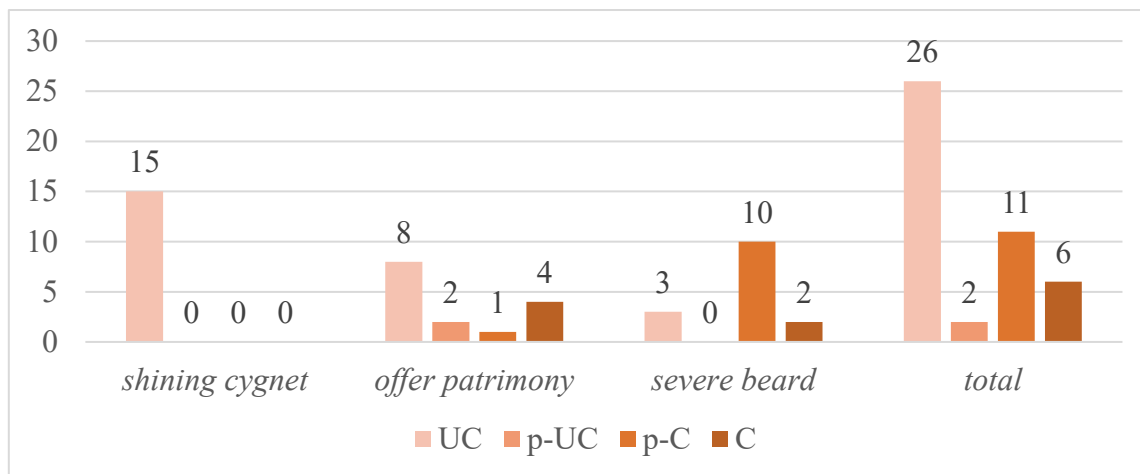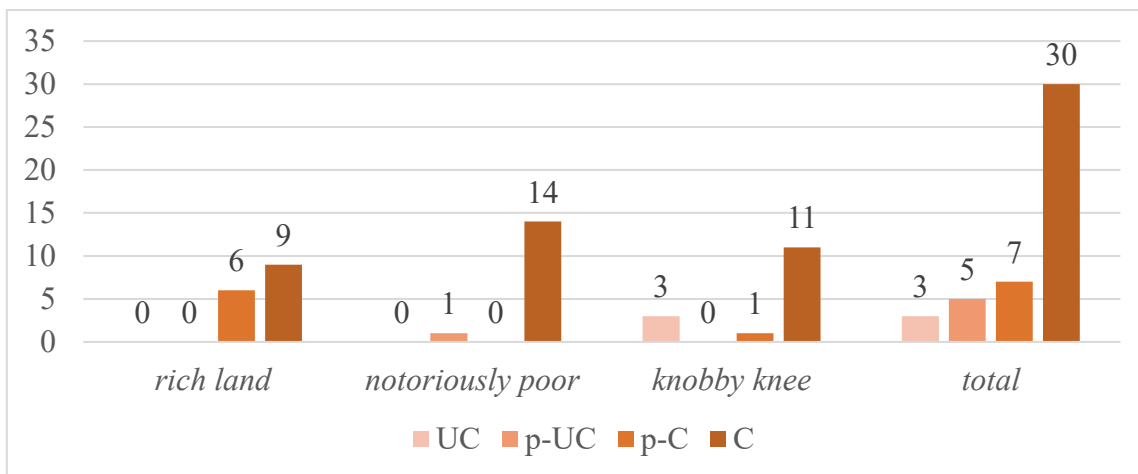
Figure 2: Absolute frequencies of renderings of unconventional SL collocation in the literary text

Overall, the unconventional SL collocations were most commonly translated with unconventional equivalents. However, there are some differences among the three collocations. The unconventional SL collocation whose renderings showed the least variety and was translated entirely unconventionally was *shining cygnet*, and the one that showed the most variety was *offer patrimony*, but still leaning towards the unconventional side with eight renderings being unconventional, two possibly unconventional, one possibly conventional and four conventional. On the other hand, renderings of s*evere beard* leaned to the opposite end, towards conventionality, with 10 renderings possibly conventional and two conventional, and only three unconventional. In this case, this means that 80% of the renderings were more conventional in the TL than their counterparts in the SL were, while overall 58% of the renderings were equally as unconventional in the TL as they were in the SL. Furthermore, the frequency distribution was right-skewed in the first two instances and overall, while it was left-skewed for *severe beard*.

7.2.3. Popular scientific text

7.2.3.1. Renderings of conventional SL collocations

The absolute frequencies of different types of renderings of conventional SL collocations in the popular scientific text is presented in Figure 3.

Figure 3: Absolute frequencies of renderings of conventional SL collocations in the popular scientific text

Similarly as in the literary text, the conventional SL collocations were predominantly translated with conventional equivalents. However, the three collocations showed different behaviour in translation, depending on the conventionality category. The frequency distribution was left-skewed across all three collocations and overall, i.e. in favour of the conventional end. However, four renderings were more unconventional than the others, one rendering of *notoriously poor* and three of *knobby knee*, with the former being possibly conventional and the latter unconventional.

7.2.3.2. Renderings of unconventional SL collocations

The absolute frequencies concerning the renderings of unconventional SL collocations with regard to the popular scientific text are presented below, in Figure 4.
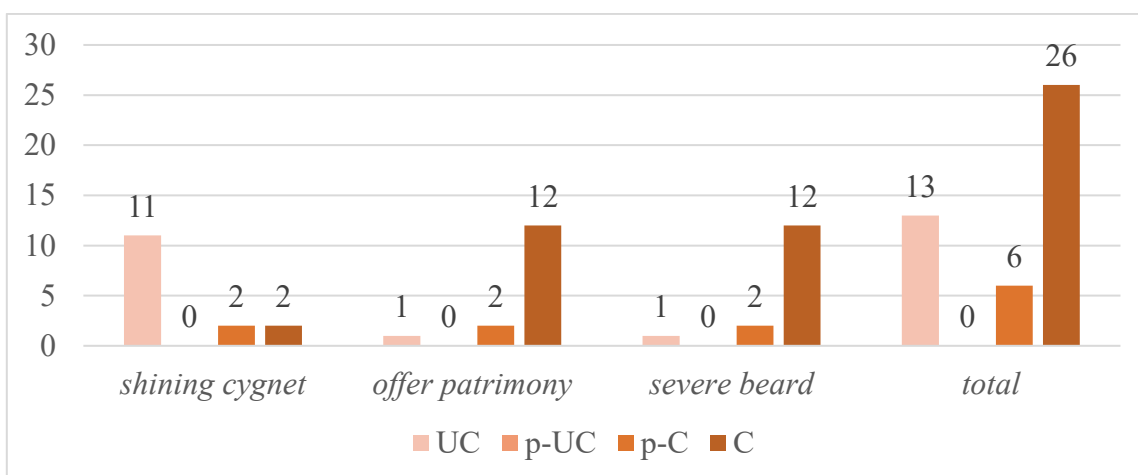

Figure 4: Absolute frequencies of renderings of unconventional SL collocations in the popular scientific text

As shown above, results regarding *shining cygnet* were the only ones right-skewed; 11 renderings were unconventional with only four more conventional than the source. On the other hand, the renderings of the other two collocations and the renderings overall in the sample were left-skewed, favouring the conventional end. Moreover, the results concerning the latter two collocations were fairly homogenous with 12 conventional renderings each, two possibly conventional and only one unconventional.

7.2.4. The difference between conventional and unconventional renderings with respect to conventionality of SL collocations

A chi-square test of independence was performed to examine the relation between the conventionality of SL collocations and the conventionality of their renderings. The above-described conventionality categories of renderings were grouped into two categories, unconventional (UC, p-UC) and conventional (p-C, C). A chi-square test of independence showed that both conventional and unconventional SL collocations were more likely to be translated conventionally than unconventionally in the TL, X^2 (1, N = 180) = 26.61, p < .0001. Furthermore, the effect size as measured by Cramer's V indicated a medium effect.

7.2.5. The difference between renderings of conventional and unconventional SL collocations with respect to genre

A chi-square test of independence was used again to examine the relation between the conventionality of renderings and genre, with renderings grouped into conventional and unconventional ones as above. A chi-square test of independence showed no significant association between the renderings of conventional SL collocations and the genre of the text, X^2 (1, N = 180) = 0, p < .1. Furthermore, concerning the association between renderings of unconventional SL collocations and the genre of the text, a chi-square test of independence showed that unconventional SL collocations were more likely to be translated unconventionally in the literary texts than in the popular scientific ones, i.e. that they were more likely to be translated conventionally in the popular scientific texts, X^2 (1, N = 180) = 8.78 , p < .003. Cramer's V, i.e. the effect size, indicated a medium effect.

7.3. Participants' perception of the SL collocations' conventionality

7.3.1. Descriptive results

As mentioned above, after the translation of each text, the participants were asked to evaluate the conventionality of both conventional and unconventional SL collocations on a Likert-type

scale from 1 to 5 (an ordinal scale), where 1 meant completely unconventional and 5 completely conventional. Their evaluations are presented in Table 4, separately for each collocation and text.

Table 4: Participants' evaluation of the SL collocations' conventionality

SL collocations	1 lit.	1 pop. scient.	2 lit.	2 pop. scient.	3 lit.	3 pop. scient.	4 lit.	4 pop. scient.	5 lit.	5 pop. scient.
rich land	0		0		1		5		24	
	0	0	0	0	0	1	2	3	13	11
notoriously poor	0		7		8		14		1	
	0	0	4	3	3	5	8	6	0	1
knobby knee	0		3		4		12		11	
	0	0	1	2	3	1	5	7	6	5
total for C	0		10		13		31		36	
shining cygnet	16		13		1		0		0	
	9	7	5	8	1	0	0	0	0	0
offer patrimony	4		13		8		5		0	
	3	1	7	6	3	5	2	3	0	0
severe beard	12		13		2		1		2	
	6	6	7	6	1	1	0	1	1	1
total for UC	32		39		11		6		2	

(Left margin labels: rows 1–3 grouped as "conventional"; rows 4–6 grouped as "unconventional".)

The participants mostly evaluated the SL collocations as attested in the SL reference corpus, i.e. conventional ones as conventional and unconventional ones as unconventional. As seen in Table 4, in 36 (40%) cases, the conventional collocations were evaluated with 5, meaning that the participants perceived them as completely conventional. Out of the rest, 31 (35%) times the collocations were evaluated with 4, 13 (14%) with 3, 10 (11%) with 2 and none were evaluated with 1, i.e. as completely unconventional. Among the collocations, *rich land* was most times evaluated with 5 with the evaluations ranging only from 3 to 5, while *notoriously poor* and *knobby knee* were in most cases evaluated with 4 with the evaluations ranging from 2 to 5; however, the former collocation was evaluated with 5 only once and the latter 11 times.

When it comes to unconventional SL collocations, the most frequently chosen number on the scale was 2; in 39 (44%) cases the participants evaluated the unconventional collocations not as completely conventional but still conventional. Out of the rest, 32 (36%) times the collocations were evaluated with 1, 11 (12%) times with 3, 6 times (7%) with 4 and 2 times (2%) with 5. Among the unconventional collocations, the evaluations for *shining cygnet* were the most homogenous; in almost all cases it was evaluated with either 1 or 2, and only once

with 3. On the other hand, *offer patrimony* and *severe beard* were most commonly evaluated with 2; however, the evaluations for the former collocation ranged from 1 to 4 while the evaluations for the latter one ranged from 1 to 5 (see Table 4).

7.3.2. Correlation between the participants' evaluation of a collocation's conventionality and their way of rendering it

Due to the skewness of the results, as presented above, Spearman's rank correlation coefficient, or Spearman's rho (ρ) was used to calculate the correlation between the conventionality category of the renderings and the participants' evaluation of the SL. The correlation test showed the following results: N = 180, ρ = 0.394, p < .001. More precisely, the results indicated a weak positive correlation on the borderline, showing a tendency that the level of a collocations' conventionalisation in translation was greater when a participant evaluated an SL collocation with a greater number on the Likert-type scale and vice-versa. Furthermore, the effect size, as measured by Spearman's rho, showed a small effect.

8. Discussion

8.1. Conventional and unconventional SL collocations in translation

The main aim of this study was to determine what happens to collocations in translation, more precisely to see if there is any evidence of conventionalisation. Subsequently, the first hypothesis proposed was that both conventional and unconventional SL collocations become conventional in the TL through translation. The evidence presented in the previous section supports the first hypothesis, but it is clear that this is not absolute. More precisely, the chi-square test of independence showed that conventional renderings of both conventional and unconventional SL collocations are more likely as opposed to unconventional renderings. However, if we take a look at the absolute frequencies of rendering types, it becomes clear that the renderings of conventional SL collocations were indeed conventional in the target texts, while the renderings of unconventional SL collocations were not as conventional but showed a move towards conventionality in the TL. Moreover, this also partly supports Kenny's (2001) findings – unconventional collocations are not necessarily conventionalised in the target literary texts, while on the other hand provides evidence of conventionalisation as reported by Kegalj (2023).

A closer look into how each of the unconventional SL collocations was translated might also shed some light on the results. The data in Figures 2 and 4 show that one unconventional SL collocation significantly stood out from the other collocations, and that is *shining cygnet*. This

collocation showed little to no variety in the target texts. Apart from the variables and their possible influence, as discussed below, this could be a result of a lack of a (more) conventional alternative. More precisely, the most common equivalent of *cygnet* in Croatian is *labudić*, which is not an uncommon word in itself, but it is uncommon in combination with other words. The *MaCoCu* corpus reveals very few collocates with very low frequencies, e.g. *bespomoćan* ("helpless")[16]. As a possible result, translators presumably opted for the simplest solution – a literal translation since there was no conventional one. Furthermore, one possible way this collocation could have been conventionalised, as exemplified by the data (see Appendix C), was to simply omit the whole collocation or only keep the node. Presumably, this and other ways of rendering the collocation more conventionally were too extreme for the participants since they deviate from the source significantly.

8.2. The level of a TL collocation's conventionality and genre of the text

The second hypothesis and its subhypotheses were concerned with the interplay of conventionality and genre, i.e. it was hypothesized that there would be more unconventional collocations in translation of the literary text since it can be described as a creative and more free genre as opposed to a popular scientific text, which is usually considered stricter in style. For that reason, the hypothesis expected that the presence of conventionalisation would be higher in renderings of the popular scientific text and lower in the literary one, which was found to be true.

The chi-square test of independence showed no notable difference between the two genres for conventional SL collocations, i.e. conventional SL collocations were translated as equally (un)conventional in both texts. On the other hand, for unconventional SL collocations, the chi-square test of independence showed they were more likely to be translated conventionally in the popular scientific text, while their renderings were more likely to be unconventional in the literary text, which means that the second hypothesis and its subhypotheses were confirmed. This is also in line with Kegalj's (2023) findings on conventionalisation in Croatian translations, even though the subsets of translations are different.

More precisely, if we take a look at Figure 1 and 3, it is clear that the conventional SL collocations were translated pretty similarly in both texts. The only notable exception is the absolute frequency value of *notoriously poor* in the literary text. Albeit statistically irrelevant, it is interesting that this is the only conventional SL collocation with low conventionality values

---

[16] The *Croatian Collocation Database* does not have any entries for *labudić*.

in both. The reason behind this might be interference, i.e. the collocation in those five cases is translated literally as *notorno loš*.

Similarly, also statistically not relevant but interesting to note is that *knobby knee* was translated in three instances completely unconventionally in the popular scientific text, while all of the instances in the translation of the literary text were the complete opposite, wholly conventional. The reason for this may be that translators in two out of those three cases felt that it is not common to talk about swans having knees, so they replaced the node with some other body part, presumably more typical of swans, e.g. *zglob* ("joint"). This could also be considered evidence of conventionalisation but on another level, not the level of collocations since this led to the collocation being unconventional.

8.3. The level of a TL collocation's conventionality and translators' perception of SL collocations

The third and final hypothesis expected that a translator's perception of an SL collocation might affect their choice of a translation equivalent. In other words, if a translator perceived an SL collocation as (un)conventional, they would be more likely to render it in the same way since this could be considered a deliberate choice by the author of the source text, especially when it comes to unconventional collocations, or in this context, better said *creative* collocations.

As shown in Table 4, the participants generally evaluated the collocations as attested in the SL reference corpus, as conventional or unconventional. Their evaluations showed a tendency, albeit small, that they were more likely to translate a collocation more (un)conventionally if they perceived it to be more (un)conventional in the SL.

It is also important to note here that, of course, correlation does not signify a causal relationship. In that sense, it may be that the results presented above for one variable are in fact a result of another variable. It may be that a collocation was translated more conventionally solely on the basis of genre conventions, and not on the basis of a participant's perception, or it even may be that both of those variables affected the translation. It should also be noted that a translator's exposure to a certain collocation while translating the first text could have affected their approach to the translation of the second text. This is also why it was useful that translators evaluate the SL collocations both times; however, the two datasets showed no notable difference.

8.4. Limitations and relevance

The main limitation of this study is its scope. This was a very small-scale study which focused solely on six collocations and their translations by 15 beginner translators. The same participants translated both texts, which is good since the data is more comparable; however, this also means that the first translation could have affected the second one, leading to skewed results. Such intervening variables, along with, e.g. a participant's lack of attention, could have also played a role. Furthermore, due to the skewness of the results, a non-parametric correlation test was used, which usually requires larger sample sizes for the results to be precise (Eddington 2015: 37). Another limitation is the corpus-based approach itself insofar as the corpus resources for Croatian are not very extensive and developed, and neither are other linguistic resources. However, *MaCoCu*, as the largest and most recent corpus of Croatian available, was used in combination with other lexicographic resources in an attempt to get the most representative data. This is all to say that the findings of this study should not be taken at face value, but should rather be interpreted as preliminary since this topic is under-researched in TS in general and especially on the English-Croatian language pair.

On the other hand, precisely because of the fact that not many studies have tackled this and only one study concerned with the language pair in question scratched its surface, the main value of this study is the new findings, which, however small-scale, do tell us something about Croatian translations from English, but also about the process of translation itself. In addition, this study hopes to contribute to the revival of the topic of translation universals, which reached its peak some twenty years ago, but still remains to this day not researched or defined enough, especially with regard to new technologies that might prove useful in its further development as a method of researching translations.

9. Conclusion

The search of regularities of translation, albeit perhaps no longer popular in the form of translation universals, has been central to descriptive translation studies since its emergence. Translation universals, not as absolute patterns, but rather as probabilistic claims that may describe particular subsets of translation and only then be generalised on a slightly higher non-universal level, are empirically very useful in combination with modern corpus tools that represent a language better than even before.

One such universal – conventionalisation – was in the focus of this study. Conventionalisation, as a tendency for translations to be in a way lexically impoverished since translators mostly prefer common TL items over creative and innovative ways of rendering uncommon SL items, was here studied on collocations as building blocks of the lexical level of language. Collocations, as understood by this study in its Firthian sense, have indeed shown a degree of conventionalisation through translation. Unconventional SL collocations were more likely to be translated conventionally in the TL. However, this was not found to be absolute, which is a possible result of TL language's restrictions, whose repertoire may hinder conventionalisation, as remarked above. Furthermore, as proposed by the second hypothesis, genre did prove to play a role, with the presence of conventionalised elements being higher in the translations of the popular scientific text – a genre that values precision of expression and does not leave as much room for creativity as the literary genre does. This is also in line with previous claims that a TL's genre may influence the translation process. In addition, the translation process may also be influenced by a translator's perception of a collocation. It was found that participants' perception of an SL collocation's conventionality correlated with the conventionality of translation solution they picked, meaning that SL conventionality is also one factor that translators consider during the translation process.

This study builds onto the previous research discussed above. Evidence was predominately found in both of conventionalisation of collocations, as also found by Bernardini (2007) and Kegalj (2023). However, some evidence against conventionalisation was also found, partly supporting Kenny's (2001) findings. However, it should be noted that, even though the findings here are closely related to the mentioned studies, no clear-cut link may be made to, e.g., Bernardini's (2007) research since she researches conventionalisation with regard to comparable target texts, and the study at hand takes a different approach by researching conventionalisation in connection to the source and target languages. Similarly, Mauranen's (2000) research may also be associated with the present study since translations have shown

peculiar combinatorial preferences; however, this is also only an indication that would require further research.

In conclusion, it is important to emphasise that this study was quite small in scope, so no further generalities can be made, as noted above. However, the evidence has pointed to the fact that conventionalisation may be found in Croatian translations of English collocations – its existence cannot be disputed but its reach still remains unclear. This means that there are still avenues of research worth exploring, such as larger-scale studies on whole texts, not just selected collocations, combined with a detailed comparison with non-translated texts, or research on collocations on different subsets of translations, especially on the English-Croatian language pair, both of which remain under-researched in translation studies. Only then can broader claims be made about the nature of Croatian translations from English in connection to translation universals and the company a translated word keeps made known.

References

Baker, M. (1993). Corpus Linguistics and Translation Studies - Implications and Applications. In Baker, M. Gilli, F. & Tognini-Bonelli, E. (eds.). *Text and Technology: In Honour of John Sinclair*. John Benjamins, pp. 233–250.

Baker, M. (1996). Corpus-based translation studies: the challenges that lie ahead. In Harold Somers, H. (ed.) *Terminology*, *LSP and Translation: Studies in Language Engineering in Honour of Juan C. Sager.* John Benjamins, pp. 175–186.

Barnbrook, G., Manson, O. & Krishnamurthy, R. (2013). *Collocation: Applications and Implications*. Palgrave Macmillan.

Bernardini, S. (2007). Collocations in Translated Language: Combining Parallel, Comparable and Reference Corpora. In Davies, M., Rayson, P., Hunston, S. & Danielsson, P. (eds.). *Proceedings of the Corpus Linguistics Conference (CL2007)*. University of Birmingham, pp. 1–16.

Chesterman, A. (2004). Beyond the particular. In Mauranen, A. & Kujamäki, P. (eds.). *Translation Universals: Do They Exist?* John Benjamins, pp. 33–49.

Eskola, S. (2004). Untypical frequencies in translated language: A corpus-based study on a literary corpus of translated and non-translated Finnish. In Mauranen, A. & Kujamäki, P. (eds.). *Translation Universals: Do They Exist?* John Benjamins, pp. 83–99.

Eddington, D. (2015). *Statistics for Linguists: A Step-by-Step Guide for Novices.* Cambridge Scholars Publishing.

Evert, S. (2007). 58. Corpora and collocations. Extended manuscript. https://lexically.net/downloads/corpus_linguistics/Evert2008.pdf [Downloaded June 30 2023].

Firth, J. R. (1957). A synopsis of linguistic theory 1930–1955. In *Studies in Linguistic Analysis*. Oxford. Reprinted 1968 in F. R. Palmer (ed.). *Selected Papers of J. R. Firth*. Longman, pp. 168–205.

Gablasova, D., Brezina, V. & McEnery, T. (2017). Collocations in Corpus-Based Language Learning Research: Identifying, Comparing, and Interpreting the Evidence. *Language Learning,* 67(1): pp. 155–179.

García Izquierdo, I. & Montalt i Resurrecció, V. (2002). Translating into Textual Genres. *Linguistica Antverpiensia,* 1: pp. 135–143. https://doi.org/10.52034/lanstts.v1i.12

Hanks, P. (2013). *Lexical Analysis: Norms and Exploitations*. MIT Press.

Hrvatski jezični portal. https://hjp.znanje.hr/

Kegalj, J. (2023). *Žanrovska analiza pomorskopravnih tekstova i ostvarenje prijevodnih univerzalija u njihovim prijevodima s engleskoga jezika.* Doctoral thesis, University of Zagreb. https://doi.org/10.17234/diss.2023.8936

Kenny, D. (1998). Creatures of Habit? What Translators Usually Do with Words. *Meta*, 43(4): pp. 515–523. https://doi.org/10.7202/003302ar

Kenny, D. (2001). *Lexis and Creativity in Translation: A Corpus Based Approach*. Routledge.

Kolokacijska baza hrvatskog jezika. http://ihjj.hr/kolokacije/

Marina, V., & Snuviškiene, G. (2005). Error analysis of scientific papers written by non-native speakers of English. *Transport*, 20(6): pp. 274–279. https://doi.org/10.3846/16484142.2005.9638031

Mauranen, A. (2000). Strange Strings in Translated Language: A Study on Corpora. In Olohan, M. (ed.). *Intercultural Faultlines. Research Models in Translation Studies 1: Textual and Cognitive Aspects.* St. Jerome Publishing, pp. 119–141.

Mauranen, A. (2007). Universal Tendencies in Translation. In Anderman, G. & Rogers, M. (eds.). *Incorporating Corpora. The Linguist and the Translator.* Multilingual Matters, pp. 32–48.

Miller, M. (2011). *The song of Achilles.* Ecco Press.

Molnár, O. (2019). Genre Analysis in the Translation Studies Paradigm Today. *CLINA,* 5(2): pp. 37–62. https://doi.org/10.14201/clina2019523762

Newmark, P. (2004). Non-literary in the Light of Literary Translation. *The Journal of Specialised Translation* 1: pp. 8–13. https://jostrans.soap2.ch/issue01/art_newmark.php (Accessed 28 June 2024).

Rychlý, P. (2008). A lexicographer-friendly association score. *Proceedings of Recent Advances in Slavonic Natural Language Processing, RASLAN*, pp. 6–9.

Schäffner, C. (2000). The role of genre for translation. In Trosborg, A. (ed.), *Analysing Professional Genres.* John Benjamins, pp. 210–224.

Scottish Wildlife Trust. (2020). *Everything you need to know about mute swans*. https://scottishwildlifetrust.org.uk/2020/04/mute-swans/ (Accessed 1 May 2024).

Silić, J. (2006). *Funkcionalni stilovi hrvatskog jezika.* Disput.

Sinclair, J. (1991). *Corpus, Concordance, Collocation*. Oxford University Press.

Sketch Engine. https://www.sketchengine.eu/

Sketch Engine. (n.d.). *CQL – Corpus Query Language.* https://www.sketchengine.eu/documentation/corpus-querying/ (Accessed 28 May 2024).

Sketch Engine. (n.d.). *enTenTen – English corpus from the web.* https://www.sketchengine.eu/ententen-english-corpus/ (Accessed 30 May 2024).

Sketch Engine. (n.d.). *MaCoCu Corpora from the web.* https://www.sketchengine.eu/macocu-corpora-from-the-web/ (Accessed 15 June 2024).

Stefanowitsch, A. (2020). *Corpus linguistics: A Guide to Methodology*. Language Science Press.

Toury, G. (2004). Probabilistic explanations in translation studies: Welcome as they are, would they qualify as universals? In Mauranen, A. & Kujamäki, P. (eds.). *Translation Universals: Do They Exist?* John Benjamins, pp. 15–32.

Toury, G. (2012). *Descriptive translation studies - and beyond*. Revised edition. John Benjamins.

Vermeer, H. J. (1989). Skopos and commission in translational action. In Chesterman, A. (ed.). *Readings in Translation Theory*. Finn Lectura, pp. 173–187.

Xia, Y. (2014). *Normalization in Translation: Corpus-based Diachronic Research into Twentieth-century English–Chinese Fictional Translation*. Cambridge Scholars Publishing, pp. 75–95.

Appendix A

Appendix A contains the texts that the participants were asked to translate. The literary text was adapted from the novel *Song of Achilles* (2011) by Madeline Miller and the popular scientific text was written mainly based on the article *Everything you need to know about mute swans* by the Scottish Wildlife Trust (2020).

Literary text (183 words)

I knew the name. Tyndareus was king of Sparta and held huge tracts of the rich southern land, the kind my father coveted. I had heard of his daughter too, rumoured to be the fairest woman in our countries. Her mother, Leda, was said to have been ravished by Zeus, the king of the gods himself, disguised as a swan. Nine months later, her womb yielded two sets of twins: Clytemnestra and Castor, children of her mortal husband; Helen and Polydeuces, the shining cygnets of the god. But gods were notoriously poor parents; it was expected that Tyndareus would offer patrimony to all.

(...)

Finally, a day came in which my father ordered me bathed and brushed. He had me change my tunic, then change again. I obeyed, though I saw no difference between the purple with gold or crimson with gold. Neither did my knobby knees. My father looked powerful and stern with a severe black beard on his face. The gift that we were presenting to Tyndareus stood ready, a beaten-gold mixing bowl embossed with the story of the princess Danae.

Popular scientific text (184 words)

Once their offspring hatch, both parents diligently care for them; even cobs offer patrimony, thereby ensuring their young's safety and teaching them essential skills for survival. Even though swans are not notoriously poor parents, instances arise when the cygnets are reluctant to leave them. In such cases, the parents may turn on them to encourage their leaving, a phenomenon observed as preparatory behaviour for the forthcoming breeding season.

Initially, young swans may present as dusky brown-grey accompanied by a grey-black bill. However, as they mature from shining cygnets to adult swans, their juvenile bills turn red and black. The adult male's bill also exhibits a pronounced swelling of the black knob at its base during the breeding season.

Swan pairs typically remain in the same area for life, occupying the rich lands with their nest site. The abundance of water provides their food, e.g. widgeon grass[1], which they can be seen

eating, giving the impression of having a severe green-brown beard. Proximity to water is also important to swans for swimming, which interestingly might reveal their knobby knees otherwise hidden by their dry plumage.

Appendix B

The literary text with its accompanying questionnaire and the popular scientific text with its accompanying questionnaire as used in the experiment, respectively.

Section A: Prijevod

U nastavku se nalazi književni tekst na engleskom. Zamislite da ste književni prevoditelj te prevedite tekst na hrvatski kao da ga prevodite za objavu. Vremenskog ograničenja nema te se slobodno koristite svim resursima i alatima kojima biste se i inače koristili. Jedino nije dopušteno koristiti se već postojećim prijevodom na hrvatski.

A1. **I knew the name. Tyndareus was king of Sparta and held huge tracts of the rich southern land, the kind my father coveted. I had heard of his daughter too, rumoured to be the fairest woman in our countries. Her mother, Leda, was said to have been ravished by Zeus, the king of the gods himself, disguised as a swan. Nine months later, her womb yielded two sets of twins: Clytemnestra and Castor, children of her mortal husband; Helen and Polydeuces, the shining cygnets of the god. But gods were notoriously poor parents; it was expected that Tyndareus would offer patrimony to all.**

(...)

Finally, a day came in which my father ordered me bathed and brushed. He had me change my tunic, then change again. I obeyed, though I saw no difference between the purple with gold or crimson with gold. Neither did my knobby knees. My father looked powerful and stern with a severe black beard on his face. The gift that we were presenting to Tyndareus stood ready, a beaten-gold mixing bowl embossed with the story of the princess Danae.

Section B: Pitanja o tekstu

U nastavku se nalazi šest engleskih kombinacija riječi iz izvornog teksta. Prema vašem mišljenju, za svaku od njih na skali od 1 do 5 procijenite koliko smatrate da su uobičajene u engleskom.

B1. **Broj 1 označava da kombinacija riječi nije uopće uobičajena, a broj 5 da je vrlo uobičajena.**

	1	2	3	4	5
rich land	☐	☐	☐	☐	☐
shining cygnet	☐	☐	☐	☐	☐
notoriously poor	☐	☐	☐	☐	☐
to offer patrimony	☐	☐	☐	☐	☐
knobby knee	☐	☐	☐	☐	☐
severe beard	☐	☐	☐	☐	☐

Section C: Demografska pitanja

Ovaj odjeljak odnosi se na demografske podatke. Ispunite odgovore u nastavku.

C1. Dob (godine izražene brojevima).

C2. Rod.

Muški ☐
Ženski ☐
Ne želim se izjasniti ☐
Ostalo ▼

37

Section A: Prijevod

U nastavku se nalazi stručni tekst na engleskom iz jednog veterinarskog zbornika. Zamislite da ste stručni prevoditelj te prevedite tekst na hrvatski kao da ga prevodite za objavu. Vremenskog ograničenja nema te se slobodno koristite svim resursima i alatima kojima biste se i inače koristili.

A1.

> Once their offspring hatch, both parents diligently care for them; even cobs offer patrimony, thereby ensuring their young's safety and teaching them essential skills for survival. Even though swans are not notoriously poor parents, instances arise when the cygnets are reluctant to leave them. In such cases, the parents may turn on them to encourage their leaving, a phenomenon observed as preparatory behaviour for the forthcoming breeding season.
>
> Initially, young swans may present as dusky brown- grey accompanied by a grey-black bill. However, as they mature from shining cygnets to adult swans, their juvenile bills turn red and black. The adult male's bill also exhibits a pronounced swelling of the black knob at its base during the breeding season.
>
> Swan pairs typically remain in the same area for life, occupying the rich lands with their nest site. The abundance of water provides their food, e.g. widgeon grass[1], which they can be seen eating, giving the impression of having a severe green- brown beard. Proximity to water is also important to swans for swimming, which interestingly might reveal their knobby knees otherwise hidden by their dry plumage.

[1] hr. "vitičava rupija"

Section B: Pitanja o tekstu

U nastavku se nalazi šest engleskih kombinacija riječi iz izvornog teksta. Prema vašem mišljenju, za svaku od njih na skali od 1 do 5 procijenite koliko smatrate da su uobičajene u engleskom.

B1. **Broj 1 označava da kombinacija riječi nije uopće uobičajena, a broj 5 da je vrlo uobičajena.**

Section C: Demografska pitanja

Ovaj odjeljak odnosi se na demografske podatke. Ispunite odgovore u nastavku.

C1. Dob (godine izražene brojevima).

C2. Rod.

Renderings of SL collocations, their conventionality categories and the participants' evaluations of SL collocations.

Conventional SL collocations

	rich land	Conv. category*	Particip. eval.**	*notoriously poor*	Conv. category*	Particip. eval.**	*knobby knee*	Conv. category*	Particip. eval.**
Literry text	plodno tlo	C	5	notorno loš	p-UC	3	kvrgavo koljeno	C	4
	plodni kraj	p-C	5	biti općepoznato + clause	C	2	kvrgavo koljeno	C	3
	bogata zemlja	C	5	(biti) na glasu kao + NP	C	3	kvrgavo koljeno	C	5
	bogata zemlja	C	5	biti općepoznato + clause	C	4	kvrgavo koljeno	C	5
	plodna zemlja	C	5	(biti) na zlu glasu kao + NP	C	4	kvrgavo koljeno	C	4
	bogata zemlja	C	5	biti na lošem glasu kao + NP	C	4	kvrgavo koljeno	C	5
	plodna zemlja	C	5	biti dobro znano + NP	C	4	koščato koljeno	C	5
	plodna zemlja	C	5	(ne biti) na glasu kao + NP	C	2	kvrgavo koljeno	C	3
	bogati posjed	p-C	5	biti poznat kao + NP	C	3	kvrgavo koljeno	C	5
	bogata zemljišta	p-C	4	svi + znati + clause	C	2	kvrgavo koljeno	C	2
	bogata zemlja	C	4	notorno loš	p-UC	4	kvrgavo koljeno	C	3
	plodna zemlja	C	5	notorno loš	p-UC	4	kvrgavo koljeno	C	3
	rodna zemlja	p-C	5	notorno loš	p-UC	4	kvrgavo koljeno	C	5
	bogati posjed	p-C	5	zloglasno loš	p-UC	2	kvrgavo koljeno	C	4
	plodna zemlja	C	5	biti na lošem glasu kao + NP	C	4	kvrgavo koljeno	C	4
Popular scientific text	bogat izvor hrane	p-C	5	na glasu kao + NP	C	5	kvrgavo koljeno	C	5
	plodna zemlja	C	5	na glasu kao + NP	C	3	kvrgavo koljeno	C	5
	bogata zemlja	C	5	sloviti kao + NP	C	3	kvrgavo koljeno	C	4
	bogato tlo	C	5	na zlu glasu kao + NP	C	2	grbavo koljeno	UC	3
	bogato područje	p-C	5	poznati kao + NP	C	4	koščato koljeno	C	5
	plodno područje	p-C	4	poznati po + NP	C	4	kvrgav zglob	UC	4
	bogat krajolik	p-C	4	[općenito] nisu loši roditelji	C	4	kvrgavo koljeno	C	5
	plodno tlo	C	5	[naročito] ozloglašen kao	C	2	kvrgavo koljeno	C	2
	plodni kraj	C	5	poznati kao + NP	C	3	kvrgavo koljeno	C	4

Ø + zemlja	C	4	poznati po + NP	C	4	kvrgavo koljeno	C	5
plodna zemlja	C	5	notorno loš	p-UC	3	kvrgavo koljeno	C	4
raskošna priroda	p-C	3	na glasu kao + NP	C	2	kvrgava noga	p-C	4
staništa bogata vegetacijom	p-C	5	poznati kao + NP	C	4	kvrgav zglob	UC	2
plodna zemlja	C	5	poznati kao + NP	C	4	kvrgavo koljeno	C	4
bogata zemlja	C	5	na glasu kao + NP	C	3	kvrgavo koljeno	C	4

Unconventional SL collocations

	shining cygnet	Conv. category*	Particip. eval.**	*offer patrimony*	Conv. category*	Particip. eval.**	*severe beard*	Conv. category*	Particip. eval.**
Literary text	blistav mladunac labuda	UC	1	učiniti [nekoga] baštinikom	p-C	2	oštra brada	p-C	2
	blistav labudić	UC	1	preuzeti odgovornost [oca]	C	2	jednostavna brada	UC	1
	sjajan mladi labud	UC	2	ponuditi ostavštinu	UC	3	oštra brada	p-C	1
	blistav labudić	UC	1	osigurati nasljeđe	p-UC	2	duga brada	C	2
	blistav labudić	UC	2	pristati biti otac	UC	2	gruba brada	p-C	3
	blistav mladi labud	UC	1	pobrinuti se o djeci	p-UC	2	oštra brada	p-C	2
	blistav labud	UC	2	preuzeti ulogu [oca]	C	1	bujna brada	C	2
	blistav labudić	UC	1	uzeti pod svoje	C	1	ozbiljna brada	UC	1
	blistav labudov ptić	UC	1	ponuditi nasljedstvo	UC	4	oštra brada	p-C	1
	blistav mladi labud	UC	1	ponuditi nasljeđe	UC	2	oštra brada	p-C	1
	sjajan labud	UC	2	ponuditi nasljedstvo	UC	4	ozbiljna brada	UC	2
	sjajan mladi labud	UC	2	ponuditi nasljeđe	UC	3	velika brada	p-C	2
	besprijekoran mladi labud	UC	3	uzeti pod svoje ime	UC	3	gruba brada	p-C	5
	blistav labudić	UC	1	pružiti nasljedstvo	UC	1	oštra brada	p-C	2
	blistav labudić	UC	1	priznati očinstvo	C	2	oštra brada	p-C	1
Popular scientific text	Ø	C	1	pružiti skrb	C	3	gusta brada	C	2
	mlad labud	p-C	2	sudjelovati u odgoju	C	3	gusta brada	C	1
	mladi labudić	UC	1	pružiti očinstvo	UC	2	gusta brada	C	2
	blistav ptić	UC	2	skrbiti	C	2	gruba brada	p-C	3
	mlad ptić	p-C	2	sudjelovati u odgoju	C	2	Ø + brada	C	1
	sjajni mladunac	UC	2	pružati zaštitu	C	4	gusta brada	C	2
	blistava mladunčad	UC	2	brinuti	C	2	gusta brada	C	2
	labudić	C	1	primiti pod svoje	p-C	3	stroga brada	UC	2

sjajan ptić	UC	2	iskazati se kao [NP]	C	3	duga brada	C	2
blistav mladi labud	UC	1	prihvatiti [očinsku] ulogu	C	2	duga brada	C	5
sjajan mladi labud	UC	2	brinuti	C	4	ogromna brada	p-C	4
blistajući labudić	UC	1	skrbiti	C	1	Ø + brada	C	1
sjajan labudić	UC	1	brinuti	C	3	čupava brada	C	1
divan labudić	UC	1	preuzimati ulogu oca	C	2	bujna brada	C	1
sjajan mladi labud	UC	2	nuditi roditeljsku brigu	p-C	4	duga brada	C	1

* Conv. category = Conventionalisation category

** Particip. eval. = Participants' evaluation

42