Chapter 3

Stylistic normalisation, convergence and cross-linguistic interference in translation: The case of the Czech transgressive

Olga Nádvorníková

Charles University

The Czech transgressive is a non-finite verb form belonging to the cross-linguistic converb category. In contrast with other converbs (e.g. Romance *gerundio* or the Russian деепричастие), the Czech transgressive has a strong stylistic mark and is very rare in contemporary language. Using a parallel (multilingual) corpus and a comparable corpus of translated and non-translated Czech, the paper investigates the differences in the frequency of the transgressive in translated and non-translated fiction and non-fiction. The data show the effect of stylistic normalisation in fiction, but not in non-fiction. The results of the potential effect of cross-linguistic interference are less conclusive, indicating that a thorough contrastive analysis of different language pairs is required first. Finally, the effect of convergence was observed neither in fiction nor in non-fiction.

1 Introduction

The Czech transgressive is part of the cross-linguistic category of converb, i.e. "a non-finite verb form whose main function is to convey adverbial subordination" (Haspelmath 1995: 3). Therefore, the Czech transgressive belongs to the same category as the Romance *gerundio*, English participial adjuncts in *-ing*, the Russian деепричастие and the Polish *imiesłów przysłówkowy*, which is also reflected in translations, as illustrated in example (1):

(1) a. Czech

"Bliju,   soudruhu   četaři,"   odpověděl
puke.1.SG.PRS comrade.VOC sergeant.VOC reply.M.SG.PST.PTCP
jsem   **opíraje**   **se**   rukou   o
be.1SG.AUX **lean.CONVERB.PS.IMPF.M.SG** REFL hand.INSTR against
zeď.
wall.ACC

'"I am puking, Comrade Sergeant," I answered, leaning with one hand
against the wall.' (Milan Kundera, *Žert*/*The Joke*, 1991/1969)

   b. English

"Puking, Comrade Sergeant," I replied, **leaning** against the wall with
one hand. (transl. David Hamblyn; Oliver Stallybrass, 1992)

   c. French

Je suis en train de dégueuler, camarade sergent, expliquai-je **en
m'appuyant** d'une main au mur. (transl. Marcel Aymonin, 1975)

   d. Polish

– Rzygam, towarzyszu plutonowy – odpowiedział em, **opierając się**
ręką o mur. (transl. Emilia Witwicka, 1999)

   e. Russian

"Блюю, товарищ сержант", – ответил я, **опираясь** рукой о стену.
(transl. Нина Шульгина, 1999)

   f. Spanish

"Vomito, camarada sargento", le respondí **apoyándome** con la mano
en la pared. (transl. Fernando de Valenzuela, 1994)

However, in contrast with the other converbs, the Czech transgressive has a
strong stylistic mark (bookish or even archaic), is used rarely and only in written
texts.

This paper aims to investigate whether translated and non-translated Czech
texts differ in the frequency of the transgressive. We assume that a higher fre-
quency of transgressives in translations in comparison with non-translated texts
may be attributed to the cross-linguistic influence (in translations from languages
using converbs). The opposite result, i.e. a lower frequency of transgressives in
translations than in non-translated texts, may be explained by the tendency to-
wards normalisation. We assume also that a higher tendency to convergence may
be manifested by a lower coefficient of variation of the frequency of the trans-
gressive in translations.

The analysis is conducted on the InterCorp parallel (multilingual) corpus; a comparable corpus of translated and non-translated Czech, named Jerome; and a large monolingual synchronic corpus of Czech (SYNv8). All the corpora are limited to fiction and non-fiction. InterCorp and Jerome are used for comparison of translated and non-translated texts; the SYNv8 corpus (restricted to non-translated texts only) serves as a reference corpus for the other two corpora under analysis.

The paper is organised as follows. §2 introduces the twofold theoretical background of the research: the morphological, semantic, syntactic and stylistic properties of the Czech transgressive in the framework of the cross-linguistic category of converbs (§2.1) and the theory of special features of translated language, namely normalisation, convergence and cross-linguistic interference (§2.2). After presenting the corpora used in this research (§3), we introduce the results (§4): the analysis of the potential influence of normalisation and convergence in translations, and subsequently, the potential impact of cross-linguistic interference. In the final part of the paper, we summarise the main outcomes of the research and suggest some open questions for future research.

2 Theoretical background

2.1 The Czech transgressive

2.1.1 The Czech transgressive as a converb

According to Nedjalkov (1998: 431), most European languages have one or two converbal forms, whereas languages outside Europe often have several converbal forms (e.g. Japanese, see Alpatov & Podlesskaya 1995).[1] According to Nedjalkov (1998: 430), polyconverb languages within Europe are Basque, Finnish and Lithuanian with six converbs each. Converb-free languages in Europe are rare, for example, Maltese and Romani (Nedjalkov 1998).

Due to their non-finite form (see the definition in §1), converbs convey the adverbial meaning in a more condensed way than the corresponding finite subordinate adverbial clause (see Vachek 1955; Nosek 1964; for Romance, for instance, Čermák et al. 2020; for Czech Bečka 1977). Because of their verbal character, they are also richer in information than complete nominalisations (verbal nouns, prepositional phrases, etc.). Moreover, in contrast with coordinate finite clauses,

---

[1]For instance, "the average number of converbs in languages spoken within the territory of Russia is more than seven converbal forms per language" (Nedjalkov 1995: 431).

converbs allow for hierarchisation of information; in other words, the event conveyed by the converb is considered secondary (see 1a).

Converbs may differ regarding their position in the system of the given language (see Nedjalkov 1995: 104–106; 1998: 433). Strict, canonical converbs are specialised in the converbal (adverbial) function (e.g. Russian and Czech converbs, the French *gérondif* and Estonian forms in *-des*, etc.). By comparison, non-strict converbs fulfil, apart from the converbal function, other functions, e.g. participial or infinitival, such as the English forms in *-ing*, German *-end* or Spanish and Italian *-ndo* (see Nedjalkov 1998: 425; for Romance, see Čermák et al. 2020).

From the syntactic point of view, it should be noted that as non-finite verb forms, converbs do not have a valence position for the subject. In some languages, the subject (controller) of the converb has to be coreferential with the subject of the main clause, like in Slavic languages, including Czech (*same-subject converbs*, see Nedjalkov 1998: 425; Komárek 1986: 153)[2] or French (*gérondif*, see Grevisse & Goosse 2016: 1252). In other languages, the controller of the converb may be non-coreferential, as in Portuguese, Spanish (see Čermák et al. 2020: 111) or in Modern Greek, Armenian and Albanian (Nedjalkov 1998: 425).

Finally, concerning the semantic interpretation of converbs, we can distinguish contextual and specialised converbs (Nedjalkov 1998: 431). Specialised converbs only have one or two specific meanings (e.g. Finnish has a specialised converb conveying manner, see Nedjalkov 1998: 443). However, most European languages, including Czech, belong to the contextual converb type, i.e. their converbs are semantically vague, the potential meanings are numerous and given by the context (for the factors influencing the semantic interpretation of converbs, see for example König 1995; König & van der Auwera 1990: 337; Dvořák 1983: 29–41 for Czech; and for French Nádvorníková 2012).[3]

The meanings conveyed by contextual converbs can be divided into two large groups: temporal (simultaneity, anteriority and posteriority) and non-temporal (manner/means, cause, concession, condition etc.). The meaning of simultaneity proper or concomitance/attendant circumstance is the default meaning: "if a European language has only one converb, then it is a mixed converb of contextual Simultaneity" (Nedjalkov 1998: 432; see a similar observation for French *gérondif* in Kleiber (2007: 117; 2009: 19)). This observation is corroborated by a diachronic trend: "If a language moves from the group of two-converb languages

---

[2]Dvořák (1970: 37–45), in his diachronic study of Czech, points out that 30% of transgressives in the 17th century were non-coreferential.

[3]Moortgat (1978: 157) considers the French gerund to be a "semantic chameleon" (see also Halmøy 2003).

to the group of mono-converb languages then the remaining converb will belong to the contextual Simultaneity type" (Nedjalkov 1998: 437).

Despite the aforementioned variety in the subtypes, all converbs share the basic definition as a "non-finite verb form conveying the adverbial subordination" (Haspelmath 1995: 3, see §1). Most European converbs share other properties, in particular, the contextual semantic interpretation and the use as a means of syntactic condensation (*he said and he smiled – he said, smiling – he said, with a smile*). More specifically, Romance and Slavic converbs are considered typical (see Haspelmath 1995: 45 for the former and Nedjalkov 1998: 422 for the latter). According to Nedjalkov (1998), Slavic converbs are prototypical for the category, especially the Russian converb. Germanic languages, conversely, make, except for English, "only parsimonious use of converbs" (König 1995: 72, and a similar statement in Kortmann (1997: 192)). According to Kortmann, in English, "free adjuncts are far from playing a minor role" and the frequency of adverbial participial clauses is five times higher in English than in German (Kortmann 1997: 192). Similar differences in the use of converbs can be observed in the Slavic and Romance language families. In Slovak, the frequency of the converb is much lower than in Russian (Brtková 2004: 25). Similarly, in French, the frequency of *gérondif* is several times lower than the frequency of the corresponding forms in Italian, Spanish and Portuguese (1,571 instances per million words (ipm) against 4,098 ipm, 4,886 ipm and 6,939 ipm respectively, see Čermák et al. 2020: 116).

The Czech transgressive displays most of the properties considered by Nedjalkov as defining the prototypical (Slavic) converb: it is syntactically strict, i.e. it may be used only in the adverbial (converbal) function; it is formally simple, i.e. its formation involves suffixes, not adpositions; it has two converbal forms, one of which is a converb of contextual Simultaneity and the other as a converb of contextual Anteriority; and it is referentially the same-subject (co-referential). However, the Czech converb shows one important particularity: it maintains a very complex, archaistic morphology, involving adjectival agreement markers (in gender and number) (see §2.1.2), and, as a consequence, it acquires a strong stylistic mark and is used very rarely in contemporary language (see §2.1.3).

2.1.2 Morphological features of the Czech transgressive

As mentioned above, the Czech transgressive has two main forms (Karlík 2017):

- the "present transgressive" (*přechodník přítomný* or *-c transgressive*), formed with imperfective verbs only and conveying the meaning of simultaneity.

- the "past transgressive" (*přechodník minulý* or *-š transgressive*), formed with perfective verbs only and conveying the meaning of Anteriority.

When influenced by contextual factors (see 2.1.1), these basic meanings may acquire other nuances, such as manner, cause, condition, etc. (Dvořák 1983: 33). Each form (present and past) has a different set of morphemes, varying according to the type of verb stem and as the consequence of agreement with the subject (controller) of the transgressive in gender and number, as mentioned above. Table 3.1 summarises this complex morphology of the Czech transgressive.

Table 3.1: Morphology of the Czech transgressive

Czech transgressive forms		Form	
		Present	Past
M.SG		-a / -e / -ě	-∅ / -v
F.SG + N.SG		-ouc / -íc	-ši / -vši
PL (M+F+N)		-ouce / -íce	-še / -vše
Aspect	Imperfective	**Present (Simultaneity)** CONV.PS.IMPF	Simultaneity/Anteriority CONV.PT.IMPF
	Perfective	Futurate CONV.PS.PF	**Past (Anteriority)** CONV.PT.PF

Table 3.1 demonstrates four possible combinations of aspectual and formal characteristics of the Czech transgressive. Conv.ps.impf and Conv.pt.pf are the dominant forms, mentioned by most Czech grammars (e.g. Komárek 1986: 154; Cvrček 2010: 148–249; or Karlík et al. 1995: 335–337).[4] For Conv.ps.impf, see 1a; for Conv.pt.pf, see 2.

(2) **Uslyšev**, že Švejka naznačil plukovník ordonancí u 11. kumpanie, zvolal: "Pomoz nám pánbůh." (Jaroslav Hašek, *Osudy dobrého vojáka Švejka za světové války*, 1921–1923/1996)
hear-CONV.PT.PF.M.SG
'**Having heard** that the colonel marked Švejk to be the ORDONANC with the 11th company, he yelled out: "Help us Lord God."' (transl. Zdenek K. Sadloň)

---

[4]Nedjalkov (1998: 437) gives for the Czech converb conveying simultaneity the endings -a/-je/-oic, which is incorrect.

Conv.pt.impf was already rare in Old Czech (Dvořák 1970: 115); in contemporary Czech, it is not in use (Komárek 1986: 154). Finally, the form of the futurate transgressive (Conv.ps.pf) is supposed to be used only to convey anteriority in the future, i.e. combined with the main verb in the future tense (see Nedjalkov 1995: 126; Komárek 1986: 154). However, these forms were replaced by past converbs (Conv.pt.pf, see Oktábec 1953: 261) and are not in use either.

Even when limited to Conv.ps.impf and Conv.pt.pf, the morphology of the Czech converb is very complex and contrasts with the converbal systems in other Slavic languages, where the converb went through a process of adverbialisation and the forms were simplified, in particular by dropping the agreement with the subject (controller). In contemporary Czech, the only non-coreferential converbs are the grammaticalised ones: as much as in other languages (see Haspelmath 1995: 27–41), Czech converbs may be reanalysed in other categories, mainly adverbs (e.g. *chtě nechtě* 'reluctantly') and prepositions (e.g. *počínaje* 'starting with' or *nemluvě* 'notwithstanding', see Komárek 1986: 156).

Some languages maintained different sets of suffixes for the past (perfective) and the present (imperfective) forms (e.g. Russian and Polish). Other languages, for instance Slovak (Dvonč et al. 1966: 487), went further in the process of simplification and use the same set of suffixes for the perfective and imperfective verbs. The same tendency can be observed in Czech dialects (Dvořák 1983: 55–56; Michálková 1963), which demonstrates that spoken, non-standard Czech also adverbialised the transgressive.

This difference between standard literary Czech and its dialects (and other Slavic languages) is caused by a normative intervention made by grammarians during the Czech National Revival movement in the first third of the 19th century. At that time, the use of the Czech language was limited, since the language of economic and cultural elite was German, and Czech was spoken mostly by the rural population and the poorer inhabitants of cities. Therefore, while choosing the norm for the Czech language to be resuscitated, the grammarians and lexicographers of the National Revival movement did not opt for the language of their time (the 19th century), which was considered unprestigious and decayed, but the norm of the flourishing period of the Czech state, culture and language at the end of the 16th century, i.e. the language that was more than 200 years old at that time.

More importantly for our topic, the newly defined norm of standard literary Czech re-introduced in the transgressive its old complex morphology of the literary norm of the end of the 16th century. Since the transgressive was mostly used in written texts, especially for its advantages as a means of syntactic condensation, the norm was respected. Nevertheless, the transgressive gradually ceased

to be part of the internalised, unconscious competence of the speakers; its frequency was in constant decline and the form acquired its stylistic mark.

2.1.3 Stylistic features and frequency of the Czech transgressive

Extensive research into transgressives conducted by Dvořák (1983: 60) demonstrates the constant decrease in the frequency of this form between 1781 and 1978 (from 6.49% of all verb forms in the period 1781–1830 to only 0.14% in 1971–1978). He observes the decrease in the frequency of the transgressive in the 18th century already (Dvořák 1970: 142), which indicates that the normative intervention during the National Revival movement may not have been the main factor triggering the decrease of the frequency of this form in Czech. Nevertheless, it is plausible to assume that in the 20th century, the archaistic morphology and stylistic mark resulting from the normative intervention contributed considerably to the retreat of this form. The most recent grammar of Czech, published in 2010 and based on corpus data, states that the transgressive is "very rare" and that it represents less than 1% of the verb forms in Czech (Cvrček 2010: 249).

It is worth noting that there is a neat difference in the frequency of the two main forms of the transgressive: Cvrček & Kováříková (2011: 130) indicate that the frequency of Conv.ps.impf is nowadays less than 0.1%, but the frequency of Conv.pt.pf is even less than 0.01% of all verb forms, which means that the ratio of the two forms is 10. A similar difference in the frequency of the present and the past transgressive was already observed by Dvořák (1983: 60): 0.34% Conv.ps.impf and only 0.04% Conv.pt.pf of all verb forms in texts published between 1960 and 1970 (ratio 8.5). Conversely, in the period of 1781–1830, Dvořák observed 4.17% of Conv.ps.impf and 1.39% of Conv.pt.pf, i.e. only a ratio of 3.[5] Even though the exact figures given by Dvořák for the different time spans may not be fully reliable, due to the lack of comparability of the sub-corpora under analysis, the tendency is clear: Conv.pt.pf is systematically less frequent than Conv.ps.impf.

The difference in frequency between the two main forms of the transgressive may be ascribed not only to the specific morphology of Conv.pt.pf (see Table 3.1), but also to the differences in the meaning of the two forms and the availability of competing forms in the language. Conv.pt.pf, conveying the meaning of anteriority, is strongly concurred by other forms, especially finite subordinate clauses of a temporal or a specific adverbial meaning (e.g. the cause, as in example 2). Conv.ps.impf, by contrast, mostly conveys a simple accompanying circumstance

---

[5]Dvořák also indicates the frequencies of the two remaining forms of the trangressive in 1781–1830: 0.03% for Conv.pt.impf and 0.9% for Conv.ps.pf (Dvořák 1983: 69).

(see example 1a or "*řekl jsem* **usmívaje se**" - CONV.PS.IMPF.M.SG 'I said **smiling**'), which cannot be expressed by a subordinate clause, but only by a coordinate one ("*řekl jsem* **a usmíval jsem se**" 'I said **and I was smiling**') or a simple SP ("*řekl jsem* **s úsměvem**" 'I said **with a smile**').

Kortmann (1997: 281) made a similar observation for most European languages: they do not explicitly encode the meaning of concomitance (by an adverbial subordinator), so this meaning is mostly conveyed by converbs or a simple juxtaposition of two finite clauses. Even though the replacement of the converb by a subordinate clause moves the form from non-finite to finite and explicates its meaning by a subordinator (see Nádvorníková 2017), in contrast with the coordinate clause, it maintains the adverbial subordination relation and hence the hierarchisation of events typical for converbs (see the definition in §1). As a consequence, the coordinate clause is a less obvious concurrent of the converb than a subordinate one, and the meaning of the accompanying circumstance is more likely to persist in this form than more specific adverbial meanings. Furthermore, as remarked by Nedjalkov (1995), the accompanying circumstance is the most frequent meaning conveyed by converbs in general (see the same observation for Romance languages in Čermák et al. (2020: 122) and for Czech in Dvořák (1983: 33)).

As mentioned above, the archaistic morphology of the transgressive is also the source of its specific stylistic mark. Most Czech grammars consider the transgressive as bookish (Conv.ps.impf) or even archaic (Conv.pt.pf), and limited to the written language (Komárek 1986: 154; Cvrček 2010: 249; Karlík et al. 1995: 337). The stronger stylistic mark of Conv.pt.pf correlates with the aforementioned lower frequency.

The bookish/archaistic stylistic mark of the transgressive also influences its frequency in different text registers. Most sources agree that the transgressive is typical for fiction (Dvořák 1983: 105; Bečka 1977: 24; Čechová et al. 1997: 102), in particular because of its ability to convey in a condensed way the accompanying circumstance in narrative sequences and introductory clauses (Dvořák 1983: 107; Bečka 1977: 19 and example 1a). The stylistic mark in fiction is also exploited in historical novels or as a means of irony or parody (Čechová et al. 1997: 102–103; Komárek 1986: 154). However, in fiction intended for children or young readers, transgressives are less frequent than in fiction for adults (see Jelínek et al. 1961: 90).[6] In non-fiction, the transgressive is considered less frequent than in fiction and conveys more specific adverbial meanings than a simple accompanying cir-

---

[6]Bečka (1977: 23) also mentions the potential influence of a specific author's idiolect (e.g. the Czech author Vladislav Vančura, 1891–1942, is known for his penchant for transgressives).

cumstance (see Dvořák 1983: 33; Bečka 1977: 21; Karlík et al. 1995: 337).[7] Dvořák (1983: 106 and 108) points out that the transgressive is more frequent in social sciences than in natural or technical sciences. Finally, in journalistic texts, the transgressive is the least frequent, in comparison with fiction and non-fiction (Dvořák 1983: 106; Jelínek et al. 1961: 90).

2.2 The transgressive in translations

To our knowledge, only a few researchers have focused specifically on the use of transgressives in translation, apart from three rather dated studies (Bečka 1977; Dvořák 1972; 1983). However, the topic is occasionally addressed in contrastive studies exploring Czech equivalents of converbs.

In his quantitative study, Dvořák (1972; 1983) analysed various Czech translations of the same source texts (four source texts in Russian, one in French and one in English). The translations were published between 1863 and 1975 and six different translations on average were analysed for each text. The results confirmed the decrease in the frequency of the transgressive observed in non-translated texts (see §2.1.3), but the normalised frequency of transgressives was almost always higher in translations than in non-translated texts from the corresponding period. For instance, in the Czech translation of Charles Dickens' *The Posthumous Papers of the Pickwick Club (chapters 1-5)* published in 1925, transgressives represent 5.5% of all verb forms, whereas the average for the given period in non-translated texts is only 1.384% (Dvořák 1983: 94).

These results suggest that in translations from languages using converbs (i.e. most European languages, see below), the effect of cross-linguistic interference (or *shining through*, see §2.2) may be expected and the frequency of transgressives may be higher in translated than in non-translated texts. However, other studies indicate the opposite conclusion.

First, in their contrastive research of Czech equivalents of Romance converbs, Čermák et al. (2015; 2020) show that in Czech translations from four Romance languages (French, Italian, Portuguese and Spanish), the transgressive represents the least frequent counterpart (from 2.0% in translations from French to 9.6% in translations from Portuguese), despite the presumed systemic equivalence. In comparison, the finite counterparts (coordinate and subordinate clauses) form about 70% of the whole. Malá & Šaldová (2015: 240) present a similar result in translations from English: the transgressive represents only 2.1% of the Czech counterparts of English adverbial participles; the overwhelming majority of counterparts being

---

[7]Karlík (2017) and Čechová et al. (1997) consider the transgressive also appropriate in highly formal, e.g. diplomatic or legal, documents.

finite verbs (73%). Finally, in translations from Russian (Kocková 2011), transgressives constitute less than 1% of the equivalents of (past) деепричастия. These results suggest that in translations, the frequency of the transgressive may be as low as in non-translations, or even lower.

Second, Bečka (1977: 26) in his (non-quantitative) analysis of the transgressive in translations points out that translators had been warned against the use of the transgressive and that they avoid it because of its stylistic mark. Similarly, Levý (2011: 51) states that the frequency of the transgressive in Czech is lower in translations than in non-translated texts, because translators are over-concerned to avoid stylistically marked features. These observations indicate that, on the contrary, the frequency of transgressives may be influenced by the effect of stylistic normalisation and, therefore, be lower than in non-translated texts.

Toury (1995) states that translations are governed by two universal laws: the law of interference and the law of growing standardisation (or *normalisation*, according to Baker (1993; 1996)). Cross-linguistic interference (or *shining through*, according to Teich (2003)) consists of transferring linguistic features of the source language into a target language (see Toury 1995: 274–279). Normalisation, by contrast, may be defined as "the tendency to conform to patterns and practices that are typical of the target language, even to the point of exaggerating them" (Baker 1996: 176–177).

Various studies have shown the effect of cross-linguistic interference in translation in various language pairs. For instance, Dai & Xiao (2011), when analysing Chinese texts translated from English, found that passive voice is more frequent in Chinese translated from English than in non-translated Chinese texts. Similarly, Cappelle (2012) shows that English texts translated from French contain fewer manner-of-motion verbs than English texts translated from German. He explains this effect by the typological differences between the two source languages: German and English are satellite-framed languages, whereas French is a verb-framed language.

As for normalisation, this is defined by the linguistic properties as well as the sociocultural norms of the target language (see Lefer & Vogeleer 2013: 17). Alongside the explicitation, the simplification and the levelling out (convergence, Laviosa 2002) it is one of the specific features of translation ("translation universals", according to Baker (1993; 1996)) that is addressed the most in literature. In Chesterman's (2004: 39) terms, it can be conceived either as an S-universal, causing differences "between translations and their source texts", or as a T-universal, giving rise to differences between translations and comparable non-translated texts in the target language. In this study, we focus on the effect of normalisa-

tion as a T-universal, by comparing non-translated Czech texts with translations in the same language. We also partially focus on the convergence (levelling out).

Normalisation as a T-universal was analysed by Delaere et al. (2012); their results confirmed the tendency to normalisation (standardisation) in translated Dutch, in comparison with non-translated texts of the same language. Similarly, Chlumská (2017) observed the effect of normalisation in translation in the choice of two forms of the verb *say* in Czech: *říci* (formal, stylistically marked form) and *říct* (standard, stylistically neutral). In her corpus, translations showed a higher frequency of the latter form than non-translations, which suggests the effect of (stylistic) normalisation (Chlumská 2017: 65). Lapshinova-Koltunski (2018), who compared translations from English into German (in six different text registers), observed that the effect of normalisation is sensitive to two factors: text register (the highest score of normalisation was in translations of fiction) and the translator's proficiency (the normalisation score was higher in student translations than in professional translations).

Levelling out (or convergence, Laviosa 2002) is sometimes considered as a subtype of normalisation. Baker (1996: 177) defines levelling out as "the tendency of translated text to gravitate around the centre of any continuum rather than move towards the fringes". Laviosa (2002: 71) is more specific and points out that the convergence implies a relatively higher level of homogeneity of translated texts concerning certain linguistic features, such as lexical density, sentence length, etc. As stated by Baker (1996: 184) and by Chlumská (2017: 104), less attention has been paid to this feature than to the other translation universals as it is more difficult to operationalise. Lapshinova-Koltunski (2015) confirmed the tendency to convergence in several translation variants in German (translated from English). Chlumská (2017: 104–121) analysed various potential indicators of the convergence (sentence length, type-token ratio, etc.) in Czech and observed its effect in translations of fiction but not in non-fiction.

2.3 Hypotheses and research questions

Our main research question aims to find out the differences in the frequency of transgressives in translated and in non-translated texts. Based on the theory of the interplay between the cross-linguistic interference and the normalisation in translation (see §2.2), we can formulate the following hypotheses (H_1 and H_2 being in opposition):

H_0 Translated and non-translated texts of the same text register do not differ in the frequency of transgressives.

H_1 Due to the effect of cross-linguistic interference, in translations from languages using converbs, the frequency of transgressives is higher than in non-translated texts of the same text register. Based on typological observations made in the literature (Haspelmath 1995, Nedjalkov 1995, see §2.1.1), we expect more transgressives in translations from Romance and Slavic languages, especially Russian, than in translations from Germanic languages (with the potential exception of English). Transgressives resulting from this interference can also be expected in translations from poly-converbal Latvian, Finnish and Japanese.

H_2 Due to the effect of (stylistic) normalisation, the frequency of transgressives is lower in translations than in non-translated texts of the same text register (independently of the source language and the text register).

Based on the theory of convergence (see §2.2), we can formulate the third hypothesis:

H_3 Due to the tendency of translations towards convergence (greater homogeneity), the coefficient of variation of the frequency of transgressives in translations is lower than in non-translated texts of the same text register.

Our second research question aims to find out what other factors influence the frequency of transgressives in translated and non-translated texts. From a strictly linguistic point of view, we expect an important difference in frequency between the two forms of the transgressive since the past form (Conv.pt.pf) is stylistically more marked than the present form (Conv.ps.impf). Among the extra-linguistic factors, we expect the greatest influences to be the date of publication of the text (the older the text, the higher the frequency) and the text register (more transgressives in fiction, exploiting its stylistic mark and the ability to convey accompanying circumstance, and fewer transgressives in non-fiction, more stylistically neutral than fiction).

3 Data and Methods

As mentioned in §2, the main source of data for our research is corpora including translated texts: the comparable corpus of translated and non-translated Czech named Jerome (Chlumská 2013), and the InterCorp parallel (multilingual) corpus (Rosen et al. 2019). The data obtained from these corpora are confronted with the

data extracted from the large monolingual synchronic corpus of Czech named SYNv8, limited to non-translated texts (Křen et al. 2019).

All these corpora were created by the Institute of the Czech National Corpus and are freely available using the same corpus interface (KonText; www.korpus.cz and http://kontext.korpus.cz). InterCorp was annotated using the POS-tagger named Morče (see http://ufal.mff.cuni.cz/morce/index.php); SYNv8 and Jerome were annotated by a hybrid system developed in-house (combining stochastic and rule-based disambiguation, see Hnátková et al. 2011).

The error rate of the transgressive POS-tagging is 3.6% for the Conv.ps.impf and 6.8% for the Conv.pt.pf (tested on a sample of 250 occurrences for each form).[8] Even though these POS-tagging errors can influence the resulting frequencies, we consider our observations reliable, since the main information for our research is not the absolute frequencies of the transgressive, but the comparison of the relative ones.

3.1 The Jerome comparable corpus of translated and non-translated Czech

Jerome (Chlumská 2013; 2017) is a monolingual corpus specifically designed for the research of translation features (in terms of T-universals, see Chesterman (2004) and §2.2), in fiction and non-fiction. The corpus comprises translated and non-translated texts in Czech, the two subcorpora being comparable in size and other relevant factors. For instance, all the texts were published between 1992 and 2009 and the same author/translator may be represented by a maximum of three texts to prevent the risk of the influence of a specific idiolect.

The representation of source languages in the translation part of the corpus reflects the situation on the publishing market in the Czech Republic, where translations from English are three times more frequent (i.e. probably three times more read) than from any other language.[9] In total, it includes 22 different source languages in fiction and 15 source languages in non-fiction. The potential interference effect can be explored using a smaller balanced subcorpus including an

---

[8]In the corpus queries, we excluded from the analysis the most frequent grammaticalised forms of the transgressive in the two sub-corpora. The resulting queries are:
[tag="V(e|m).*" & word!="((N|n)e)?(C|c)ht(ě|íc)|(N|n)emluvě|(P|p)očínaj(e|íc)|(K|k)onč(e|íc)|(N|n)evyjímaj(e|íc)|(T|t)ak říkajíc|(S|s)oudě"] for fiction and
[tag="V(e|m).*" & word!="((N|n)e)?(C|c)ht(ě|íc)|(N|n)emluvě|(P|p)očínaj(e|íc)|(K|k)onč(e|íc)|(N|n)evyjímaj(e|íc)|(T|t)ak říkajíc|(S|s)oudě|(V|v)ycházej(e|íce)|(N|n)epočítaj(e|íc)"] for non-fiction.

[9]According to the Czech National Library statistics of translated books, available (in Czech) at http://text.nkp.cz/sluzby/sluzby-pro-sluzby-pro-vydavatele/vykazy, for more details, see Cvrček & Chlumská (2015: 313).

equal amount of texts translated from 14 different languages in fiction and 6 languages in non-fiction. Table 3.2 summarises the composition of the Jerome comparable corpus:

Table 3.2: Composition of the Jerome comparable corpus of translated and non-translated Czech

Jerome corpus	N° of tokens (incl. punctuation)		Source languages in translation
	translated	non-translated	
Fiction (non-balanced)	26,617,523	26,551,540	da nl en fi fr de el(new) he hu is it ja no pl pt ru sr sk sl es sv
Non-fiction (non-balanced)	15,946,319	15,949,930	ar en fr de el(old) hu it la pl ro ru sr sk es sv
Fiction (balanced)	1,765,433	1,768,079	da nl en fi fr de is it ja pl pt ru es sv
Non-fiction (balanced)	774,610	779,288	en fr de it pl ru

In the whole (non-balanced) corpus, translations from English represent 69% of the subcorpus of translations (by the number of tokens). The other languages represented by at least 500,000 tokens are French and German (8% each) and Russian and Polish (3% and 2% respectively). In non-fiction, the composition is similar: translations from English represent 55% of the sub-corpus, followed by German (25%) and French (8%). The other languages usually do not exceed 1%.

As mentioned above, all the texts included in the Jerome corpus were published between 1992 and 2009; nevertheless, some of them were first published earlier. Since the frequency of the transgressive is highly likely to be sensitive to the date of the creation of the text (see §2.1.3), we eliminated these texts from our corpus research (14 texts in translated fiction, 24 texts in non-translated fiction, one text in translated non-fiction and three texts in non-translated non-fiction). Thus, in translated fiction, we excluded, e.g. the Czech translation of William Faulkner´s novel *The Wild Palms*, first published in 1960. In non-translated fiction, the set of eliminated texts includes not only texts first published before 1992, for instance four novels by Vladislav Vančura,[10] first published in the 1920s, but

---

[10] *Amazonský proud, Pekař Jan Marhoul, Pole orná a válečná* and *Poslední soud.*

also texts first published in the given interval (1992–2009), but written earlier, e.g. a posthumous edition of a novel by Lev Blatný (*Servus, Ser-vá-ci*), an author who died in 1930 and the memoirs of a Habsburg Empire army officer in the Great War (*Z turecké armády do britského zajetí*).

It is very likely that, especially in translations, new editions of texts published earlier were revised and adapted by editors in publishing houses. However, our data show that the inclusion of these texts in the research may have skewed the results, if not excluded. For instance, all the aforementioned eliminated texts show the highest normalised frequencies of the transgressive in our subcorpora of fiction, which corroborates the hypothesis of the influence of the date of creation/publication of the text on the frequency of the transgressive.

Furthermore, to maintain the comparability with the InterCorp parallel corpus (see below), we limited the Jerome corpus to novels and short stories in fiction and scientific (SCI) and popular (POP) texts in non-fiction (eliminating e.g. textbooks and encyclopaedias, not included in InterCorp). For these reasons, the size of the Jerome sub-corpora introduced in the results of our research (see §4) is smaller than that given in Table 3.2.

3.2 The InterCorp multilingual corpus

InterCorp (https://wiki.korpus.cz/doku.php/cnk:intercorp:verze12) is a large multilingual (parallel) corpus currently involving 41 languages, with Czech as pivot language (Čermák & Rosen 2012; Nádvorníková 2016). The corpus is composed of the so-called core, which comprises fiction and partially non-fiction, and collections (movie subtitles, the Bible, journalistic texts, Acquis communautaire and EuroParl). Our research exploits only the core of the corpus, because, in contrast with the collections, the quality of translations is higher in the core texts and all the metadata necessary for research in translation studies is available (date of publication, source language, name and sex of the author/translator, different text sizes in tokens, etc.).

The main advantage of the InterCorp parallel corpus, in comparison with the Jerome comparable corpus, is its larger size, i.e. the larger number of texts and different authors/translators, which reduces the risk of the influence of a specific text style or an author's/translator's idiolect. This is also the reason why we do not use the Jerome corpus in the interference hypothesis testing (H_1), as it is limited to one to three texts per language, but instead, we use the InterCorp parallel corpus (see §2.1.3).

However, the translated and non-translated sub-corpora of the InterCorp corpus are not comparable, neither in size nor composition. As can be seen in Ta-

ble 3.3, the size of the non-translation sub-corpora in the InterCorp corpus is quite small, which is due to the limited availability of translations from Czech into foreign languages. This issue is even more pronounced in non-fiction than in fiction. In addition, the sub-corpus of non-translations in InterCorp is not limited solely by size, but also by composition, as foreign publishing houses particularly choose texts by well-known and established authors for translations from Czech. As a consequence, the non-translation subcorpora of InterCorp are not a reliable source of data for real language use in Czech.

For this reason, the data for the comparison of translations with non-translated texts were not extracted from InterCorp, but from the largest corpus of contemporaneous Czech – SYNv8 (Křen et al. 2019; Hnátková et al. 2014), limited to non-translated fiction (novels and short stories) and non-fiction (scientific and popular texts). Table 3.3 demonstrates the resulting size of the sub-corpora.

Table 3.3: Composition of the InterCorp parallel corpus and the SYNv8 reference corpus of Czech

	Corpus	InterCorp		SYNv8
		translated	non-translated	non-translated
Fiction	texts (n)	1,179	286	496
	tokens (n)	107,375,278	19,208,622	30,527,709
	SLs (n)	32	–	–
Non-fiction	texts (n)	80	13	650
	tokens (n)	6,803,832	881,833	33,878,274
	SLs (n)	5 (de,it,fr,en,sv)	–	–

The overwhelming majority of texts in the InterCorp parallel corpus and SYNv8 were published after 1950, with the majority after 1980. However, some texts were first published much earlier, e.g. *Osudy dobrého vojáka Švejka* by Jaroslav Hašek (1921–1923, see example 2) and the Czech translation of *The Jungle Book* by Rudyard Kipling (1911).

To maintain the comparability with the results obtained on the Jerome corpus and to reduce the influence of the date of publishing, we limited the InterCorp parallel corpus and the SYNv8 to texts (first) published after 1992. The whole corpora, including older texts, are only used to analyse the evolution of the frequency of the transgressive (see Figures 3.1 and 3.2). The normalisation, convergence and cross-linguistic hypotheses are thus tested only on the texts published

after 1992 (inclusive). Similar to the Jerome corpus, the resulting sub-corpora provided in §4 are smaller than those in Table 3.3. Furthermore, as much as in the Jerome corpus, even in the sub-corpora limited to texts published after 1992, we identified and eliminated some of the texts first published or written earlier (e.g. the novel *Nesmrtelnost/Immortality* by Milan Kundera, written in 1987–1988 and the Czech translation of *Les Mots* by Jean-Paul Sartre, first published in 1967).

As for the source languages, the whole fiction/non-fiction sub-corpus of InterCorp involves 31 different source languages: Arabic (ar), Belarussian (be), Bulgarian (bg), Catalan (ca), Croatian (hr), Danish (da), Dutch (nl), English (en), Finnish (fi), French (fr), German (de), Hindi (hi), Hungarian (hu), Italian (it), Japanese (ja), Lithuanian (lt), Latvian (lv), Macedonian (mk), Norwegian (no), Polish (pl), Portuguese (pt), Romany (rn), Romanian (ro), Russian (ru), Slovak (sk), Slovene (sl), Serbian (sr), Spanish (es), Swedish (sv), Turkish (tr) and Ukrainian (uk).[11] The most represented languages are German and English (more than 30 million tokens each, i.e. more than 10% of the corpus each). The source languages representing between 5% and 10% of the corpus (i.e. more than 20 million tokens) are Polish, Spanish, Croatian and French (for detailed information, see https://wiki.korpus.cz/doku.php/en:cnk:intercorp:verze12).

In the corpus limited to texts published after 1992 (inclusive), the number of source languages is only twenty (da, de, en, es, fi, fr, hr, it, ja, lv, nl, no, pl, pt, ro, ru, sk, sl, sr, sv). Translations from English prevail (36% of the sub-corpus), followed by German, Spanish and Swedish (see Table 3.5 for more details). It can be observed that all the languages included in this sub-corpus belong to the European area (except for Japanese) and except for Finnish and Latvian, they all belong to one of the three prevailing language families in Europe (Romance, Slavic and Germanic). The corpus thus allows to test the normalisation and convergence hypotheses (§4.1) and investigate the potential cross-linguistic interference effect (§4.2).[12]

---

[11]Nine source languages are available in collections only and not in the core of the corpus: Greek (el), Estonian (et), Hebrew (he), Icelandic (is), Malay (ms), Maltese (mt), Albanian (sq), Chinese (zh) and Vietnamese (vi).

[12]The number of source languages in the Jerome corpus is higher than in the InterCorp parallel corpus, because InterCorp includes only source languages for which source texts are really available in the corpus, whereas the Jerome corpus simply includes all translated texts available in Czech.

4 Analysis

Even though our main analysis focusses on the potential effects of normalisation, convergence and cross-linguistic interference in translation (see Sections 4.1 and 4.2), we will first briefly examine the evolution of the frequency of the transgressive, in translated and non-translated texts. By doing so, we intend to verify the soundness of the limitation of the data for our analysis of the texts published after 1992 (inclusive). As mentioned above (§3.1), the sub-corpus of translations includes all the texts in the translated sub-corpus of InterCorp (limited to fiction and non-fiction), regardless of the date of publication or the source language. The non-translated texts are extracted from the reference corpus SYNv8.
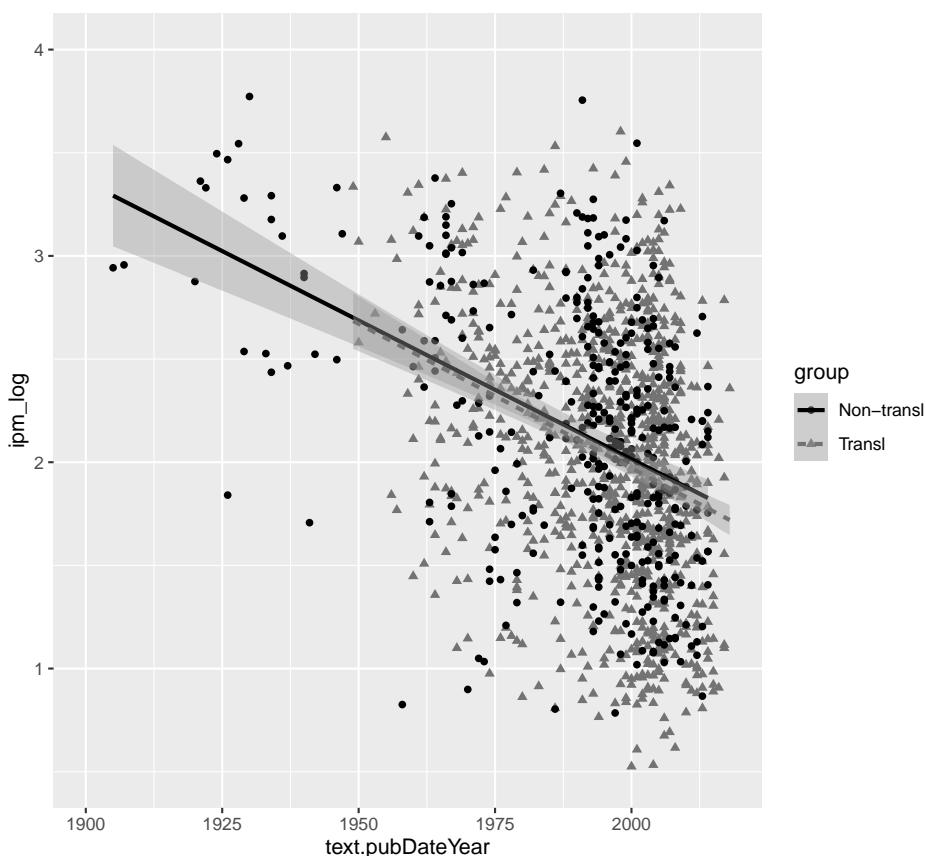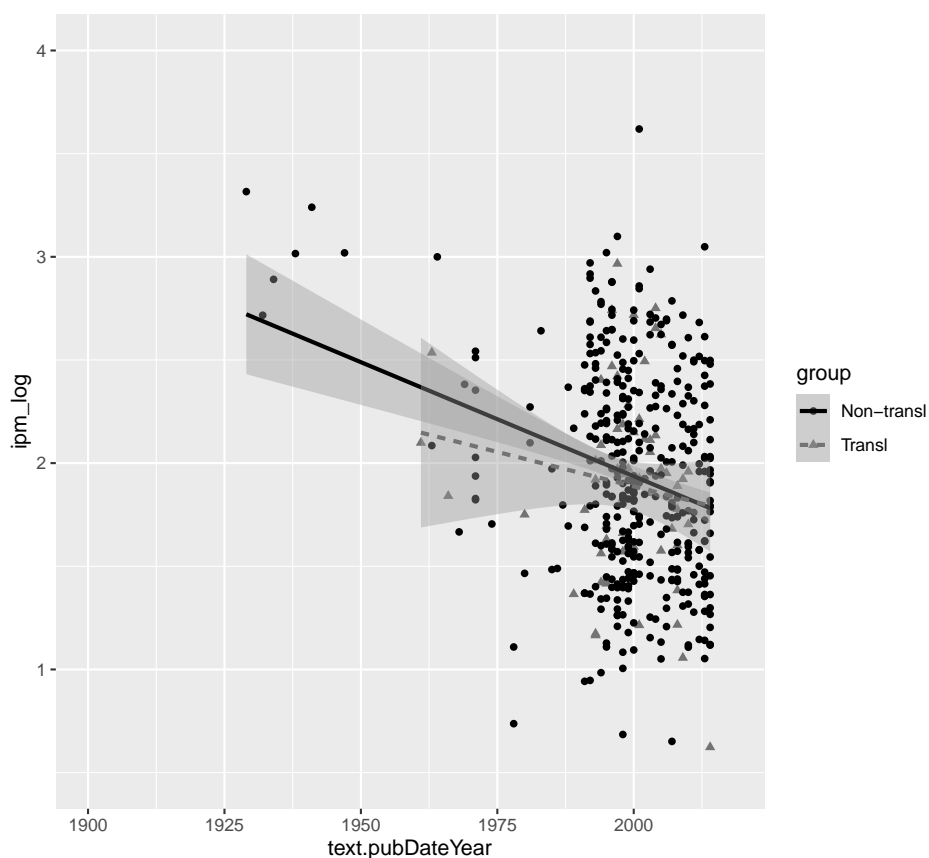

Figure 3.1: Normalised frequency of the transgressive in translated and non-translated fiction (InterCorp vs. SYNv8)

As shown in Figures 3.1 and 3.2, the time span for non-translated texts is larger than that of translations: the first texts in non-translated sub-corpora were published at the beginning of the 20th century in fiction (*Pověsti vyšehradské* by Popelka Biliánová, 1905) and in the 1920s in non-fiction. The first translations, by contrast, start in 1949 in fiction (Jorge Amado's novel *Suor*, a translation from Portuguese) and in non-fiction (*Wstęp do semantyki* by Adam Schaff, a Polish author) in 1963. Since the language of translations becomes obsolete faster than that of non-translated texts, this difference is expected.

Figure 3.2: Normalised frequency of the transgressive in translated and non-translated non-fiction (InterCorp vs. SYNv8)

Figures 3.1 and 3.2 fully confirm the tendency observed in previous research (see §2.1.3): the normalised frequency of the transgressive constantly decreases in both translated and non-translated texts. It is worth noting that the decrease

is more pronounced in fiction than in non-fiction and that the frequency of the transgressive is lower in non-fiction than in fiction. It is also necessary to point out that the actual decrease may be less dramatic than suggested by these figures, since the form of the graph is influenced by the few texts at the beginning of the observed period showing very high frequencies of the transgressive.

The data also indicate that the decrease in the frequency of the transgressive also continues after 1992, which suggests that there may be differences due to the time factor between texts within the time span of the limited corpora used in the main research. However, a further limitation of the corpus to more recent texts would have reduced the reliability of the results; hence we maintain the 1992 limit.

As for the differences between the translated and non-translated texts, Figures 3.1 and 3.2 suggest that they are only very slight, with a tendency to differentiation in the latest years in fiction and to a similarity in non-fiction. In what follows, we will investigate the statistical significance of these differences.

4.1 Normalisation and convergence in translation

Tables 3.4 and 3.5 show the absolute (n) and normalised (f) frequencies in instances per million words of the transgressive in translated and non-translated fiction and non-fiction, and the standard deviation (SD) and the coefficient of variation (CV, (SD/f)*100) for all the subcorpora. Even though the coefficient of variation is mostly higher in non-translations than in translations, with the exception of the fiction part of the Jerome corpus, the differences are very slight. This means that the convergence hypothesis (H_3, see §2.2) is not confirmed by our data, and with regard to the frequency of the transgressive, translations do not show more homogeneity than non-translated texts.

Table 3.4: Frequencies of the transgressive (both forms) in fiction (n = absolute frequency, f = normalised frequency in instances per million words, CV = coefficient of variation)

Fiction	corpus	texts	tokens	n	f	SD	CV
transl	Jerome	380	23,301,169	2,538	108.92	228.23	**209.54**
non-transl	Jerome	247	15,692,373	2,795	**178.11**	**368.50**	206.89
transl	InterCorp	774	71,063,940	9,268	130.42	262.94	201.61
non-transl	SYNv8	328	20,663,102	3,090	**149.54**	**343.43**	**229.66**

As for the normalisation hypothesis, Tables 3.4 and 3.5 show that the normalised frequency of the transgressive is indeed higher in non-translations than in translations, regardless of the corpus (Jerome translated/non-translated or InterCorp/SYNv8) and the text register (fiction or non-fiction). However, the differences in the normalised frequency of the transgressive are statistically significant in fiction only (p<.0001), as demonstrated in Figures 3.3(a)–3.3(d). The differences observed in non-fiction are not significant even at p<.05. This means that the normalisation hypothesis is confirmed in fiction, but not in non-fiction. From a methodological point of view, this result also indicates that the investigation of specific features of translation may be strongly text-type dependent.

The difference between the two text registers (fiction and non-fiction) regarding the tendency to normalisation may be due to various factors, especially because in fiction, translators are more likely to exploit the stylistic mark of the transgressive, which may increase their awareness about the pitfalls of this form and cause stylistic normalisation. Among the texts at the top of the frequency list of the transgressive in fiction are (in both translated and non-translated subcorpora) texts overtly exploiting the archaistic stylistic mark of the transgressive, in particular historical novels and fantasy stories (e.g. Andrzej Sapkowski's fantasy novel *Miecz przeznaczenia* tops the list of translations in InterCorp – 2,652.29 ipm). A similar motivation is found in translations of old source texts, first published in the 19th century. Even if they do not belong to the category of historical novels and the translations are recent, these texts show a high normalised frequency of the transgressive (e.g. Eça de Queiros' novel *A Cidade e as Serras*, 1,900.2 ipm, and two novels by Honoré de Balzac – *Le colonel Chabert* and *Gobseck*, 2026.87 ipm). Especially in non-translated texts at the top of frequency lists, the transgressive is used also in order to create a humoristic or ironic effect.[13] In some texts, the transgressive reflects a specific, syntactically complex style of the author of the source text, e.g. in the translation of the novel *Trans-Atlantyk* by the Polish author Witold Gombrowicz (2,196.12 ipm) and in a collection of short texts by the Belgian (French-writing) author Jean-Philippe Toussaint *Autoportrait* (1,817.00 ipm).

Nevertheless, a much more thorough analysis of the types of the use of transgressives (in translated as well as in non-translated texts) is needed. For instance,

---

[13]For instance, at the very top of the frequency list in non-translated fiction, we find a short text by Michal Šanda (*Obecní radní Stoklasné Lhoty vydraživší za 37 Kč vycpaného jezevce pro potřeby školního kabinetu* ['Municipal councillors of Stoklasná Lhota having auctioned a stuffed badger for 37 CZK for the use of the school science collection']), with 3,517.69 ipm of the normalised frequency of the transgressive.

Table 3.5: Frequency of the transgressive (both forms) in non-fiction (n = absolute frequency, f = normalised frequency in instances per million words, CV = coefficient of variation)
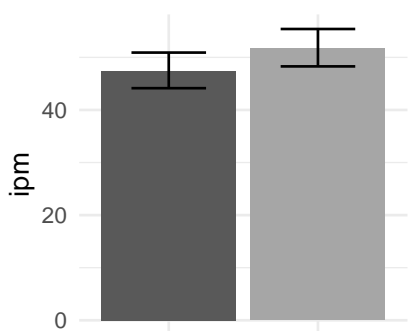
| Non-fiction	corpus	texts	tokens	n	f	SD	CV
transl	Jerome	221	15,904,500	754	47.41	113.73	239.89
non-transl	Jerome	242	15,719,462	813	**51.72**	**126.67**	**244.91**
transl	InterCorp	78	6,591,970	720	109.22	160.85	147.27
non-transl	SYNv8	592	30,988,911	3,447	**111.23**	**166.90**	150.05
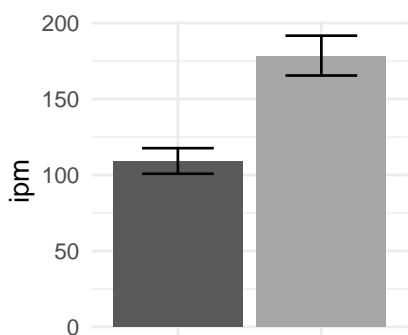
(a) Jerome translated vs. non-translated (fiction)
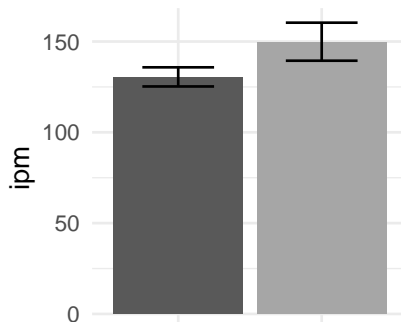
(b) InterCorp vs. SYNv8 (fiction)

(c) Jerome translated vs. non-translated (non-fiction)
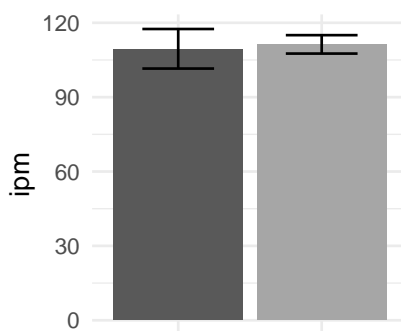
(d) InterCorp vs. SYNv8 (non-fiction)

various factors triggering the use of transgressives may combine in one text,[14] and in texts in the middle or at the bottom of the frequency list the use of the trangressive may be less motivated by its stylistic properties than by its use as a means of syntactic condensation. However, the aforementioned types extracted from the top of the frequency lists indicate that the frequency of the transgressive in fiction is probably closely related to the specific style of individual texts and authors.

In contrast, in non-fiction, not only is the overall frequency of the transgressive lower than in fiction (see Tables 3.4 and 3.5), but it appears more governed not by the specific stylistic norm of the text sub-type but by the individual style of texts and authors. Most texts containing transgressives in non-fiction sub-corpus belong to the domain of social sciences (both in translated and non-translated texts), especially philosophy and religious studies (Radim Palouš *Totalismus a holismus*, 756.93 ipm or *Cogitata metaphysica* by Benedict de Spinoza, 779.64 ipm), literary studies (e.g. Roland Barthes' *Mythologies*, 632.16 ipm) and history (e.g. *Každodennost renesančního aristokrata* by Marie Šedivá, 717.52 ipm or Ferdinand Seibt's *Deutschland und die Tschechen*, 552.14 ipm). In technical and natural science books, by contrast, the transgressives are much less frequent or even completely absent.[15]

It is important to point out that in non-fiction, the proportion of texts containing zero transgressives is higher than in fiction (one quarter of texts have no transgressive at all in the latter and one third in the former). More importantly for our topic, in both corpora (Jerome and InterCorp), more texts show zero transgressives in translations than in non-translated texts, and the maximum frequencies are higher in non-translated texts than in translations.[16]

Figures 3.3 and 3.4 show density plots of the normalised frequencies of the transgressive in the fiction part of InterCorp/SYNv8 (Figure 3.3) and the Jerome

---

[14]In the *Autoportrait*, for instance, the high frequency of the transgressive may be the result of a combination of complex syntax and irony in the source text (personal communication with Jovanka Šotolová, the Czech translator of the text). The age and personal style of the author (in non-translated texts) and the translator (in translations) may also come into play.

[15]This difference, already observed in previous studies (Dvořák 1983: 106 and 108, see §2.1.3), may also explain the difference in the normalised frequency of the transgressive in the non-fiction sub-corpora of Jerome on the one hand, and SYNv8 on the other hand: the former is a mix of various text register sub-types, whereas the latter contains more books from the domain of humanities.

[16]By contrast, in all the subcorpora, regardless of the text register or the translated/non-translated distinction, about a quarter of texts show the normalised frequency of the transgressive to be higher than the average of the whole sub-corpus (25% in SYNv8 fiction and 25% in all the other sub-corpora).

corpus (Figure 3.4) in translated and non-translated texts. It can be seen that in both corpora, even though the number of texts showing higher normalised frequencies of the transgressive is higher in non-translated texts than in translations, the differences are not extensive. Thus, the main difference between the translated and non-translated texts consists mainly in "category zero": the number of texts containing no transgressives at all is higher in translations than in non-translated texts. This is also the main cause of the normalisation effect in translations.
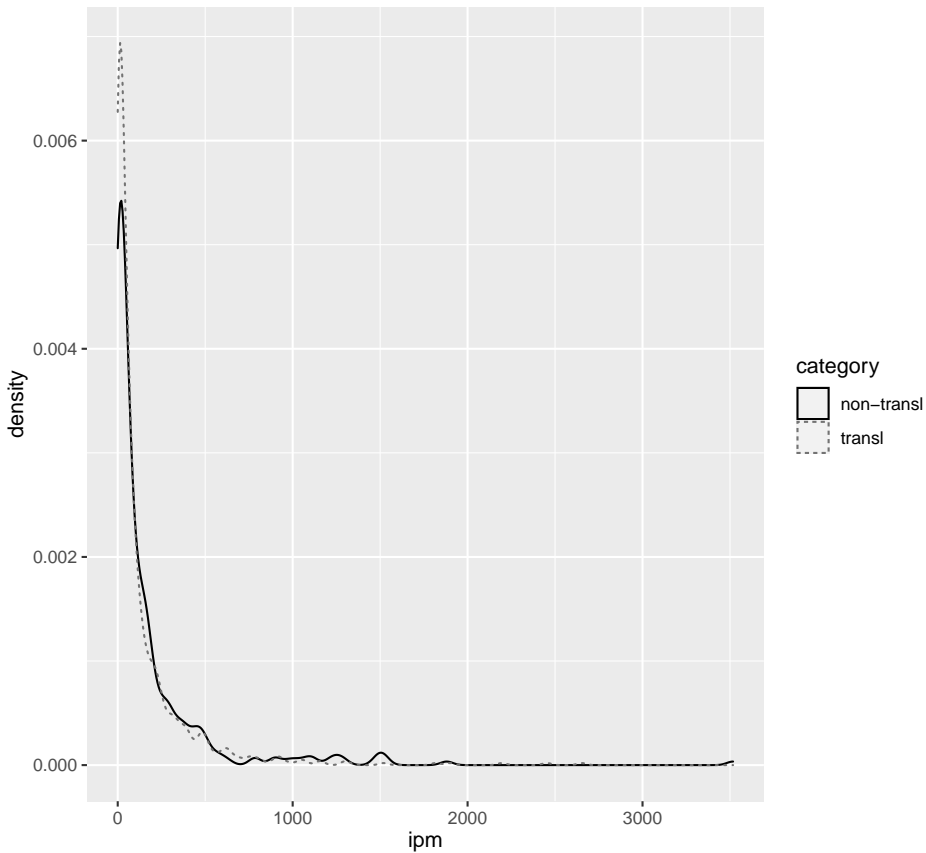

Figure 3.3: InterCorp/SYNv8 translated vs. non-translated density plot (fiction)

Figures 3.3 and 3.4 suggest that if translators decide to use transgressives, they do so in a way similar to non-translated texts. However, more translators than authors of original Czech texts decide not to use transgressives at all. In the Jerome
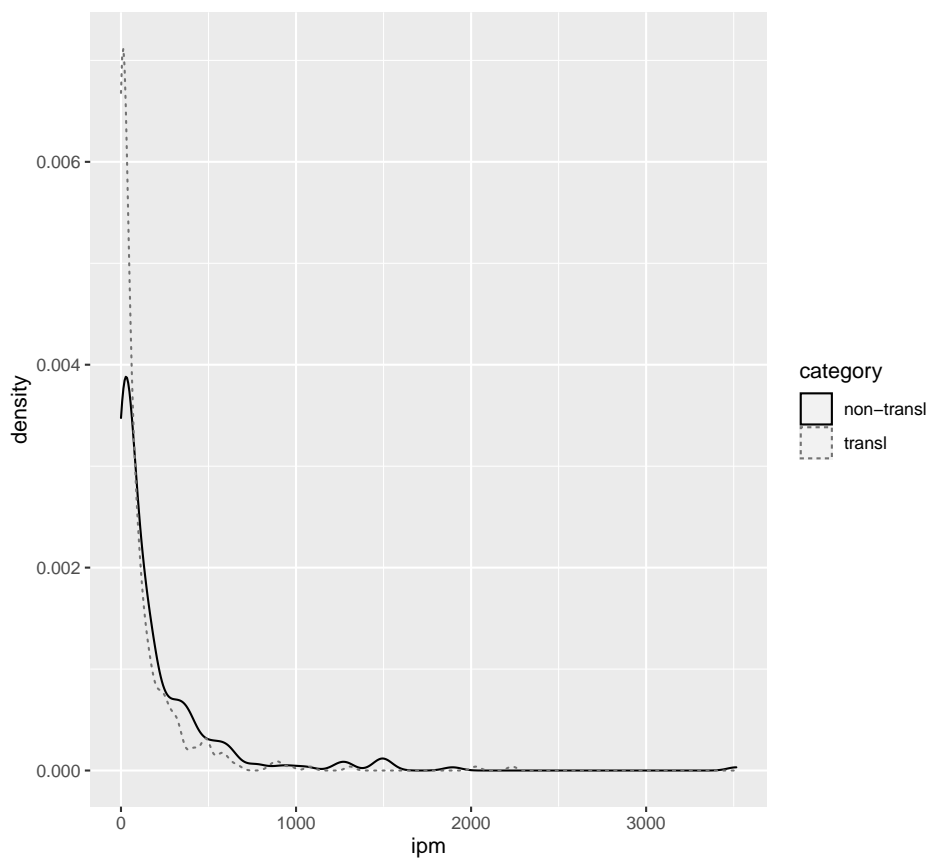
Figure 3.4: Jerome translated vs. non-translated density plot (fiction)

corpus, for instance, 31% of translations do not contain any transgressive, and 13% only one, i.e. 44% of texts have an extremely low frequency of the transgressive. In the non-translated texts, only 20% of texts show no transgressive and 11% only one occurrence, i.e. only 31% of texts without (or almost without) transgressives. These results suggest that translators could use more transgressives without being afraid to violate the norm of the target language (with respect to the style of the source text, of course).

It is also worth noting that in fiction, normalisation and convergence are more pronounced in the past transgressive forms (Conv.pt.pf) than in the present forms (Conv.ps.impf). As expected, the frequency of the past transgressive is much lower than that of the present form (Conv.pt.pf represents 6% of all transgressives in translations and 14% in non-translations, see Table 3.6). However, the

rate of the difference between translated and non-translated texts is higher in Conv.pt.pf than in Conv.ps.impf (3.24 and 1.52 respectively). The tendencies are similar in both text registers and both corpora (Jerome and InterCorp/SYNv8); therefore, we illustrate these with the numbers for the fiction part in the Jerome corpus only, in Table 3.6.

Table 3.6: Frequency of the transgressive (present and past form) in Jerome (fiction) (n = absolute frequency, f = normalised frequency in instances per million words, CV = coefficient of variation)

Jerome corpus	form	n	f	SD	CV
translated	Conv.ps.impf	2,376	108.92	101.97	93.62
non-translated	Conv.ps.impf	2,441	**155.55**	**282.66**	**181.72**
translated	Conv.pt.pf	162	6.95	39.74	**571.80**
non-translated	Conv.pt.pf	354	**22.56**	**118.03**	523.18

All the differences between translated and non-translated texts observed in Table 3.6 are statistically significant (p<.0001), and the comparison of Figures 3.5(a) and 3.5(b) demonstrates that the difference is more pronounced in Conv.pt.pf (Figure 3.5(b)) than in Conv.ps.impf (Figure 3.5(a)).
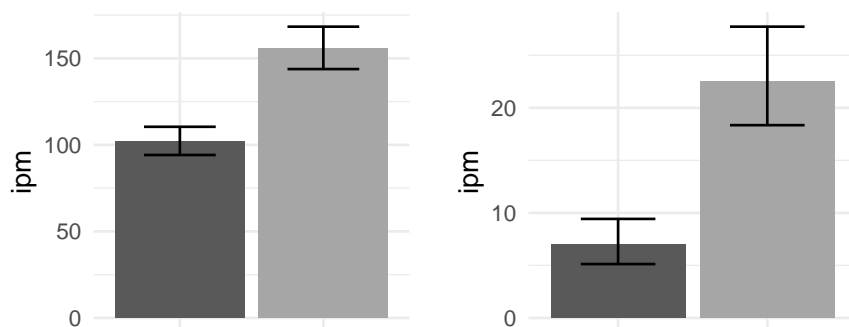
(a) Jerome translated vs. non-translated (fiction) frequency of Conv.ps.impf

(b) Jerome translated vs. non-translated (fiction) frequency of Conv.pt.pf

The greater tendency to normalisation of Conv.pt.pf is due to the more important stylistic mark of this form, in comparison with Conv.ps.impf. We can recall that Conv.ps.impf is considered bookish, whereas the Conv.pt.pf is assigned an archaistic stylistic mark. Since translators normalise, it is natural that they tend to avoid the form manifesting a stronger stylistic mark.

4.2 Cross-linguistic interference in translations

Since normalisation is considered a universal phenomenon, in §4.1 we analysed its potential effect in translations for all the source languages together. Conversely, cross-linguistic interference is intrinsically related to the linguistic properties of the different source languages. Concerning converbs, hypothesis H_1 expects more transgressives in translations from Romance and Slavic languages than in translations from Germanic languages. As stated in §3.2, we conducted this analysis on the fiction part of the InterCorp parallel corpus (texts published after 1992 including), which contains a larger number of texts than the Jerome comparable corpus and the non-fiction sub-corpus of InterCorp.

Table 3.7 shows the absolute and the normalised frequencies of the transgressive (both forms together) in translations from the 20 source languages available

Table 3.7: Frequency of the transgressive (present and past form) in different source language sub-corpora of InterCorp (fiction)

Rank	src.lang	positions (n)	texts (n)	abs.fq.	ipm
1	pl	2,436,840	35	891	365.64
2	pt	1,250,080	16	398	318.38
3	sr	366,940	6	108	294.33
4	ro	372,404	5	95	255.10
5	es	8,393,499	101	1,762	209.92
6	fr	5,009,729	73	988	197.22
7	hr	1,242,178	19	209	168.25
8	sk	994,572	16	165	165.90
9	de	8,920,552	91	1,154	129.36
10	ru	1,306,704	11	154	117.85
11	en	25,810,495	226	2,597	100.62
12	it	1,044,540	14	103	98.61
13	fi	1,355,134	23	124	91.50
14	nl	1,657,687	23	151	91.09
15	lv	228,997	5	17	74.24
16	sl	835,792	11	37	44.27
17	da	1,023,334	9	44	43.00
18	sv	6,604,972	69	207	31.34
19	no	1,498,553	16	46	30.70
20	ja	710,938	5	18	25.32
	total	71,063,940	774	9,268	130.42

in the sub-corpus of fiction translated into Czech in the InterCorp parallel corpus. At first sight, the results confirm the H_1, since Slavic and Romance source languages are grouped at the top of the frequency list (except for Italian in Romance and Slovenian in Slavic), whereas the Germanic languages are found mostly in the lower part of the table (except for German, which is ranked 9 in the table). English, considered exceptional among the other Germanic languages, is found in the middle of the list. It is important to note that only eight source language sub-corpora show a normalised frequency higher than 140.54 ipm, i.e. the frequency in the reference non-translated fiction corpus in SYNv8 (see Table 3.4). This confirms the tendency to normalisation observed in §4.1.

However, upon closer examination, the results introduced in Table 3.7 appear much less reliable. For instance, it is true that within the group of Romance languages, the lower frequency of the transgressive in translations from Italian may be explained by the lower frequency of the Italian converb (*gerundio*, see Čermák et al. 2020) in comparison with Portuguese and Spanish (ranking second and fifth). The French *gérondif*, however, is even less frequent than the Italian *gerundio* (ibid.), but translations from French rank 6th, just after Spanish. This brief observation reveals the first methodological pitfall of the analysis of the potential effect of cross-linguistic interference based only on frequencies: without understanding the *valeur* of the converb in the system of the source language and without a detailed analysis of parallel concordances in the individual language pairs, all the cross-linguistic observations are potentially unreliable.

Similarly, a closer look at the group of Slavic languages reveals other discrepancies of the purely frequential approach to cross-linguistic interference. Polish, for instance, using its two converb forms extensively, is likely to be found at the top of the list, which is the case in Table 3.6. However, the position of Russian in Table 3.6 is surprising: even though its converb is considered prototypical (see §2.1.1) and its two converb forms are well attested, Russian only ranks 10th, even after Slovak, making only very limited use of its converb (see §2.1.1 and Brtková (2004: 25)). By its ranking, Russian is placed even below German, which is considered to make only "parsimonious" use of converbs (see §2.1.1 and König (1995: 72)). Similarly, polyconverb Finnish, Latvian and Japanese surprisingly only rank 13th, 15th and even 20th.

The reliability of the results for the different language sub-corpora introduced in Table 3.7 is undermined by the same (external) factors as in the analysis of normalisation: the frequency of the transgressive may be influenced by the specific style and topic of the text, by the individual preferences of the translators, and even by the date of publication of the source text. Moreover, since the corpus is divided into 20 sub-corpora, the risk of systematic bias is higher than in the

normalisation testing. For example, in the small subcorpus of translations from Slovak, we find two fantasy novels showing very high frequencies of the transgressive, which may influence the results for the whole sub-corpus, containing only 16 texts. Similarly, the normalised frequency of the transgressive in translations from Portuguese is skewed by one translation of a text first published in the 19th century (Eça de Queiros´ novel *A Cidade e as Serras*) and showing the normalised frequency of the transgressive more than 12 times higher than in the reference corpus SYNv8. In the subcorpus of translations from Romanian, it is not possible to say whether the sub-corpus reflects cross-linguistic interference or the personal preferences of the translator because all the five texts in this sub-corpus were translated by the same translator (Jiří Našinec).

Figure 3.5 summarises the tendencies in the frequency of the transgressive and the limitations of the reliability of the data extracted from our corpus (the confidence intervals).

We can see that for Danish, Japanese, Norwegian and Serbian, the data extracted from our corpus are not reliable. The rest of the data confirm the tendencies observed in Table 3.7, i.e. a higher frequency of transgressives in translations from Slavic and Romance languages (except for Slovenian, and partly Slovak and Russian) and a lower frequency in translations from Germanic languages.

Nevertheless, the analysis of the potential effect of the cross-linguistic interference between the converb in the source language and the Czech transgressive necessitates a thorough contrastive examination of individual language pairs. Subsequently, there needs to be a detailed analysis of the occurrences in parallel concordances, which takes into account the linguistic factors of the use of the transgressive (and its counterpart(s) in the source language), and the potential influence of the style of the text, the translators' idiolects and other factors.

5 Conclusion

The Czech transgressive is a specific case of the cross-linguistic category of converb. On the one hand, it shows most properties of the prototypical converbs: it is strict, has two forms (present and past transgressive), is referentially same-subject (i.e. coreferential with the controller of the main clause) and, as with most European converbs, its semantic interpretation is contextual (with the prevailing meaning of accompanying circumstance). On the other hand, it has an archaistic morphology, requiring agreement with the controller in number and gender and a strong stylistic mark: bookish for the present transgressive and archaistic for the past transgressive. Because of this stylistic mark, the transgressive is used rarely in contemporary language, and only in written texts.
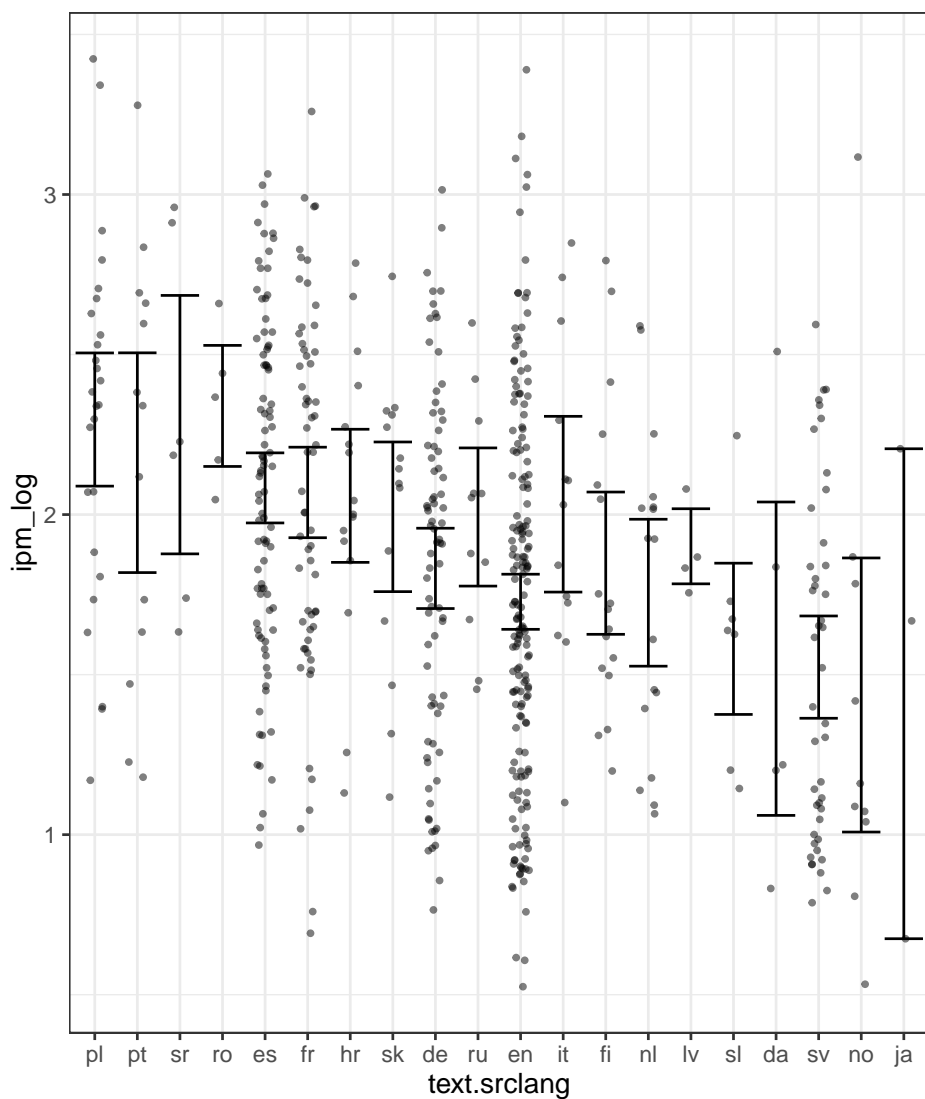
Figure 3.5: Normalised frequency of the transgressive in 20 different source language sub-corpora in the InterCorp corpus (fiction)

In this study, we investigated the potential impact of these double-face characteristics of the Czech converb on translations of fiction and non-fiction in Czech. Our preliminary frequential analysis confirmed the constant decrease in the frequency of the transgressives in both text registers and both translated and

non-translated texts during the 20th and 21st centuries. This observation also justified the limitation of our corpora to texts published after 1992 (inclusive). In line with expectations, the frequency analysis revealed the strong dominance of the present transgressive over the past form, which corroborates the diachronic trend suggested for Czech by Nedjalkov (1995); Czech appears to be moving from a bi-converbal language to a mono-converbal one.

The main findings of our study are the confirmation of the normalisation effect in translations of fiction (but not in those of non-fiction), the absence of convergence in translations in comparison with non-translated texts, and the necessity of a thorough contrastive analysis of converbs before investigating the potential effect of the cross-linguistic interference.

As for the normalisation, the difference in the frequency of the transgressive between translated and non-translated fiction is not extensive but is statistically significant. Of greater interest, a detailed analysis of the distribution of the frequencies revealed that this normalisation effect is caused especially by the number of texts using zero transgressives: in translations 31%, in non-translated texts only 20% of the texts. This means that more translators decided to avoid transgressives than the authors of the original texts. Finally, the normalisation impact is stronger in the past transgressive, showing a stronger stylistic mark, than in the present transgressive. These results suggest that if translators decided to use more transgressives – with respect to the style of the source text, of course – they would not violate the norm of the target language.

In non-fiction, the effect of normalisation was not observed. This text-register difference may be explained either by the overall lower frequency of the transgressive in non-fiction than in fiction or precisely by the stylistic mark of the transgressive. In fiction, the authors and translators appear to exploit this characteristic of the transgressive, e.g. the use as a means of irony or parody (mainly in non-translated texts), as the reflection of a specific, very complex style and syntax of the source text in translations or to create the archaistic effect in historical novels or in fantasy stories. This last use was also observed in translations of source texts first published in the 19th century, even if the actual translation was recent. In non-fiction, the use of the transgressive appears to be governed not by the individual style of the text or the author, but by the norms of the text register sub-types. In line with observations in previous studies, the transgressive is more frequent in humanities (philosophy, history, literary studies, etc.) than in natural and technical sciences. Nevertheless, all these observations require a more thorough analysis of individual texts and concrete occurrences of transgressives in context.

Pertaining to the convergence hypothesis, based on the analysis of the coefficient of variation, it was observed neither in fiction nor in non-fiction. This means that translations are as heterogeneous in the frequency of the transgressive as non-translated texts. However, both in translations and in non-translated texts, the coefficient of variation is higher in the past form of the transgressive, considered archaistic, than in the present form, considered only bookish. This result indicates that the effect of convergence may vary according to the stylistic mark of the linguistic feature under investigation.

The results for the cross-linguistic hypothesis are the least conclusive. The comparison of the normalised frequency of the transgressive in twenty source language subcorpora showed a higher frequency of transgressives in translations from Slavic and Romance languages, where the converbs are considered prototypical, and a lower frequency in translations from Germanic languages, supposedly to make very limited use of converbs (except for English). However, several partial results were not consistent with the hypotheses. In the Slavic languages, for instance, translations from Slovak show a higher frequency of transgressives than translations from Russian, although converbs in Slovak are rare but abundant in Russian. Similarly, translations from French contain more transgressives than those from Italian despite the much lower frequency of the French *gérondif* than the Italian *gerundio*.

These inconsistencies reveal two important pitfalls of the purely frequential analysis of the cross-linguistic interference effect in translations. First, since the use of the transgressive is intrinsically linked to its stylistic mark, the results are extremely sensitive to the composition of the different source language subcorpora and the style of the texts they contain. Second, and more importantly, these results reveal the necessity of a thorough contrastive analysis of the different language pairs, taking into account the frequency and the *valeur* of the different converbs in the language systems, and their specific uses in context.

Future research may provide not only a more fine-grained contrastive analysis of converbs in different language pairs but also a deeper understanding of the motivations of the normalisation and convergence in translation and various factors coming into play in the process of translation and the translation workflow. It is worth investigating, for instance, the potential effect of the translator's proficiency (do experienced translators use the transgressive more than translators in the early stage of their career? What is the role of translators' training in their attitude to the transgressive? cf. Lapshinova-Koltunski 2018), the sex of the translator (preliminary results indicate female translators use transgressives less than their male colleagues; see the impact of the gender factor in Magnifico & Defrancq 2018), the target audience (is there a difference between translated and

non-translated literature intended for children and young readers, with regard to the use of transgressives? cf. e.g. Čermáková 2017), and the attitude of text revisers in publishing houses to the transgressive and the impact of their interventions on its frequency in (translated as well as non-translated) texts (see also Bisiada 2017; 2018; 2019; Kruger 2018). Only this complex approach may help to fully conceive of translation as a socially contexted behaviour and understand the norms to which the translator is supposed to adhere to.

Acknowledgements

References

Alpatov, Vladimir M. & Vera Podlesskaya. 1995. Converbs in Japanese. In Martin Haspelmath & Ekkehard König (eds.), *Converbs in cross-linguistic perspective: Structure and meaning of adverbial verb forms – adverbial participles, gerunds*, 465–487. Berlin: Mouton de Gruyter.

Baker, Mona. 1993. Corpus linguistics and translation studies: Implications and applications. In Mona Baker, Gill Francis & Elena Tognini-Bonelli (eds.), *Text and technology: In honour of John Sinclair*, 233–250. Amsterdam: John Benjamins.

Baker, Mona. 1996. Corpus-based translation studies: The challenges that lie ahead. In Harold Somers (ed.), *Terminology, LSP and translation: Studies in language engineering in honour of Juan C. Sager*, 175–186. Amsterdam: John Benjamins.

Bečka, Josef Václav. 1977. Přechodníky jako překladatelský problém [Transgressives as a translation problem]. *Zprávy Kruhu přátel českého jazyka* 5. 17–27.

Bisiada, Mario. 2017. Universals of editing and translation. In Silvia Hansen-Schirra, Oliver Czulo & Sascha Hofmann (eds.), *Empirical modelling of translation and interpreting*, 241–275. Berlin: Language Science Press.

Bisiada, Mario. 2018. The editor's invisibility: Analysing editorial intervention in translation. *Target* 30(2). 288–309.

Bisiada, Mario. 2019. Translated language or edited language? A study of passive constructions in translation manuscripts and their published versions. *Across Languages and Cultures* 20(1). 35–56.

Brtková, Kamila. 2004. K otázke prekladu ruských prechodníkov do slovenčiny [On translation of Russian transgressives into Slovak]. *Opera Slavica* XIV(2). 20–31.

Cappelle, Bert. 2012. English is less rich in manner-of-motion verbs when translated from French. *Across Languages and Cultures* 13(2). 173–195.

Čechová, Marie, Jan Chloupek & Eva Minářová. 1997. *Stylistika současné češtiny [Stylistics of contemporary Czech]*. Prague: ISV – nakladatelství.

Čermák, František & Alexandr Rosen. 2012. The case of InterCorp, a multilingual parallel corpus. *International Journal of Corpus Linguistics* 17(3). 411–427.

Čermák, Petr, Dana Kratochvílová, Olga Nádvorníková & Pavel Štichauer. 2020. *Complex words, causatives, verbal periphrases and the gerund: Romance languages versus Czech (a parallel corpus-based study)*. Praha: Karolinum.

Čermák, Petr, Olga Nádvorníková, Leontýna Bratánková, Štěpánka Černikovská, Jan Hricsina, Jiří Jančík, Jaroslava Jindrová, Dana Kratochvílová, Zuzana Krinková, Petra Laufková, Daniel Petrík, Eliška Třísková & Pavel Štichauer. 2015. *Románské jazyky a čeština ve světle paralelních korpusů [Romance languages and Czech in the light of parallel corpora]*. Praha: Karolinum.

Čermáková, Anna. 2017. Translating children's literature: Some insights from corpus stylistics. *Ilha do Desterro – A Journal of English Language, Literatures in English and Cultural Studies* 71(1). 117–134.

Chesterman, Andrew. 2004. Beyond the particular. In Anna Mauranen & Pekka Kujamäki (eds.), *Translation universals – do they exist?*, 33–49. Amsterdam: John Benjamins.

Chlumská, Lucie. 2013. *Jerome corpus – comparable corpus of translated and non-translated Czech*. Praha: Institute of the Czech National Corpus. https://kontext.korpus.cz.

Chlumská, Lucie. 2017. *Překladová čeština a její charakteristiky [Translated Czech and its features]*. Praha: Nakladatelství Lidové noviny.

Cvrček, Václav. 2010. *Mluvnice současné češtiny 1 [Grammar of contemporary Czech]*. Praha: Karolinum.

Olga Nádvorníková

Cvrček, Václav & Lucie Chlumská. 2015. Simplification in translated Czech: A new approach to type-token ratio. *Russian Linguistics* 39(3). 309–325.

Cvrček, Václav & Dominika Kováříková. 2011. Možnosti a meze korpusové lingvistiky [Opportunities and limitations of corpus linguistics]. *Naše řeč* 94(3). 113–133.

Dai, Guangrong & Richard Xiao. 2011. SL "shining through" in translational language: A corpus-based study of Chinese translation of English passive. *Translation Quarterly* 62. 85–108.

Delaere, Isabelle, Gert de Sutter & Koen Plevoets. 2012. Is translated language more standardized than non-translated language? Using profile-based correspondence analysis for measuring linguistic distances between language varieties. *Target* 24(2). 203–224.

Dvonč, Ladislav, Gejza Horák, František Miko, Jozef Mistrík, Ján Oravec, Jozef Ružička & Milan Urbančok. 1966. *Morfológia slovenského jazyka [Morphology of the Slovak language]*. Bratislava: Vydavtelstvo Slovenskej akadémie vied.

Dvořák, Emil. 1970. *Vývoj přechodníkových konstrukcí ve starší češtině [Evolution of transgressive constructions in old Czech]*. Praha: Universita Karlova.

Dvořák, Emil. 1972. Přechodníkové konstrukce v překladech beletrie do češtiny [Transgressive constructions in translation of fiction in Czech]. *Slavica Pragensia* 14(2–4). 101–114.

Dvořák, Emil. 1983. *Přechodníkové konstrukce v nové češtině [transgressive constructions in modern Czech]*. Praha: Univerzita Karlova.

Grevisse, Maurice & André Goosse. 2016. *Le bon usage*. Louvain-la-Neuve: De Boeck Supérieur.

Halmøy, Odile. 2003. *Le gérondif en français*. Paris: Ophrys.

Haspelmath, Martin. 1995. The converb as a cross-linguistically valid category. In Martin Haspelmath & Ekkehard König (eds.), *Converbs in cross-linguistic perspective: Structure and meaning of adverbial verb forms – adverbial participles, gerunds*, 1–57. Berlin: Mouton de Gruyter.

Hnátková, Milena, Michal Křen, Pavel Procházka & Hana Skoumalová. 2014. The SYN-series corpora of written Czech. 160–164.

Hnátková, Milena, Vladimír Petkevič & Hana Skoumalová. 2011. Linguistic annotation of corpora in the Czech National Corpus. In Viktor Zacharov (ed.), *Proceedings of the international conference "Corpus Linguistics – 2011")*, 15–20. St. Petersburg: St. Petersburg State University.

Jelínek, Jaroslav, Josef Václav Bečka & Marie Těšitelová. 1961. *Frekvence slov, slovních druhů a tvarů v českém jazyce [Frequency of words, word classes and word forms in Czech]*. Praha: Státní pedagogické nakladatelství.

Karlík, Petr, Marek Nekula & Zdenka Rusínová. 1995. *Příruční mluvnice češtiny [Reference grammar of Czech]*. Praha: Nakladatelství Lidové noviny.

Karlík, Petr. 2017. *Přechodník [Transgressive]*. https : / / www . czechency . org / slovnik/P%C5%98ECHODN%C3%8DK (12 August, 2020).

Kleiber, Georges. 2007. La question temporelle du gérondif: Simultanéité ou non ? In Frédéric Lambert, Catherine Moreau & Jean Albrespit (eds.), *Les formes non finies du verbe 2*, 109–123. Rennes: Presses Universitaires de Rennes.

Kleiber, Georges. 2009. Gérondif et relations de cohérence: Le cas de la relation de cause. In Elena Comes & Florica Hrubaru (eds.), *Relations de discours II – actes du XVe séminaire de didactique universitaire*, 9–24. Cluj: Editura Echinox.

Kocková, Jana. 2011. Ekvivalenty ruských přechodníků sloves dokonavého vidu v češtině na základě paralelních korpusů [Equivalents of Russian perfective transgressives in Czech: A corpus-based study]. In František Čermák (ed.), *Korpusová lingvistika praha*, 251–261. Praha: NLN.

Komárek, Miroslav. 1986. *Mluvnice češtiny 2 [Grammar of Czech]*. Praha: Academia.

König, Ekkehard. 1995. The meaning of converb constructions. In Martin Haspelmath & Ekkehard König (eds.), *Converbs in cross-linguistic perspective: Structure and meaning of adverbial verb forms – adverbial participles, gerunds*, 57–97. Berlin: Mouton de Gruyter.

König, Ekkehard & Johan van der Auwera. 1990. Adverbial participles, gerunds and absolute constructions in the languages of Europe. In Johannes Bechert, Giuliano Bernini & Claude Buridant (eds.), *Toward a typology of European languages*, 57–95. Berlin: Mouton de Gruyter.

Kortmann, Bernd. 1997. *Adverbial subordination: A typology and history of adverbial subordinators based on European languages*. Berlin: Mouton de Gruyter.

Křen, Michal, Václav Cvrček, Tomáš Čapka, Anna Čermáková, Milena Hnátková, Lucie Chlumská, Tomáš Jelínek, Dominika Kováříková, Vladimír Petkevič, Pavel Procházka, Hana Skoumalová, Michal Škrabal, Petr Truneček, Pavel Vondřička & Adrian Zasina. 2019. *Korpus SYN, version 8 from 12/12/2019*. Praha: Institute of the Czech National Corpus. https://kontext.korpus.cz/.

Kruger, Haidee. 2018. The effects of editorial intervention: Implications for studies of the features of translated language. In Gert De Sutter, Marie-Aude Lefer & Isabelle Delaere (eds.), *Empirical translation studies*, 113–157. Berlin: Mouton de Gruyter.

Lapshinova-Koltunski, Ekaterina. 2015. Variation in translation: Evidence from corpora. In Claudio Fantiunoli & Federico Zanettin (eds.), *New directions in corpus-based translation studies*, 93–114. Berlin: Language Science Press.

Lapshinova-Koltunski, Ekaterina. 2018. Text classification for detection of translationese in novice and professional translations. In Sylviane Granger, Marie-Aude Lefer & Laura Aguiar de Souza Penha Marion (eds.), *Book of abstracts: Using corpora in contrastive and translation studies conference*, 108–110. Louvain: CECL/Université catholique de Louvain.

Laviosa, Sara. 2002. *Corpus-based translation studies: Theory, findings, applications*. Amsterdam: Rodopi.

Lefer, Marie-Aude & Svetlana Vogeleer. 2013. Interference and normalization in genre-controlled multilingual corpora. *Belgian Journal of Linguistics* 27. 1–21.

Levý, Jiří. 2011. *The art of translation*. Amsterdam: John Benjamins.

Magnifico, Cédric & Bart Defrancq. 2018. Norms and gender in simultaneous interpreting: A study of self-repairs. In Sylviane Granger, Marie-Aude Lefer & Laura Aguiar de Souza Penha Marion (eds.), *Book of abstracts: Using corpora in contrastive and translation studies conference*, 119–120. Louvain: CECL/Université catholique de Louvain.

Malá, Markéta & Pavlína Šaldová. 2015. English non-finite participial clauses as seen through their Czech counterparts. *Nordic Journal of English Studies* 1. 232–257.

Michálková, Věra. 1963. K polovětným konstrukcím v nářečí [On semi-predicative constructions in dialects]. *Naše řeč* 46(3). 136–145.

Moortgat, Bernadette. 1978. *Participe et gérondif: Étude de l'opposition entre la présence et l'absence de EN devant la forme en -ant*. Université de Metz: Thèse de 3e cycle.

Nádvorníková, Olga. 2016. Le corpus multilingue InterCorp et les possibilités de son exploitation. In Éva Buchi, Jean-Paul Chauveau & Jean-Marie Pierrel (eds.), *Actes du XXVIIe congrès international de linguistique et de philologie romanes (Nancy, 15–20 juillet 2013)*, 223–237. Strasbourg: Société de linguistique romane/ÉLiPhi.

Nádvorníková, Olga. 2017. Parallel corpus in translation studies: Analysis of shifts in the segmentation of sentences in the Czech-English-French part of the InterCorp parallel corpus. In Joseph Emonds & Markéta Janebová (eds.), *Language use and linguistic structure*, 445–461. Olomouc: Palacký University Olomouc.

Nádvorníková, Olga. 2012. *Korpusová analýza faktorů sémantické interpretace francouzského gérondivu [A corpus-based analysis of factors influencing semantic interpretation of the French gérondif]*. Univerzita Karlova v Praze. (Doctoral dissertation).

Nedjalkov, Igor V. 1995. Some typological parameters of converbs. In Martin Haspelmath & Ekkehard König (eds.), *Converbs in cross-linguistic perspective:*

Structure and meaning of adverbial verb forms – adverbial participles, gerunds, 97–137. Berlin: Mouton de Gruyter.

Nedjalkov, Igor V. 1998. Converbs in the languages of Europe. In Johan van der Auwera (ed.), *Adverbial constructions in the languages of Europe*, 420–455. Berlin: Mouton de Gruyter.

Nosek, Jiří. 1964. Notes on syntactic condensation in modern English. *Travaux linguistiques de Prague* 1. 281–288.

Oktábec, Vojtěch. 1953. O překládání ruských přechodníkových vazeb [On translation of Russian transgressive constructions]. In Jaroslav Moravec (ed.), *Kniha o překládání: Příspěvky k otázkám překladu z ruštiny*, 258–279. Praha: Nakladatelství Československo-sovětského institutu.

Rosen, Alexandr, Martin Vavřín & Adrian Zasina. 2019. *InterCorp multilingual corpus – Czech, version 12 from 12/12/2019*. Praha: Institute of the Czech National Corpus. https://kontext.korpus.cz/.

Teich, Elke. 2003. *Cross-linguistic variation in system and text: A methodology for the investigation of translations and comparable texts*. Berlin: Mouton de Gruyter.

Toury, Gideon. 1995. *Descriptive translation studies and beyond*. Amsterdam: John Benjamins.

Vachek, Josef. 1955. Some thoughts on the so-called complex condensation in modern English. *Sborník prací Filozofické fakulty Brněnské univerzity* 4. 63–77.