
**ENGLISH *IT*-CLEFTS WITH FOCUSED SUBJECT
FROM THE PERSPECTIVE OF FSP
AND THEIR CZECH TRANSLATION EQUIVALENTS**

ANNA KUDRNOVÁ

ABSTRACT

The paper offers a brief survey of relevant theory on English *it*-clefts, with an emphasis on information structure. In the practical part, it compares *it*-clefts with focused subject in original English written texts and their Czech translation counterparts, based on a sample research in the parallel multilingual corpus InterCorp. Due to the different means of expressing FSP in the two languages, Czech naturally uses other ways to convey the desired meaning apart from the cleft construction, such as word-order, focalizing particles, and other devices. To get a better idea about the extent to which the translation may be influenced by the surface structure of the original, another additional sample research was carried out using the same queries to find English translations containing *it*-clefts and their original Czech source sentences. The preliminary research suggests that at least in the chosen area, the source language does not seem to significantly interfere in the surface structure of the resultant translation.

Keywords: syntax, *it*-cleft, functional sentence perspective, translation

1. Introduction

The objective of this paper is to examine English *it*-clefts, in particular those with focused subject, in the light of their Czech equivalents, using a parallel corpus, and mainly with regard to information structure. The first part presents the theoretical issues concerning *it*-clefts and related FSP aspects, while the second half concentrates on the research: its methods, advantages and problems, and puts forward some preliminary findings. Beside the most obvious and necessary line of research, which involves finding Czech translation equivalents of English *it*-clefts, other supplementary strands of search have been executed (or are to be executed), in order to gain more insight into some translation-related issues, such as interference from the source language into the target language.

2. Motivation

There are several reasons that make a comparative, corpus-based approach a convenient method. First, as regards FSP, the *it*-cleft has one advantage: it has a relatively fixed, recognizable syntactic structure, thus it can be searched for in the corpus; whereas in most “unmarked” constructions the information structure is expressed by a combination of means, usually not overtly identifiable. Researchers are therefore often confined to manual search and comparison. The surface structure of *it*-clefts, on the other hand, enables more extensive search and a greater variety of results. Inevitably, the corpus method has its disadvantages, too, which will be discussed in more detail in the second part of the paper.

As regards the comparative approach, with two languages as typologically different as English and Czech (analytical and inflectional, respectively), translation can serve as a useful tool; rendering a sentence into another language with different FSP means often helps to see the information structure more clearly. At the same time it needs to be emphasized that the equivalents found in a parallel corpus are translations proper rather than linguistic tools, with all that this entails. In other words, the choices made in the translation may be influenced by other factors than the form of the original structure that is being analysed (for example by the tendency to avoid repeating words/constructions, maintain the cohesion of the resulting text, or other translation-related factors).

3. Theoretical part

3.1 *It*-clefts in English

The cleft construction in English is a fairly common syntactical means, one of whose primary functions is to highlight the element that carries the information focus. Since English word order is for the most part fixed, which limits the operational scope of the FSP factor of linearity, information structure must sometimes be made explicit by this secondary construction, which puts the most dynamic element in focus. Nevertheless, the *it*-cleft, like many other linguistic means, is not restrained to a single function; it often has a textual role, such as introducing a new topic, summarizing etc. (cf. Dušková, 2010: 42). In order to determine the information structure of a given *it*-cleft, it is always necessary to consider other FSP factors, too, most importantly the context (Greenbaum and Quirk, 1990: 412)

As regards the information structure of the two constituent clauses of the cleft-sentence, Prince (1978) distinguishes two types of *it*-clefts; the dependent clause of the first type contains “old” or “knowable” information, which can sometimes even be omitted. The second type, called informative-presupposition *it*-clefts, on the other hand, actually presents new information in the dependent clause (Prince, 1978: 896–898). As a result, either of the constituent clauses may contain the rheme.

3.2 Cleft sentences in Czech

Czech as an inflectional language with free word-order relies heavily on linearity to indicate FSP structure, much more than English. It follows that the occurrence of cleft constructions will be markedly lower in Czech. However, the difference is not absolute and cleft constructions are indeed used in Czech as well, although they are less frequent and likely to be perceived as more marked. Their status in the language system seems to be slightly peripheral, as they have been seen by some linguists as a syntactical borrowing from other languages, such as English or French, and its excessive and unjustified use has been frowned upon as a mark of stylistic awkwardness (Filipec, 1955). What is interesting is that in spite of the lower frequency of Czech *it*-clefts, both the types identified by Prince (stressed-focus and informative presupposition) occur; the latter type serves mainly as a stylistic device expressing attitude or involvement, or is used to emphasize a repeated structure, even though the rheme is in the dependent clause (Filipec, 1955).

3.3 Comparative perspective: advantages and challenges

Leaving purely abstract linguistic aspects aside – what are the grounds on which translation equivalents are decided on? We may quite safely assume that translators, given the amount of time that they spend close-reading the text, do in most cases determine the FSP structure of the original correctly. What is more difficult to answer is the question how in particular they choose from the several options that the target language offers, and whether the original syntactic structure affects this decision; the factors involved in the choice are numerous. In a simplified perspective: there will probably be instances in which the translator simply copies the syntactic structure of the source without considering other options. On the other hand, some translators who are aware of the fact that *it*-clefts are much rarer in Czech than in English may avoid the construction altogether in an effort to prevent undesirable interference. The tendency to opt for the more conventional or more central solutions in the target language is by some theoreticians considered to be one of the translation universals (cf. for example the discussion in Chlumská, 2015: 33). We are thus faced with two opposing trends. Needless to say that this is also a matter of personal preference and idiolect, as it is in writing. To find out the extent to which these factors influence the resulting translation, an analysis could be carried out, determining the frequency of *it*-clefts in original texts and their translations, as well as in different texts translated by the same translator, or, alternatively, in two different translations of one source text. Needless to add, the choice may also be arbitrary or influenced by the surrounding translated text rather than by the original.

4. Research

Since the subject matter is rather extensive, only a sample research was carried out. Its main goal, besides obtaining preliminary results, was to observe the main tendencies, so that the following, larger-scale research could be adjusted if necessary. Since the concordances found in the corpus are preceded by a limited number of words, it is not always easy

(or reliable) to determine the type of the *it*-cleft according to Prince's classification. Hence the strictly comparative nature of the present research. The absence of context remains the greatest obstacle to any qualitative corpus-based analysis.

4.1 Sources

The corpus used for the sample research was InterCorp, a Czech parallel multilingual corpus. In the main line of research, a subcorpus of original English texts was used, along with the relevant Czech translations. More specifically, the field was limited to fiction; the main reason for this choice being that there is a large amount of material which is at the same time very diverse (with respect to authors, the complexity of the structures, style etc.). As opposed to unpublished texts, translations of fiction are supposed to undergo some sort of editorial process, which should to some degree lower the number of mediocre translation equivalents not reflecting the difference between the two languages.

4.2 Queries and results

CQL queries were formulated with the aim of covering as many instances of *it*-clefts as possible, including interrogative sentences, modal structures, *it*-clefts whose dependent clause is introduced by zero relative (juxtaposed) instead of the more usual relatives such as *that* or *who*. Some spaces were allowed for optional elements like adverbs, predeterminers, determiners, adjectives or pronouns. In the end, three queries were formulated, as follows:

- ▶ Query 1: declarative *it*-clefts and polar interrogatives:
 ([[] containing [word="(?)it"] containing [word="(?(i)(is|was)"])[tag="RB"]?
 [tag="PDT"]? [tag="DT"]? ([tag="JJ"]*)? [tag="PP"]? ([tag="N.*"]+|[tag="PP"])
 [word="that|who|which"]?[tag="MD"]?[tag="V.*"]+
It is / it was / is it / was it + (adverb)(predeterminer)(determiner)(one or more adjectives)(possessive pronoun) + one or more **nouns/personal pronoun** + (*that/who/which*)
 + (modal verb)(one or more **verb forms**)
- ▶ Query 2: *Wh*-questions:
 [word="(i?)(what|who)"] [word="is|was"] [word="it"] [word="that"]? [tag="V.*"]+
What / who is / was it that (one or more verb forms)
- ▶ Query 3: Declarative and interrogative *it*-clefts containing a modal verb:
 ([[] containing [word="(?)it"] containing [tag="MD"] [word="have"]?[word="(be|been)"]
 [tag="RB"]? [tag="PDT"]? [tag="DT"]? ([tag="JJ"]*)? [tag="PP"]?
 ([tag="N.*"]+|[tag="PP"])[word="that|who|which"]?[tag="MD"]?[tag="V.*"]+
It + modal + (have) be/been (including inversion) (adverb)(predeterminer)(determiner)
(one or more adjectives)(possessive pronoun) + one or more nouns/personal pronoun +
(that/who/which) + (modal verb)(one or more verb forms)

Since the queries cover rather a broad spectrum of sentences, they had to be sorted manually, and many of them were found irrelevant. Most commonly, the "false positives" with an identical surface structure were relative clauses, e.g. *It was a trick swim-*

mers used to re-oxygenate their blood between tightly scheduled races. The table below illustrates the efficiency of the search in both lines of research, showing how many of the results turned out to be genuine *it*-clefts with a focused subject. Query 1 was the most productive, as could be expected: 1886 concordances were found, 200 of which were randomly selected for the analysis. The remaining two queries proved to be much less productive. In the second, additional research, Query 1 yielded 364 results (which were then limited to a random sample of 200 lines); Queries 2 and 3 displayed the same trend as in the first case.

	1 st line of research (English > Czech)		2 nd line of research (Czech > English)	
	total	relevant	total	relevant
Query 1	200	90	200	74
Query 2	7	4	4	0
Query 3	75	14	15	4

4.3 Control search

It can be assumed that in spite of all effort, the queries cannot and will not cover all *it*-clefts present in the texts. To find out how efficient the queries are, I decided to carry out a manual search in a fiction book, namely George Orwell's *Animal Farm*. This control search yielded exactly the same number of *it*-clefts as the combination of all three queries. However, *Animal Farm* is short, stylistically rather monotonous and consists mainly of shorter, not very complex sentences; the *it*-clefts found in the novel are, as a rule, less complex than many other from different sources. While this proves that the queries are probably quite well-suited for such simple structures, it does not necessarily mean they will be absolutely reliable in syntactically elaborate texts. A similar result of automatic vs. manual search is described in Dušková (2015). The automatic search thus does not seem to be significantly inferior to manual search.

It may be noted in passing that this additional sub-study exemplifies the advantages of a corpus-based research, as the frequency and type of *it*-clefts clearly differs from text to text, at least in a genre as diverse as fiction.

5. Lines of research

5.1 First line of research: English to Czech

The core of the analysis involves searching for English *it*-clefts and considering both them and their Czech counterparts. In the sample research, the three queries introduced in the previous section were used to obtain material for the analysis, with the number of results specified in section 4.2. The English *it*-clefts were consequently classified according to the criteria presented by Prince, and their Czech counterparts according to the means used to convey the given information structure. At this point it must be added that the chosen approach proved to be somewhat limited in the classification of English

it-clefts – the corpus often does not provide sufficient context to safely determine whether the content of the subordinate clause is given or new. For this reason, there is some degree of uncertainty in the less obvious cases.

As regards the typology of Czech translation equivalents of English *it*-clefts, as could be expected, Czech translations proved to use a variety of means to express the information structure indicated by an English cleft sentence. The following types of syntactic means were found in the sample analysis:

- Cleft construction

(1) It was her voice that had betrayed them. (London)

Byl to její hlas, který je prozradil.

The Czech syntactic equivalent of the English *it*-cleft; arguably perceived as more marked than in the original.

- Word-order (the rhematic element is placed at the end of the clause)

(2) It was papa made me leave. (Hemingway)

To mi poručil táta.

This solution makes use of the Czech word-order which is primarily governed by the FSP principle. It is possibly less stylistically marked than an *it*-cleft. (Filipec 1955)

- Focalizing particles (*to, právě, zrovna...*)

(3) I read a book about Goya, which said it was his deafness that made him into a major artist. (Lodge)

Četl jsem knihu o Goyovi, kde se tvrdilo, že právě hluchota z něj udělala významného umělce.

Another fairly frequent solution using lexical means; focalizers or rhematizers are expressions whose primary function is to signalize categories of information structure and which are typically placed in front of the emphasized unit. Rhematizers can have various specifying meanings (Panevová et al., 2014: 221–222).

- Combination (WO + focalizer)

(4) Bobby finally grasped that it was he who was being beckoned and slowly made his way over. (Jackson)

Bobby nakonec pochopil, že ten muž mává právě na něj, a pomalu k němu zamířil.

The rheme is indicated by word-order and further emphasized by a focalizer, which makes the information structure more evident.

- Pronoun

(5) It wasn't I who killed you, Albus. (Rowling)

Já jsem tě nezabil, Albusi.

In Czech, a personal pronoun in the role of subject is normally not expressed, as verbal categories are evident from the inflectional ending of the verb. When the pronoun does appear, it is emphatic or marked in some other way (Grepš et al. 2012: 412); thus being a valid equivalent for a cleft construction.

- Other

(6) It was the telephone that was tantalizing Starling. (Harris)

Starlingová měla celou tu dobu chuť skočit po telefonu.

This category contains examples in which the whole clause has been reformulated in some way.

- None

(7) It was Pop who started the song, which was We Ain't Got a Barrel o' Money, and everyone took it up in shrill voices. (Bates)

Taťka Larkin první zanotoval Peněz tak mít jako želez a všichni se k němu pronikavými hlasy přidali.

This group includes instances in which no apparent syntactic means are used to increase the communicative dynamism of the element that has been emphasized by the English cleft; they are mostly informative-presupposition *it*-clefts. These include some peripheral examples, such as:

(8) It may be a wise child that knows its own father, but it is a laughing child that knows its own mother. (Morris)

Chytré dítě může poznat svého otce, ale smějící se dítě jistě poznává svou matku.

This particular type has been mentioned in literature as a special case, a “proverbial” sentence with a predicational meaning instead of specificational, which is typical of *it*-clefts (Patten, 2012: 4). The construction clearly does not serve the usual function of (primarily) expressing information structure, and the translation reflects that. Or similarly:

(9) It wasn't every man who could afford her, so she'd get the same type again and again. (Palahniuk)

Všichni muži si ji nemohli dovolit, a tak vídala stále dokola tytéž typy.

The above example is not proverbial, but the *it*-cleft serves a similar function: it is a stylistic device that increases markedness rather than a means to highlight information structure.

5.2 Second line of research: Czech to English

Apart from the main research on English texts translated into Czech, another, secondary search was conducted, this time on Czech sentences which were translated into English as *it*-clefts. The queries were therefore identical (and performed on English text); the only difference was the direction of translation. The purpose of this research was to find out whether the equivalent pairs in the opposite direction (from Czech to English) tend to differ significantly.

For instance, if the use of *it*-clefts in Czech translations were mostly based on the source structure, it could be expected that the frequency of Czech *it*-clefts in this direction would be much lower. The types of equivalents proved to be more or less the same as in the first study, namely:

- Cleft construction

(10) But it is she who does the instructing!

Ale je to ona, kdo poučuje! (Kundera, *Nesmrtelnost*)

In this case, the choice of the English equivalent is quite self-evident. In fact, it would be interesting to search for the construction in the Czech section of the parallel corpus and see if there are any English translations that do *not* use cleft construction.

- Word-order

(11) I've thought that since I was a kid, but I realized that back then it was you who denied it.

Myslím si to od malička, spíš mi utkvělo, že jste to tenkrát popírala vy. (Kohout)

The information structure expressed by the principle of linearity in Czech is translated by an *it*-cleft.

- Focalizing particles (*to, právě, zrovna...*)

(12) It is the second tear that makes kitsch kitsch.

Teprve ta druhá slza dělá z kýče kýč. (Kundera, *Nesnesitelná lehkost bytí*)

- Combination (WO + focalizer)

(13) And it was this obedience mingled with shyness that drove me wild; when I approached her, she shrank back and covered her crotch with her hands....

A mne vzrušovala k šílenství právě ta oddanost smíšená s ostychem; když jsem k ní přistoupil, skrčila se a zakryla si rukama klín ... (Kundera, *Žert*)

- Pronoun

(14) One of my physicians had to leave the clinic, and it was I who asked for his resignation.

Jeden z mých lékařů musel z kliniky odejít, sám jsem ho k tomu vyzval. (Stýblová)

- None

(15) And so it was not the flu that kept him from sleeping with Tereza on her first visit.

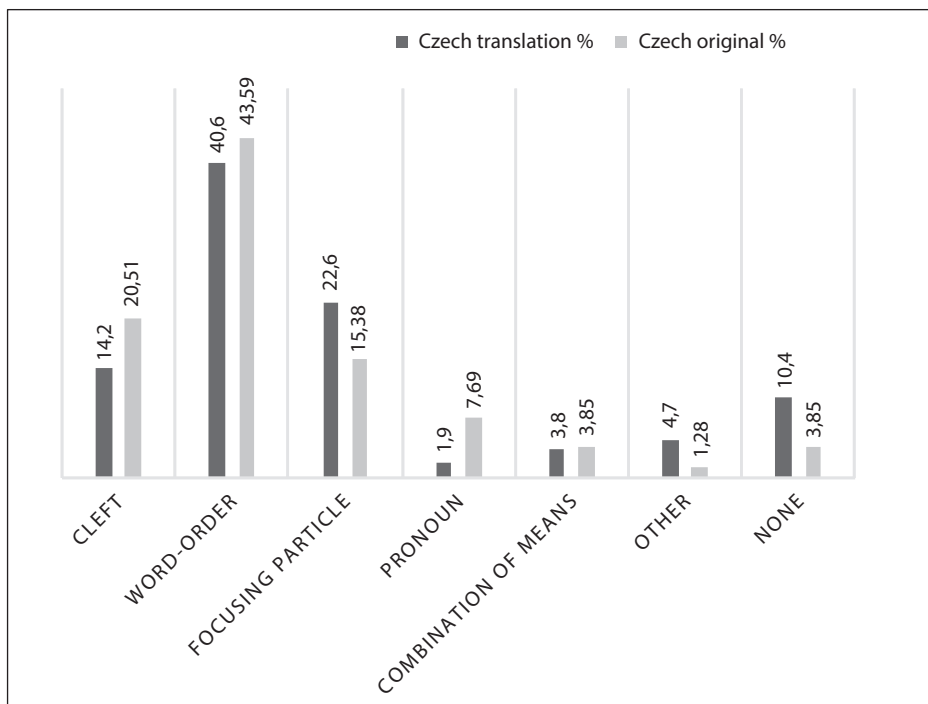
Ostatně když u něho byla Tereza poprvé s chřipkou, nikdy s ní nespál. (Kundera, *Nesnesitelná lehkost bytí*)

In some cases, no apparent syntactic device in the Czech sentence seems to require the *it*-cleft; the sentence is generally reformulated, similarly to example (9). The cleft seems to express a certain markedness.

6. Results

As explained in section 4.3, the classification of *it*-clefts was somewhat complicated due to limited context provided by the concordances. What can, however, be quantified without doubt are the proportions of FSP means used in the translation equivalents. Figure 1 presents a comparison of Czech equivalents of English *it*-clefts and original Czech constructions that were translated as *it*-clefts in English. Although the results may not be strictly comparable (the size of the randomly selected sample does not significantly differ, but no special attention was paid to the year in which the texts were written, their length and other possibly relevant factors), there are some tendencies to observe.

Figure 1.



The figure reveals several evident findings: in both directions of translation, the most common Czech equivalent of an *it*-cleft is a clause in which the rhematicity of the subject is indicated by its final position (FSP principle of linearity). Focalizers and cleft constructions have approximately equal frequencies, followed by the remaining types. What is remarkable is that the differences between the two directions are not radical and the distribution of the individual categories is roughly similar. There is a lower frequency of emphatic personal pronouns in translations from English, suggesting that this way of expressing the information structure of English *it*-clefts may not be an obvious choice for Czech translators. On the other hand, focalizers/rhematizers are more frequent in this direction.

7. Conclusion

The presented preliminary study suggests that as regards fiction, translators of both languages seem to be quite reliable in choosing the appropriate FSP means characteristic of the target language, at least statistically. In other words, the translation equivalent pairs consisting of English *it*-clefts and their Czech counterparts do not display fundamental differences depending on the direction of translation. Nevertheless, this sample research in itself is not statistically very relevant, due to the lower number of examples; more equivalents will have to be analysed, and ideally with respect to other variables.

As regards further research, the results of this preliminary study show that there are still many issues that need to be addressed. The first condition of a quantitatively relevant result is that the research should be carried out on a larger scale. That, however, brings the aforementioned problem of insufficient context, which is a significant complication for the classification. For further research it will be necessary to find a suitable method to deal with this problem.

To have a better understanding of the role of cleft constructions in Czech, it may be useful to compare their occurrence in original Czech texts and those translated from English, both in terms of frequency and typology. Such a research would indicate whether there tends to be any interference from the source language, quantitative or qualitative.

REFERENCES

- Dušková, L. (2010) Syntactic construction, information structure and textual role: An interface view of the cleft sentence. *Brno Studies in English* 36/1, 29–45.
- Dušková, L. (2015) Textual roles of two forms of rhematic subjects: initial rhematic subjects vs. subjects rhematized by *it*-clefts. *Linguistica Pragensia* 25, 49–66.
- Filipec, J. (1955) Byl to můj přítel, který... [It was my friend who...]. *Naše řeč* 38/7–8, 193–198. Available at: http://nase-rec.ujc.cas.cz/archiv.php?art=4433#_ftn1 [2016-09-26].
- Greenbaum, S. and R. Quirk (1990) *A Student's Grammar of the English Language*. Harlow: Longman.
- Grepl, M. et al. (2012) *Příruční mluvnice češtiny*. Prague: Lidové noviny.
- Chlumská, L. (2015) *Překladová čeština a její charakteristiky*. Prague: Faculty of Arts.
- Panevová, J. et al. (2014) *Mluvnice současné češtiny 2: Syntax češtiny na základě anotovaného korpusu*. Prague: Karolinum.
- Patten, A. (2012) *The English It-Cleft: A Constructional Account and a Diachronic Investigation*. Berlin/Boston: Mouton de Gruyter.
- Prince, E. (1978) A comparison of *wh*-clefts and *it*-clefts in discourse. *Language*, 54/4, 883–906.

Sources

Český národní korpus – *InterCorp*. Praha: Ústav Českého národního korpusu FF UK. Available at: <http://www.korpus.cz>.

- Bates, H.E. *The Darling Buds of May*
Harris, T. *The Silence of the Lambs*
Hemingway, E. *The Old Man and the Sea*
Jackson, M. *Five Boys*
Lodge, D. *Deaf Sentence*
London, J. *At the Rainbow's End*
Morris, D. *The Naked Ape*
Orwell, G. *Animal Farm*
Palahniuk, C. *Choke*
Rowling, J. K. *Harry Potter and the Deathly Hallows*

- Kohout, P. *Sněžím*
Kundera, M. *Nesmrtelnost*
Kundera, M. *Nesnesitelná lehkost bytí*
Kundera, M. *Žert*
Stýblová, V. *Skalpel, prosím*

ANGLICKÉ KONSTRUKCE *IT-CLEFTS* S VYTČENÝM PODMĚTEM Z POHLEDU AČV A JEJICH ČESKÉ PŘEKLADOVÉ EKVIVALENTY

Resumé

Příspěvek představuje stručný přehled teoretických aspektů anglických vytýkacích konstrukcí s důrazem na aktuální členění větné. V praktické části se srovnávají vytýkací věty s vytčeným podmětem s jejich českými překladovými protějšky na základě vzorku získaného z paralelního vícejazyčného korpusu InterCorp. Vzhledem k tomu, že každý z obou jazyků používá jiné aktuálněčlenské prostředky, čeština požadovaný význam kromě vytýkací konstrukce nutně vyjadřuje i jinak, například slovosledem, zdůrazňovacími částicemi atd. Pro lepší představu o tom, nakolik může být překlad ovlivněn povrchovou strukturou originálu, byl proveden další vzorový výzkum se stejnými dotazy, pomocí nichž byly vyhledány anglické překlady obsahující vytýkací konstrukci a jim odpovídající české originály. Tyto předběžné výzkumy ukazují, že přinejmenším ve zvolené oblasti zřejmě nenastává výrazná interference zdrojového jazyka do výsledného textu.

Anna Kudrnová

Department of English Language and ELT Methodology

Faculty of Arts, Charles University

anna.kudrnova@gmail.com