L. Chlumská

THE EQUIVALENTS OF ENGLISH PRESENT PERFECT TENSE IN CZECH ACCORDING TO PARALLEL CORPORA

1. English Present Perfect Tense

1.1. According to grammar textbooks

English present perfect is a compound tense formed by combining a present-tense form of the auxiliary verb "to have" with the past participle of the main verb. As for its function, recently published corpus-based Real Grammar[1] for students of English gives a definition that "the present perfect tense is used for actions or states that happened or started at an unspecified time in the past" and "the action or state continues into the present or is still important in present."

Longman Grammar of Spoken and Written English[2] has a slightly different approach and emphasizes the semantic, aspectual point of view. That is why authors of this grammar rather call this grammatical combination the perfect aspect present tense instead of the present perfect tense and compare it with progressive and simple aspects. The perfect aspect is much more common in British English than in American English.

1.2. Translation of present perfect tense into Czech

In contemporary Czech, there is no such tense or a single specialized verb form as the English present perfect. Czech has a simpler set of tenses: past, present and future. However, it has a special category of aspect, which means that every verb is either perfective, or imperfective (with the exception of bi-aspectual verbs). Most verbs in

---

[1] *Conrad S., Biber D.* Real Grammar. A Corpus-Based Approach to English. New York, 2009. P. 7.

[2] *Biber D. et al.* Longman Grammar of Spoken and Written English. Harlow, 1999. P. 460–468.

Czech form pairs of one perfective and one imperfective verb with generally the same meaning. The aspectual distinctions exist on the lexical level – there is no unique method of forming a perfective verb from a given imperfective one (or vice versa).

When conveying the English present perfect tense into Czech, the translator has to choose from the given repertoire of three tenses and combine the chosen tense with perfective or imperfective aspect (provided there is an aspectual pair of verbs with the desired meaning). Generally, the English present perfect tense is said to translate into Czech as past tense (without any remark about aspect) and sometimes as present tense.

One of the few authors providing some information about the aspectual point of view is Dušková in her Grammar of Contemporary English with Reference to Czech[1]. According to this grammar, the present imperfective tense is used when the English present perfect has this inclusive meaning, e.g. describes an action which began in the past and continues into the present. It is usually the case when time markers, such as *how long*, *since, for* or *so far,* are used.

There has not been any detailed corpus research done concerning the use of aspect (or *vid* in Czech), the role of time markers, such as *since, up to now, recently, yet* etc., or other possible ways of translation. This paper will try to provide some answers.

2. Parallel Corpus in Translation Studies

2.1. Parallel corpus – pros and cons

Corpus linguistics and parallel corpora in particular, have revolutionized translation and contrastive studies. Thanks to large language corpora, linguists finally have a chance to explore and compare languages and their specifics from different points of view.

---

[1] *Dušková L. et al.* Mluvnice současné angličtiny na pozadí češtiny (Grammar of Contemporary English with Reference to Czech). Praha, 2003. P. 220–226.

Research can be based on solid results from a reliable source, and what is even more important, from a source showing actual usage. In translation studies, such a descriptive approach can be extremely useful, since it can reveal hidden patterns in language or unknown tendencies in both original and translated texts.

It is indisputable that parallel corpora or more specifically translation corpora (consisting of an original text and its translations into other languages) have many advantages. However, they may also have some drawbacks which every researcher has to take into account.

First, the compilation of parallel corpora is indeed a challenging task. It is very difficult to find a balanced and representative sample of original texts and their translations of different genres, therefore parallel corpora often limit themselves to one text type, such as fiction or news. Secondly, the quality of translation can be an issue. The question is how, or whether at all, the quality of translated texts should be examined. Finally, smaller languages may find it difficult to find enough translations from their language into another one; the opposite direction of translation is usually much more prolific.

2.2. Project InterCorp

This research is based on the English-Czech parallel corpus from the InterCorp database[1]. InterCorp is a large parallel synchronic corpus covering a number of European languages (currently 22) with Czech being the pivot language. The corpus is compiled mostly by teachers and students of the Faculty of Arts, Charles University in Prague, and by other collaborators of the Institute of the Czech National Corpus.

After registration, the corpus is accessible to public via a web interface called Park. InterCorp mainly contains fiction, but there is also a smaller amount of news texts available in some languages. InterCorp is a continuous project, which means that the amount of tokens for every language has been growing.

---

[1] Czech National Corpus – InterCorp, available at:
http://www.korpus.cz/intercorp/

3. Research Results

3.1. Data

The English-Czech part of InterCorp includes 34 texts which were originally written in English and were translated into Czech. For the purpose of this research, only 25 texts have been chosen, the criteria being the text type (fiction) and the translation date (after 1990). Also the translators have been taken into account: there are no two books written by the same author *and* translated by the same translator. There is one translator occurring twice in this subcorpus, but each time translating a different author[1].

The total number of tokens in this subcorpus is 1 909 782 in the Czech part and 2 250 130 in the English part (as of 24th February 2011). The corpus query was designed to look up all the instances of present perfect, except for the forms *have/'ve been* or *has/'s been* and *have/'ve got* or *has/'s got*, as they are the two most common verb forms used in present perfect (the latter with clear possessive meaning) and they would require another specific research. The focus has been on lexical verbs in active voice (*have done, have seen, have decided* etc.).

In total, 2 541 instances of present perfect tense have been found, 660 out of which (i.e. 26%) included a time marker, such as *yet, already, since, ever, before, never, all the time, lately* etc. Since the size of the subcorpus itself is not very representative in the era of large corpora, all of the found instances were used as a sample. For this reason, it was necessary to manually analyze and tag the translation equivalents of the present perfect tense and the time markers (if present) in both English and Czech texts.

The time markers in Czech did not always correspond with the time markers in English, e.g. Czech *už* ("already" in English) was used very often, even though there was no marker in the original sentence.

---

[1] Full bibliographic data of all the texts used in this research are available at request. Due to the paper length restrictions, it was not possible to include them here.

12

This would, however, require an individual research.

3.2. Time markers

Several time markers were found in the texts with the present perfect tense. The following frequency table summarizes the most common ones:

Time markers

Time marker	Number of instances (%)
already	65 (9,9)
always	50 (7,6)
before	28 (4,2)
ever	66 (10)
for	32 (4,8)
in (years, weeks)	25 (3,8)
just	46 (7)
never	113 (17,1)
since	42 (6,4)
yet	42 (6,4)

Among other, less frequent time markers are *now* and *all (year, day etc.)* with 11 instances, *the first time* (9 instances), *so far* (8), *how long* (5 instances), *lately* (3) etc.

3.3. Translation equivalents

In order to analyze the translation equivalents, the following categories were established: MN (past imperfective), MD (past perfective), PN (present imperfective), BD (future perfective, since perfective verbs do not technically express the present), LEX (lexical equivalent, e.g. non-verbal or idiomatic), MOD (modal verb) and other (non-standard translations, omissions, conditionals etc.).

The categories are distributed in the subcorpus as follows (in %): MN (23,5), MD (56,3), PN (6,6), BD (3,4), LEX (6,2), MOD (1,2) and other (2,8).

It is obvious that the past perfective tense in Czech is prevailing, most common verbs being *říci* ("say")*, udělat* ("do")*, najít* ("find")*, rozhodnout se* ("decide")*, stát se* ("become", "happen")*, vzít* ("take")*, setkat se* ("meet") etc. The second most prominent category is the past imperfective tense, represented mainly by verbs such as *vidět* ("see")*, slyšet* ("hear")*, mluvit* ("talk")*, vyprávět* ("tell stories", "talk") or *čekat* ("wait"). Present imperfective, suggested by some linguists as a possible equivalent, is not as frequent as we may have expected.

3.4. Influence of time markers on translation equivalents

When we look at how the most frequent markers influence the translation equivalent, certain patterns seem to appear. The time marker *already* seems to require the past perfective (48 instances of MD, only 13 MN, 3 LEX and other). Sentences with the time marker *ever,* on the other hand, translate mostly as the past imperfective (41 MN out of 66, 16 MD, 6 LEX and other). Similar pattern is typical for the marker *never* (61 MN, 35 MD, 5 MOD and other), which translates into Czech as *nikdy* ("never")*, v životě* ("in my life")*, ještě* ("yet"), and *nikdy předtím* ("never before"). The marker *always* also seems to require the past imperfective (32 MN, 6 MD, 5 LEX, 4 MOD and other), whereas sentences with *just* are translated into Czech using the past perfective tense (32 MD, 6 MN and other).

4. Conclusion

The paper has demonstrated that parallel corpora offer us many a possibilities for research, not only in the area of translation studies. The corpus-driven approach can reveal many interesting information not only about translation equivalents, but also about the languages themselves. In this case, the results have shown that the present perfect tense has several translation equivalents in Czech which are also influenced by the time markers, such as *always*, *never* or *already*. The findings may serve as a basis for further research in this area.