Is Czech more effective in loanword adaptation than Polish? Score assignment as a method for measuring loanword adaptation

Mirosław Bańko
Faculty of Polish Studies
University of Warsaw

Diana Svobodová
Pedagogical Faculty
University of Ostrava

Objective

The aim of this study was to measure the degree of loanword adaptation in Polish and Czech and to test the hypothesis that in Czech, lexical loans are more strongly integrated with the spelling, pronunciation and grammar of the language than they are in Polish. To this end, a scoring system was devised in which lexical loans were assigned points according to certain predefined criteria. The total score of a loan is therefore a measure of its adaptation.

The study was carried out between 2012 and 2015 and included a hundred English words borrowed into both Polish and Czech. Strictly speaking, then, the study was meant to measure and compare the degree of adaptation of English lexical loans in these two languages. The scoring system was devised accordingly, with English as the donor language and Polish and Czech as the recipient languages in mind. However, the method developed for this study can be adjusted to other language pairs, as needed.

Hypothesis

The observation that the Czech language is more effective than Polish in the adaptation of English loans, i.e. that it integrates them more strongly with its spelling, pronunciation and morphological system, has occasionally been made before, e.g. Bogusławska (1994), Siatkowska (1997), Szczepańska (2004), Labocha (2006). The productivity of the Czech derivational system was particularly emphasized and contrasted with Polish word-formation techniques which rely on analytic constructions to a greater extent. The difference was explained in terms of the different histories of these two languages. In Czech, during the period of National Revival, loanwords, especially those of German origin, were expelled from the language and replaced with native coinages, which contributed to the development of Czech derivational system. In Polish, on the other hand, strong influences of the French language from the 17th to the 19th centuries favored the use of analytic constructions, i.e. multi-word units (Damborský 1977).

Although rich in empirical material, the studies referred to above were unsystematic and based on ad-hoc examples with which their authors tried to demonstrate their point. We do not mean to claim that the authors' intuitions were wrong, but we know of no earlier attempt to actually measure the degree of English loan adaptation in Polish and Czech. To put it differently, we are unaware of any quantitative approach to the question of whether the tendency to integrate English loans with Czech spelling, pronunciation and morphology is indeed stronger than it is in Polish.

An anecdotal example that seems to support the other authors' observations is an advertisement one of the present authors saw in the Stromovka park in Prague. The ad, pinned to a tree trunk by a desperate dog owner whose pet had got lost, was an appeal for help to find it. A photo of the dog was printed and its appearance was described as follows: "velký je jak zlatý retrívr a vypadá jak kokršpaněl" ('it has the size of a golden-retriever, but looks like a cocker-spaniel'). The Polish translation would be: "ma wielkość golden retrievera, a wygląda jak cocker spaniel", with two loans from English unassimilated graphically, unlike in Czech.

Different spellings have different effects on the perception of words, e.g. the same dog may seem emotionally closer and nicer to readers when its name is well assimilated in their language than when the name clearly exhibits its foreign origin. Though interesting, such differences will not be dealt with here, because they are the subject of our study in another part of the present project, see *Comparative analysis of synonyms and variants in Polish and Czech*. In this report we focus on formal aspects of loanword adaptation, leaving out the issue of the semantic integration of loans with the recipient language.

Method

In order to test the above hypothesis, a hundred English words, borrowed into both Polish and Czech, were selected and care was paid to make the selection diversified in many ways, above all with respect to the adaptation processes the words had undergone. Then a scoring system was devised in which a loanword is assigned points if it meets some predefined criteria concerning its adaptation in the recipient language. The total score of a word can be used as a measure of its adaptation and the overall score of all words selected for the study can be an index of the language's ability to adapt lexical loans. By comparing the scores of two languages, one can see which language is more effective in loanword adaptation.

The criteria we used pertain to the spelling of lexical loans, their pronunciation, inflection and their derivational potential. The criteria are independent of one another, so that not only the total scores of particular words can be compared but also their partial scores, corresponding to selected criteria. In addition, partial scores of all words investigated can be added up according to a selected criterion, thus giving an idea of whether Polish, for instance, is equally effective as Czech in assimilating the spelling of loans, or whether Czech is ahead of Polish with respect to their derivational potential.

Before we go into the details, two general points must be made. First, the scoring procedure was intended to evaluate quantitatively the depth of loanword integration, not the speed of this process. The diachronic dimension was not taken into account.

Second, the procedure was devised with lexical loans proper in mind. Loan translations (also called calques) and semantic loans (also called semantic shifts) were excluded from the present study. The same holds for half-translations, e.g. Polish *dwuklik* (from English *double-click*) or Czech *spolulídr* (from English *co-leader*) were not included.

The English words selected, the loans, their meanings and all the other information used in the scoring procedure was recorded in two Excel spreadsheets, one for Polish and one for Czech, see appendix.

Details of scoring

The scoring rules were intentionally very simple. This was so as to minimize the danger of subjectivity in decision making and to reduce the amount of work needed to assign scores to words.

As stated above, only formal aspects of loanword adaptation were taken into account. Some authors claim that loan integration should be based on the textual extension of loans, i.e. their frequency in different language varieties. However, we could not take stylistic distribution under consideration, because a large share of the loans we studied are not frequent enough in corpora to provide reliable statistical data.

The scores assigned to loans were devised in such a way as to balance the different factors affecting the integration of loans with the recipient language. The maximum score for a word was 10 points, which made it easy to calculate percentages.

Most of the italicized subtitles below correspond to the column names in the spreadsheets in which the data were recorded.

Meaning

The meaning of a loan and in particular its relation to the meaning of the corresponding English word had no influence on the loan's score. Nonetheless, each loan was provided with a short meaning description so that it was easier to interpret the other information about it.

The meaning descriptions are not always full dictionary-like definitions. Sometimes only a synonym word is given, a more general term or a description of a more general concept. In addition, different senses of the same word are separated with a semicolon only, without paying attention to how close they are, e.g. from the viewpoint of metaphorical and metonymical processes they originated from. In sum, the meaning descriptions are only clues for readers and whenever they prove to be insufficient, readers are advised to consult a dictionary.

NOTES

> As it was not the goal of this study to compare the meanings of loans with those of the corresponding English words, it can only be generally stated that the two meanings overlap, but are not always identical. For instance, *chill-out* is an adjective in English, used in phrases such as *chill-out music* or *chill-out room*, whereas in Polish and Czech it is a noun, denoting a kind of music and a room, e.g. in a club to listen to it (the latter meaning only in Czech). Another example is *skate*, in English a verb and a noun, but in Polish and Czech a noun only. Indeed, while in English *skate* refers to an ice skate, a roller skate or a skateboard, in Czech it is used only with reference to a skateboard and in Polish only to a skateboarder. The latter word can illustrate the well-known fact that loanwords are often adopted in only some of the meanings they have in the donor language.

Spelling

When a loanword is used in the original English spelling exclusively, it was assigned no points. When a loan's spelling is different from the original spelling only in the use of some inflectional or derivational affixes which have been appended to the basic form (e.g. the nominative singular for nouns) in the recipient language, the loan scored no points either. Examples of words which received no points are *DVD* in Polish (an indeclinable word used only in this spelling), *surf* in both Polish and Czech (used only in this spelling, but declinable), and *logovat* in Czech (*logować się* in Polish) which differ from English *log (in)* only in the use of derivational and inflectional affixes.

When a loanword is used in the original form, but also has an assimilated spelling, it was given one point. Whether the original spelling is more frequent than the assimilated one or not had no influence on scores, but affected the order of presentation: as a rule, more frequent spellings were listed before the less frequent ones.

When a loanword is used in an assimilated spelling exclusively, it was assigned two points.

NOTES

1. Loanwords with two or more assimilated spellings were treated as if they had only one. The reason is that the abundance of spellings is no proof of a loan's better integration with the recipient language. On the contrary, spelling variance is symptomatic of early stages of loan adaptation.
2. Cases of morphological variance in spelling were treated as any other examples of spelling variance, i.e. they were no exception to the above rule, cf. *celebryta* (with *-a* ending) and *celebryt* (with zero ending), both coming from English *celebrity*.
3. Cases in which one variant (usually the original one) has a double consonant letter and another variant has a single letter instead, being otherwise identical, were treated as normal examples of spelling variance and were subject to the above rules, cf. *fitness* and *fitnes*.
4. Some cases of spelling variance were ignored for scoring purposes: those in which variants differ exclusively in the use of capital vs. small letters (cf. *Internet* and *internet*) and those in which variants are spelled as one word, two words or with a hyphen (cf. Polish *pornobiznes* or *porno biznes*, or *porno-biznes*, outside the list of words selected for this study). Variant spellings of this sort were included in the spreadsheets, but did not affect the process of scoring.
5. Spellings which are not attested in standard dictionaries were preceded with an asterisk.
6. Variant spellings were separated by a slash mark.

Loans which preserve the original spelling and have no assimilated spelling variants are not all alike, cf. *surf* and *golf*, both borrowed to Polish and Czech from English (the latter as a sports term). While *surf* could, in principle, be spelled differently, according to its pronunciation, i.e. *serf*, the word *golf* cannot be spelled another way, because its pronunciation accords with its spelling in both Polish and Czech. Based on this observation, one could argue that *golf* is better assimilated graphically than *surf* and should be assigned a higher score in both languages. However, we do not think there is enough reason to do so. After all, the users of Polish or Czech have not done anything special about the English word *golf*, they have adopted it in its original form, because the original form fits their own languages well. If the lexical score is to reflect the work done on a certain loan to integrate it with the recipient language, then there should be no difference in scores between *surf* and *golf*, because no work (or not enough work) has yet been done on *surf* and no work has to be done on *golf*.

Pronunciation

The differences between the phonological system of English on the one hand and those of Polish and Czech on the other hand are so great that practically each loan has to adjust its pronunciation to the phonology of the receiving language. It was decided therefore that each loan should be assigned one point for its phonological assimilation.

Loans which exhibit variance in pronunciation were normally treated as if they were pronounced only one way. However, if one variant is close to the original pronunciation and another one is less similar, the loanword scored one extra point. For instance, *notebook*, pronounced [n'oʊ̯tbuk] or [n'otbuk] in Polish, and *gay*, pronounced [gej] or [gaj] in Czech, received two points each, because in each pair the former pronunciation is closer to the way these words are pronounced in English. Likewise, *interview* scored two points in the Polish spreadsheet, because it is stressed either the English way, i.e. ['interwju], or like in French, i.e.

[interwj'u] (the latter pronunciation being recommended in standard dictionaries). A similar example from Czech is *bodyguard*, pronounced with a long vowel in the third syllable, i.e. [b'odygárd], like in English, or with a short vowel, i.e. [b'odygard].

On the other hand, *spam* received only one point in the Czech spreadsheet, because neither of its two pronunciations: [spem] and [spam] can be judged as closer to the original [spæm] (the vowel [æ] is not used Czech). Likewise, the Polish word for *celebrity*, spelled either *celebryta* or *celebryt*, scored only one point, because there are two pronunciations here, but each spelling variant is pronounced only one way. The same holds for Czech *deadline*, for instance, spelled also *deadlajna* and pronounced in two ways accordingly: [d'edlajn] or [d'edlajna].

NOTES

1. Only modern variants were taken into consideration. In Polish, *komputer* (from English *computer*) was once pronounced [kompj'uter], now the only standard pronunciation is [komp'uter]. At some time in the past both pronunciations must have co-existed, but this fact was ignored for scoring. The word scored only one point.
2. The spreadsheets are intended above all for scholars dealing with Slavonic studies, where IPA is not very popular and the Slavistic transcription is more common. Therefore, we decided not to use IPA to record the pronunciation of words, but to employ a notation close to the Slavistic phonetic alphabet.
3. The back semivowel closing, e.g. the English word *show*, was marked with the /u̯/ sign, cf. *Web* in Polish or *toast* in Czech. Word stress was marked with an accent (') before the stressed vowel. A glottal stop was marked with the ʔ sign.

Inflection

When a loan has no inflections in the receiving language, it was assigned no points.

When a loan inflects in a very limited way, appearing in the English singular form when used with the singular or plural reference and, optionally, in the English plural form when used with the plural reference, it was assigned one point, cf. *talkshow*, used in both Polish and Czech in this form in all cases in the singular and replaced optionally with *talkshows* in the plural. Although this is not, strictly speaking, an instance of inflecting a loan in the recipient language, but only a case of using the morphological means of the donor language, it reflects the need to integrate the loan with the grammar of the receiving language.

When a loan is given inflectional morphemes of the recipient language, it was assigned two points. The number of syncretic forms in its paradigm was not relevant. Essentially, the presence of two different forms with a native word ending was enough to give a loan two points.

NOTES

1. If a loan has spelling variants, it scored points for only one of them – the one which yielded more points.
2. If a loan has more than one meaning and inflects only in some of them, it nevertheless scored points for inflection. However, it did not score points for each of its meaning separately.
3. If a loan inflects both the English way and according to the inflectional patterns of the receiving language (e.g. *bodyguard* in Czech, having *bodyguardi* or *bodyguards* in the plural), it scored two points, just as if it had the inflection of the receiving language only.

Gender assignment

As gender assignment is obligatory for nouns in Polish and Czech, a noun scored no extra points for it. Nouns that happen to have variant genders were given no extra points either. The reason

is that gender variance is no proof of a noun's better integration with the receiving language. It is only a sign that for some reason the noun is difficult to incorporate into gender categories.

However, when the process of gender assignment of a noun involves the use of morphemes of the receiving language, whether inflectional or derivational, the noun scored one point, see *Suffixation* below.

Suffixation

When a loan in its base form (or dictionary form) includes some inflectional or derivational non-zero morphemes which are absent from the donor language, the loan was assigned one point. Examples are Polish *celebryta* and Czech *celebrita*, both coming from English *celebrity* and both formed by means of the inflectional ending *-a*. Polish *logować się* and Czech *logovat*, both from English *log (in)*, represent another example.

NOTES

> The extra point was assigned only when the extra morphemes were appended in the receiving language. They need not be of native origin, but they must not belong to the original word. For instance, numerous Polish loans are based on the English plural, e.g. *fotos*, from English *photos*. When they inflect, they are given Polish endings, both in singular (cf. the genitive *fotosu*) and in plural (cf. the nominative *fotosy*). However, as the base form *fotos* has no non-zero morphemes appended in Polish, the word would be assigned no extra points (if it was included in our study).

Derivatives

When a loan has no derivatives in the receiving language, it was assigned no points. When a loan has just one derivative, it was assigned one point. When a loan has two derivatives, it was assigned two points. When a loan has three or more derivatives, it was assigned three points.

The maximum score was limited to three points in order to maintain a balance between various factors affecting the loanword adaptation. If a loan has more than three derivatives, they were all included in the appropriate spreadsheet, but no more than three points were assigned.

NOTES

1. Derivatives were understood broadly, including simple formations, compounds, acronyms, blends, etc., but excluding analytic forms, i.e. multi-word units.
2. Only direct derivatives were counted, i.e. those formed directly from the loanword. Indirect derivatives, i.e. those coming from a derivative of a loan, were ignored.
3. In case of doubt whether a derivative is a direct one or not, the decision was made to the loan's disadvantage, i.e. the doubtful derivative was ignored.
4. If there was doubt whether a word is a direct derivative of a loan or an independent borrowing, the decision was made to the loan's disadvantage, i.e. the doubtful word was ignored.
5. Variant spellings were recorded, but did not count for scoring: if a derivative is spelled in more than one way, it was counted only once.
6. Variant spellings were ordered according to their frequency and separated by a slash mark.
7. Spellings which are not attested in standard dictionaries were preceded with an asterisk.

Evidence

As the lexicographic treatment of new Anglicisms in both Polish and Czech dictionaries is unsatisfactory, a decision was made to assign scores on the basis of corpus data: the National

Corpus of Polish (full version, approx. 1 billion words) and Czech SYN Corpus (approx. 2 billion words). The necessary condition for a loanword to be defined as inflecting, having a particular spelling or a particular derivative was set to a minimum of three different attestations per billion words in at least three different sources, at least two of which were printed books or magazines, or their online versions.

The condition that at least two sources were books or magazines was laid down in order to eliminate words which are attested only in low-quality texts, not subjected to professional proofreading, e.g. transcripts of conversations, forum posts or computer chats. By 'different sources', different titles and different authors were meant. For example, two issues of the same newspaper were not different in this sense, neither were two novels by the same author.

NOTES

1. Examples of usage in which a given word is only cited for some purpose, e.g. in a discussion about whether it is correct or not, were ignored. Proper names were ignored too, unless the loan itself is a proper name.
2. Words attested in dictionaries but not in the above-mentioned corpora were ignored.
3. Words attested in corpora but absent from dictionaries were preceded with an asterisk. Asterisked words are not meant to be understood as incorrect: some of them are well-formed and might have been overlooked by lexicographers, some others may be ill-formed, but the difference is not relevant here.

Discussion

In all categories, the Czech language received higher scores than Polish. However, the differences are not equally significant, see below:

	Spelling	Pronunciation	Inflection	Suffixation	Derivatives	Total
Polish	73	107	176	3	142	501
Czech	74	116	178	4	186	558

The scores for spelling and inflection differ by less than 2 per cent, so the difference can be neglected. The relative difference in suffixation is higher, but there were only a few cases of suffixation among the words studied, so again no generalizations can be made. Pronunciation deserves more attention, because its score is almost 10 per cent higher for Czech than for Polish (an issue we will revisit later on). But the most striking difference is concerned with derivatives, their score being over 30 per cent higher for Czech than for Polish. It is mainly because of derivatives that the total score of the Czech language is over 10 per cent higher than that of Polish.

Overall, the results are in only partial agreement with former studies, based on intuition and unsystematic observations. It was claimed before that the Czech language was more productive in derivation, but it was also claimed that it was ahead of Polish as far as the ease of spelling adaptation was concerned, which does not follow from the table above.

In order to gain deeper insight into how well Polish and Czech adapt English loans, let us make a closer inspection of data within particular categories.

Spelling

The number of spelling variants for Polish is 171 (34 of them asterisked, i.e. absent from standard dictionaries), while for Czech it is 204 (68 of them marked with an asterisk). Although a higher number of spelling variants need not be proof of a loan's better integration with the recipient language, the difference observed between Polish and Czech suggests that the Czech language exhibits a stronger tendency to seek variants that would best fit its spelling.

Pronunciation

The number of variants in pronunciation is 112 for Polish compared with 133 for Czech. Again, the tendency to assimilate the pronunciation is stronger in Czech, the more so that the highest score of 2 points was assigned to 16 Czech loans and only 7 Polish loans. Interestingly, the two languages use partly different means to assimilate the pronunciation of English words. In Czech, which has long and short vowels, variants may differ in how well they reproduce the length of the English sounds, cf. [b'odygárd] (with long [á], hence closer to English) and [b'odygard] (with short [a]). In Polish, which has no vowel length distinctions but more freedom in the positioning of stress, variants may differ in how well they reproduce the original English stress, cf. [s'upermen] and [sup'ermen] (the latter pronunciation less similar to the English one).

Inflection

The distribution of data in the Inflection column of our spreadsheets deserves analysis, see below:

	Native	Native/English	English	NO	Total
Polish	85	1	4	10	100
Czech	74	14	2	10	100

The higher number of Native/English tags for Czech, accompanied by lower number of Native tags, indicates that the Czech language is no better than Polish in adopting English loans to its inflectional paradigms. On the contrary, it more often retains the English *-s* plurals instead of abandoning them to the advantage of native inflections. In both languages, the same number of NO tags was recorded and the uninflected words tagged this way largely overlap, cf.

in Polish: *CD*, *cool*, *CV*, *DVD*, *interview*, *OK*, *online*, *outdoor*, *party*, *PC*

in Czech: *Barbie*, *CD*, *cool*, *CV*, *DVD*, *OK*, *online*, *party*, *PC*, *SMS*

Suffixation

Only three loans in the Polish spreadsheet were assigned points for suffixation, these are *celebryta*, *kornfleksy*, *logować się*. In the Czech spreadsheet, four loans scored points in the same category: *celebrita*, *kornfleksy*/*kornfleky*, *dedlajna*, *logovat*. With one exception, the same words were suffixed in both languages.

Derivatives

The 186 to 142 point advantage of Czech over Polish is not the only indication that the Czech language is more productive in derivational processes. Another indication is the total number of derivatives: 245 for Czech and only 154 for Polish (variant spellings excluded). Moreover, there are only 7 loans in the Czech spreadsheet with no derivatives compared to 23 loans in the Polish spreadsheet. Examples of English words which were borrowed to Polish and Czech and served as a base for derivational processes in Czech, but not in Polish, are *Barbie*, *bodyguard*, *briefing*, *CD*, *coach*, *disc jockey*, *display*, *drive*, *DVD*, *interview*, *jeep*, *loser*, *make-up*, *party*, *piercing*, *remake*, *rocker*, *Skype*, *smiley* (actually, some of these words do have derivatives in Polish, but the words derived are so rare that they failed to meet our frequency criteria and were not recorded in the spreadsheets). The reverse situation – i.e. a loan having derivatives in Polish, but the corresponding Czech loan having no derivatives in Czech – is rare, with only two cases being attested in our data: *OK* and *puzzle*.

Conclusions

This study was, to our knowledge, the first attempt of this kind, aimed at measuring the degree of loanword adaptation in Polish and Czech. We are unaware of similar studies made for other languages, so we had to devise our methodology from scratch. All assumptions were explained in detail, so that similar studies could be modeled on ours, whether for Polish and Czech or for other related or unrelated languages.

The study has shown that earlier reports based on unsystematic observations were only partly true. Though the Czech language is generally more effective than Polish in the adaptation of English loans, it owes its advantage mainly to its more productive derivational system, much less to the way it adopts the pronunciation of English words. Its advantage over Polish in the adaptation of spelling is debatable (manifesting itself only in a higher number of spelling variants), while in terms of inflection Czech proved no better than Polish in the adaptation of English loans.

This study has some limitations which could be overcome in later work. First, the number of English words inspected was relatively small compared to the around 3000 Anglicisms present in both Polish and Czech (not counting purely technical or slang words). Secondly, the decisions on how many points a word should score for its spelling, pronunciation and morphology are inherently subjective to some extent. The details of the scoring system could be changed and it would be interesting to see how much that would affect the overall picture of loanword adaptation in the languages compared.

References

Bogusławska, E. (1994). *Adaptacja fonetyczna i graficzna pożyczek angielskich w języku polskim i czeskim*, „Poradnik Językowy", no 8, 1–15.

Český národní korpus - SYN. Ústav Českého národního korpusu FF UK, Praha, http://www.korpus.cz.

Damborský, J. (1977). *Zestawienia dwuwyrazowe w języku polskim*, in: J. Damborský, *Studia porównawcze nad słownictwem i frazeologią polską i czeską*. Warszawa: PWN, 21–29.

Labocha, J. (2006). *Nowe słownictwo w języku polskim i czeskim*, in: K. Ożóg, E. Oronowicz-Kida (eds.), *Przemiany języka na tle przemian współczesnej kultury*. Rzeszów: Wydawnictwo Uniwersytetu Rzeszowskiego, 66–70.

Narodowy Korpus Języka polskiego, http://www.nkjp.pl.

Siatkowska, E. (1997). *Sposoby adaptacji morfologicznej, przekształcenia semantyczne, przenikanie do żargonów najnowszych anglizmów w języku polskim i czeskim*, in: T. Z. Orłoś, J. Damborský (eds.), *Konsekwencje sąsiedztwa polsko-czeskiego dla rozwoju języka i literatury*. Wrocław: Centrum Badań Śląskoznawczych i Bohemistycznych UWr, 59–62.

Szczepańska, E. (2004). *Internacjonalizacja a anglicyzmy w języku czeskim i polskim*, „Bohemistyka", no 2, 96–104.

Published on 5/12/2015

This study was carried out within a Polish-Czech project: *Adaptation, perception and reception of verbal loans in Polish and Czech – linguistic, psychological and historic-cultural factors*, supported with a grant from the National Science Centre, Poland, no DEC-2011/03/B/HS2/02279.