N+N Borrowings from English: A New Stress Pattern in Czech?

Tomáš Duběda

Institute of Translation Studies, Charles University, Prague, Czech Republic

dubeda@ff.cuni.cz

Abstract

In this paper, I investigate English N+N and Adj+N borrowings in Czech (e.g. *body building* and *Bloody Mary*). N+N Anglicisms provoke prosodic tension, as native Czech prosody does not, under normal conditions, allow the destressing of the last word in a noun phrase. An analysis of recorded N+N and Adj+N borrowings shows that Czech speakers adopt the foreign stress pattern without difficulty, treating most N+N phrases as compounds. This pattern also spills over to Adj+N phrases. Spelling reflects pronunciation in the sense that hyphenated or single-word spellings are more common in Czech than in English and guarantee a better matching between spelling and pronunciation. Lexical frequency does not seem to have an influence on stressing, but phrase length is correlated with stress probability. The imported prosodic pattern seems to have gained an important position in Czech loanword phonology, despite quite considerable inter-speaker variability in its usage.

Index Terms: stress, primary stress, noun phrases, Czech, English, loanwords, Anglicisms

1. N+N structures in English and their equivalents in Czech

Expressions like *trade union*, *orange juice* and *couch surfing* are noun phrases with a nominal pre-modifier [1]. Unlike Adj+N phrases (*European Union*, *cold juice*, *easy surfing*), N+N structures mostly have a left-headed stress pattern [2], [3], which is a feature typical of English compounds [4]. The close relationship between N+N phrases and compounds is also reflected by spelling variability (*babysitter/baby-sitter*, *air conditioner/air-conditioner*). The prosodic left-headedness of these structures is to be considered the default pattern [2], despite numerous exceptions (*Iron Curtain, London Bridge*). Under focus, the stress position is primarily determined by information structure [5]: e.g. *orange JUICE, not orange CAKE.*

Juxtaposed nominal premodifiers are not found in native Czech grammar. The most typical grammatical structures that correspond to English N+N phrases are the following:

1) Adj+N: *Prague Castle – Pražský hrad* (pre-modifying adjective);
2) N+N$_{gen}$: *faculty staff – zaměstnanci fakulty* (post-modifying genitive);
3) N+PrepP: *car keys – klíče od auta* (post-modifying prepositional phrase);
4) Compound: *speed skating – rychlobruslení.*

All these structures are, under normal conditions, stressed on every lexical word. When placed at the end of a prosodic phrase, the nuclear accent is realised on the last lexical word [6]: e.g. *Seznámil se s několika zaměstnanci FAKULTY*. This is not the case in the corresponding English sentence, where the nuclear accent is, for grammatical reasons, on the penultimate word: *He met some of the FACULTY staff.*

The aforementioned differences imply prosodic tension in cases where the English N+N structure interferes with Czech grammar. This can happen in three cases:

1) An N+N phrase is borrowed from English as a whole (*body building*);
2) An N+N phrase is borrowed and partly translated (*shopping centrum*);
3) A new N+N phrase is constructed with a proper name as the first element (*Sconto nábytek* 'Sconto furniture').

The structure under 3), which is an alternative to the more regular structure with the proper name in second position (*nábytek Sconto*), is criticised as grammatically intrusive [7].

The Phonological Database of Anglicisms in Czech [8] contains almost 8% of items longer than one word (367 out of a total number of 4,689; only entries containing at least one space were counted). This set is structured as follows:

1) N+N phrases: 197   (*pet shop, melting pot*)
2) Adj+N phrases: 99   (*happy hour, sexual harassment*)
3) Other structures: 71   (*voice over, take away, no comment, fish and chips*)

These lexical borrowings often have variable spelling: *second hand/ second-hand/ secondhand, cash flow/cash-flow/ cashflow, up to date/up-to-date*.

No empirical evidence is available so far as to the stress pattern of N+N borrowings from English. Informally, it can be assumed that the English stressing rule (i.e. treating the phrase as a compound) is commonly used, despite the aversion of Czech prosody to nominal phrases with a destressed final word.

2. Aim and hypotheses

The purpose of this paper is to describe the stress patterns for N+N borrowings in Czech. The following eight hypotheses will be tested:

1) Czech speakers accept the left-headedness of N+N phrases to a certain degree, thus importing a foreign prosodic pattern.

2) They follow the general tendency of English stressing, treating N+N phrases as compounds, and Adj+N phrases as two-word structures.

3) They are not able to detect exceptions to the general pattern (e.g. *Iron Courtain*).

4) There is correlation between pronunciation and spelling in the sense that hyphenated or single-word spellings are more likely to be pronounced with the second element destressed.

5) Czech speakers are sensitive to frequency in the sense that frequent phrases are more likely to behave like compounds (i.e. have the second element destressed).

6) They are sensitive to word length in the sense that phrases containing longer words are less likely to behave like compounds (i.e. have the second element destressed).

7) There is considerable individual variability among speakers with regard to second-element destressing.

8) Speakers with a better knowledge of English are more likely to follow the English stress pattern.

3. Data and recording

N+N and Adj+N phrases (197 and 99 respectively, cf. Section 3) were filtered from the Phonological Database of Czech Anglicisms. The database contains lexical items retrieved from Czech dictionaries, with all orthographic forms attested. Two orthographic categories were established: a) two-word spelling only (e.g. *action painting, bloody Mary*) b) variable spelling, i.e. two-word + hyphenated, two-word + single-word, or two-word + hyphenated + single-word (e.g. *border kolie/ borderkolie, black box/black-box/blackbox*).

The combination of two grammatical structures and two spelling patterns gives four different categories of entries. For each category, eight entries were selected in alphabetical order (starting from the letter "a") until the required number was reached. Repetition of the first element in several entries (e.g. *black box, black list, black metal*) was avoided. The list of entries is given in Table 1. Two phrases have partly nativised orthography (*bendlington teriér* 'Bedlington Terrier' and

border kolie 'Border Collie'), one phrase is a pseudo-Anglicism (*babybox* 'baby hatch'), and one has undergone grammatical conversion (*air-condition* is used as a noun in Czech).

The stress patterns of the original English phrases were primarily obtained mostly from dictionaries of British English [9], [10], [11]. In nine cases, the phrases were not found in any of the dictionaries, and their stress patterns were thus obtained from a native speaker of British English. The frequency of each phrase was established in the SYN Corpus of written Czech (version 5; 3,836 million words; [12]) through a manual search for all possible spelling variants in all inflected forms.

Each entry was embedded in a short sentence and always placed in final position. For entries with variable spelling only the two-word spelling variant was used. The 32 sentences were supplemented by 8 distractors containing one-word Anglicisms. The following examples illustrate the sentences used:

Nevěděl, co znamená air mail. 'He didn't know what air mail meant.' (N+N, two-word spelling only)

Nikdy neodmítla sklenku bloody Mary. 'She never refused a glass of Bloody Mary.' (Adj+N, two-word spelling only)

Dává si záležet na své image. 'He/she cares about his/her image.' (distractor)

The final set of 45 sentences was randomised and recorded by 20 native speakers of Czech (students of Translation Studies, aged 18–24) whose knowledge of English ranged from B1 to C2 (according to CEFRL [13]).

I classified each of the 640 items (32 phrases x 20 speakers) either as left-headed (compound-like: i.e. with the second element destressed), or as right-headed (compliant with the default stressing rule), on a perceptual basis.

Table 1: *List of tested entries including all spelling forms found in consulted dictionaries. Each entry is provided with the stress pattern of the original English phrase (L – left-headed, R – right-headed), and with information about frequency in the SYN Corpus (version 5).*

Structure	Orthography	
	Two-word only	Variable
N+N	*action painting* (L; 36)	*air condition, air-condition* (L or R; 94)
	air mail (L; 27)	*army shop, army-shop* (L; 326)
	ballad opera (L; 15)	*art director, art-director, artdirector* (L; 1,039)
	beat generation (L; 744)	*baby box, baby-box, babybox* (L; 18,663)
	bedlington teriér (R; 21)	*beach soccer, beachsoccer* (R; 462)
	bone bed (L; 0)	*body building, bodybuilding* (L; 254)
	brain drain (L; 62)	*border kolie, borderkolie* (R; 1,249)
	brass band (R; 1 258)	*break dance, break-dance, breakdance* (L; 3,531)
Adj+N	*Basic English* (R; 29)	*acid jazz, acid-jazz* (R; 547)
	bloody Mary (R; 489)	*big beat, big-beat, bigbeat* (L; 4,226)
	blue beat (L; 8)	*black box, black-box, blackbox* (R; 421)
	circular pitch (R; 0)	*common law, common-law* (R; 180)
	closed shop (R; 0)	*fair play, fair-play, fairplay* (R; 17,650)
	common sense (R; 137)	*fast back, fastback* (L; 127)
	compact disc (R; 46)	*free jazz, free-jazz, freejazz* (R; 1,097)
	cool jazz (R; 137)	*good bye, good-bye, goodbye* (R; 2,794)

4. Analysis of the results

4.1. Hypothesis 1: Tolerance towards the imported prosodic pattern

64% of the recorded occurrences were realised with the second word destressed (74% in N+N phrases, and 54% in Adj+N phrases). This indicates that the imported prosodic pattern is anything but unusual. It prevails in N+N phrases, and its frequency is also remarkably high in Adj+N phrases.

4.2. Hypothesis 2: Difference between N+N phrases and Adj+N phrases

The difference between the aforementioned percentages (74% vs. 54%) is significant (t-test, p = 0.03).

Of the 16 analysed N+N phrases (cf. Figure 1), seven have a stable prosodic structure typical of compounds (*air condition, brass band, air mail, body building, army shop, break dance, baby box*), seven have a 50–80% likeliness to be treated as compounds (*art director, ballad opera, action painting, beach soccer, brain drain, border kolie, bone bed*), and two are treated mostly as sequences of two words (*beat generation, bedlington teriér*).
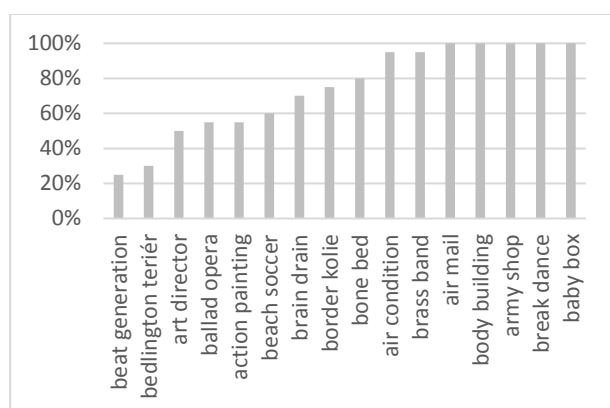

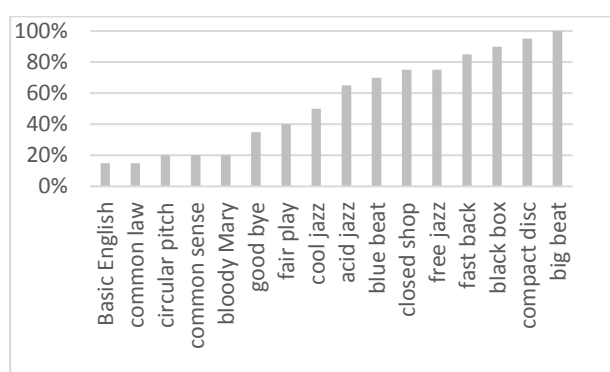Figure 1: *Tendency to second-word destressing in N+N phrases.*

Figure 2: *Tendency to second-word destressing in Adj+N phrases.*

In Adj+N phrases (cf. Figure 2), eight of the 16 items are more than 50% likely to be realised with the second word destressed. This observation is surprising, as only three of sixteen phrases are left-headed in English (cf. Table 1). It

seems that despite the anomalous nature of second-word destressing in Czech, there is tolerance towards this stress pattern in borrowings from English, and that this pattern spills over from N+N phrases, where it is habitual in English, to Adj+N phrases.

4.3. Hypothesis 3: Compliance with unusual stressing in English

Most of the tested N+N phrases are left-headed in English, four of them are described as right-headed (*Bedlington Terrier, brass band, beach soccer, Border Collie*), and one as either left-headed or right-headed (*air-condition*). I tested the difference in stressing between these two subsets in corresponding Czech loanwords by means of a t-test. Items that are right-headed in English also have a slightly lower percentage of left-headed realisations in Czech (71% vs. 76%), but this difference is not significant (p = 0.37). It seems that Czech speakers are not sensitive to patterns that deviate from the general rule. This is not surprising because prosodic right-headedness in English N+N phrases is triggered by fine semantic criteria [2] and is influenced by analogy.

In English Adj+N phrases, the right-headed pattern is the norm. Three phrases were identified as left-headed in the sample (*blue beat, big beat, fast back*). These items also show a stronger tendency to right-headedness in Czech (85% vs. 47%), which is statistically significant (p = 0.03). It is questionable whether this tendency is due to the influence of the source language; another plausible explanation is the influence of spelling and word length (cf. Sections 4.4 and 4.6). In addition, because of the low number of observations it is not possible to make more general conclusions.

4.4. Hypothesis 4: Correlation with spelling

All items were presented as non-hyphenated two-word structures for the recording. However, some items have alternative spelling, either hyphenated or single-word (cf. Table 1). Since hyphenation and single-word spelling may reflect the tendency of the word to behave like a compound, and language users possibly stock both kinds of information – phonetic and graphic – in their memory, I tested the propensity to second-word destressing for the two spelling types. This propensity is 55% in phrases with two-word spelling only and 74% in phrases with alternative spelling, and the difference was statistically significant (t-test, p = 0.03).

4.5. Hypothesis 5: Sensitivity to frequency

Next I tested whether frequent collocations are more likely to be treated like compounds. The correlation coefficient (Pearson) between the frequency of each phrase and the tendency of the second element to be destressed is 0.11, which is a very weak positive correlation.

4.6. Hypothesis 6: Sensitivity to word length

The last tested parameter was word length. I correlated the syllabic length of the first word, the second word, and the entire phrase with the tendency of the second element to be destressed, and arrived at the following correlations:

Length of the first word:	−0.37
Length of the second word:	−0.27
Length of the entire phrase:	−0.41

All three measures are weakly correlated with stressing: the higher the number of syllables in an expression, the lower the probability of the second element being destressed.

4.7. Hypothesis 7: Variability among speakers

Figure 3 shows the overall frequency of second-word destressing for individual speakers. The mean is 64% (cf. section 1.1), and the percentages are rather evenly distributed between the two extremes (41% and 94%). The range is 53%, which indicates considerable inter-speaker variability.
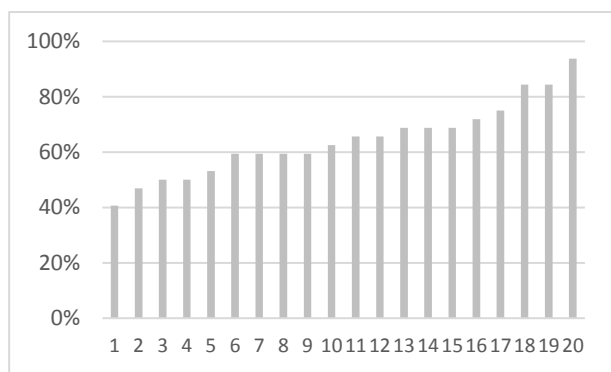

Figure 3: *Tendency to second-word destressing in individual speakers.*

4.8. Hypothesis 8: Influence of proficiency in English

For each speaker, I calculated the number of phrases that were stressed in compliance with English stressing, and correlated this number with the speaker's declared proficiency in English (B1–C2). The correlation coefficient (0.53) indicates a moderate positive correlation.

5. Conclusions

The question of stress in N+N borrowings from English was addressed from several complementary angles. First, I demonstrated that second-element destressing in borrowed N+N phrases is rather frequent (74% of the recorded occurrences), although not a part of native prosody, and should thus be considered an imported prosodic pattern.

The difference between borrowed N+N and Adj+N phrases, while being statistically significant, is much less clear-cut than in English. Czech speakers are less sensitive to the grammatical structure of the phrase, and tend to collapse both structures into a large category of "two-word English borrowings". The 54% part of Adj+N phrases with the second word destressed indicates a prosodic spillover from the N+N category, which means that the imported prosodic pattern spreads to other structures than those of which it is characteristic in the donor language.

For phrases that have exceptional stress patterns in English (right-headed N+N and left-headed Adj+N), I was not able to make any convincing conclusions due to the low number of observations.

Spelling is correlated with second-word destressing in the sense that hyphenated and single-word spelling favours the perception of the phrase as a compound, and thus increases the probability of second-word destressing. Czech spelling conventions seem to be more variable than in English: apart from two-word spelling, which was a necessary condition for the phrase to be included in the sample, many entries, both in the N+N and Adj+N categories, are used with hyphenated and single-word spelling, unusual in English (e.g. *borderkolie, fairplay*). This orthographic normalisation increases the regularity of the spelling-pronunciation interface.

While lexical frequency does not seem to be correlated with stressing, syllabic length turned out to be a relevant factor: longer phrases are more likely to be pronounced with the second word stressed.

The analysis of inter-speaker variability showed that the tendency to second-word destressing in borrowed N+N and Adj+N is not uniform across speakers. This observation is not surprising, as it has been shown that the phonological subsystem of Czech Anglicisms – and loanwords in general – is less stable than the native phonological system [14], [15]. Furthemore, it can be argued that stress distribution is partly influenced by the speakers' knowledge of English, as there is a positive, though not strong, correlation between their proficiency in English and the tendency to transfer English stress patterns into Czech.

7. References

[1]   S. Greenbaum, *The Oxford English Grammar*. Clarendon Press, 1996.

[2]   M. Liberman and R. Sproat, "The Stress and Structure of Modified Noun Phrases in English," *Lexical Matters*, CSLI, pp.131–181, 1992.

[3]   P. Roach, *English Phonetics and Phonology*, CUP, 2006.

[4]   H. van der Hulst, B. Hendriks, and J. van de Weijer, "A survey of word prosodic systems of European languages," *Word Prosodic Systems in the Languages of Europe*, Berlin – New York: Mouton de Gruyter, pp. 425–475, 1999.

[5]   D. R. Ladd, *Intonational Phonology*, CUP, 2008.

[6]   T. Duběda, *Études de prosodie contrastive : le cas du français et du tchèque*, Prague: Karolinum, 2012.

[7]   R. Adam, "Anglicizace češtiny a českého komunikačního prostoru," *Přednášky z 55. běhu LŠSS*, pp. 21–36, 2012.

[8]   T. Duběda, "The phonology of anglicisms in Czech: Exploring extralinguistic factors," *12th Slavic Linguistics Society Annual Meeting. Book of Abstracts*, Ljubljana: ZRC SAZU, pp. 74–75, 2017.

[9]   D. Jones et al., *Cambridge English Pronouncing Dictionary*, 18th edition, CUP, 2011.

[10] Oxford Advanced Learner's Dictionary, https://www.oxfordlearnersdictionaries.com

[11] Cambridge Dictionary (English): https://dictionary.cambridge.org/dictionary/english

[12] M. Hnátková, M. Křen, P. Procházka, and H. Skoumalová, "The SYN-series corpora of written Czech," *Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC'14)*, Reykjavík: ELRA, pp. 160–164, 2014.

[13] *Common European Framework of Reference for Languages*. https://www.coe.int/t/dg4/Linguistic/Source/Framework_EN.pdf

[14] M. Romportl et al., *Výslovnost spisovné češtiny. Výslovnost slov přejatých*, Praha: Academia, 1978. [Pronunciation of standard Czech: Loanwords]

[15] T. Duběda, "Empirické mapování výslovnostního úzu u cizích slov," *Slovo a slovesnost 77*, pp. 123–142, 2016. [An Empirical Study of the Pronunciation of Foreign Words in Czech]