



Nestandardní adjektivní přirovnání v korpusových datech

Jaroslav Emmer (České Budějovice)

NON-STANDARD ADJECTIVAL SIMILES IN CORPUS DATA

The adjectival simile is a well-established multiword expression (MWE) type. Although it occurs in many languages, it remains an understudied phenomenon within phraseology. The relatively universal existence of this idiom (of comparison) indicates that it represents an extralinguistic concept. However, lexical representations of one concept may differ across languages, potentially giving rise to loan forms that may appear foreign (at least initially) in some languages.

This study investigates non-standard adjectival similes in Czech, specifically their occurrence in parallel corpus data where the source texts are not originally Czech. Firstly, using general and lemma-specific queries, the adjectival similes were mined from translated data in a parallel corpus (InterCorp v16 — Czech). Secondly, the extracted adjectival similes were searched for in a reference corpus (Syn v12) to check their frequencies in Czech texts.

The results show that around 10% (27 out of 262) of all Czech adjectival similes retrieved from InterCorp v16 are non-standard, which can be attributed to two reasons. Firstly, one of their components (an adjective or a noun) is replaced by a synonym, or the whole simile represents a fusion of two unique MWEs. Secondly, the lexical components indicate a foreign influence, usually through loan translation.

Corpus data typically constitute empirical evidence for the existence of MWEs and justify their inclusion in dictionaries. However, the present study of adjectival similes shows that frequency alone is insufficient evidence. A careful and methodical approach to the study of language data and the presentation of our findings is necessary to ensure we do not hastily institutionalise foreign MWEs without proper scrutiny.

KEYWORDS

adjectival simile, corpus data, transformation, standardisation, multiword expression

KLÍČOVÁ SLOVA

adjektivní přirovnání, korpusová data, transformace, ustálenost, víceslovné lexikální jednotky

DOI

<https://doi.org/10.14712/23366591.2025.2.2>

1. ÚVOD

Tato studie se věnuje nestandardním adjektivním přirovnáním v korpusových datech a důvodům jejich existence. Hlavním cílem tohoto výzkumu je poskytnout přehled nestandardních adjektivních přirovnání objevujících se ve zkoumaných korpusech a nabídnout interpretaci jejich možného původu. Mezi dílčí cíle patří sledování užitečnosti korpusových dat při sestavování slovníků, kde důležitou roli hrají povaha zdrojových textů, korpusová anotace a způsob jejího využití, reprezentativnost zvolených korpusů a zkoumané jazykové jevy.



Adjektivní přirovnání (např. *slепý jako patrona*) patří mezi ustálené víceslovné lexikální jednotky (dále VLJ), které jsou mnohdy specifické pro daný jazyk z hlediska dílčích komponentů. Nežádka ale dochází k tvorbě a užití neustálených forem, což bývá způsobeno dvěma hlavními faktory.

Prvním z důvodů existence nestandardních forem adjektivních přirovnání jsou atypické vnitrojazykové lexikální transformace (Čermák, 2007). Pokud budeme považovat srovnání (či porovnání) dvou entit za nadjazykový koncept, jeho manifestace může mít různé jazykové formy (i z hlediska struktury), jejichž míra ustálenosti se často liší. Některé formy dokonce nemusí být v daném jazyce standardizované, v takových případech zpravidla pozorujeme nahrazení původních komponentů synonymy, případně splynutí dvou různých VLJ.

Nestandardní formy však nemusí být pouze důsledkem řídkých transformací, nýbrž se může jednat o vliv cizích jazyků. Lexikální odlišnosti často vychází z historie a kulturních specifik daného jazyka, nicméně v dnešní době otevřené mezinárodní komunikace a mísení kultur (a jazyků) mnohem více dochází ke stírání jazykových hranic, a zejména mezinárodně používané jazyky jako angličtina nebo španělština do značné míry pronikají do ostatních jazyků ve formě výpůjček, kalků aj. To se do velké míry projevuje také v oblasti frazeologie, kde však může být cizí vliv zprvu méně nápadný, neboť citlivost rodilých mluvčích vůči jednoslovným výpůjčkám bude zpravidla vyšší než vůči těm v podobě VLJ. To se týká i adjektivních přirovnání, u nichž vliv cizích jazyků potenciálně dává vzniknout nestandardním formám, které obsahují komponenty pro češtinu netypické (např. *slепý jako netopýr*). Ač v žádném případě nelze vliv cizích jazyků omezit pouze na angličtinu, jedná se z hlediska vlivu o dominantní jazyk. Proto je jedním z důležitých faktorů v diskusi o nestandardních formách obecná otevřenost k užívání anglicismů v češtině (Klégr a Bozděchová, 2024, s. 29), jelikož sem lze zahrnout i toleranci struktur anglického původu, např. *mějte hezký den*.

2. PŘIROVNÁNÍ JAKO VLJ

Přirovnání (angl. *simile*) se vyskytují v mnoha jazycích, což naznačuje lidskou tendenci připodobňovat a srovnávat. Čermák (2007, s. 385) uvádí podobnost jako významný sémantický vztah (viz synonyma). Z hlediska konstrukční gramatiky (Langacker, 1987; Goldberg, 1995; Croft, 2007) jsou přirovnání ustálené idiomatické konstrukce. Název „přirovnání“ vyplývá z jejich sémantiky, protože představují přirovnání jedné entity ke druhé.

Norrick (1986, s. 46) vysvětluje, že přirovnání vznikají z potřeby popisovat entity analogicky, protože doslovný popis může být obtížný nebo nevystihuje záměr mluvčího. Používá termíny ‚digital‘ pro doslovný popis a ‚analogical‘ pro srovnání. Čermák (2007, s. 385) tyto jevy označuje jako ‚přímé‘ a ‚nepřímé‘ pojmenování. Přirovnání jsou nepřímá pojmenování, protože využívají analogie pro popis referenta. Liší se od běžných srovnávacích konstrukcí, kde pravá složka představuje prototypického nositele určité vlastnosti, která může být implikována nebo přímo vyjádřena.

Petkevič et al. (2020, s. 43) definují přirovnání takto: „ustálená VLJ, typicky tvořená slovesnou nebo adjektivní frází obsahující komparátor.“ Na základě formální



přítomnosti vlastnosti (*tertium comparationis*) rozdělujeme přirovnání na implicitní a explicitní (Čermák, 2007, s. 397). V anglické literatuře se někdy používají označení ‚opení‘ (otevřená) a ‚closed‘ (uzavřená), což vychází z „otevřenosti“ přisuzovaného *tertia*. Přirovnání obsahující explicitně vyjádřenou vlastnost jsou označována za adjektivní (*closed*), zatímco přirovnání s implikovanou vlastností patří do kategorie verbálních (*open*). Toto anglické rozdělení vychází z povahy centrálního komponentu frazému.

2.1 ADJEKTIVNÍ PŘIROVNÁNÍ

Typologické označení ‚adjektivní přirovnání‘ tradičně pojmenovává formální VLJ nebo konstrukci ve tvaru ADJ + *jak(o)* + SUBST (tzv. *jako*-forma). Adjektivní přirovnání vždy explicitně vyjadřují konkrétní přisuzovanou vlastnost, kterou u verbálních přirovnání není vždy snadné určit, jelikož se může jednat o celý soubor vlastností. Hanks (2004, s. 37) však uvádí, že žádná konkrétní vlastnost není pro pochopení nevyjádřených vlastností potřeba, jelikož při každodenní komunikaci rutinně spoléháme na vágní vyjadřování a rychlé, povrchní interpretace.

Z kognitivního hlediska pak lze na adjektivní přirovnání nahlížet jako na koncept, který může být reprezentován různými VLJ. V takovém případě se jedná o těsné spojení nositele a jemu přisuzované vlastnosti, např. [břitva] + [ostrý], na konceptuální úrovni. Tento koncept se pak objevuje ve více konstrukcích, např. *mít jazyk (ostrý) jako břitvu, být (ostrý) jako břitva* nebo *ostrý jako břitva* (srov. ‚superlemma‘ Petkevič et al., 2020, s. 40; ustálené varianty Čermák, 2007, s. 389). Prototypickou formou adjektivních přirovnání v češtině je bezesporu *jako*-forma (v angličtině *as*-form). V angličtině se navíc objevuje tzv. ‚složená forma‘, např. *razor sharp* [břitva ostrý], což je dáno morfologickou typologií angličtiny — v češtině takové formy neexistují.

Variantnost nemá obecnou platnost, různé koncepty adjektivních přirovnání tak mohou disponovat různými paradigmaty, resp. objevují se ve více transformacích s různou mírou ustálenosti (Čermák, 2007, s. 389). I z toho důvodu mohou být některá adjektivní přirovnání mluvčím známá, přestože se v jazyce vyskytují ve své prototypické *jako*-formě vcelku ojediněle (Moonová, 2008, s. 33).

Zmiňme také jeden problém týkající se kategorizace přirovnání. Rozlišení mezi verbálním a adjektivním přirovnáním v češtině zpravidla vychází z (ne)přítomnosti adjektiva. Čermák (2007, s. 387) adjektivní přirovnání (*lehký jako pírkó*) považuje za redukované formy obecnějších verbálních struktur (*být lehký jako pírkó*). Otázkou však zůstává, zda je v rámci *usage-based* modelu příhodné nahlížet na obecnější verbální struktury jako na prototypické formy přirovnání. Redukce explicitních přirovnání na implicitní, např. (*být*) *paličatý jako mezek* na *být jako mezek*, není v češtině problematická, jelikož existence *jako*-formy představuje výchozí interpretaci přisuzované vlastnosti ve verbální struktuře. V angličtině však mají standardní verbální (*look like a clown*) a adjektivní přirovnání (*sharp as a razor*) odlišné komparátory a redukované formy typu **be as a razor* jsou nestandardní, stejně tak se běžně neobjevují ani formy jako **sharp like a razor*. Z toho důvodu nelze v angličtině adjektivní přirovnání považovat za pouhou redukci verbálních forem, což naznačuje, že se přinejmenším z kognitivního hlediska jedná o dva různé fenomény.

2.2 FUNKCE AJDEKTIVNÍCH PŘIROVNÁNÍ

Adjektivní přirovnání plní několik funkcí. Zaprvé se jedná o funkci srovnávací, která vychází ze sémantického vztahu podobnosti (Čermák, 2007, s. 385). Moonová (2011) však poukazuje na fakt, že nemalý počet přirovnání srovnává věci značně odlišné. Tento jev lze pozorovat i u metafory, kde nezřídka dochází k (implicitnímu) srovnání dvou velmi odlišných entit, které pouze disponují jednou podobnou vlastností, jež však ani u jedné z entit mnohdy nepatří k vlastnostem prominentním nebo dokonce danou entitu definujícím (Barnden, 2016).

V druhé řadě plní adjektivní přirovnání funkci intenzifikační. Moonová (2008, s. 7) této funkci přisuzuje pragmatickou povahu, na místě jsou však také úvahy o gramatikalizaci komparátoru s komparátem. Pravá složka přirovnání, typicky ve formě *jak(o) + SUBST*, bývá ustálená do takové míry, že srovnávací funkce zůstává upozaděna, zejména pak v případech typu *drahý jako prase*, které reprezentují pokročilý stupeň gramatikalizace a skutečné srovnání už nepřipadá v úvahu (Emmer, 2020).

Mezi další funkce pak patří humor nebo ironie (Norrick, 1986, s. 44; Moonová, 2008, s. 9; Veale, 2013), ale také např. funkce poetická (Gargani, 2016) a specifikace (Norrick, 1986, s. 47). Obecně pak můžeme konstatovat, že (adjektivní) přirovnání patří k ustálenému typu VLJ, jejichž užití ozvláštňuje promluvu (příp. ji aktualizuje). Přirovnání bude téměř vždy vnímáno jako stylisticky příznakové, ať už se jedná o (spíše) knižní výrazy typu *bledý jako kámen* nebo více hovorové případy typu *blbej jak tágo*.¹

3. METODOLOGIE

Pro *usage-based* model představuje práce s korpusem výchozí metodologii, jelikož se jedná o přístup zakládající si na autentických datech. Těžba dat proběhla ve formě vyhledávání specifické ustálené *jako*-konstrukce.² Způsob vyhledávání byl komplexní, od obecných strukturních dotazů (např. [tag="A.*"] [lemma="jako"] [tag="N.*"]) pro získání seznamu adjektivních přirovnání po specifické vyhledávání lemmat. Tímto způsobem bylo dosaženo skutečné frekvence jednotlivých adjektivních přirovnání, jelikož u obecnějších dotazů nezbyvá než se spoléhat na přesnost morfologické anotace. Pro těžbu adjektivních přirovnání byl použit paralelní korpus InterCorp v16 — čeština (dále jen InterCorp), který po odfiltrování původně českých textů obsahuje 491 019 657 tokenů psaného jazyka. Spodní hranice absolutní frekvence byla stanovena na $n=10$ (relativní frekvence 0,2 i.p.m.) z důvodu relevance jednotlivých typů z hlediska výskytu i praktičnosti při zpracování dat.

Výsledný seznam adjektivních přirovnání považuje každý token za samostatnou jednotku (při kvantitativním zpracování), a to i v případech, kdy se tokeny liší pouze

1 Přirovnání zpravidla nebude označováno za stylově neutrální.

2 V angličtině je takovýto top-down druh korpusového zkoumání označován termínem ‚corpus-based‘.



formou komparátoru (*jako/jak*). Formální varianty způsobené flexí substantivní složky a shodou gramatického rodu adjektiva s referentem nejsou považovány za odlišné varianty.

3.1 POSUZOVÁNÍ NESTANDARDNOSTI ADJEKTIVNÍCH PŘIROVNÁNÍ

Posuzování nestandardnosti vytěžených adjektivních přirovnání vychází z jejich frekvencí v referenčním korpusu Syn v12 obsahujícím 6 063 263 030 tokenů (dále jen Syn), kde byl použit pozitivní filtr pro původně české texty. Vyhledávání adjektivních přirovnání v Synu proběhlo individuálně pomocí lemmatických dotazů vycházejících z tokenů vytěžených v InterCorpu. Syn představuje robustní korpus, který sice není reprezentativní (dominují publicistické texty), představuje však jeden z největších dostupných korpusů psané češtiny. Pro potřeby zkoumání méně frekventovaných VLJ, mezi něž adjektivní přirovnání bezesporu patří, je nezbytné pracovat s většími daty. Navíc, účelem této studie není představit reprezentativní minimum (případně optimum) adjektivních přirovnání, vliv publicistických textů tak nepůsobí kontraproduktivně (srov. Kopřivová a Šichová, 2023, s. 95). Původně byla zvažována také data z mluveného korpusu ORAL v1, nicméně tento korpus o velikosti šesti milionů tokenů neobsahuje dostatečný počet adjektivních přirovnání.

Za nestandardní adjektivní přirovnání jsou považována ta, která mají v InterCorpu statisticky významně vyšší absolutní frekvenci než v Synu. Pro výpočet statistické významnosti byl u každého přirovnání původně použit chí-kvadrát test. Avšak s přihlédnutím k faktu, že adjektivní přirovnání jsou obecně mnohem četnější v InterCorpu než v Synu, byl následně použit alternativní výpočet poměru mezi frekvencemi v InterCorpu a Synu. Posuzování (ne) poměru frekvencí je velmi konzervativní (inkluzivní) a vychází ze dvou stanovených kritérií: (i) přirovnání má více než dvojnásobně vyšší frekvenci v InterCorpu; (ii) absolutní frekvence v Synu nepřesahuje $n=10$. Nestandardnost vyplývající z těchto kritérií může být způsobena různými faktory (viz Sekce 4 a 5).

Adjektivní přirovnání s komparátorem *jak* jsou považována za nestandardní pouze tehdy, pokud jsou nestandardní i jejich prototypické *jako*-formy. Formy s *jak* jsou často expresivnější a vyskytují se více v hovorovém jazyce. Korpusová data ukazují, že *jak* je obecně méně časté než *jako*, ale jeho výskyt naznačuje ustálenost přirovnání v češtině. Existence *jak*-formy může tedy být považována za doklad standardizace přirovnání. Frekvence *jak*-forem jsou však obecně nižší a nelze je samostatně považovat za směrodatné kritérium při diagnostikování ustálenosti.

4. DATA

Seznam přirovnání vytěžených z InterCorpu čítá $n=8313$ (16,93 i.p.m.) tokenů, reprezentovaných 262 unikátními typy. Jak ilustruje Tabulka 1, tentýž soubor přirovnání má v Synu frekvenci $n=19041$ (3,14 i.p.m.), což poukazuje na výrazně vyšší frekvence těchto adjektivních přirovnání v InterCorpu s přihlédnutím k velikostem korpusů.

Korpus	Absolutní frekvence	Relativní frekvence (i.p.m.)
InterCorp v16	8313	16,93
Syn v12	19041	3,14

TABULKA 1. Frekvence adjektivních přirovnání v InterCorp v16 a Syn v12

Frekvenční seznamy obou korpusů ukazují, že existuje určitý soubor ustálených adjektivní přirovnání s vysokou frekvencí v poměru k ostatním typům. Můžeme tedy hovořit o jakémisi minimu adjektivních přirovnání (Emmer, 2024, s. 34), se kterými se mluvčí relativně běžně setkávají v současné češtině. Takové minimum pak jistě patří do výuky češtiny, ať už jako mateřštiny nebo cizího jazyka.

jako-forma	InterCorp v16	Syn v12
ostrý jako břitva	209	989
černý jako uhel	207	334
bledý jako stěna	194	177
jasný jako facka	193	299
volný jako pták	192	395
tvrdý jako kámen	191	323
bílý jako sníh	168	313
pevný jako skála	164	448
studený jako led	156	92
bílý jako stěna	154	252
zdravý jako řípa	152	625
lehký jako pírkó	150	371
jasný jak facka	149	168
silný jako býk	124	73
bílý jako křída	122	76

TABULKA 2. Seznam nejčastějších přirovnání v InterCorp v16 a jejich frekvence v Syn v12.

Tabulka 2 obsahuje seznam patnácti nejfrekventovanějších adjektivních přirovnání v InterCorpu spolu s jejich frekvencemi v Synu. Vyšší frekvence v InterCorpu mohou naznačovat vyšší míru expresivity daných adjektivních přirovnání, což může být jeden z důvodů jejich častějšího výskytu v datech obsahujících filmové a seriálové titulky. Publicistické texty (většina dat v Syn) taktéž zahrnují stylisticky příznakové výrazy, avšak míra expresivity bude do jisté míry omezená, což může značně redukovat frekvence hovorových přirovnání typu *blbej jak tágo* nebo *zlitej jak dobytek*.

4.1 NESTANDARDNÍ PŘIROVNÁNÍ

Celkem 27 položek (10,31 %) z vytěženého seznamu adjektivních přirovnání z InterCorpu bylo označeno za nestandardní. Jedná se o tyto formy:



bílý jako duch / bledý jako mrtvola / bledý jako vosk / černý jako smůla / čistý jako slza / hořký jako žluč / krásný jako slunce / kulatý jako měsíc / lehký jak facka / lehký jako facka / nehybný jako kámen / paličatý jako mezek / rovný jako šíp / rudý jako oheň / silný jako kůň / slabý jako kotě / slepý jak netopýr / slepý jako netopýr / světlý jako len / svěží jako sedmikráska / tuhý jako špalek / tvrdohlavý jako osel / velký jako Texas / volný jako ptáček / vysoký jako strom / zdravý jako kůň / ztuhlý jako prkno

Obecně lze pozorovat dva hlavní důvody nestandardnosti, které zjednodušeně označme za jazykově interní (tj. variantnost v rámci jednoho jazyka bez cizích vlivů) a externí (tj. formy motivované vlivem cizího jazyka).

4.2 ŘIDŠÍ TRANSFORMACE

Řidší transformace přirovnání se řadí mezi jazykově interní fenomén. Standardní adjektivní přirovnání mnohdy disponují variabilitou, především ve formě synonymních alternativ v adjektivním slotu (*vysoký/velký jako hora*), případně také ve slotu nominálním (*rychlý jako blesk/vítr*), s různou mírou ustálenosti. Takového alternace mohou mít odlišné významy, případně preferovat konkrétní referenty, zpravidla se však jedná o synonymní VLJ.

Pak jsou tu však případy, kde dochází k nahrazení původního adjektiva synonymem, které vyústí v nestandardní VLJ. V případě adjektivního slotu (*tertia*) např. původní formu *jednoduchý jako facka* vystřídá méně ustálená forma *lehký jako facka*. Tato transformace pravděpodobně nebude působit cize pro rodilé mluvčí češtiny, jelikož se kognitivně jedná o ustálený koncept (řekněme např. „dát facku je snadné“ nebo „každý chápe význam facky“), pouze jeho lexikální reprezentace může být vnímána jako řidší. Dalším příkladem tohoto typu je *paličatý jako mezek*. Data opět naznačují, že se jedná o nestandardní transformaci přirovnání *tvrdohlavý jako mezek*, kde dochází k nahrazení ustáleného adjektiva řidší formou (*tvrdohlavý* → *paličatý*). V tomto případě také přispívá ustálená metafora *mezek paličatý*. Nabízí se také pohled, že přirovnání *paličatý jako mezek* představuje ustálenou, stylově nižší variantu. Tuto interpretaci nelze ve zkoumaných datech ověřit, ale vychází z jazykové intuice rodilých mluvčích češtiny.

K transformaci dochází také v nominálním slotu (komparátu), např. ustálené *volný jako pták* nahradí forma s deminutivem *volný jako ptáček*. Dalším příkladem je *tuhý jako špalek*, kde data naznačují, že jde o řidší transformaci ustáleného *tuhý jako prkno*. Nakonec sem patří i forma *černý jako smůla*, přičemž ustálenější variantou tohoto přirovnání na základě dat je *černý jako smola*. Zde dochází k nahrazení původního substantiva *smola* častější synonymní formou *smůla*. V posledním případě je opět důležité vzít v potaz povahu zkoumaných dat (a do jaké míry odráží jazykovou realitu), neboť *smola* působí v současně češtině zastarale. Varianta *černý jako smůla* tak může být novou prototypickou formou.

Některé výše uvedené případy mohou do jisté míry souviset s kontaminací frazémů (srov. Kopřivová, 2023, s. 34–35), respektive jejich fúzí. Obecně však není u adjektivních přirovnání snadné posoudit, do jaké míry je nová (neustálená) forma výsledkem fúze a kdy se jedná spíše o transformaci prostřednictvím synonyma,



např. *tvrdohlavý jako osel*. Data naznačují, že ustálená forma tohoto přirovnání je *tvrdohlavý jako mezek*. Vzhledem k taxonomickému spojení mezi oslem a mezkem se jedná o velmi blízká pojmenování zvířat, která navíc nemusí být pro běžného člověka snadno rozeznatelná. Na tvrdohlavost můžeme také nahlížet jako na negativní vlastnost charakteru. Výrazy *chovat se jako osel* nebo jednoduché označení *osel* taktéž implikují negativní vlastnost(i), což může vyústit v kontaminaci (*chovat se jako osel + tvrdohlavý jako mezek* → *tvrdohlavý jako osel*), neboť tvrdohlavost můžeme vnímat jako formu (příp. manifestaci) hlouposti. I v případech *tuhý jako špalek* se nabízí fúze dvou různých forem, tedy *spát jako špalek* a *tuhý jako prkno*.

4.3 CIZÍ VLIV

Cizí vliv představuje externí důvod pro existenci nestandardního přirovnání. Nesporný vliv zkušenosti a kvality práce překladatele na výsledný překlad není třeba dlouze diskutovat.³ Některé případy nestandardních adjektivních přirovnání jistě budou výsledkem nekvalitního překladu ve formě kalkování, nicméně důvody existence těchto forem mohou být komplexnější. Za jeden z důležitých faktorů lze považovat překladatelovu znalost (cizího) zdrojového jazyka a vlastní mateřštiny. Vedle očekávané kvality překladu zde také může hrát roli zvýšená tolerance ke konstrukcím ve zdrojovém jazyce, jelikož se intuitivně jedná o „známou“ konceptualizaci (např. dnes již poměrně ustálené výrazy *mějte hezký den* nebo *diskutovat problém*). Nutno však konstatovat, že translátologické otázky a téma kvality překladu se všemi relevantními aspekty jsou nad rámec této studie a vliv jednotlivých faktorů by zajisté byl předmětem samostatného zkoumání. Z toho důvodu zde cizí vliv představuje velmi obecný důvod existence nestandardních forem, potenciálně zahrnující mnoho různých dílčích jevů.

Do této „kategorie“ můžeme zařadit hned několik přirovnání z vytěženého vzorku. V první řadě se jedná např. o formu *slepý jako netopýr*. Komparátum *netopýr* se vyskytuje v angličtině v běžně užívaném přirovnání *blind as a bat*, pro češtinu jsou však standardními formami *slepý jako patrona*, případně *slepý jako krtek*. Dalším případem cizího vlivu je *velký jako Texas*. Přirovnání s adjektivem *velký* obecně představují problém, jelikož není vždy jednoznačné, zda se jedná o ustálené přirovnání, nebo o doslovné ad hoc srovnání (srov. např. *velký jako hrách*, *velký jako ruka* nebo *velký jako dům*). Data z anglických korpusů ukazují, že v angličtině se v případě *velký jako Texas* opravdu jedná o ustálené přirovnání (*big as Texas*), kde skutečná velikost Texasu nehraje roli (Emmer, 2023). V češtině je tato forma obsahující anglické proprium zjevně nestandardní a vychází z doslovného překladu. V neposlední řadě sem patří také *rovný jako šíp* (pravděpodobně z anglického *straight as an arrow*). Standardní česká přirovnání mají formu *rovný jako pravitko* nebo *rovný jako svíčka* (příp. i *svíce*).

³ Tento problém je čím dál více upozadován s nástupem mechanického překladu, zejména ve formě překladů generativními modely.



5. DISKUSE

Data z InterCorpu a Synu ukazují značný nepoměr celkových frekvencí adjektivních přirovnání, což pramení ze tří skutečností. Zaprvé, data ve srovnávaných korpusech nejsou homogenní. Mnoho textů v InterCorpu jsou titulky k filmům a seriálům, jedná se tedy o původně mluvený jazyk. Některá adjektivní přirovnání mohou být typičtější pro mluvený jazyk, což by vysvětlovalo jejich vyšší frekvenci v InterCorpu. To však nelze označit za obecný trend, jelikož adjektivní přirovnání se obecně vyskytují spíše v psaných textech pro širší veřejnost. Moonová (2008, s. 22) dokonce přímo zmiňuje úzkou souvislost jako-forem s publicistickými texty, které tvoří většinu dat v Synu. Z toho vyplývá, že pokud se konkrétní přirovnání v Synu vyskytuje s výrazně nižší relativní frekvencí než v InterCorpu, dané přirovnání pravděpodobně není úplně standardní.

Zadruhé, frekvence v Synu se týká pouze přirovnání vytěžených z InterCorpu, nejde tudíž o skutečnou frekvenci adjektivních přirovnání (jako typu VLJ) v celém korpusu. To však platí i pro InterCorp, jelikož byl použit cutoff ($n \geq 10$). Cílem této studie není vyčerpávající výčet adjektivních přirovnání v InterCorpu, nýbrž poukázat na vyšší frekvenci nestandardních forem v tomto korpusu, k čemuž poslouží i takto limitovaný vzorek.

A zatřetí, několik typů adjektivních přirovnání v InterCorpu nejsou standardními formami v češtině, jak naznačují data ze Synu. V některých případech se jedná o překlady z cizích jazyků, v jiných lze pozorovat nestandardní lexikální transformace, případně pak kontaminace frazémů. K (ne)standardnosti je však potřeba přistupovat velmi citlivě. Adjektivní přirovnání, až na pár výjimek, obecně disponují nižšími frekvencemi, přestože jsou v jazyce pevně ukotvena (Moonová, 2008, s. 29). Korpusová data nám mohou poskytnout informace o četnosti konkrétních typů adjektivních přirovnání, jejich absenci v korpusových datech však nelze vždy jednoznačně interpretovat jako nestandardnost, nebo dokonce úplné vymizení ze současného jazyka.

Tématem je taktéž autenticita češtiny, což úzce souvisí s texty v korpusu Syn. Většina textů je sice označena jako původně české (v metadatech), u beletrie jde ale často o překlady (viz evidentní původ jednotlivých děl). Publicistické texty jsou však problematictější, jelikož překlad u nich není deklarovaný, přesto frazeologismy v mnoha publicistických textech naznačují (někdy velmi silně), že se jedná a nepůvodní české texty. Mezi potenciální doklady tohoto jevu patří právě i nestandardní adjektivní přirovnání. Schematické přejímky (tj. cizí konstrukce s českými lexikálními komponenty) budou zpravidla méně nápadné než jednotlivá slova v podobě (pseudo)anglicismů.

Důležitým tématem je také role korpusových dat v lexikografii. Dnešní (online) slovníky nestaví pouze na jazykové intuici autorů a primárně se spoléhají na rozsáhlá data z jazykových korpusů, přičemž odpadá omezený rozsah tištěné podoby, což umožňuje inkluzivnější přístup. Nepochybně má každý kvalitní slovník svá kritéria pro to, z jakých dat čerpá, přesto zde přetrvává problém „míry“ deskripce a spolehlivosti uplatňovaných kritérií. Jakkoliv se může zdát volba InterCorpu pro tvorbu slovníku češtiny jako nevhodná, v případě Synu už to není tak jednoznačné. Přesto data na adjektivních přirovnáních ukazují, že ani Syn nutně neobsahuje přirozenou češtinu. Spoléhat čistě na frekvenci v korpusech jakožto na ukazatel ustálenosti VLJ

nestačí. Ryze deskriptivní přístup musejí doplnit další kontrolní mechanismy, aby se rozlišovalo mezi přirozeným vývojem (tj. postupným přejímáním) a prostou kumulací nekvalitních překladů.

6. ZÁVĚR

Tato studie ukázala, že zhruba každé desáté adjektivní přirovnání v InterCorpu je ne-standardní. Nestandardnost těchto přirovnání způsobují řidší transformace ustálených forem způsobené synonymními náhražkami dílčích komponentů, případně vlivem cizích jazyků v podobě lexikálních kalků. Výsledky této studie poukazují na úskalí spojená se sestavováním slovníkových seznamů na základě korpusových dat a na nutnost taková data důsledně analyzovat a interpretovat. Korpusové frekvenční seznamy nezřídka obsahují ne-standardní adjektivní přirovnání, a zejména překládané texty mohou vyvolávat dojem, že dochází k přirozenému zvýšenému výskytu nových variant. Z tohoto pohledu je původní jazyk textu naprosto zásadním kritériem. Zůstává otázkou, zda vůbec využívat data z překládaných textů při tvorbě slovníků daného jazyka, jelikož je riziko výskytu ne-standardních konstrukcí nesporně vyšší.

LITERATURA

- BARNDEN, J. A. (2016): Metaphor and simile: Categorizing and comparing categorization and comparison. In: GOLLA, E. — F. ERVAS (Eds.), *Metaphor and Communication*, s. 25–46. John Benjamins.
- ČERMÁK, F. (2007): *Frazeologie a idiomatika česká a obecná / Czech and General Phraseology*. Praha: Nakladatelství Karolinum.
- CROFT, W. (2007): Construction grammar. In: D. GEERAERTS — H. CUYCKENS (eds.) *Handbook of Cognitive Linguistics*, s. 463–508. Oxford: Oxford University Press.
- EMMER, J. (2020): Boring as hell: a corpus study of intensifying post-modification of predicative adjectives in the ADJ as NOUN frame. *Linguistica Pragensia*, 30, 2, s. 127–137.
- EMMER, J. (2023): *English Idioms of Comparison in Corpus Data*. Disertační práce. Praha: FF UK.
- EMMER, J. (2024): Anglická adjektivní přirovnání: srovnání korpusového vzorku s výběrem ve standardní příručce idiomů. *Časopis pro moderní filologii*, 106, 1, s. 28–43.
- GARGANI, A. (2016): Similes as poetic comparisons. *Lingua*, 175–176, s. 54–68.
- GOLDBERG, A. E. (1995): *Constructions: A Construction Grammar Approach to Argument Structure*. Chicago, IL: University of Chicago Press.
- HANKS, P. (2005): Similes and Sets: the English Preposition like. In: R. BLATNÁ — V. PETKEVIČ, V. (eds.) *Jazyky a jazykověda (Sborník k 65. narozeninám prof. PhDr. Františka Čermáka, DrSc.)*, s. 31–44. Praha: FF UK.
- KLÉGR, A. — BOZDĚCHOVÁ, I. (2024): *Lexikální anglicismy v češtině*. Praha: FF UK — Nakladatelství Karolinum.
- KOPŘIVOVÁ, M. (2017): Contribution Towards a Corpus-Based Phraseology Minimum. In: R. МИТКОВ (ed.) *Computational and Corpus-Based Phraseology*, s. 220–231. London/Berlin/Heidelberg: Springer Verlag.
- KOPŘIVOVÁ, M. — ŠICHOVÁ, K. (2023): Proverbs in Contemporary Czech. Corpus Probe into Written Texts. *Journal of Linguistics*, 74, 1, s. 92–99.
- LANGACKER, R. W. (1987): *Foundation of Cognitive Grammar (Vol. 1). Theoretical Prerequisites*. Stanford: Stanford University Press.



- MOON, R. (2008): Conventionalized as-similes in English: a problem case. *International Journal of Corpus Linguistics*, 13, 1, s. 3–37.
- MOON, R. (2011): Simile and dissimilarity. *Journal of Literary Semantics*, 40, 2, s. 133–157.
- NORRICK, N. R. (1986): Stock Similes. *Journal of Literary Semantics*, 15, 1, s. 39–52.
- PETKEVIČ, V. — KOPŘIVOVÁ, M. — HNÁTKOVÁ, M. — JELÍNEK, T. — KOPŘIVA, P. — ROSEN, A. — SKOUMALOVÁ, H. — VONDŘIČKA, P. (2020): Typologie víceslovných jednotek v češtině a frekvenční zastoupení jejich hlavních vlastností v žánrově vyváženém korpusu. In: *Studie z aplikované lingvistiky*, 11, 2, s. 37–62.
- VEALE, T. (2013): Humorous similes. In: T. Ford (ed.), *Humor: International Journal of Humor Research*, 26, 1, s. 3–22.

UŽITÉ KORPUSY

- KLÉGR, A. — KUBÁNEK, M. — MALÁ, M. — ROHRAUER, L. — ŠALDOVÁ, P. — ŠEBESTOVÁ, D. — VAVŘÍN, M. — ZASINA, A. J.: *Corpus InterCorp — English, version 16 from 12th October 2023*. Ústav Českého národního korpusu FF UK, Praha 2023. Dostupné z <http://www.korpus.cz>
- ROSEN, A. — VAVŘÍN, M. — ZASINA, A. J.: *Corpus InterCorp — Czech, version 16 from 12th October 2023*. Ústav Českého národního korpusu FF UK, Praha 2023. Dostupné z <http://www.korpus.cz>

PŘÍLOHA 1

InterCorp v16 — kompletní seznam vytěžených přirovnání (262 položek)

bílý jak sníh / bílý jak stěna / bílý jako duch / bílý jako křída / bílý jako lilie / bílý jako mléko / bílý jako papír / bílý jako plátno / bílý jako sníh / bílý jako stěna / bílý jako vápno / bílý jak tágo / bílý jako stěna / bílý jako duch / bílý jako křída / bílý jako mrtvola / bílý jako smrt / bílý jako stěna / bílý jako vosk / bílý jako bratr / černý jak noc / černý jak uhel / černý jako bota / černý jako eben / černý jako havran / černý jako inkoust / černý jako noc / černý jako saze / černý jako smola / černý jako smůla / černý jako uhel / černý jako uhlí / červený jako krev / červený jako rak / chladný jako kámen / chladný jako led / chladný jako ocel / chytrý jak opice / chytrý jako liška / chytrý jako opice / čilý jako rybička / čistý jak lilie / čistý jak lilium / čistý jako křišťál / čistý jako lilie / čistý jako lilium / čistý jako sklo / čistý jako slovo boží / čistý jako slza / čistý jako sníh / dřevý jako cedník / dřevý jako řeseto / drzý jak opice / drzý jako opice / hebký jako hedvábí / hebký jako samet / hezký jak obrázek / hezký jako obrázek / hladký jako hedvábí / hladký jako sklo / hladký jako úhoř / hladký jako zrcadlo / hladový jak vlk / hladový jako vlk / hluchý jak pařez / hluchý jak poleno / hluchý jako pařez / hluchý jako poleno / hořký jako žluč / hubený jak lunt / hubený jak tyčka / hubený jako lunt / hubený jako tyčka / jasný jak den / jasný jak facka / jasný jako den / jasný jako facka / jasný jako slunce / jednoduchý jak facka / jednoduchý jako facka / jemný jako hedvábí / kluzký jako had / kluzký jako úhoř / krásný jako anděl / krásný jako obrázek / krásný jako slunce / křehký jako sklo / krotký jako beránek / kulatý jako měsíc / kulatý jako míč / lehký jak facka / lehký jak pírkó / lehký jako facka / lehký jako peřičko / lehký jako pírkó / lehký jako pták / lehký jako vzduch / lehoučký jako pírkó / líný jako veš

/ mazaný jako liška / měkký jako máslo / mírný jako beránek / mírný jako ovečka / modrý jako moře / modrý jako nebe / modrý jako obloha / mokrý jako myš / mrštný jako kočka / mrtvý jako kámen / nafouklý jako balon / nahý jako Adam / namačkaný jako sardinka / napjatý jak struna / napjatý jako luk / napjatý jako struna / napjatý jako tětiva / napnutý jako struna / nehybný jako kámen / nehybný jako socha / němý jako ryba / obezřetný jako had / opilý jak zákon káže / ošklivý jak noc / ošklivý jako noc / ostrý jak břitva / ostrý jako břitva / ostrý jako dýka / ostrý jako jehla / ostrý jako meč / ostrý jako nůž / ostrý jako žiletka / paličatý jako mezek / pěkný jako obrázek / pevný jak skála / pevný jako kámen / pevný jako ocel / pevný jako skála / pilný jako včelička / pilný jako včelka / plochý jako prkno / podobný jako vejce vejci / přesný jak hodinky / přesný jako hodinky / přirozený jako dýchání / průhledný jako sklo / pyšný jako páv / rovný jako pravítko / rovný jako šíp / rovný jako svíce / rovný jako svíčka / rudý jak rak / rudý jako krev / rudý jako krocan / rudý jako oheň / rudý jako rak / rychlý jako blesk / rychlý jako šíp / rychlý jako vítr / silný jako býk / silný jako dub / silný jako kůň / silný jako lev / silný jako medvěd / silný jako paže / silný jako smrt / slabý jako kotě / slabý jako moucha / sladký jako med / slepý jak netopýr / slepý jak patrona / slepý jako krtek / slepý jako netopýr / slepý jako patrona / slizký jako had / snadný jak facka / snadný jako facka / špinavý jako prase / starý jak Metuzalém / starý jako čas / starý jako lidstvo / starý jako Metuzalém / starý jako svět / šťastný jak blecha / šťastný jako blecha / šťastný jako dítě / štíhlý jako proutek / studený jak led / studený jako kámen / studený jako led / studený jako ryba / suchý jak trouď / suchý jako trouď / světlý jako len / svěží jako rybička / svěží jako sedmikráska / svobodný jako pták / temný jako noc / tenký jako list (papíru) / tenký jako nit / tenký jako papír / těžký jako cent / těžký jako kámen / těžký jako olovo / tichý jako myška / tlustý jako prase / tuhý jak prkno / tuhý jako prkno / tuhý jako špalek / tupý jak poleno / tupý jako poleno / tvrdohlavý jako beran / tvrdohlavý jako mezek / tvrdohlavý jako osel / tvrdý jak kámen / tvrdý jak skála / tvrdý jako diamant / tvrdý jako hřebík(y) / tvrdý jako kámen / tvrdý jako křemen / tvrdý jako ocel / tvrdý jako skála / tvrdý jako šutr / tvrdý jako železo / tvrdý jako žula / utahaný jako kotě / utahaný jako kůň / utahaný jako pes / velký jako auto / velký jako dlaň / velký jako dům / velký jako hlava / velký jako hora / velký jako hrách / velký jako kolo od/u vozu / velký jako kůň / velký jako meloun / velký jako ořech / velký jako pes / velký jako pěst / velký jako slon / velký jako talíř / velký jako Texas / velký jako velryba / věrný jako pes / volný jako ptáček / volný jako pták / volný jako vítr / vysoký jako dům / vysoký jako hora / vysoký jako strom / zářící jako slunce / zdravý jak řípa / zdravý jako buk / zdravý jako kůň / zdravý jako řípa / zdravý jako ryba / zdravý jako rybička / zdravý jako rybka / zelený jako tráva / žlutý jako citron / zpcený jako myš / ztuhlý jako prkno