

Automatická měřítka čitelnosti pro česky psané texty¹

Josef Šlerka

Ústav informačních studií a knihovnictví, FF UK, Praha;

Ataxo Interactive, Praha

josef.slerka@gmail.com

Filip Smolík

Psychologický ústav AV ČR, Praha

smolik@praha.psu.cas.cz

Abstract:

The study reviews the fundamental assumptions of various readability measures and provides a case study demonstrating the properties of some readability measures when applied to Czech. Based on the analogy with the measures of production (MLU, IPSyn), it is argued that automatic readability measures provide a valid estimate of text reading difficulty. Different ways of validity testing for the readability indices are proposed. The case study shows that different measures based on average sentence and word length are very similar, while measures taking into account the token-type ratio (repetition index) behave differently. Readability indices for a series of textbooks show gradual increase in estimated difficulty with increasing grade level. The estimates for other texts also fit the expectations based on the text purpose and style.

Key words:

readability, reading comprehension, Flesch-Kincaid Grade Level, MLU, syntactic complexity

1. Historie

Označením *automatická měřítka čitelnosti* překládáme v tomto textu anglický termín *readability formulas*, používáme jej tedy pro vzorce, které se snaží kvantifikovat stupeň obtížnosti textu pro čtení. Podle Heidi Anne E. Mesmerové (2008) pocházejí první automatická měřítka čitelnosti z dvacátých let minulého století. Publikovali je Harry Kitson v roce 1921 a Bertha Livelyová a Sidney Pressey v roce 1923, tj. skoro před devadesáti lety. Od té doby tato měřítka prodělala řadu proměn. V roce 1980 jich bylo již něco kolem 200 různých typů. Svoje uplatnění nacházejí v mnoha oblastech – od novinářství přes oblast zdravotnictví a práva až k průmyslu. Své vlastní testy čitelnosti připravila například i americká armáda. Testy čitelnosti jsou dnes běžnou součástí amerických verzí textových editorů včetně MS Office (<http://office.microsoft.com/en-us/word/HP051863181033.aspx>) nebo Google Docs.

Mezi hlavní výzkumníky v oblasti automatických měřítek čitelnosti bývají řazeni Rudolf Flesch, Edgar Dale a Jeanne Challová. V bývalém Československu se problémům čitelnosti

1 Vznik této studie byl zčásti podpořen grantem GA AV ČR č. KJB700250801 *Zpracování větné shody a pádových tvarů u dětí* uděleným druhému autorovi.

věnoval především slovenský lingvista Jozef Mistrik, který pro výpočet čitelnosti navrhl vlastní vzorec a publikoval jej v prvním vydání své práce *Štylistika slovenského jazyka* (Mistrik, 1970).

2. Teoretické pozadí

2.1 Psychologická dimenze měřítek čitelnosti

Základem většiny indexů čitelnosti je intuitivně pochopitelná myšlenka, že existují charakteristické rozdíly mezi snadno čitelným a obtížně čitelným textem. Nejstarší a nejrozšířenější měřítka se přitom opírají zejména o předpoklad, že příznakem snadněji čitelného textu jsou kratší věty s kratšími slovy. Průměrná délka slov a vět je rozhodující součástí neznámějších indexů čitelnosti, jako je Fleschův-Kincaidův vzorec, vzorec ARI apod. Krátké věty a jednoduchá slova používáme, mluvíme-li s dětmi (Fraser – Roberts, 1975; Snow, 1977) nebo s osobami, u nichž předpokládáme sníženou míru porozumění, např. s cizinci (Snow – Van Eeden – Muysken, 1981). Krátké věty se vyskytují i v textech složitějších, ale čím méně pokročilému čtenáři je text určen, tím více takových krátkých vět bude obsahovat. Podobně většina textů obsahuje snadno srozumitelná slova i slova obtížnější, např. dlouhá nebo zřídka se vyskytující. V textech určených nezkušeným čtenářům se však obtížných slov bude objevovat mnohem méně nebo nebudou přítomna vůbec. Obtížné texty ovšem budou stále obsahovat mnoho snadno srozumitelných slov.

Souvislost délky vět se srozumitelností je do značné míry určena tím, že při porozumění textu zapojujeme obsahově omezenou pracovní paměť (Daneman – Carpenter, 1980; Just – Carpenter, 1992). Delší věty více zatěžují pracovní paměť, protože materiál ze začátku čtené věty se musí integrovat s materiálem, k němuž se čtenář dostane později. V mezidobí musí být původní materiál udržován v pracovní paměti a u delších vět je třeba udržovat více materiálu po delší dobu. Tím je ovšem část kapacity pracovní paměti vyčerpána, a pracovní paměť proto nemůže být efektivně využívána k dalším účelům, jako je vyhledávání významu slov nebo udržování koherence ve větším textovém celku. Důležité je, že pracovní paměť dokáže efektivně ukládat větší shluky informací jako jednoduché obsahy, pokud čtenář tyto celky dobře zná. Tak jako je sled písmen VODA zapamatovatelný snadněji než sled písmen BPFS (viz např. Murdock, 1961; Peterson – Peterson – Miller, 1961), dokážou zkušební čtenáři v pracovní paměti efektivněji ukládat celé skupiny slov. Není proto překvapivé, že zkušenější čtenáři dokážou bez problémů číst delší věty.

Souvislost mezi délkou slov a srozumitelností textu je dvojího typu. Jednak je to percepční, vjemová souvislost. Delší slova se z pochopitelných důvodů čtou delší dobu (viz např. Just – Carpenter, 1980). Tento fakt je dobře doložen experimentálně – delší doba zpracování informace znamená rovněž větší úsilí. Jednak mezi srozumitelností a délkou slov existuje nepřímý vztah zprostředkovaný frekvencí. Platí totiž, že vysoce frekventovaná slova jsou v jednotlivých jazycích v průměru kratší (Zipf, 1935), což znamená, že frekventovaná slova jsou snadněji srozumitelná (viz např. Scarborough – Cortese – Scarborough, 1977). Kratší slova tedy budou v průměru snadněji srozumitelná i proto, že jsou v průměru frekventovanější.

Další proměnnou, která se objevuje v některých indexech čitelnosti, je frekvence slov. Její automatické získávání je složitější, a proto se neobjevuje v nejtradičnějších měřítkách, jako je Fleschův vzorec. Lze však předpokládat, že její zahrnutí by vedlo k přesnějším odhadům

srozumitelnosti. Text psaný krátkými větami a krátkými slovy bude jistě hůře čitelný, bude-li v textu mnoho zřídka se vyskytujících slov.

Lze uvažovat o řadě dalších měřítek, která by mohla zpřesnit odhad čitelnosti textu. Četnost podřadických spojek může například poukazovat na syntaktickou složitost, která by nejspíše byla spojena s horší srozumitelností. Některé náměty pro rozšíření indexů zmiňujeme v následujícím oddíle, v němž se zabýváme indexy syntaktické komplexnosti dětské produkce.

2.2 Validita čili platnost měřítek čitelnosti

Ačkoli je intuitivně srozumitelné, proč indexy čitelnosti založené na délce slov a vět fungují, lze se ptát, jak spolehlivě. Intuice může klamat, a proto je třeba se ohlížet po nezávislém způsobu ověření platnosti indexů srozumitelnosti. Existují v principu dva zdroje dokladů o tom, že indexy skutečně odrážejí náročnost textu. Jedním z těchto zdrojů je analogie s měřítky jazykové produkce. Druhým jsou pak více či méně systematické studie validity měřítek čitelnosti vzhledem k externím kritériím.

2.2.1 Čitelnost a měřítka produkce

Ve studiu osvojování jazyka se jako jedno ze základních měřítek pokročilosti jazykového vývoje používá index průměrné délky věty, tzv. MLU (viz např. Brown, 1973). Průměrná délka věty roste s věkem a souvisí s dalšími měřítky pokročilosti jazykového vývoje (Leadholm – Miller, 1992; Rice et al., 2010). Je samozřejmé, že průměrná délka věty závisí rovněž na kontextu produkce – v řadě situací i dospělí mluví výhradně v krátkých větách. Proto je MLU spolehlivým měřítkem jen při hodnotách menších než 4 až 6 slov, vyšší hodnoty jsou ve spontánní konverzaci vzácné i u dospělých.

Protože průměrná délka věty je silně ovlivněna kontextem, vznikly složitější indexy určené k hodnocení jazykové komplexnosti dětských produkcí. K neznámějším patří index produktivní syntaxe čili IPSyn (Scarborough, 1990). Tento index v daném vzorku řeči skóruje výskyt 56 jazykových prostředků, jako jsou pomocná slovesa, způsobová slovesa, souvětí, podřadná souvětí atd. Index tak bere v úvahu nikoliv pouhou délku vět, ale skutečnou syntaktickou složitost vyjadřování dítěte. Podle studií je IPSyn lepším indikátorem jazykového vývoje než MLU, což není překvapivé. Větší přesnost indexu je ovšem vykoupena větší obtížností jeho výpočtu.

MLU, IPSyn a další indexy se používají ke zjišťování pokročilosti jazykového vývoje, ale lze je použít i ke sledování opačných procesů ve stáří a při demencích. Susan Kemperová se svými spolupracovníky doložila, že ke konci osmé dekády života dochází k nápadnému snížení gramatické komplexnosti mluvených vět (Kemper – Marquis – Thompson, 2001). Gramatická komplexnost byla měřena měřítkem D-level, které odráží zejména míru používání vedlejších vět.

Index produktivní syntaxe a další měřítka používaná pro hodnocení jazykové produkce ukazují, jakým způsobem by indexy srozumitelnosti textu mohly zahrnovat podrobnější informace o jeho obtížnosti. Do celkových indexů by bylo možné promítnout míru výskytu určitých funkčních slov, která charakterizují složitější texty. To mohou být například již zmiňované podřadící spojky, ale také méně frekventované předložky. Pomocí průměrného počtu spojek ve větě by bylo možné odhadovat, jak často se v textu objevují souvětí.

2.2.2 Experimentální a psychometrické studie validity čtenářských indexů

Metodologicky nejspolehlivější, avšak také nejnáročnější způsob ověření validity měřítek spočívá v tom, že větší skupině čtenářů prezentujeme texty, které se liší co do sledovaných měřítek. U týchž osob pak otázkami zjišťujeme, jak zadaným textům porozuměly. Tímto způsobem můžeme například zjistit, jaká složitost textu podle měřítek čitelnosti odpovídá čtenářským dovednostem typickým pro určitý věk. Problém této metody spočívá v tom, že je vysoce finančně a časově náročná. Chceme-li například zjistit, jaká úroveň měřítek čitelnosti odpovídá schopnostem typickým pro děti ve čtvrté třídě základní školy, musíme vzorku nejméně několika desítek dětí dát přečíst řadu textů, které se liší co do měřítek čitelnosti. Zároveň musíme pro každý text provést test porozumění, tj. zeptat se dětí na skutečnosti z textu. Chceme-li získat přehled o tom, jaká úroveň měřítek obtížnosti je průměrná různému věku, musíme podobnou proceduru opakovat ve všech věkových skupinách, které nás zajímají. Taková standardizační procedura je náročná a drahá, na druhou stranu však poskytuje velmi podrobné informace a může sloužit jako pevný základ pro hodnocení textů, které jsou určeny dětem různého věku.

Validitu měřítek obtížnosti čtení můžeme ověřit i jednoduššími způsoby. Zajímá-li nás obtížnost textu v souvislosti s vývojem dětských čtenářských dovedností, můžeme testy porozumění nahradit odhadem učitelů a dalších profesionálů, kteří mají zkušenost s prací s dětmi různého věku. Pokud nás nezajímá kalibrace měřítek čitelnosti vzhledem k věku či nějakému očekávanému výkonu, ale pouze jako nástroj pro srovnávání různých textů, můžeme se opřít i o experimentální měřítka, zejména o měření rychlosti čtení. Pokud indexy obtížnosti skutečně odrážejí náročnost čtení, měly by být obtížnější texty čteny pomaleji.

2.3. Strukturnělingvistická dimenze automatických měřítek čitelnosti

Kromě psychologické dimenze měřítek čitelnosti můžeme teorii opřít i o pozice Pražského lingvistického kroužku. Čitelnost můžeme chápat jako soulad s určitými pravidly užívání psaného projevu, samozřejmě přiměřený myšlence, která je textem vyjádřena.

Ve shodě s českým strukturalismem můžeme pokročit od pravidel k normě, která je ve skutečnosti předmětem měření čitelnosti. Norma sama je realizací funkce a v tomto kontextu je funkcí míra snadné čitelnosti textu. Nenechme se zmást tím, že v českém prostředí nemá tato funkce určitá jasně formulovaná pravidla. Pravidla jsou totiž pouhou kodifikací normy. Neříkají nám však nic o povaze normy jako takové. Navíc k pohybu ke kodifikaci nemusí vůbec dojít. V českém stylistickém prostředí se s tak ostře formulovanými pravidly pro srozumitelné psaní nesetkáme, to však neznamená, že zde neexistuje norma nekodifikovaná. Jak upozorňuje J. Mukařovský (2000) ve studii *Estetická funkce, norma a hodnota jako sociální fakty*, lingvisté vědí o existenci jazykových norem, které neprošly gramatickou kodifikací, a které jsou přesto jazykovými kolektivy dodržovány.

V tomto ohledu můžeme přijmout Mukařovského (2000, s. 149–150) charakteristiku normy jako regulujícího energetického principu: „[Norma] dává pocítit svou přítomnost jedináčímú individuu jako omezení jeho akce, pro individuum, které hodnotí, je silou řídicí jeho úsudek.“ Právě toto propojení produkce na jedné straně a hodnocení (a tak nutně i předchozí spotřeby) na straně druhé je to, co umožňuje, aby měřítka čitelnosti mohla fungovat a měřit tak funkční čitelnost textu.

Jinými slovy, pokud jsme schopni tvrdit, že významnou složkou čitelnosti textu je dodržování určitých – byť ne nutně pozitivně formulovaných – norem textové produkce, pak zároveň říkáme, že je možné takovéto normy měřit.

3. Metody

Pro předvedení principů fungování automatických měřítek čitelnosti jsme vybrali vzorce Flesch reading ease score (fres) a jeho variantu Flesch-Kincaid Grade Level (fkg), vzorec LIX (lix), Automated readability index (ari), Gunning fog index (fog), SMOG index (smog) a dvě varianty Mistříkova vzorce.

a) Flesch reading ease score (fres) a Flesch-Kincaid Grade Level (fkg)

Různé varianty tohoto vzorce patří mezi automatickými měřítka čitelnosti k nejrozšířenějším. Jsou dostupné i jako součást anglické verze Microsoft Office i onlinového textového editoru Google Docs (opět v anglické verzi). Ve všech verzích vzorec zohledňuje tři kritéria: celkový počet vět, celkový počet slov a celkový počet slabik. Jeho různé kalibrace vedou k různým škálám.

Flesch reading ease score

Vzorec:

$$\text{fres} = 206,835 - 1,015 \cdot (\text{počet slov} / \text{počet vět}) - 84,6 \cdot (\text{počet slabik} / \text{počet slov})$$

Vysoká čísla ukazují na snadnou srozumitelnost, nízká naopak na srozumitelnost obtížnou. Klíč pro interpretaci číselných hodnot uvádí anglická verze internetové encyklopedie Wikipedia (viz http://en.wikipedia.org/wiki/Flesch-Kincaid_readability_test). Je třeba vzít v úvahu, že numerické konstanty uvedené ve vzorci byly nalezeny empiricky a jsou – stejně jako interpretační tabulka – specifické pro angličtinu.

Skór	Úroveň textu
90.0–100.0	snadno srozumitelné pro průměrného jedenáctiletého žáka
60.0–70.0	snadno srozumitelné pro průměrné třinácti- až patnáctileté žáky
0.0–30.0	porozumění bude nejlepší u absolventů univerzity

tab. 1

Flesch-Kincaid Grade Level (školní ročník)

Vzorec:

$$\text{fkg} = (0,39 \cdot \text{počet slov}) / \text{počet vět} + (11,8 \cdot \text{počet slabik}) / \text{počet slov} - 15,59$$

Jedná se o jinou verzi předcházejícího vzorce, v níž jsou konstanty nastaveny tak, aby výsledek odpovídal školnímu ročníku, pro nějž je daný text přiměřeně obtížný. Například výsledek 8,2 by odpovídal dokončené osmé třídě základní školy. Podobně jako v předchozím případě platí, že hodnoty konstant ve vzorci je třeba kalibrovat pro každý jazyk jinak. Je rovněž otázkou, zda přiměřený index čitelnosti pro různý věk roste lineárně, jak by tento vzorec naznačoval. Rozdíl v měřítku čitelnosti mezi texty zvládanými v první a ve druhé třídě může být zcela jiný, než je rozdíl mezi texty zvládanými v sedmé a v osmé třídě. Proto je třeba převody na „čtenářský věk“ chápat velice opatrně.

b) LIX (lix)

Metoda LIX pochází z roku 1968 a jejím autorem je švédský jazykovědec C. H. Björnsson. Metoda zohledňuje počet vět, počet slov a počet složitých slov. Za složitá se v původním vzorci považují slova, která jsou delší než dvě slabiky.

Vzorec:

$$\text{lix} = \text{počet slov} / \text{počet vět} + \text{počet složitých slov} \cdot 100 / \text{počet slov}$$

c) Automated readability index (ari)

Vzorec Automated readability index byl vytvořen R. J. Senterem a E. A. Smithem v roce 1967 pro potřeby americké armády. V roce 1970 byl kalibrován na technických manuálech (DuBay, 2004). Místo počtu slabik se opírá o počet písmen, což usnadňuje jeho automatický výpočet.

Vzorec:

$$\text{ari} = (4,71 \cdot \text{počet písmen}) / \text{počet slov} + (0,5 \cdot \text{počet slov}) / \text{počet vět} - 21,43$$

d) Gunning fog index (fog)

Měřítka pochází z roku 1952 a jeho autorem je Robert Gunning. Značnou popularitu si získalo díky tomu, že bylo kalibrováno pro pasáž o 100 slovech. Nebylo tedy příliš obtížné spočítat výsledek manuálně. V našich testech ovšem vždy používáme údaje z celého textu. Vzorec pracuje s počtem slov, počtem vět a s počtem složitých slov, která jsou definována jako slova delší než tři slabiky (http://en.wikipedia.org/wiki/Gunning_Fog). Vzorec posloužil i jako východisko pro vytvoření výše uvedeného indexu LIX.

Vzorec:

$$\text{fog} = 0,4 \cdot (\text{počet vět} / \text{počet slov} + (100,0 \cdot \text{počet složitých slov}) / \text{počet slov})$$

e) SMOG index (smog)

SMOG (Simple Measure of Gobbledygook) je vzorec z roku 1969, jehož autorem je G. Harry McLaughlin. Posloužil jako zjednodušení Gunningova fog indexu, protože pracuje pouze se dvěma kritérii: s počtem vět a s počtem víceslabičných slov. V angličtině se jako víceslabičná

chápu slova o třech a více slabikách (v našich testech jsme pro účely češtiny zvolili jiné kritérium). Původní vzorec je kalibrován tak, aby výsledek zhruba odpovídal dosaženému ročníku vzdělání.

Vzorec:

$$\text{smog} = \sqrt{(\text{počet složitých slov} \cdot 30 / \text{počet slov}) + 3}$$

f) Mistríkův vzorec lemmatizovaný (mistrík), Mistríkův vzorec (mistrík_uplny)

Podle slovenského jazykovědce Jozefa Mistríka nezohledňuje většina klasických měřítek čitelnosti index opakování slov. Index opakování slov je poměr celkového počtu slov v textu a počtu různých slov. Je tedy tím nižší, čím méně se v textu opakují jednotlivá slova. V anglické literatuře je inverzní měřítko známo jako *type-token ratio*. Nižší index opakování podle Mistríka (1970, s. 50) ukazuje na bohatost slovníku, rostoucí index opakování je příznakem rostoucí předvídatelnosti textu, a tím i rostoucí srozumitelnosti: „Pri prvom stretnutí so zriedkavejším slovom sa spomalí tempo vnímania, ale pri opakovaných stretnutiach sa spomaľovanie znižuje, až sa úplne stráca.“ Mistrík proto navrhuje vlastní vzorec, který obsahuje i index opakování.

Vzorec:

$$\text{mistrík} = 50 - (\text{průměrná délka slov ve slabikách} \cdot \text{průměrný počet slov ve větě}) / (\text{počet slov} / \text{počet různých slov})$$

Pro tento vzorec platí, že čím nižší je výsledné číslo, tím je text obtížnější.

3.1 Použitá kritéria

Použitá měřítka zahrnují celkem osm různých kritérií: počet slov, počet různých slov, počet vět, průměrný počet slabik ve slově, průměrný počet slov ve větě, počet slabik v textu, počet složitých slov a počet znaků. Tabulka 2 přehledně shrnuje, které z indexů využívají kterých kritérií.

		počet slov	počet jedinečných slov	počet vět	průměrný počet slabik ve slově	průměrný počet slov ve větě
Flesch-Kincaid Grade Level	fkg1	x		x	x	
LIX	lix	x		x		
Automated readability index	ari	x		x		
Gunning fog index	fog	x		x		
Flesch reading ease score	fres	x		x		
Mistríkův vzorec (lemmatizovaný)	mistrík	x	x		x	x
Mistríkův vzorec	mistrík_uplny	x	x		x	x
SMOG index	smog			x		

tab. 2

3.2 Procedury a postupy

Pro testování jsme vyvinuli vlastní knihovnu Rullus pro jazyky Ruby. V případě Mistríkova vzorce jsme zvolili dvě varianty pro určení indexu opakování; v první jsme vycházeli z nelemmatizovaných tvarů slov, ve druhé z lemat. V první verzi se tedy jednalo o poměr celkového počtu slov a počtu různých slovních tvarů, ve druhém případě šlo o poměr celkového počtu slov a počtu různých slov (bez ohledu na tvar). Pro lematizaci textu využívá knihovna Rullus volně dostupný slovník z programu OpenOffice. Pro stanovení počtu slabik, které je důležité i pro identifikaci složitých slov, jsme využili jednoduchý postup spočítání samohlásek ve slově, které jsme brali jako identické s počtem slabik. Slova neobsahující žádnou nebo jen jednu samohlásku byla chápána jako slova jednoslabičná. Lemmatizace a určení počtu slabik jsou pouze přibližné, ale vzhledem k tomu, že se opírají o rozsáhlé zdroje dat, lze jejich chybu zanedbat. Za složitě slovo jsme v češtině považovali takové, které obsahuje čtyři a více slabik.

3.3 Materiály

Pro testování jsme zvolili tři různé skupiny textů:

- čítanky pro 5.–9. třídu základní školy zbavené ukázkami z poezie a dramatu;
- dva díly učebnice *Dějiny země Koruny české*;
- část prozaického díla Karla Čapka.

S jednotlivými skupinami textů jsme vážali následující hypotézy:

- a) Indexy čitelnosti by měly poukazovat na větší obtížnost u čítanek pro vyšší ročníky.
- b) *Dějiny zemí Koruny české* jsou schválenou učebnicí pro střední školy určenou pro přípravu k maturitě. Kniha by tedy měla vykazovat výrazně vyšší náročnost než čítanek pro základní školy.
- c) Prozaické dílo Karla Čapka zabírá stylově širokou oblast od dětské literatury až po oblast fejetonu, testy by tuto šíři měly reflektovat. Indexy čitelnosti by se měly měnit v závislosti na žánru a cílovém publiku.

3.4 Výsledky testování

Výsledné hodnoty vybraných indexů čitelnosti jsou shrnuty v tabulce 3.

Ročník:	fkgl	lix	ari	fog	fres	mistrik	mistrik_uplno	smog
5	11,7800359	20,0959059	7,28787205	13,8672585	26,8729745	45,5215511	43,4167067	8,50046503
6	14,7997533	25,1023655	9,6366488	18,5008349	6,33004113	43,5271252	41,242707	9,82203625
7	11,3689904	19,9856216	6,90071001	14,246249	28,3675484	45,802203	43,9595974	8,46828428
8	12,1926053	21,6504478	7,72292064	15,0845398	24,9162206	45,7334105	43,6525647	8,92886731
9	12,9546692	22,3935069	8,18182557	15,6981759	19,8368026	42,4753995	40,4625688	9,13267936
Soubor:	fkgl	lix	ari	fog	fres	mistrik	mistrik_uplno	smog
dejiny01	18,5151424	35,8099443	14,2777781	26,6663728	-12,761354	44,0927865	40,3907604	12,6976617
dejiny02	19,7733194	39,8392911	15,0894258	30,4613005	-21,08597	43,6338233	39,7855729	13,654451
Soubor:	fkgl	lix	ari	fog	fres	mistrik	mistrik_uplno	smog
apokryfy	10,9297822	19,0071929	6,63463367	12,4143651	34,0499319	45,9400895	44,3923384	8,14144204
bozimuka	10,5047801	17,626184	6,05268119	11,3838548	36,056525	46,3190726	44,6316568	7,74784091
dasenka	13,4920682	24,6391653	9,54658503	13,9923568	27,7138971	39,9159213	37,074551	9,0582281
devateropohadek	11,2677633	20,2419666	7,15115034	11,7663634	37,2158036	45,8929371	44,0482341	8,09099505
hordubal	9,81850102	15,4298384	5,31555505	9,52717448	39,9684993	47,0738313	45,9121335	7,06251655
krakatit	11,0972321	19,2689111	6,64767684	12,392319	33,6917562	46,7579157	45,0686913	8,18130826
kritikaslov	16,0404444	30,0240934	12,22503	18,7790867	11,2190887	39,3962863	35,6690045	10,9646355
obycejnyzivot	11,0979516	19,2705095	6,64858895	12,393017	33,6892706	46,7574313	45,0679545	8,18167988
zarivehlabiny	15,2266343	29,2788281	11,3793718	17,5296952	18,5501932	40,9768599	37,3426728	10,5475372

tab. 3

Obecně vzato, příklady čítanek a učebnice *Dějiny zemí Koruny české* ukazují, že většina automatických měřítka čitelnosti je v souladu s předpoklady o vzrůstající složitosti textů přiměřeně věku. Největší citlivost projevuje fog index, naopak nejméně citlivý byl v tomto kontextu Mistrikův vzorec bez ohledu na varianty. Numerický rozdíl mezi nejnižšími a nejvyššími výslednými hodnotami nemusí ovšem ještě poukazovat na skutečnou citlivost, záleží rovněž na míře chybové variability, pro jejíž odhad jsme neměli dostatek údajů.

V případě analýz díla Karla Čapka se však Mistríkův vzorec jeví jako citlivější. Zohlednění indexu opakování se ukazuje jako užitečné, pravděpodobně proto, že indikuje rozdíl mezi odborným a uměleckým textem. Výsledné hodnoty měřítek čitelnosti se převážně shodují s intuitivně předpokládaným odhadem čitelnosti. Srovnáme-li však Čapkovy díla kupříkladu s učebnicí *Dějiny země Koruny české*, ukazuje většina indexů na to, že by fejetonistická *Kritika slov*, případně raná poetická práce *Zářivé hlubiny*, měla být čitelná lépe než učebnice. Pouze Mistrík je zařazuje jako výrazně složitější než učebnice. Na dílech Karla Čapka se ukazuje, že „čitelnost“ textu není pravděpodobně obecně příznaková pro sémantické gesto autora, ale že je daleko těsněji spjata se zvoleným stylem a žánrem.

U souboru čítanek je překvapivé výrazné vychýlení v čítance pro 6. třídu základní školy. Částečně lze toto vychýlení vysvětlit velkým zastoupením textů, které mají povahu vyprávění (Dášeňka, Rumcajs, Maxipes Fík), protože tyto texty jsou složeny z řady dlouhých souvětí. Extrémem je v tomto případě Dášeňka, která vychází jako prakticky nejobtížnější text i z použitého korpusu děl Karla Čapka.

V případě *Dějin země Koruny české* je dobře vidět závislost vzorce na konkrétním jazyku. Ačkoli trendy jsou při srovnání různých textů zřetelné, v případě vzorce Flesch reading ease score se české texty dostávají snadno do záporného čísla.

3.5 Korelace měřítek

Tabulka 4 obsahuje vzájemné korelace jednotlivých měřítek čitelnosti (Pearsonův korelační koeficient) vypočtené ze vzorku textů v tabulce 3. Korelační koeficient může nabývat hodnot od -1 (lineární nepřímá úměrnost) do 1 (lineární přímá úměrnost). Hodnota 0 znamená statistickou nezávislost proměnných. Vyšší absolutní hodnota koeficientu tedy poukazuje na větší spřízněnost mezi sledovanými indexy. Tabulka zřetelně ukazuje, že indexy založené na počítání průměrné délky slov a vět a počtu dlouhých slov jsou si vzájemně velmi podobné – korelace nad $0,9$ se v práci se sociálními daty dá považovat za identitu. Tabulka rovněž ukazuje, že Mistríkův vzorec se od ostatních indexů liší, neboť s nimi koreluje výrazně méně. Na druhou stranu je korelace Mistríkova vzorce s jinými měřítky čitelnosti stále nezanedbatelná, což potvrzuje, že všechna měřítka odrážejí podobné vlastnosti textů.

	fkgl	lix	ari	fog	fres	mistrík	mistrík_ uplňy	smog
fkgl	1	0,99	0,99	0,96	-0,96	-0,6	-0,69	0,99
lix	0,99	1	0,99	0,96	-0,95	-0,56	-0,66	0,99
ari	0,99	0,99	1	0,94	-0,93	-0,62	-0,72	0,98
fog	0,96	0,96	0,94	1	-0,99	-0,38	-0,49	0,98
fres	-0,96	-0,95	-0,93	-0,99	1	0,43	0,54	-0,97
mistrík	-0,6	-0,56	-0,62	-0,38	0,43	1	0,99	-0,5
mistrík_ uplňy	-0,69	-0,66	-0,72	-0,49	0,54	0,99	1	-0,6
smog	0,99	0,99	0,98	0,98	-0,97	-0,5	-0,6	1

tab. 4

4. Závěr

Testy v naší pilotní studii ukazují, že aplikace měřítek čitelnosti na české texty poskytuje intuitivně srozumitelné trendy, které odpovídají předpokladům. Je samozřejmé, že pro adaptaci na české podmínky je třeba změnit konstantní členy ve vzorcích tak, aby aplikace měřítek na české texty vyústila v pohodlně interpretovatelné číselné hodnoty.

Naše studie ověřila, že kvantitativní indexy srozumitelnosti textu poskytují v češtině smysluplné údaje. Tím se otevírá možnost pro jejich cílené využití při výběru a hodnocení textů pro praktické účely, například při přípravě učebnic. Takové praktické využití ovšem předpokládá validizaci konkrétních indexů pro české prostředí. Ta by stanovila konkrétní hodnoty indexů, které odpovídají různým úrovním čtenářské vyspělosti. V rámci validizace by rovněž bylo možné ověřit, zda měřítka zahrnující index opakování, jako je Mistríkův vzorec, představují citlivější měřítka skutečné čtenářské obtížnosti, či zda jsou údaje o délce slov a vět dostatečné. Rovněž by bylo záhodno otestovat další měřítka, která budou pracovat s průměrnou frekvencí používaných slov nebo s množstvím málo frekventovaných slov v textu. Knihovna Rullus, která byla využita pro zpracování textů a výpočet indexů čitelnosti v této práci, představuje vhodný nástroj pro další rozvíjení a adaptaci měřítek obtížnosti čtení pro češtinu.

Literatura:

- BROWN, Roger A. (1973): *First Language: The Early Stages*. Cambridge, MA: Harvard University Press.
- DANEMAN, Meredith – CARPENTER, Patricia A. (1980): Individual differences in working memory and reading. *Journal of Verbal Learning and Verbal Behavior*, 19, s. 450–466.
- DUBAY, William (2004): *The Principles of Readability*. Costa Mesa, CA: Impact Information.
- FRASER, Colin – ROBERTS, Naomi (1975): Mothers' speech to children of four different ages. *Journal of Psycholinguistic Research*, 4, s. 9–16.
- JUST, Marcel Adam – CARPENTER, Patricia A. (1980): A theory of reading: From eye fixations to comprehension. *Psychological Review*, 87, s. 329–354.
- JUST, Marcel Adam – CARPENTER, Patricia A. (1992): A capacity theory of comprehension: Individual differences in working memory. *Psychological Review*, 99, s. 122–149.
- KEMPER, Susan – MARQUIS, Janet – THOMPSON, Marilyn (2001): Longitudinal change in language production: effects of aging and dementia on grammatical complexity and propositional content. *Psychology of Aging*, 16, s. 600–614.
- LEADHOLM, Barbara – MILLER, Jon F. (1992): *Language Sample Analysis: The Wisconsin Guide*. Milwaukee: Wisconsin Department of Instruction.
- MESMER, Heidi Anne E. (2008): *Tools for Matching Readers to Text: Research-Based Practices*. New York, NY: Guilford Press.

MISTRÍK, Jozef (1970): *Štylistika slovenského jazyka*. Bratislava: Slovenské pedagogické nakladateľstvo.

MUKAŘOVSKÝ, Jan (2000): *Studie I*. Brno: Host.

MURDOCK, Bennet B. (1961): The retention of individual items. *Journal of Experimental Psychology*, 62, s. 618–625.

PETERSON, Lloyd R. – PETERSON, Margaret Jean – MILLER, Arthur (1961): Short-term retention and meaningfulness. *Canadian Journal of Psychology / Revue canadienne de psychologie*, 15, s. 143–147.

RICE, Mabel L. et al. (2010): Mean length of utterance levels in 6-months intervals for children 3 to 9 years with an without language impairments. *Journal of Speech, Language, and Hearing Research*, 53, s. 333–349.

SCARBOROUGH, Hollis S. (1990): Index of Productive Syntax. *Applied Psycholinguistics*, 11, s. 1–22.

SCARBOROUGH, Don L. – CORTESE, Charles – SCARBOROUGH, Hollis S. (1977): Frequency and repetition effects in lexical memory. *Journal of Experimental Psychology: Human Perception and Performance*, 3, s. 1–17.

SNOW, Catherine (1977): The development of conversation between mothers and babies. *Journal of Child Language*, 4, s. 1–22.

SNOW, Catherine – VAN EEDEN, Roos – MUYSKEN, Pieter (1981): The interactional origins of foreigner talk: Municipal employees and foreign workers. *International Journal of the Sociology of Language*, 28, s. 81–91.

ZIPF, George Kingsley (1935): *The Psychobiology of Language*. New York, NY: Houghton Mifflin.