

# Žánrová preference v narrativech dětí mladšího školního věku

SILVIE CINKOVÁ – ANEŽKA NÁHLÍKOVÁ

## Genre preferences in the narratives of younger school children

**ABSTRACT:** In this paper, we present an analysis of hypothesized sex-specific genre preferences in spontaneous narratives in a corpus of 286 transcribed recordings of younger school children (aged 6–11). We draw upon Nicolopoulou's conception of narratives as a symbolic activity with sex-specific tendencies: girls prefer the "family" genre and boys prefer the "heroic-agonistic" one. We operationalized Nikolopoulou's genre classification into 10 thematic narrative features and annotated their respective presence or absence in each narrative. The application of a combination of statistical methods to our sample revealed that the sex-correlated narrative preferences are so weak that they have no implications, e.g. for school practice. The most prominent result, in fact, even counters the a priori assumption of sex-based narrative preferences: no matter what sex or age, children show a strong preference for "mutual aid and cooperation".

**Key words:** narrative, interpretive sociocultural approach, family genre, heroic-agonistic genre, younger school-aged children

**Klíčová slova:** narativum, interpretativní sociokulturní přístup, rodinný žánr, hrdinský žánr, děti mladšího školního věku

## Úvod

Spontánní produkce narativ<sup>1</sup> je důležitým zdrojem informací o kognitivním, psychickém a jazykovém vývoji dětí. Tyto informace nacházejí praktické využití například v diagnostice učebních poruch nebo při tvorbě učebních materiálů.

V naší studii zkoumáme možný vliv věku a pohlaví na výběr narativních prvků spontánních fiktivních narativ u mladších školních dětí. Na dětská spontánní narativa zde pohlížíme jako na prostředek k uspořádání a interpretaci lidské zkušenosti, jehož výsledný tvar je ovlivňován prostředím, v němž se vyprávěč pohybuje, a navazujeme tak především na studie A. Nicolopoulou (1994, 2008a, 2008b). Zaměřujeme se na tematické narativní prvky, protože dostupné studie strukturálních prvků dětských narativ vliv pohlaví vyprávěče nepotvrzují (srov. Peterson – McCabe, 1983; Harčariková – Klimovič, 2011), zatímco studie zaměřené na tematické prvky naznačují, že pohlaví vyprávěče určitou roli hraje. Dívkám se přisuzuje preference přátelských a nápomocných postav, zatímco chlapci preferují agresivní hrdiny (Libby – Aries, 1989, vzorek 3–5letých dětí).

---

<sup>1</sup> Vyprávění neboli narativum chápeme jako „jakoukoli sekvenci vět, které obsahují přinejmenším jedno časové spojení“ (Labov – Waletzky, 1967, s. 28). Použití termínu *narativum* jsme v české odborné literatuře nezaznamenali. Ve slovenské literatuře věnující se dětské řeči je však tento termín upřednostňován (srov. Harčariková – Klimovič, 2011 aj.) před pojmem *narativ*, kterého se užívá v naratologii. Protože v našich výzkumech, co se týče terminologie, ze slovenské literatury vycházíme, budeme jej zde (vedle pojmu *vyprávění*) také používat. Od těchto pojmů odlišujeme termín *příběh* (příběh = to, o čem se mluví; narativum/vyprávění = písemné či ústní ztvárnění příběhu; viz Müller – Šidák, 2012, s. 413 a 559).

A. Nicolopoulou u 3–5letých amerických dětí konstatuje výrazné rozdíly mezi narativy chlapců a narativy dívek, která žánrově charakterizuje jako *hrdinská* (*heroic-agonistic*) vs. *rodinná* (*family genre*).

Studie je členěna následovně: vymezení hrdinského a rodinného žánru; charakteristika datového souboru s ukázkami textů; stručný přehled použitých analytických metod; statistická analýza dat se závěrečnou diskusí.

## 1. Hrdinský vs. rodinný žánr

Hrdinský žánr je v kontrastu k rodinnému žánru charakterizován zaměřením na *chaos*, *napětí* a *rozvrat* (Nicolopoulou, 2008a, s. 317). Hlavní postavou je obvykle jeden hrdina, který prožívá různá dobrodružství a často se dostává s někým do konfliktu. Jeho cílem je stát se sebevědomou a respektovanou osobností, která zvítězí nad nepřítelem, nebo si svým hrdinským činem vysloužit obdiv (Nicolopoulou, 2008b, s. 248). Vztahy mezi postavami obvykle nejsou příliš pevné a není jim věnována velká pozornost, tj. narativa pomocí nich ve většině případů nejsou strukturována.

Ústředním bodem rodinného žánru je skupina postav (obvykle rodina či skupina přátel). Tato skupina často vystupuje jako celek. Členové této skupiny jsou propojeni různými vztahy, které strukturují celé narativum (Nicolopoulou, 2008a, s. 312) – vztahy se mohou vyvíjet, přetvářet, vznikat nebo zanikat. Cílem jednání postav bývá zajistit, uchovat nebo obnovit stabilitu dané skupiny, hájit její zájmy a zájmy jejích členů. Na závěr je většinou dosaženo stability a harmonie ve vztazích. Rodinný žánr lze nejlépe vystihnout slovy *stabilita* a *řád*.

## 2. Data

Používání tematických narativních prvků jsme zkoumaly na korpusu narativ od náhodného vzorku 286 rodilých českých mluvčích z prvních až pátých tříd osmi základních škol ve Středočeském a Východočeském kraji (Náhliková, 2015).

Narativa byla shromažďována v období od září 2013 do března 2016. Úkolem všech dětí bylo vytvořit vyprávění na základě následujícího pokynu: „Vyprávěj mi pohádku. Může to být pohádka jakákoli, jakkoli dlouhá, ale musíš si ji sám/sama vymyslet. Má to být tedy pohádka, kterou jsi nikde nikdy neslyšel/a, neviděl/a ani nečetl/a.“

Sběr dat probíhal obvykle v budově základní školy, kterou dané dítě navštěvovalo. Vyprávění byla zaznamenávána na diktafon. Se zadáním nebyly děti seznámeny předem, na přípravu před spuštěním diktafonu však každé z nich dostalo tolik času, kolik potřebovalo. Jako motivační faktor byl použit příslib malé odměny na závěr. Při nahrávání postupoval explorátor metodou řízené narační aktivity dítěte (srov. Harčariková – Klimovič, 2011, s. 62): do vyprávění nijak nezasahoval, pouze v průběhu přirozeně reagoval (smíchem, údivem, přitakáním), čímž se snažil navodit příjemnou atmosféru a ukázat dítěti, že to, co vypráví, je pro něj zajímavé.

Nahrávky byly přepsány podle pravopisných pravidel, zachyceny jsou však všechny nespisovné slovní tvary a některé odchylky od standardního vyjadřování. Velkými

písmeny jsou označeny pouze iniciály vlastních jmen. V závorkách je zaznamenána řeč explorátora (podle Náhlíkové, 2015, s. 119–120).<sup>2</sup>

Korpus sestává z 286 narativ, která můžeme považovat za autorská. Celkem 26 narativ jsme předtím vyřadily, protože šlo o evidentní převyprávění známých pohádek nebo epizod z dětských pořadů. Rozložení narativ podle pohlaví a ročníků ukazuje tabulka 1.

**Tabulka 1:** Rozložení ročníků a pohlaví v korpusu narativ

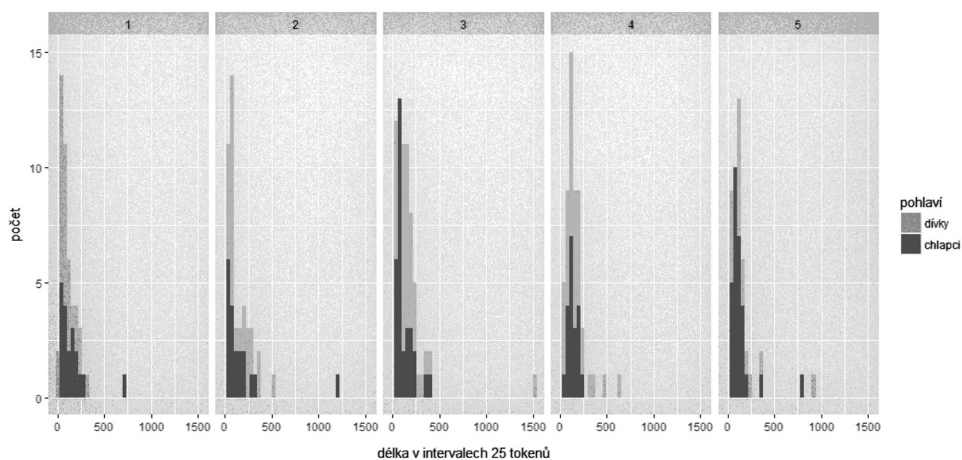
| Ročník | Dívky | Chlapci | Součet |
|--------|-------|---------|--------|
| 1      | 28    | 19      | 46     |
| 2      | 27    | 19      | 46     |
| 3      | 49    | 31      | 80     |
| 4      | 34    | 20      | 54     |
| 5      | 30    | 29      | 59     |

Délka narativ se typicky pohybuje mezi 50 a 250 tokeny napříč ročníky i pohlavími. Nejkratší narativum má 11 a nejdelší narativum 1514 tokenů. Na obrázku 1 vidíme histogramy délek narativ u dívek vs. chlapců ve fazetách (dílčích grafech) nazvaných podle jednotlivých školních ročníků. Na ose X je zobrazena délka jednotlivých narativ v počtu tokenů. Na ose Y je počet dívek, resp. chlapců, jejichž narativa měla danou délku (v rozmezí 25 tokenů). Například světlešedý sloupec ve fazetě 5 nejdále vpravo vyjadřuje, že jedna dívka (viz osa Y) z páté třídy (název fazety) vyprodukovala pohádku o délce 950–975 tokenů (viz osa X). Všechna rozložení délek jsou tzv. *šikmá doprava*. To ukazuje, že delší narativa mají rozmanitější délky.

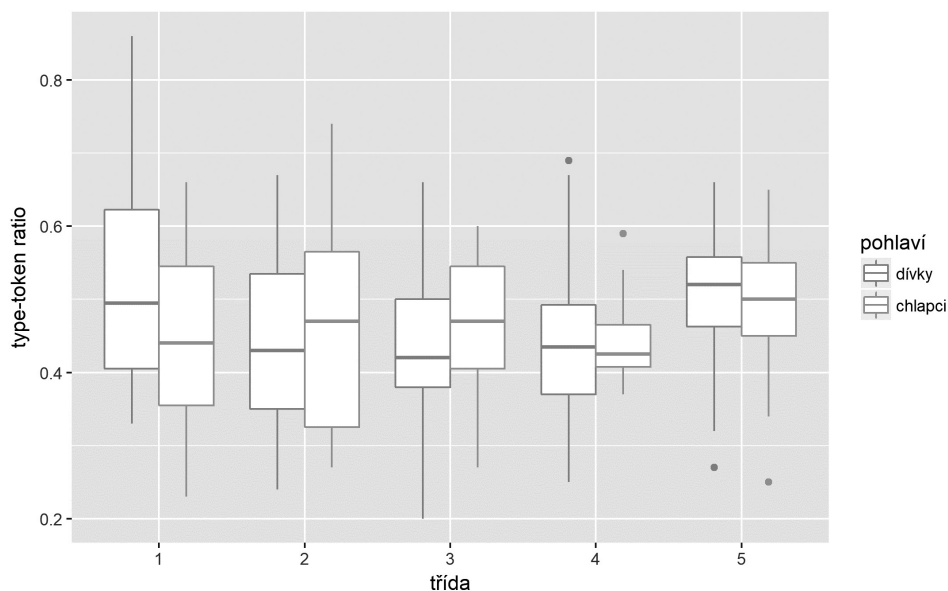
Velmi přibližně jsme také změřily *lexikální diverzitu* narativ mírou *type-token ratio* (TTR). TTR je obecně známá míra definovaná jako poměr mezi počtem typů a počtem tokenů v textu. Pohybuje se mezi nulou a jedničkou (někdy se také přepočítává na procenta). Problémem TTR je citlivost na délku textu. Čím delší text je, tím více v něm autor opakuje funkční slova, která pak TTR snižují. Naopak u textů o jednom slově vychází maximální možná hodnota. Přes tento nedostatek se TTR používá pro hrubé porovnání přibližně stejně dlouhých textů. TTR se používá i při orientačním hodnocení slovní zásoby u dětí (Wiggins – Marshall – Friel, 2007).

Počet typů jsme určily jako počet lemmat, která jsme získaly nástrojem UDPipe (Straka – Hajič – Straková, 2016; Wijffels, 2018). Obrázek 2 představuje rozložení lexikální diverzity podle ročníku a pohlaví ve formě tzv. *krabicových diagramů* (*boxplot*). Každá „krabice“ obsahuje tu typičtější polovinu všech měření, tj. tu okolo střední hodnoty. Jako střední hodnota zde figuruje *medián*, tj. hodnota, která je přesně uprostřed, pokud jednotlivá měření seřadíme podle velikosti. Na rozdíl od aritmetického průměru není medián citlivý na odlehlé hodnoty. Medián je vyznačen jako pruh v krabici. „Vousy“ čnicí z krabic označují rozpětí odlehlejších hodnot. Zcela extrémní odlehlé hodnoty jsou

<sup>2</sup> Na základě tohoto korpusu již bylo publikováno několik převážně kvalitativních studií zabývajících se strukturálními i tematickými aspekty dětského vyprávění také z naratologického hlediska (Náhlíková, 2015, 2016a, 2016b, 2017). Korpus byl vytvořen v rámci doktorské práce A. Náhlíkové a bude zveřejněn po její obhajobě.



**Obrázek 1:** Histogramy délek narativ v tokenech podle ročníků a pohlaví



**Obrázek 2:** Lexikální diverzita narativ podle ročníků a pohlaví; vlevo v každém páru dívky, vpravo chlapci

vyznačeny individuálně jako body. Podle obrázku 2 se medián lexikální diverzity napříč pohlavími i ročníky pohybuje mezi cca 0,45 a 0,55 a není patrná žádná tendence pro věk ani pro pohlaví kromě snížení rozptylu mezi 25. a 75. percentilem od třetí třídy výš. Extrémně vysoké hodnoty TTR v první třídě (dlouhý horní „vous“) jsou způsobeny silným zastoupením extrémně krátkých textů (do 50 slov) a můžeme je zanedbat.

### 3. Operacionalizace hrdinských vs. rodinných narativních prvků

K rozlišení obou žánrů jsme pro účely naší studie stanovily deset kategorií, z nichž pět jsme a priori asociovaly s rodinným žánrem a pět s hrdinským. V závorkách uvádíme jejich zkrácené názvy, které se objevují v diagramech v pozdějších sekcích.

Kategorie rodinného žánru:

1. jasně definovaná skupina (skuphrd);
2. skupina je důležitější než jednotlivec (kol);
3. důraz na vztahy (vzth);
4. vzájemná pomoc mezi členy skupiny / pomoc postavě stojící mimo skupinu (pomoc);
5. cíl: vyřešení/navázání vztahů, spokojenost členů skupiny (c\_vzth).

Kategorie hrdinského žánru:

1. osamocený hrdina (osamhrd);
2. boj/násilí (boj);
3. zranění/smrt (zrsmrt);
4. hrdinské činy (hrd);
5. cíl: vítězství hrdiny (c\_vit).

Ačkoli kritéria připomínají opačné póly pěti škál, jde o deset binárně hodnocených vlastností, které se navzájem nevylučují. Takto pojaté kategorie se osvědčily lépe než škály, protože v roli hlavního hrdiny se na malé ploše může vystřídat několik postav nebo postava se skupinou jako v příkladech (1) a (2).

- (1) dívka, 1. třída:  
byl jeden zub a ten zub ten si ten se dlouho nečistil a jednou byl jeden kluk ktorej měl špinavý zuby protože si je nečistil a a jednou mu maminka řekla že si musí vyčistit zuby a tak von si je vyčistil a už nikdy ho nebolely zuby
- (2) chlapec, 4. třída:  
em byl jednou jeden kůň a ten pořád oral a oral a jednou, když ho pán dal do stáje, tak ten pán zapomněl zavřít ty dveře, no takový ty dveře, od stáje a kůň utek do lesa. v tom lese si zanedlouho našel spoustu kamarádů. medvěda, lišku, srnku, veverku a vlka. a jednou si těch pět, nebo šest, pět, řeklo, že pudou na výpravu a šli a najednou uslyšeli výstřely, em tak a liška, medvěd říkal, že em to sou hony na zvířata, a tak všichni utíkali a em a em ten kůň, a doběhli potom na opuštěnej statek, na kterým ten kůň našel em klisnu, tak oni jí potom vzali sebou a vrátili se, potom se vrátili oklikou do své jeskyně, kde bydleli všichni, a zazvonil konec a pohádky byl konec.

Anotaci prováděla jedna anotátorka, která předem znala celý korpus narativ a při přiřazování vlastností brala v potaz ostatní narativa ve vzorku.

Pro ilustraci zde uvedeme ještě jedno typicky hrdinské (3) a jedno typicky rodinné (4) narativum.

- (3) hrdinské narativum – chlapec, 2. třída, výrazně zkráceno:  
[...] že jsem jel ve vesmíru v obrovské létající lodi ale do mé lodi se dostali nepřátelé pomalu protože jsem si vlastně neuvědomil že sem spal že sem omylem nechal zapnutý most protože sme dělali přepadení na cizí loď ta byla samozřejmě nepřátelská do mé místnůstky se dostali vojáci a protože tam bylo zrovna okno kde kde se přecházelo vyhodili mě ve vesmíru ven tak sem tam chvilku lítal lítal až mě nabrala náhák cizí loď menovala se Gajd Pep Kort bojování světelné žluté meče tak sem si řek tam du protože sem už teďka věděl že teď že už to nebude facka že teď to bude loď bojová vytáhnul sem svůj meč [...] protože zezádu na mě chtěl úto-

čit majitel lodi a já sem a já sem jenom udělal krok a a von spáchal sebevraždu [...] pak se stala pěkná věc uchů letěl sem ňák dólu točil točila se mi hlava dopad sem a zjistil sem že sem v klasickém světě a že už nejsem robot ale sem člověk a že a zjistil sem že sem na normální zeměkouli na pláži a řekl sem si že tohle je prostě ráj a to je konec

- (4) rodinné narativum – dívka, 3. třída (celé):

byly tři sestry který se pořád hádaly a: jednou přišel její jejich táta a: pořád jim povídal aby se nehádaly a oni ho pořá a oni ho neposlouchaly a hádaly se dál a když se jejich máma vrátila z návštěvy z cizí země no tak jim říkala aby se už přestaly pořád hádat a oni se přestaly hádat protože jim máma byla bližší než táta a měly a poslouchaly víc mámu než tátu protože jí víc věřily pak se přestaly hádat a měly se rádi a to je konec

#### 4. Přehled použitých statistických metod

Pro studii žánrových preferencí dětí z našeho vzorku jsme upřednostnily „pohled na data z výšky“ před detailními rozborů jednotlivých textů. Prvním krokem byla manuální anotace tematických narativních prvků popsáná v předchozím oddíle. Na základě této anotace jsme pak hodnotily celý vzorek.

Naše výzkumná otázka má dvě části:

1. Vyskytují se v našem vzorku narativní prvky, které považujeme za „hrdinské“, častěji u chlapců než u dívek a narativní prvky, které považujeme za „rodinné“, častěji u dívek než u chlapců?
2. Jsou případné rozdíly mezi žánrovými preferencemi každého pohlaví při dané velikosti vzorku natolik zřetelné, aby se výsledek dal generalizovat na celou populaci českých mladších žáků?

První část výzkumné otázky zodpovídají metody deskriptivně-exploratorní statistiky. Druhou část zodpovídají metody inferenční statistiky. V praxi, kterou v našem případě reprezentují implementace statistických metod v knihovnách (balíčcích) statistického programovacího jazyka R, použitím určité metody někdy získáme jak exploratorní, tak inferenční statistiku, proto oba typy metod prezentujeme společně. Představujeme zde následující statistické metody:

- *Pearsonův chí-kvadrátový test* a *Cramérovo V*;
- *shluková analýza* (sémantické podobnosti tematických narativních prvků pomocí *vektorového prostoru*);
- *korespondenční analýza* (tematických narativních prvků a skupin dětí podle věku a pohlaví);
- *binomiální logistický regresní model* (závislosti pohlaví na tematických narativních prvcích).

*Chí-kvadrátový test* se používá pro zjištění asociace mezi dvěma kategoriálními proměnnými. Porovnávají se skutečně pozorované počty kombinací kategorií obou proměnných s počty těchto kombinací, které by vyšly, pokud by mezi proměnnými nebyla asociace vůbec žádná. Počítá se pravděpodobnost, že pozorované rozdíly způsobuje běžná náhodná variabilita dat. Pokud je tato pravděpodobnost, tzv. *p-hodnota*, dostatečně malá (v naší studii jsme stanovily hranici 5 procent), nepovažujeme rozdíl mezi očekávanými a pozorovanými hodnotami za náhodný a usuzujeme, že mezi proměnnými



existuje vzájemná závislost. V našem případě jde o kombinace kategorií proměnných *pohlaví* vs. *narativní prvek*. *Cramérovo V* měří sílu vzájemné závislosti proměnných.

*Shluková analýza* je jednou z metod určování vzájemné podobnosti prvků. V naší studii podobnost modelujeme jako souvřasy s ostatními prvky. Pro každý možný pár prvků tedy nejprve počítáme, v kolika narativech se spolu sešly. Každý prvek potom reprezentujeme *mnohorozměrným vektorem*, jehož každý rozměr je vyjádřen počtem jeho souvřasy s jedním dalším prvkem. Pro každý pár vektorů změříme vnitřní úhel, který spolu svírají, tzv. *kosinovou podobnost*. Čím podobnější si vektory-prvky jsou, tím je jejich úhel menší. Tyto vektory pak vstupují do shlukové analýzy. Algoritmus vyžaduje, abychom předem zadaly počet požadovaných shluků. Pak rozdělí datový prostor na tento zadaný počet shluků tak, aby spolu byly v jednom shluku navzájem nejpodobnější prvky, přičemž každý prvek smí patřit pouze do jednoho shluku a shluk může být tvořen i jediným prvkem. Kvalita shluků se validuje statistikou nazývanou *průměrná šířka siluety*; tu jsme zde použily. Algoritmus shlukové analýzy můžeme spouštět opakovaně, experimentovat s počtem shluků a nakonec vybrat počet, který vytvoří nejkvalitnější shluky podle kritéria dané shlukovací metody.

Pro shlukovou analýzu existuje řada různých metod založených na různých algoritmech. Každá metoda definuje členství prvků trochu jinak, například vzájemnou podobností navzájem si nejméně podobných prvků nebo naopak vzájemnou podobností navzájem si nejpodobnějších prvků či průměrnou podobností prvků ve shluku. My jsme zvolily tzv. shlukování kolem *medoidů*. Každý shluk má jeden *medoid*, tj. svého nejtypičtějšího zástupce. Algoritmus najde předurčený počet navzájem si co nejméně podobných prvků a kolem nich shlukuje jim nejpodobnější prvky.

Pomocí shlukové analýzy odhalíme shluky navzájem si nejpodobnějších tematických narativních prvků. Tohoto poznání využijeme k redukci množiny narativních prvků a tím k zahuštění dat. Zahuštění dat může pomoci k dosažení statistické signifikance výsledků. Redukovanou množinu narativních prvků podrobíme *korespondenční analýze*.

*Korespondenční analýza* kromě vztahů mezi kategoriemi dvou různých kategoriálních proměnných navíc zobrazí vztahy mezi *kategoriemi uvnitř každé proměnné* pomocí redukce mnohorozměrného prostoru. Protože však vyžaduje, aby každá z proměnných měla více než dvě kategorie, vytvoříme proměnnou z kombinace pohlaví a třídy (např. *chlapec z první třídy*). Výsledky odečítáme na tzv. *podvojném grafu (biplot)* v různých perspektivách.

*Logistický regresní model* ožřejmuje vztahy mezi závislou kategoriální proměnnou a jednou či více nezávislými proměnnými. U *binomiálního* modelu může závislá kategoriální proměnná nabývat právě dvou hodnot. Model srovnává *poměry šancí (odds ratio)* mezi různými hodnotami nezávislých proměnných (prediktorů), že nastane druhá z hodnot závislé proměnné, s šancí u jedné referenční kombinace hodnot závislých proměnných. Pokud má regresní model silnou předpovědní hodnotu, může se použít k předpovídání výsledku u neznámých dat (srov. oddíl 9.2).

Všechny statistické metody použité v tomto článku jsou popsány v mnoha učebnicích statistiky. My jsme čerpaly z učebnic pro čtenáře z humanitních oborů (Baayen, 2008; Gries, 2009; Levshina, 2015). Pro všechny statistické analýzy jsme používaly programo-

vací jazyk R (R Core Team, 2017). Pro obecné úpravy tabulárních dat a jejich vizualizaci jsme použili rodinu knihoven R tidyverse (Wickham, 2017). Pro většinu statistických metod jsme použily specializované knihovny, jejichž autory také citujeme.

## 5. Rozložení narativních prvků podle pohlaví

Výskyty výše vyjmenovaných deseti narativních prvků podle pohlaví jsme porovnali pomocí testu nezávislosti (Pearsonův chí-kvadrátový test). Pro pohlaví bez ohledu na ročník nám vyšla statisticky signifikantní závislost  $\chi^2 = 50,699$  při 9 stupních volnosti;  $p < 0,05$ . Vzorek dětí tedy můžeme považovat za reprezentativní. Výsledky chí-kvadrátového testu zobrazuje tabulka 2. Sloupec „očekáváno“ obsahuje počty dívek, resp. chlapců, u kterých by se při daném počtu dětí daný narativní prvek vyskytl, kdyby žádné preference podle pohlaví neexistovaly.<sup>3</sup>

Podrobnosti výsledků chí-kvadrátového testu velmi přehledně ilustruje mozaikový diagram na obrázku 3. Mozaikový diagram obsahuje řadu sloupců, přičemž každý sloupec zastupuje jeden narativní prvek. Každý sloupec je vodorovně rozdělen na dva obdélníky. Sloupce tedy vlastně tvoří dvě vodorovné, avšak nepravidelné řady obdélníků nad sebou. Horní řada obdélníků reprezentuje dívky, spodní řada chlapce. Výška každého obdélníku v tomto páru vyjadřuje podíl daného pohlaví na výskytu příslušného narativního prvku: čím vyšší je jeden obdélník, tím nižší musí být druhý; společně dávají součet 100 % dětí, v jejichž narativech se daný narativní prvek vyskytl. Šířka každého sloupce (tj. vertikálního páru obdélníků) znázorňuje počet pozorování daného narativního prvku ve vztahu k ostatním narativním prvkům.

Kdyby byla u každého narativního prvku obě pohlaví zastoupena rovnoměrně, všechny obdélníky by byly stejně vysoké. Stejně tak platí, že kdyby byly všechny narativní prvky u dětí stejně oblíbené, všechny obdélníky by byly stejně široké. Mozaikový graf je pro vizualizaci výsledků chí-kvadrátového testu velmi vhodný, neboť na rozdíl od tabulkového zobrazení zde i necvičené oko může okamžitě porovnat rozdíly v zastoupení prvků i pohlaví.

Nejdůležitější informace však mozaikový graf zobrazuje zbarvením a stylem obrysové čáry jednotlivých obdélníků. Hladká čára znamená kladnou preferenci daného pohlaví pro daný narativní prvek, čerchovaná čára znamená zápornou preferenci. Škála odstínů pak ukazuje sílu dané preference vypočítanou jako tzv. *standardizovaná rezidua* v chí-kvadrátovém testu.

Chí-kvadrátový test jsme musely použít proto, že ve vzorku nebyl stejný počet dívek jako chlapců. U sloupců, kde mají nižší podíl chlapci, by z prostého procentuálního zastoupení nevplývalo, zda se chlapci danému narativnímu prvků skutečně vyhýbají, nebo jestli je to náhoda, protože ve vzorku je chlapců obecně méně. Chí-kvadrátový test abstrahuje od rozdílu v zastoupení jednotlivých kategorií a reprezentuje pozitivní nebo negativní asociace jednotlivých kombinací pomocí tzv. *standardizovaných reziduí*. Čím

---

<sup>3</sup> Intuitivně velmi srozumitelné vysvětlení chí-kvadrátového testu podává např. Walker (2010, s. 108).



vyšší je absolutní hodnota standardizovaného rezidua pro dané pohlaví a narativní prvek, tím silnější je preference daného pohlaví pro tento narativní prvek nebo proti němu.

Podívejme se nyní na frekvenci jednotlivých narativních prvků a na rozložení pohlaví v každém z nich. Na základě tabulky 2 (sloupec „rezidua“) a mozaikového grafu na obrázku 3 konstatujeme, že k závislosti nejvíce přispívá prvek „cíl: vítězství“, který chlapci výrazně preferují a dívky se mu výrazně vyhýbají. Dalšími výraznými prvky jsou „hrdinské činy“, které rovněž výrazně preferují chlapci, a „vztahy“, kterým se chlapci naopak výrazně vyhýbají.

Tento výsledek sice zdánlivě podporuje apriorní představu o typicky ženských a typicky mužských tématech, avšak tyto prvky se vyskytují vzácně, a to i u chlapců. U frekvencovaných narativních prvků dosahují rezidua tak nízkých absolutních hodnot (mezi 0 a 1,96), že ani nedosahují hladiny signifikance 0,05, podle níž se v této studii řídíme. Proto nepřekvapuje, že celková asociace pohlaví a výběru narativních prvků je pro celý soubor narativních prvků velmi slabá – Cramérovo  $V^4$  vychází 0,239.

Nejvíce zastoupenými prvky jsou „osamělý hrdina“ a „skupinový hrdina“. Každé narativum totiž stojí na některém typu hrdiny, někdy i na obou. Pohlavní preference zde nejsou statisticky signifikantní, i když se zdá, že dívky podle očekávání preferují skupinového hrdinu, zatímco chlapci preferují osamělého hrdinu.

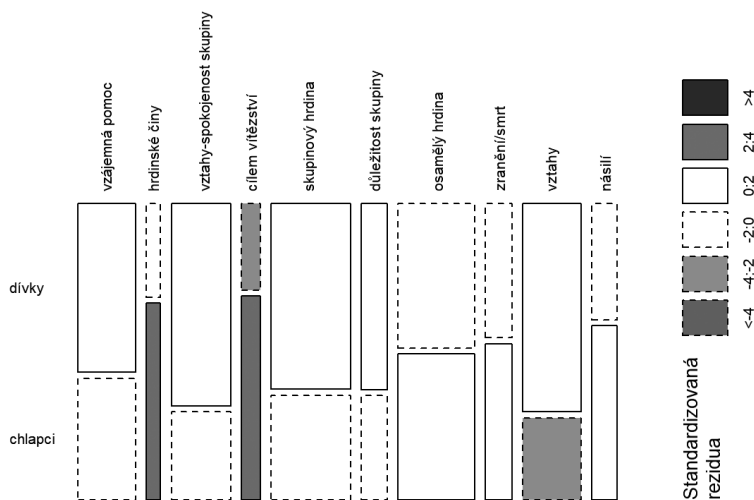
**Tabulka 2:** Pozorované a očekávané hodnoty a rezidua

|                               | Dívky      |           |         | Chlapci    |           |         |
|-------------------------------|------------|-----------|---------|------------|-----------|---------|
|                               | pozorováno | očekáváno | rezidua | pozorováno | očekáváno | rezidua |
| Vzájemná pomoc                | 65         | 65        | 0,006   | 47         | 47        | -0,007  |
| Hrdinské činy                 | 9          | 16        | -1,801  | 19         | 12        | 2,120   |
| Cíl: vztahy, spokojenost      | 80         | 67        | 1,616   | 35         | 48        | -1,902  |
| Cíl: vítězství                | 11         | 21        | -2,263  | 26         | 16        | 2,663   |
| Skupinový hrdina              | 107        | 96        | 1,141   | 58         | 69        | -1,343  |
| Skup. důležitější n. jednotl. | 32         | 29        | 0,549   | 18         | 21        | -0,647  |
| Osamělý hrdina                | 85         | 95        | -0,994  | 78         | 68        | 1,170   |
| Zranění/smrt                  | 24         | 30        | -1,128  | 28         | 22        | 1,328   |
| Vztahy                        | 81         | 66        | 1,897   | 32         | 47        | -2,333  |
| Násilí                        | 20         | 29        | 1,677   | 30         | 21        | 1,974   |

Prvky týkající se typu hrdiny jsou přítomné ve všech narativech – protože bez hrdiny není vyprávění vůbec možné. Ostatní prvky jsou pro vypravěče volitelné, a proto jsou všechny zastoupeny méně než typ hrdiny. Zajímavé je velmi silné a přitom pohlavně nespécifické zastoupení prvku „vzájemná pomoc“ oproti minimálnímu zastoupení typicky mužských prvků „hrdinské činy“ a „cíl: vítězství“. Ve sto osmnácti narativech chlapců se „hrdinské činy“ objevily pouze devatenáctkrát a „cíl: vítězství“ šestadvacetkrát.

Je tedy namístě se ptát, jak čisté jsou vůbec hrdinský a rodinný žánr mezi dětmi zastoupeny. Na tuto otázku odpovídá bodový diagram na obrázku 4. Na ose X je vynes

<sup>4</sup> Použili jsme implementaci v R balíčku lsr (Navarro, 2015).

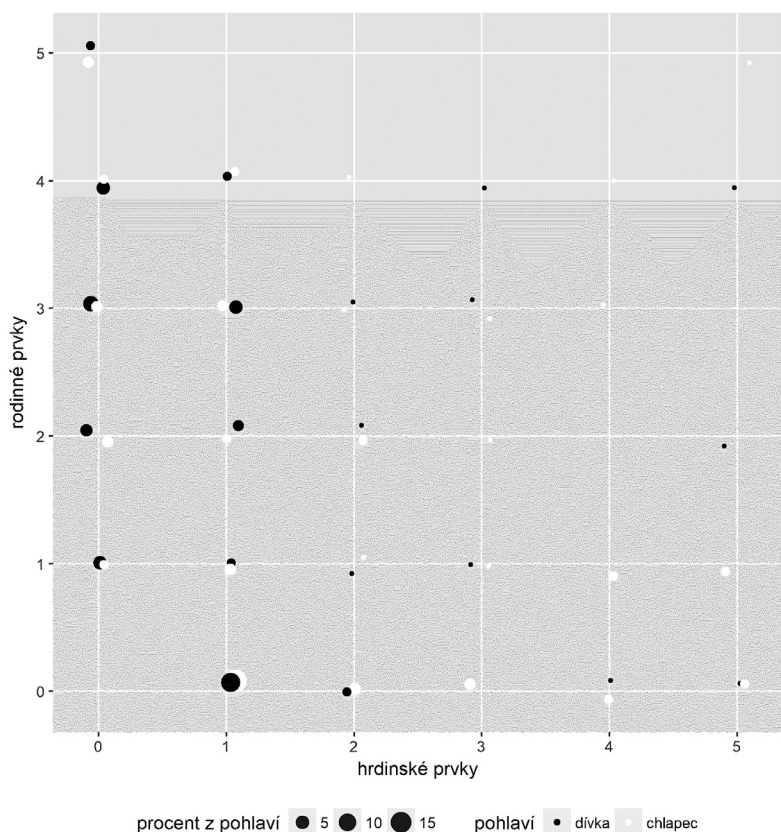


**Obrázek 3:** Zastoupení narativních prvků v korpusu a jejich preference podle pohlaví

počet hrdinských prvků a na ose Y počet rodinných prvků. V soustavě souřadnic se nacházejí dvě sady bodů rozlišené odstíny. Velikost každého bodu odpovídá procentu chlapců, resp. dívek, jejichž narativum obsahovalo kombinaci hrdinských a rodinných prvků vyjádřenou pozicí v souřadnicovém systému. Diagram prozrazuje, že žánr je u značného podílu chlapců i děvčat indikován pouze jedním prvkem, což, jak víme z obrázku 3, je nejčastěji skupinový nebo osamělý hrdina. Více a větších bodů se koncentruje v levé polovině diagramu, tj. u nízkých hodnot hrdinského žánru. Z tematicky bohatších narativ jsou nejvíce zastoupena narativa s nejvýše jedním hrdinským prvkem a aspoň jedním rodinným prvkem, přičemž nezanedbatelný podíl chlapců nepoužil žádný hrdinský prvek, ale zato všech pět rodinných. V pravé dolní, tj. čistě hrdinské, části diagramu převládají chlapci nad dívkami, ale jde o velmi malé počty, jak bylo zřejmé už na obrázku 3 a v tabulce 2.<sup>5</sup>

Můžeme tedy sice říci, že k čistě hrdinskému žánru inklinují téměř výhradně chlapci, zároveň však musíme konstatovat, že takových chlapců je menšina. Pro dívky naopak platí, že inklinují k rodinnému žánru, stejně jako velká část chlapců, kteří použili aspoň dva narativní prvky. V existenci rozdílu v preferencích se náš náález shoduje s Nicolopoulou (1994): „typické“ chlapecké hrdinské a dívčí vyprávění vypadají stejně jako příklady Nicolopoulou. Také poměry pohlaví u kritérií „skupina“ a „násilí“ zhruba odpovídají. V našem vzorku ale (kvalifikovaným odhadem) převažují vyprávění, která takto polarizovaná nejsou a podle příkladů z Nicolopoulou (1994) bychom je považovaly za neutrální nebo lehce vychýlená k rodinné straně škály. Nicolopoulou ve svých pracích o pohlavně podmíněných žánrových preferencích bohužel neuvádí přesná kritéria rozhodování ani žádné další kvantitativní charakteristiky svého vzorku, o něž bychom se mohly při porovnání žánrové čistoty obou vzorků opřít.

<sup>5</sup> Počty dětí jsme normalizovali jako procento dívek z celkového počtu dívek a procento chlapců z celkového počtu chlapců.



**Obrázek 4:** Kombinace hrdinských vs. rodinných narativních prvků u každého pohlaví

## 6. Souvýskyt jednotlivých narativních prvků v shlukové analýze

### 6.1. Postup

Z obrázku 4 je patrné, že většina narativ chlapců i dívek obsahuje prvky z obou žánrů. Abychom zmapovaly jejich souvýskyt v narativech, provedly jsme jeho *shlukovou analýzu* (*cluster analysis*), bez rozlišení pohlaví vypravěče.

Nejprve jsme z anotace vytvořily matici, jejíž sloupce reprezentovaly jednotlivé narativní prvky a jejichž řádky odpovídaly jednotlivým narativům. V této výchozí matici se nacházely prosté počty výskytů jednotlivých prvků v jednotlivých narativech (vždy buď 0, nebo 1). Ukázka je v tabulce 3.

**Tabulka 3:** Prosté souvýskyty narativních prvků – ukázka formátu tabulky

| docID | osamhrd | boj | zrsmrt | hrd | c_vit | skuphrd | kol | vzth | pomoc | c_vzt |
|-------|---------|-----|--------|-----|-------|---------|-----|------|-------|-------|
| d001  | 0       | 0   | 0      | 0   | 0     | 1       | 0   | 1    | 0     | 0     |
| d002  | 1       | 0   | 1      | 0   | 0     | 0       | 0   | 0    | 0     | 0     |
| d003  | 1       | 0   | 0      | 0   | 0     | 0       | 0   | 0    | 0     | 0     |

Na základě těchto hodnot jsme vypočítaly tzv. *pointwise mutual information scores* (PMI) (Levshina, 2015, s. 326–331; Levshina, 2014). Tuto úpravu jsme zvolily, protože míra PMI citlivě zachycuje i málo frekventované souvýskyty, což bylo potřeba vzhledem k malému zastoupení některých narativních prvků. PMI je definována takto:

$$PMI(x, y) = \log_2 \frac{O_{xy}}{E_{xy}}$$

Proměnné  $x$  a  $y$  jsou zde dokument a narativní prvek – např. dokument *d001* a narativní prvek „osamělý hrdina“, v tabulce 3 značený jako *osamhrd*.  $O$  je jejich pozorovaná hodnota (0) a  $E$  je jejich hodnota očekávaná v případě, že by spolu navzájem výskyty prvků vůbec nesouvisely. Matici očekávaných hodnot jsme z výchozí matice získaly pomocí Pearsonova chí-kvadrátového testu<sup>6</sup> (viz ukázkou v tabulce 4).

**Tabulka 4:** Očekávané hodnoty souvýskytu narativních prvků – ukázka formátu tabulky

| docID | osamhrd | boj   | zrsmrt | hrd   | c_vit | skuphrd | kol   | vzth  | pomoc | c_vzt |
|-------|---------|-------|--------|-------|-------|---------|-------|-------|-------|-------|
| d001  | 0,368   | 0,113 | 0,118  | 0,063 | 0,084 | 0,373   | 0,113 | 0,255 | 0,253 | 0,26  |
| d002  | 0,368   | 0,113 | 0,118  | 0,063 | 0,084 | 0,373   | 0,113 | 0,255 | 0,253 | 0,26  |
| d003  | 0,184   | 0,056 | 0,059  | 0,032 | 0,042 | 0,186   | 0,056 | 0,128 | 0,127 | 0,13  |

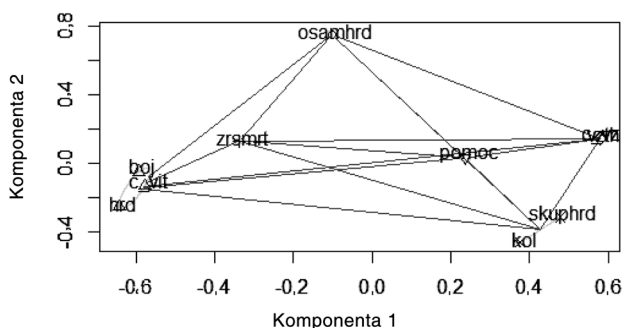
Hodnoty PMI jsme transformovaly na PPMI (*positive pointwise mutual information*) tak, že jsme hodnoty menší než nula nahradily nulami (srov. Levshina, 2015, s. 327). Výsledná matice PPMI představuje vektorový prostor, kde je každý prvek mnohorozměrným sémantickým vektorem, jehož jednotlivými rozměry jsou jednotlivé texty. Standardní mírou pro určování podobnosti vektorů je *kosinová podobnost*. Z matice PPMI jsme tedy spočítaly matici kosinových podobností (v tabulce 5 jsou hodnoty zaokrouhleny).

**Tabulka 5:** Matice kosinových podobností narativních prvků<sup>7</sup>

|         | osamhrd | boj  | zrsmrt | hrd  | c_vit | skuphrd | kol  | vzth | pomoc | c_vzt |
|---------|---------|------|--------|------|-------|---------|------|------|-------|-------|
| osamhrd | 1,00    | 0,17 | 0,18   | 0,04 | 0,14  | 0,04    | 0,01 | 0,13 | 0,16  | 0,15  |
| boj     | 0,17    | 1,00 | 0,32   | 0,45 | 0,38  | 0,05    | 0,10 | 0,04 | 0,17  | 0,04  |
| zrsmrt  | 0,18    | 0,32 | 1,00   | 0,25 | 0,15  | 0,12    | 0,09 | 0,10 | 0,16  | 0,05  |
| hrd     | 0,04    | 0,45 | 0,25   | 1,00 | 0,47  | 0,04    | 0,04 | 0,01 | 0,19  | 0,02  |
| c_vit   | 0,14    | 0,38 | 0,15   | 0,47 | 1,00  | 0,11    | 0,03 | 0,02 | 0,07  | 0,05  |
| skuphrd | 0,04    | 0,05 | 0,12   | 0,04 | 0,11  | 1,00    | 0,38 | 0,35 | 0,31  | 0,34  |
| kol     | 0,01    | 0,10 | 0,09   | 0,04 | 0,03  | 0,38    | 1,00 | 0,21 | 0,20  | 0,21  |
| vzth    | 0,13    | 0,04 | 0,10   | 0,01 | 0,02  | 0,35    | 0,21 | 1,00 | 0,29  | 0,64  |
| pomoc   | 0,16    | 0,17 | 0,16   | 0,19 | 0,07  | 0,31    | 0,20 | 0,29 | 1,00  | 0,33  |
| c_vzt   | 0,15    | 0,04 | 0,05   | 0,02 | 0,05  | 0,34    | 0,21 | 0,64 | 0,33  | 1,00  |

<sup>6</sup> V tomto případě nám nemusí překážet hodnoty menší než 5, jako by tomu bylo při provádění testu nezávislosti.

<sup>7</sup> Vysvětlivky: osamhrd = „osamělý hrdina“, boj = „boj a násilí“, zrsmrt = „zranění a smrt“, hrd = „hrdinské činy“, c\_vit = „cílem je vítězství hrdiny“, skuphrd = „skupinový hrdina“, kol = „důležitější je skupina než jednotlivci“ (tj. „kolektiv“), vzth = „vztahy“, pomoc = „vzájemná pomoc“, c\_vzt = „cílem je navázání nebo udržení vztahu, nebo spokojenost členů skupiny“.



**Obrázek 5:** Shlukování narativních prvků bez ohledu na pohlaví; horizontální a vertikální dimenze reprezentované osami X a Y dohromady vysvětlují 58,96 % variability dat

Tuto podobnostní matici jsme transformovali do distanční matice s normalizací podle nejvyšší naměřené podobnosti kromě 1 a tuto matici použily na vstupu shlukovacího algoritmu funkce *pam* balíčku *cluster* programovacího jazyka R.

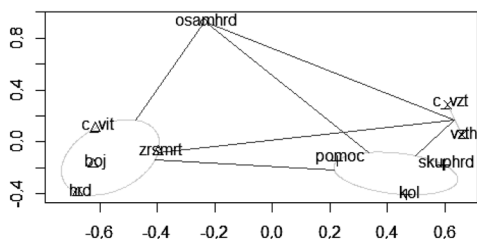
## 6.2. Výsledek

Protože shlukovacímu algoritmu je nutné předem zadat počet výsledných shluků, experimentovaly jsme s jejich počtem. Použily jsme metodu shlukování kolem medoidů (PAM, Maechler et al., 2018) a pokusy jsme validovaly pomocí průměrné šířky siluety pro celý soubor.

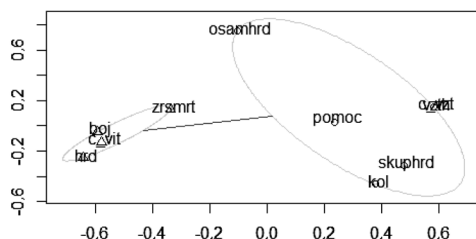
Nejúspěšnější byla šestice shluků kolem těchto prvků – medoidů (obrázek 5):

1. osamělý hrdina (osamhrd);
2. hrdinské činy (hrd) (a kolem něj: cíl: vítězství, boj a násilí (c\_vit, boj));
3. zranění a smrt (zranrt);
4. kolektiv – skupina důležitější než jednotlivec (kol) (a kolem něj: skupinový hrdina (skuphrd));
5. vztahy (vzth) (a kolem něj: cíl: navázání/udržení vztahů, spokojenost skupiny (c\_vzth));
6. vzájemná pomoc (pomoc).

Obrázek 5 zobrazuje vzájemnou vzdálenost jednotlivých narativních prvků při redukci mnohorozměrného prostoru mezi nimi na dvě nejdůležitější dimenze, tj. ty, které vysvětlují největší část variability dat, zde necelých 60 %. První (horizontální) dimenze od sebe nejdál vidí shluk tvořený medoidem „hrdinské činy (hrd)“ zcela vlevo a shluk kolem medoidu „vztahy (vzth)“ zcela vpravo. Druhá (vertikální) dimenze od sebe nejdál vidí „osamělého hrdinu (osamhrd)“ zcela nahoře a „důležitost kolektivu (kol)“ zcela dole. Zároveň k sobě mají blízko jak důležitost kolektivu se vztahy (vpravo a vpravo dole), tak hrdinské činy, vítězství, boj a násilí se zraněním a smrtí. Elipsy znázorňují intervaly podobnostních hodnot, které jsou kritériem pro členství v daném shluku. Na našem obrázku se ani náznakem nepřekrývají, což je dobré znamení, že shluky jsou od sebe opravdu snadno odlišitelné.



**Obrázek 6:** Shluky narativních prvků u chlapců; horizontální a vertikální dimenze reprezentované osami X a Y dohromady vysvětlují 54,93 % variability dat



**Obrázek 7:** Shluky narativních prvků u dívek; horizontální a vertikální dimenze reprezentované osami X a Y dohromady vysvětlují 58,96 % variability dat

Analýza hlavních komponent, která je součástí algoritmu PAM, podle obrázku 5 tedy potvrzuje, že naše narativní prvky se shodují s lidskou intuicí, ale že typ hrdiny (individuum vs. skupina) se žánrem příliš nespojují. Vodorovný rozměr v diagramu jasně patří pólům hrdinský – rodinný, zatímco svislý rozměr je charakterizován protikladem „osamělý hrdina“ – „skupinový hrdina“.

Výsledky shlukové analýzy mají oporu i v našem kvalitativním hodnocení narativ. Například prvek „zranění/smrt“, který je umístěn mnohem blíže středu než shluk s čistě hrdinskými prvky zcela vlevo, se hojně vyskytoval v souvislosti s nepřízní osudu, které postava často ani nijak aktivně nečelila (zvíře bylo nečekaně uloveno a zabito, dítěti saje krev z prstů klavír, dítě se zranilo při hře, dítě navštěvuje těžce nemocnou matku v nemocnici atp.). Prvek „osamělý hrdina“ je téměř přesně vprostředku, protože jím může být jak hrdina aktivní a bojovný, tak hrdina pasivní, citlivý a smířlivý. Jak jsme již uvedli, řada narativ má hrdinu individuálního i skupinového. Jde typicky o narativa s rodinnými prvky. Pro ty je typický i čistě skupinový hrdina.

### 6.3. Shluková analýza narativních prvků podle pohlaví

Stejný postup jsme uplatnili na subkorpus narativ chlapců, resp. dívek (obrázky 6 a 7). Medoidy u chlapců byly: „osamělý hrdina“, „hrdinské činy“, „skupinový hrdina“, „cíl: navázání/udržení vztahů, spokojenost skupiny“. U dívek to byly „boj/násilí“ a „cíl: navázání/udržení vztahů, spokojenost skupiny“. Na obou obrázcích jsou elipsy k sobě navzájem blíže, ale ani tady se nepřekrývají, takže shluky jsou jeden vůči druhému stále výrazně odlišeny.

## 7. Rozložení narativních prvků podle věku a pohlaví

Z tabulky 2 v oddíle 5 je zřejmé, že zastoupení některých narativních prvků v datech je natolik nízké, že při vyšším počtu kategorií (pět ročníků nebo dokonce pět ročníků krát dvě pohlaví) bychom čelily nedostatku dat. Proto jsme se opřely o výsledky shlukové analýzy a sloučily „hrdinské činy“, „boj/násilí“, „cíl: vítězství“ a „zranění/smrt“ do prvku s názvem „hrdinské“, „vztahy“ s „cíl: navázání/udržení vztahů, spokojenost



skupiny“ do druhého prvku nazvaného „vztahové“ a „skupina důležitější než jednotlivci“ se „skupinovým hrdinou“ do třetího prvku, který jsme pojmenovaly „skupinové“.

Novou množinu tedy tvořily tyto narativní prvky:

- hrdinské (hrdinské činy, boj/násilí, cíl: vítězství, zranění/smrt);
- vztahové (vztahy, cíl: návazání/udržení vztahů, spokojenost skupiny);
- skupinové (skupinový hrdina, skupina důležitější než jednotlivci);
- vzájemná pomoc;
- osamělý hrdina.

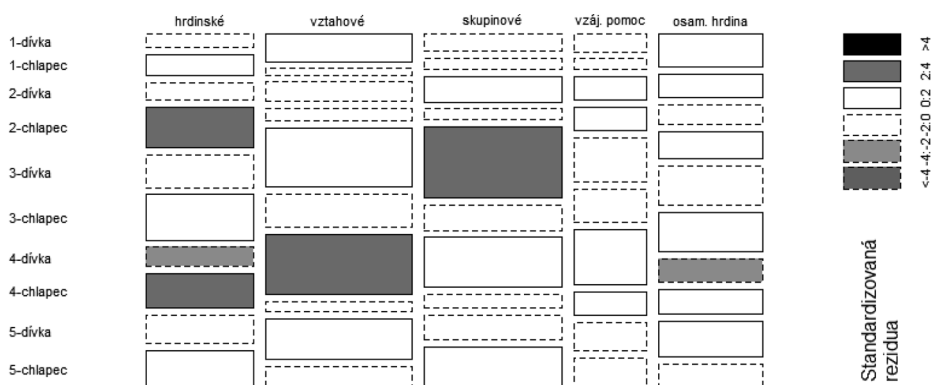
V tabulce 6 najdeme Pearsonův chí-kvadrátový test s obvyklými sloupci pozorovaných a očekávaných hodnot a reziduí (srov. tabulku 2 v oddíle 5), který ukazuje signifikantní souvislost mezi narativními prvky a kategoriemi kombinujícími věk a pohlaví ( $\chi^2 = 92,871$  při 36 stupních volnosti;  $p < 0,05$ ). Celková inerce ( $\chi^2$ /součet všech hodnot tabulky) je 0,104939. Celková inerce popisuje množství informace v tabulce. Odmocněním její hodnoty získáme tzv. *stopu*, kterou můžeme interpretovat jako korelační koeficient závislosti mezi řádky a sloupci tabulky. Stopa pro tuto tabulku je cca 0,3, přičemž stopa vyšší než 0,2 indikuje statisticky signifikantní závislost (Bendixen, 2003, s. 23). Cramérovo V ovšem hodnotou 0,166 naznačuje, že tato závislost, i když je statisticky signifikantní, je extrémně slabá – slabší než u kompletní sady narativních prvků.

**Tabulka 6:** Rozložení shlukovaných narativních prvků po pohlaví a ročnících; názvy řádků obsahují číslo školního ročníku a pohlaví dětí v dané skupině

|           | Hrdinské |     |       | Vztahové |     |       | Skupinové |     |       | Vzájemná pomoc |     |       | Osamělý hrdina |     |       |
|-----------|----------|-----|-------|----------|-----|-------|-----------|-----|-------|----------------|-----|-------|----------------|-----|-------|
|           | Poz.     | Oč. | Rez.  | Poz.     | Oč. | Rez.  | Poz.      | Oč. | Rez.  | Poz.           | Oč. | Rez.  | Poz.           | Oč. | Rez.  |
| 1-dívka   | 8        | 13  | -1,39 | 22       | 18  | 1,00  | 13        | 17  | -0,92 | 7              | 9   | -0,59 | 19             | 13  | 1,76  |
| 1-chlapec | 12       | 8   | 1,45  | 5        | 11  | -1,77 | 8         | 10  | -0,69 | 4              | 5   | -0,57 | 13             | 8   | 1,89  |
| 2-dívka   | 10       | 12  | -0,60 | 15       | 16  | -0,37 | 19        | 16  | 0,88  | 9              | 8   | 0,32  | 11             | 12  | -0,23 |
| 2-chlapec | 23       | 12  | 3,07  | 10       | 17  | -1,65 | 8         | 16  | -1,96 | 9              | 8   | 0,27  | 15             | 12  | 0,88  |
| 3-dívka   | 19       | 29  | -1,92 | 46       | 40  | 0,92  | 52        | 38  | 2,29  | 17             | 20  | -0,62 | 22             | 29  | -1,26 |
| 3-chlapec | 26       | 20  | 1,34  | 26       | 27  | -0,25 | 19        | 26  | -1,33 | 13             | 13  | -0,11 | 22             | 20  | 0,56  |
| 4-dívka   | 11       | 24  | -2,70 | 47       | 33  | 2,39  | 37        | 31  | 1,01  | 21             | 16  | 1,16  | 13             | 24  | -2,21 |
| 4-chlapec | 20       | 12  | 2,50  | 8        | 16  | -1,95 | 10        | 15  | -1,25 | 9              | 8   | 0,46  | 14             | 11  | 0,82  |
| 5-dívka   | 16       | 18  | -0,50 | 31       | 25  | 1,26  | 18        | 23  | -1,10 | 11             | 12  | -0,33 | 20             | 18  | 0,55  |
| 5-chlapec | 22       | 18  | 0,86  | 18       | 25  | -1,40 | 31        | 24  | 1,53  | 12             | 12  | -0,08 | 14             | 18  | -0,91 |

Tabulku 6 opět doprovází mozaikový diagram (obrázek 8; pro vysvětlení vizuální gramatiky mozaikového diagramu srov. obrázek 3). Tabulka 6 a obrázek 8 odhalují, že rozdíl v oblíbenosti hrdinských narativních prvků mezi pohlavími byl nejnápadnější ve čtvrté třídě. Ve čtvrté třídě byl největší rozdíl v oblíbenosti vztahových narativních prvků. Skupinové narativní prvky byly nejpreferovanější u dívek ve třetí třídě, kde se také vyskytly nejčastěji. U dívek ve čtvrté třídě pozorujeme odklon od osamělého hrdiny.

Poměr oblíbenosti hrdinských i vztahových narativních prvků mezi pohlavími s věkem tedy kolísá, ale nevidíme žádnou tendenci. Znovu se přitom ukazuje, že prvek „vzájemná pomoc“ je vlastní všem dětem obojího pohlaví napříč ročníky.



Obrázek 8: Mozaikový graf rozložení shlukovaných narativních prvků po pohlavích a ročnících

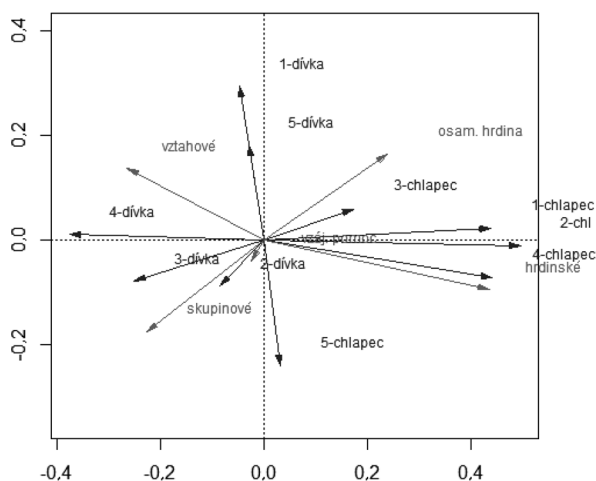
## 8. Korespondenční analýza dětí podle věku a pohlaví na základě jejich preferencí narativních prvků

### 8.1. Jednoduchá korespondenční analýza

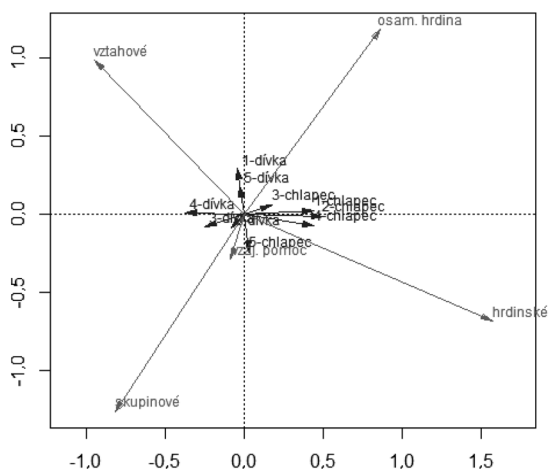
Když jsme zkoumaly preference narativních prvků u dětí podle pohlaví, sledovaly jsme zastoupení jednotlivých prvků a jejich souvšly izolovaně pomocí mozaikového grafu a shlukové analýzy. Rozdělením dětí navíc podle ročníků jsme dosáhly počtu kategorií vyššího než 2 u obou proměnných, takže můžeme použít korespondenční analýzu (Nenadic – Greenacre, 2007). Korespondenční analýza redukuje mnohorozměrný prostor vztahů mezi jednotlivými kategoriemi.

Výsledky korespondenční analýzy se nejlépe odečtou z grafu (obrázek 9). Graf je vyneseno na dvou osách, které, podobně jako u shlukové analýzy, reprezentují dva rozměry, jež dohromady vysvětlí největší podíl variace v datech. Tento typ grafu se nazývá *symetrický dvojný graf* (biplot). Na grafu na obrázku 9 vidíme deset dětských kategorií (dvě pohlaví krát pět ročníků) a šest kategorií narativních prvků z množiny prvků redukované na základě shlukové analýzy. Rozměry podél osy X a podél osy Y dohromady vysvětlují 91,3 % variability dat. Jednotlivé kategorie proměnných jsou vyjádřeny vektory ukazujícími do různých směrů od středu grafu. Osu X definuje jednak protiklad mezi dívkami (od středu doleva) a chlapci (od středu doprava), jednak protiklad mezi skupinou a vztahy (doleva) a osamělým hrdinou a hrdinskými narativními prvky (doprava). Osu Y, která ovšem vysvětluje menšinu variability dat, neumíme smysluplně interpretovat. Osa Y nevyjadřuje, jak bychom doufaly, věk dětí, protože nahoře i dole jsou zastoupeny všechny ročníky s výjimkou první a druhé třídy.

Vzájemnou asociaci každého narativního prvku s každou kategorií dětí můžeme odečíst z velikosti vnitřního úhlu, který spolu svírají. Čím více se úhel liší od pravého úhlu, tím silnější asociaci pozorujeme. Je-li úhel ostrý, je asociace pozitivní; je-li úhel tupý, je asociace negativní. Nejsilnější negativní asociace je v přímém úhlu. Záleží ovšem také na vzdálenosti bodu od průsečíku os. Čím dále je bod od průsečíku os, tím silněji definuje osu, tj. tím je důležitější pro interpretaci osy.

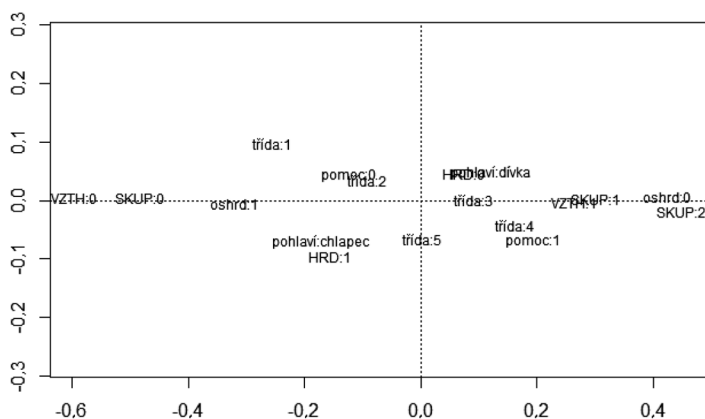


**Obrázek 9:** Symetrický dvojný graf narativních prvků vs. „věkopohlaví“ dětí; horizontální dimenze vysvětluje 73 % variability dat, vertikální dimenze vysvětluje 18,3 % variability dat



**Obrázek 10:** Normalizovaná vzdálenost „věkopohlaví“ dětí; horizontální dimenze vysvětluje 73 % variability dat, vertikální dimenze vysvětluje 18,3 % variability dat

Obrázek 9 například potvrzuje pozorování z mozaikového grafu ohledně asociace dívek se vztahovými narativními prvky: nejmenší a ostrý úhel s nimi svírá kategorie dívek ze čtvrtého ročníku (silná pozitivní asociace); lehce tupý úhel s nimi svírá kategorie dívek z druhého ročníku (slabá negativní asociace). Velmi silnou pozitivní asociaci vidíme také mezi hrdinskými narativními prvky a chlapci ze čtvrté, druhé a první třídy. Velmi silnou negativní asociaci pozorujeme mezi osamělým hrdinou a dívkami ze třetí a čtvrté třídy. Podle mozaikového grafu a tabulky 6 existuje nejsilnější negativní asociace s osamělým hrdinou u dívek ze čtvrté třídy a druhá nejsilnější u dívek ze třetí třídy.



**Obrázek 11:** Vícenásobná korespondenční analýza narativních prvků vs. věku vs. pohlaví dětí; obě dimenze dohromady vysvětlují 78,9 % variability dat; osa X vysvětluje 76,9 % a osa Y pouze 2 % variability dat

Podle velikosti úhlu by nejsilnější negativní asociaci s osamělým hrdinou měly dívky ze třetí třídy, třebaže z tabulky vyplývá nejsilnější negativní asociace osamělého hrdiny s dívkami ze čtvrté třídy; zde je třeba vzít v úvahu vzdálenost každé z dívčích kategorií od středu: dívky ze čtvrté třídy jsou dál. Posoudit vliv rozdílů vzdálenosti od středu a rozdílů ve velikosti úhlu pouhým okem neumíme, a proto musíme zkontrolovat původní data.

Ze symetrického grafu ovšem nevyčteme skutečné vzdálenosti mezi kategoriemi jedné proměnné. K tomu slouží graf na obrázku 10, kde jsou vzdálenosti mezi jednotlivými kategoriemi dětí normalizovány (pozor, za cenu toho, že si neodpovídají vzdálenosti narativních prvků mezi sebou). Tento graf nám říká vlastně totéž co obrázek 4 – totiž že čistě hrdinský žánr je doménou chlapců, ale je i mezi chlapci výjimečný – s doplněním, že nejtypičtější je pro chlapce z první, druhé a čtvrté třídy (přesné pořadí je opět nutné odečíst z reziduí v tabulce 6, protože úhly jsou velmi podobné a vzdálenosti od středu rozdílné). Zároveň je na obrázku 10 dobře vidět, že rozdíly mezi kategoriemi věku a pohlaví jsou celkově malé.

## 8.2. Vícenásobná korespondenční analýza

Jednoduchá korespondenční analýza neukázala, že by jedním z dvou nejinformativnějších rozměrů byl věk (tj. školní ročník). Není ovšem schopna zachytit věk jako samostatnou proměnnou. Proto jsme provedly ještě vícenásobnou korespondenční analýzu, jejímž výsledkem je podvojný graf na obrázku 11. Ročník zachycujeme samostatnou proměnnou „třída“, která je ovšem také kategoriální, protože korespondenční analýza s kvantitativními proměnnými nepracuje.

Tento pohled na data napovídá, že narativa dětí z 1. a 2. třídy se liší od narativ starších dětí. To souhlasí s naším kvalitativním hodnocením. Důvodem je zřetelně nižší kvalita narativ – u mnohých jsme nebyly schopné anotovat jiné znaky než typ hrdiny (srov. obrázek 4). S chlapci je nejsilněji asociován hrdinský žánr (HRD:1) a hned potom

absence žánru rodinného (VZTH:0), zatímco u dívek je tomu naopak. To odpovídá pozorování úhlů z obrázku 9: úhly mezi dívkami 2.–4. ročníku a vztahy jsou mnohem větší než úhly mezi hrdinskými prvky a chlapci vyjma snad 5. ročníku, přičemž chlapci jsou také zpravidla dál od středu než dívky.

Výsledek vícenásobné korespondenční analýzy sice v grafu budí dojem, že ročníky jsou vertikálně seřazeny sestupně, ale jejich vertikální rozptyl je velice malý v porovnání s horizontálním. Navíc celá svislá osa vysvětluje pouze 2,7 % variability dat. Vodorovná osa sice vysvětluje trochu více variability než vodorovná osa modelu založeného na jednoduché korespondenční analýze, ale dohromady vysvětlí jenom 78,9 %, zatímco jednoduchá korespondenční analýza vysvětlila 91,3 %. Celková inerce tohoto modelu je 0,100039, což je trochu méně než celková inerce modelu založeného na jednoduché korespondenční analýze (0,104939). Jeho stopa je cca 0,31. Oddělení ročníku coby samostatné proměnné tedy nepřineslo více informace.

## 9. Diskuse

### 9.1. Shrnutí

Všechny předchozí analýzy v souladu s naším intuitivním předpokladem naznačují, že chlapci tíhnou k prvkům, jako je *vůle k vítězství, boj a násilí, hrdinské činy a zranění či smrt*, zatímco dívky se jim vyhýbají trochu silněji, než je chlapci preferují. Naopak dívky preferují narativní prvky zaměřené na vztahy, ale méně, než chlapci preferují hrdinské prvky. Chlapci jako by svými preferencemi obsáhli prvky obou žánrů, zatímco dívky se vyhýbají hrdinskému žánru.

Na tomto místě bychom ale chtěly zdůraznit, že korespondenční analýza sice přehledně naznačuje, u kterých narativních prvků se u (různě starých) dívek a chlapců liší preference relativně nejvíce a u kterých nejméně, ale není z ní na první pohled zřejmé, jak silné ty preference ve skutečnosti jsou. Při interpretaci podvojných grafů na obrázcích 9–11 musíme mít stále na paměti, že i ty nejsilnější naměřené preference jsou pro praktické účely zanedbatelné. Vzorek je statisticky signifikantní, takže ho (při pětiprocentní pravděpodobnosti chyby) můžeme generalizovat na celou populaci českých mladších školních dětí.

### 9.2. Ještě jedna ukázka zanedbatelnosti preferencí narativních prvků podle pohlaví

Jak slabě preference narativních prvků souvisí s pohlavím vypravěče, nakonec ukážeme na binomiálním logistickém regresním modelu, který tipuje pohlaví vypravěče podle zvolených narativních prvků. Závislou proměnnou je v tomto modelu pohlaví, vysvětlujícími proměnnými jsou:

- skupina hrdinských prvků (HRD);
- skupina vztahových/rodinných prvků (VZTH);
- osamělý hrdina, skupinový hrdina (oshrd);
- vzájemná pomoc (pomoc).

Tabulka 7 zobrazuje efekty přítomnosti jednotlivých proměnných, jejich směrodatnou chybu a jejich statistickou signifikanci.<sup>8</sup> Efekty jednotlivých prediktorů ve sloupci „Regresní koeficient“ (*estimate*) se srovnávají s „Referenční hodnotou“ (*intercept*).<sup>9</sup> Tou je v tomto případě kombinace vypravěče „dívka“ a tematického prvku „kol“, tj. „kolektivní hrdina“. Počet hvězdiček v posledním sloupci ilustruje statistickou signifikanci. Nula znamená, že výsledek není statisticky signifikantní. Jedna hvězdička znamená, že výsledek se vejde do hladiny signifikance 0,05. Čím vyšší počet hvězdiček, tím nižší je p-hodnota, tj. tím více statisticky signifikantní výsledek je.

Statisticky signifikantní, a proto důvěryhodné, jsou pouze skupina hrdinských prvků HRD a skupina vztahových/rodinných prvků VZTH. Koeficient (síla efektu) HRD zvyšuje pravděpodobnost předpovědi „chlapec“, zatímco VZTH je negativní (přítomnost VZTH pravděpodobnost předpovědi „chlapec“ snižuje). Oba výsledky jsou v souladu s našimi předchozími analýzami.

**Tabulka 7:** Binomiální regresní model předpovídající pohlaví z narativních prvků

|                    | Regresní koeficient | Směrodatná chyba | z-hodnota | p-hodnota |    |
|--------------------|---------------------|------------------|-----------|-----------|----|
| Referenční hodnota | -0,3175             | 0,4511           | -0,704    | 0,481582  |    |
| HRD                | 0,6992              | 0,2690           | 2,600     | 0,009334  | *  |
| VZTH               | -1,3565             | 0,4056           | -3,344    | 0,000825  | ** |
| oshrd              | 0,3316              | 0,3844           | 0,863     | 0,388336  |    |
| SKUP               | 0,4393              | 0,2993           | 1,468     | 0,142087  |    |
| pomoc              | 0,3204              | 0,2756           | 1,163     | 0,245028  |    |

Předpovědní síla modelu je ovšem nepatrná; jinak řečeno, efekty statisticky signifikantních prediktorů jsou příliš slabé. Předpovědní sílu modelu odečítáme z ukazatele *reziduální odchylka* (*residual deviance*, tedy celková vzdálenost pozorovaných bodů od přímky vytvořené modelem, zde 387,69 na 285 stupňů volnosti), jehož hodnotu porovnáваме s hodnotou *nulové odchylky* (zde 359,78 na 280 stupňů volnosti). Čím podobnější si tyto dvě hodnoty jsou, tím je model slabší.

Zanedbatelně lépe by fungoval model, který by na základě pohlaví vypravěče a přítomnosti ostatních prvků předpovídal přítomnost skupiny hrdinských prvků: jeho reziduální odchylka se více lišila od nulové odchylky (341,48 na 280 stupňů volnosti oproti 362,21 na 285 stupňů volnosti). Zde má pohlaví „chlapec“ signifikantní kladný vliv na pravděpodobnost, že žánr bude HRD (síla efektu je 0,70137), spolu s přítomností vzájemné pomoci (síla efektu je 0,57329).

Se značným odstupem nejlépe by fungoval model, který by na základě pohlaví vypravěče a přítomnosti ostatních prvků předpovídal přítomnost VZTH: největší rozdíl

<sup>8</sup> V R implementováno jako `glm(formula = pohlaví ~ HRD + VZTH + oshrd + skup + kol + pomoc, family = „binomial“, data = alsex)`. Referenční hodnotou je dívka. Diskutované hodnoty reziduální a nulové odchylky, stupně volnosti a AIC jsou součástí výstupu této funkce.

<sup>9</sup> Pozor, hodnoty regresních koeficientů jsou v logistických regresních modelech transformované funkcí `logit`.



mezi nulovou a reziduální odchylkou (353,06 na 285 stupňů volnosti minus 112,05 na 280 stupňů volnosti). Statisticky signifikantní je negativní vliv pohlaví „chlapec“ (-1,2170) a pozitivní vliv vzájemné pomoci (1,6212).

Stejně vychází porovnání těchto tří modelů pomocí AIC (*Akaike's information criteria*). AIC je měřítko relativní kvality statistických modelů pro stejný soubor dat. Čím nižší AIC v tabulce 8, tím lepší je model (uvažují se hlavně pravděpodobnosti hodnot vs. počet prediktorů).

**Tabulka 8:** Porovnání regresních modelů

| Model     | AIC    |
|-----------|--------|
| “Pohlaví“ | 371,80 |
| “HRD”     | 353,02 |
| “VZTH”    | 124,05 |

## 10. Závěr

Na základě přepisů spontánních narativ dětí mladšího školního věku jsme zkoumaly předpoklad, že chlapci preferují hrdinský žánr a dívky rodinný. Ve studii jsme zohlednily i rozdílný věk dětí. Dospěly jsme k závěru, že tento předpoklad je naplněn ve velmi omezené míře, která je pro praktické použití (např. zadávání inspirativních slohových prací) irelevantní. Tento výsledek, tj. že mezi chlapci a děvčaty neexistují žánrové preference, považujeme za stejně zajímavý, jako kdybychom zjistily, že žánrové preference se u pohlaví liší. Za nejhodnotnější výsledek naší studie konečně považujeme zjištění, že všechny děti bez rozdílu věku a pohlaví tematizovaly vzájemnou pomoc.

Čtení textů nás přivedlo na další zajímavé tematické prvky, které jsme však z časových důvodů nezkoumaly, například:

- Agentivita hlavního hrdiny / hlavních hrdinů: vytváří hrdina nové situace, nebo spíše reaguje na nastalé situace?
- Míra „lidskosti“ hlavního hrdiny / hlavních hrdinů: řeší lidské problémy, nebo problémy specifické pro svůj „druh“ (např. oprýskaný lak na staré lodi, kuklení u housenky)?

Domníváme se proto, že data v našem souboru ještě zdaleka nejsou dostatečně vyčerpána a skýtají prostor pro další výzkum.

### Poděkování

Tato studie byla podpořena grantem Grantové agentury České republiky GA ČR 19-19191S. Také jsme použily nástroje a data výzkumné infrastruktury LINDAT/CLARIAH CZ (MŠMT, LM2018101). Děkujeme recenzentům za pečlivost a podnětné připomínky. Zvláštní dík patří Zdeňku Hlávkovi z Katedry pravděpodobnosti a matematické statistiky MFF UK za hodiny trpělivých konzultací v začátcích práce. Chyby v dalším průběhu studie padají výhradně na naši hlavu.

## LITERATURA

- BAAYEN, H. (2008): *Analyzing Linguistic Data: A Practical Introduction to Statistics Using R*. Cambridge: Cambridge University Press.
- BENDIXEN, M. (2003): A practical guide to the use of correspondence analysis in marketing research. *Marketing Bulletin*, 14, Technical Note 2. Dostupné online na adrese: <[http://marketing-bulletin.massey.ac.nz/V14/MB\\_V14\\_T2\\_Bendixen.pdf](http://marketing-bulletin.massey.ac.nz/V14/MB_V14_T2_Bendixen.pdf)>.
- GRIES, S. T. (2009): *Statistics for Linguistics with R: A Practical Introduction*. Berlin – New York: Mouton de Gruyter.
- HARČARIKOVÁ, P. – KLIMOVIČ, M. (2011): *Naratíva v detskej reči*. Prešov: Prešovská univerzita v Prešove, Pedagogická fakulta, Katedra komunikačnej a literárnej výchovy, Kabinet výskumu detskej reči a kultúry.
- LABOV, W. – WALETZKY, J. (1967): Narrative analysis: Oral versions of personal experience. In: J. Helms (ed.), *Essays in the Verbal and Visual Arts*. Seattle: University of Washington Press, s. 12–44.
- LEVSHINA, N. (2014): *Rling: A Companion Package for How to Do Linguistics with R. R package version 1.0*. John Benjamins. Dostupné online na adrese: <<https://benjamins.com/sites/z.195/content/package.html>>.
- LEVSHINA, N. (2015): *How to Do Linguistics with R: Data Exploration and Statistical Analysis*. Amsterdam – Philadelphia: John Benjamins.
- LIBBY, N. M. – ARIES, E. (1989): Gender differences in preschool children's narrative fantasy. *Psychology of Women Quarterly*, 13 (3), s. 293–306.
- MAECHLER, M. – ROUSSEEUW, P. – STRUYE, A. – HUBERT, M. – HORNIK, K. (2018): *cluster: Cluster Analysis Basics and Extensions. R package version 2.0.7-1*.
- MÜLLER, R. – ŠIDÁK, P. (2012): *Slovník novější literární teorie: glosář pojmů*. Praha: Academia.
- NÁHLÍKOVÁ, A. (2015): Analýza struktury a koherence dětských narativ (se zaměřením na rozdíl mezi narativy chlapců a dívek). *Didaktické studie*, 7 (2), s. 116–135.
- NÁHLÍKOVÁ, A. (2016a): Konflikt v narativech dětí mladšího školního věku. In: J. Wiendl – L. Saicová Římalová – I. Vaňková (eds.), *Lidský život a každodennost v jazyce/literatuře*. Praha: Univerzita Karlova, Filozofická fakulta, s. 144–153.
- NÁHLÍKOVÁ, A. (2016b): Zápletka v mluvených narativech dětí mladšího školního věku. *Jazyk a kultura*, 7 (25–26), s. 90–103.
- NÁHLÍKOVÁ, A. (2017): Hlavní postava v narativech dětí mladšího školního věku. *Jazykovědné aktuality*, 54 (1–2), s. 44–50.
- NAVARRO, D. (2015): *Learning Statistics with R: A Tutorial for Psychology Students and Other Beginners (Version 0.5)*. Adelaide: University of Adelaide. Dostupné online na adrese: <[www.fon.hum.uva.nl/paul/lot2015/Navarro2014.pdf](http://www.fon.hum.uva.nl/paul/lot2015/Navarro2014.pdf)>.
- NENADIC, O. – GREENACRE, M. (2007): Correspondence Analysis in R, with two- and three-dimensional graphics: The ca package. *Journal of Statistical Software*, 20 (3), s. 1–13.
- NICOLOPOULOU, A. (1994): Gender differences and symbolic imagination in the stories of four-year-olds. In: H. Dyson – C. Genishi (eds.), *The Need for Story: Cultural Diversity in Classroom and Community*. Urbana: National Council of Teachers of English, s. 102–123.
- NICOLOPOULOU, A. (2008a): The elementary forms of narrative coherence in young children's storytelling. *Narrative Inquiry*, 18 (2), s. 299–325.
- NICOLOPOULOU, A. (2008b): Rethinking character representation and its development in children's narratives. In: J. Guo – E. Lieven – N. Budwig – S. Ervin-Tripp – S. Ozcaliskan – K. Nakamura (eds.), *Crosslinguistic Approaches to the Psychology of Language: Research in the Tradition of Dan Isaac Slobin*. New York: Taylor and Francis, s. 241–251.
- PETERSON, C. – MCCABE, A. (1983): *Developmental Psycholinguistics: Three Ways of Looking at a Child's Narrative*. New York: Plenum.
- R CORE TEAM (2017): *R: A Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing. Dostupné online na adrese: <<https://www.R-project.org/>>.

- STRAKA, M. – HAJIČ, J. – STRAKOVÁ, J. (2016): UDPipe: Trainable pipeline for processing CoNLL-U files performing tokenization, morphological analysis, POS tagging and parsing. In: N. Calzolari – K. Choukri – T. Declerck – S. Goggi – M. Grobelnik – B. Maegaard – J. Mariani – H. Mazo – A. Moreno – J. Odijk – S. Piperidis (eds.), *Proceedings of the 10<sup>th</sup> International Conference on Language Resources and Evaluation (LREC 2016)*. Portorož: European Language Resources Association, s. 4290–4297.
- WALKER, I. (2010): *Výzkumné metody a statistika*. Praha: Grada.
- WICKHAM, H. (2017): *tidyverse: Easily Install and Load the 'Tidyverse'*. R package version 1.2.1. Dostupné online na adrese: <<https://CRAN.R-project.org/package=tidyverse>>.
- WIGGINS, A. K. – MARSHALL, K. A. – FRIEL, S. A. F. (2007): Assessing classroom language and literacy richness. In: K. L. Pence (ed.), *Assessment in Emergent Literacy*. New York: Plural Publishing, s. 3–51.
- WIJFFELS, J. (2018): *udpipe: Tokenization, Parts of Speech Tagging, Lemmatization and Dependency Parsing with the 'UDPipe', 'NLP' Toolkit*. R package version 0.5. Dostupné online na adrese: <<https://CRAN.R-project.org/package=udpipe>>.

Ústav formální a aplikované lingvistiky MFF UK  
Malostranské nám. 25, 118 00 Praha 1  
<[cinkova@ufal.mff.cuni.cz](mailto:cinkova@ufal.mff.cuni.cz)>

Ústav českého jazyka a teorie komunikace FF UK  
nám. Jana Palacha 2, 116 38 Praha 1  
<[Anezka.Nahlikova@seznam.cz](mailto:Anezka.Nahlikova@seznam.cz)>