



# Baterie sémantických rysů: data získaná od dětí z třetích tříd základních škol<sup>1</sup> — zpráva z pilotního sběru dat

Kristýna Konečná — Lucie Večeřová

## ABSTRACT:

**The Semantic Features Bank: Data Collected from Children Attending Their Third Year of Primary School — A Report Compiled from the Pilot Data Collection.** The present text is focused on the pilot data collection which is the first part of the research on language acquisition in children. This acquisition research is aimed especially at semantic feature production and the acquisition of depth of meaning. The goal of the research is to collect a substantial amount of data to create a battery of semantic features in children. This battery is to become a tool for examining children's vocabulary. The test group consists of children aged eight to nine. In a pilot data collection, children's preferences for linear or nonlinear record of semantic features were tested. The second form (which is nonlinear and requires children to write semantic features in columns) is widely used in both Czech and foreign research environments. Three approaches to the data collection were tested. For the purpose of our research, some specific code rules were devised and designed to register the semantic information obtained. Some statistical calculations were made as well, which allows to create a notion about feature representations and also about relations between features and concepts.

## KLÍČOVÁ SLOVA / KEY WORDS:

akvizice mateřského jazyka, hloubka významu, kategorizace, sémantický rys  
children vocabulary, concept, depth of meaning, featural representation, first language acquisition, semantic feature

## 1. TEORETICKÝ ÚVOD

Téma představovaného výzkumu spadá do oblasti vývojové psycholingvistiky. Dětské jazykové projevy totiž mohou pomoci odhalit důležité údaje o tom, jak funguje lidská řeč a jak se učíme jazyk, ať už mateřský, či cizí (srov. např. Průcha, 2011; Bloom, 2000).

Akvizice jazyka a dětská řeč jsou v odborném diskurzu vývojové psycholingvistiky zkoumány a popisovány z hlediska všech jazykových rovin. Náš výzkum se soustředí výhradně na osvojování lexikální a sémantické kompetence u českých dětí, přesněji na osvojování hloubky významu. Hloubkou významu rozumíme rozsah sémantické reprezentace slov, který je součástí tzv. hloubky slovníku (*vocabulary depth*): „Researchers usually distinguish between two aspects of an individual's vocabulary knowledge: breadth and depth. Breadth of vocabulary refers to the size of the mental lexicon and is the number of words that have some level of meaning to the individual.

---

<sup>1</sup> Zpracování a vydání článku bylo umožněno díky účelové podpoře specifického vysokoleškovského výzkumu udělené roku 2014 Univerzitě Palackého Ministerstvem školství, mládeže a tělovýchovy ČR.



[...] Depth refers to the richness of knowledge that the individual possesses about the words that are known“ (Tannenbaum — Torgesen — Wagner, 2006, s. 383).

Obvykle je dětský slovník zkoumán z hlediska rozšiřování slovní zásoby, jak ale uvádí např. Průcha (2011, s. 72), „[...] kvalitativní stránka procesu, tj. osvojování slov jakožto sémantických kategorií, je objasněna mnohem méně“. Proto považujeme za významný přínos vznik databáze, která mapuje slovní zásobu především do hloubky.<sup>2</sup>

Data pro databázi vytvářenou v rámci našeho výzkumu jsou získávána od dětí třetích tříd základních škol, jde tedy o děti ve věku 8–9 let. Výběr této věkové skupiny má několik důvodů: (i) snažily jsme se obsáhnout některé z pozdějších období akvizice jazyka, (ii) volba byla limitována obtížností zadávaných úkolů, (iii) ve zvoleném věku probandů dochází k zásadnímu vývoji v oblasti pojmového myšlení.

Ad (i): Velké množství studií se věnuje předškolnímu vývoji dětské řeči, což je bezesporu dáno tím, že v tomto období se odehrávají nejzásadnější změny ve vývoji řeči — dítě postupně začíná komunikovat pomocí jazyka a s nástupem do školy by mělo být víceméně kompetentní ve všech jazykových rovinách. Přestože se v raném dětství slovní zásoba dítěte rozrůstá závratným tempem, i během školních let se děti nadále učí významné množství nových slov a dochází také ke zpřesňování znalostí významů již známých slov. Proto je jisté na místě zkoumat i pozdější stadia vývoje dětské řeči a na úrovni lexikální (sémantické) roviny není potřeba výzkum omezovat na předškolní věk.

Ad (ii): Zkoumaná věková skupina ve věku 8–9 let byla vybrána také z pragmatických důvodů. Participanti v tomto věku dokáží prezentovat své myšlenky dostatečně systematicky. Navíc jsou schopni splnit požadovaný úkol, tj. dokáží popsat pojmy, které jim jsou v rámci výzkumného úkolu zadány, i v případě písemné varianty úlohy.

Ad (iii): Jak již bylo uvedeno výše, je zvolené období definováno jako důležité ve vývoji pojmového myšlení: „Celkový vývoj poznávacích procesů se projeví i změnou způsobu, jakým děti uvažují o slovech, co pro ně určitý slovní výraz (např. pomeranč) znamená. Malí školáci obvykle definují ve vztahu ke konkrétnímu kontextu (máme je vždycky u babičky), popisem jeho vnějších znaků nebo aktivitou, která je s ním spojena (jsou k jídlu). Postupně se charakteristika slova stává obecnější, v průběhu školního věku se zvyšuje tendence užít k takové charakteristice nadřazenou kategorii, tj. dedukce (ovoce), nebo definovat slovo pomocí vztahu tohoto objektu k jiným objektům, resp. pojmům stejné úrovně, tj. analogií (např. pomeranč je jako citron, event. protikladné analogie, např. kyselý je opak sladkého)“ (Vágnerová, 2001, s. 110).

Aby získaná data (tj. probandy uvedené významy zadaných pojmů) mohla být zpracována systematickým způsobem, byla použita metoda, ve které jsou významy rozloženy na tzv. sémantické rysy a následně jsou veškerá získaná data zpracována do

2 V získaných datech je rovněž obsažena určitá informace o probandově šířce slovníku (*vocabulary breadth*). Proband mohl popsat pojem v nesouladu s běžně uznávanou definicí daného pojmu, což by naznačovalo, že pojem nemá dostatečně osvojený, tj. ještě nezná jeho přesný význam, proband mohl nechat místo pro popis pojmu prázdné (zde ale není jasné, zda pojem nestihl vyplnit, nebo zda nezná jeho význam). Viz také o úspěšnosti odpovědí níže.



obsáhlé databáze. Tato metoda je obdobou tzv. komponentové analýzy uplatňované v lingvistice (sémiotice), která vychází z myšlenky, že sémantický rys je nejmenší významovou jednotkou, přičemž je vnímána problematičnost rozboru významu: „Složková analýza (též *komponentová analýza*) je jeden z postupů sémantického rozboru, jímž se snažíme nalézat minimální významové jednotky. [...] Podstatou složkové analýzy je napodobení postupu, který byl použit dříve ve fonologii. Fonémy na rozdíl od významových jednotek mají pouze jednu složku, a to formu. Jednotky významové pak nesou kromě formy i obsah. [...] použití [této metody; KK a LV] pro mnohem hůře uchopitelnou významovou složku je ale problematické. Další problém vzniká tím, že realizovaná významová jednotka je do značné míry nesamostatná a že její význam vyplývá až ze součinnosti s dalšími jednotkami“ (Černý — Holeš, 2004, s. 119).

Rovněž v kognitivních vědách (jako jsou kognitivní lingvistika či kognitivní psychologie) se tradiční teorie věnující se mentální (sémantické, konceptuální) reprezentaci pojmů a kategorizaci pojmů zakládají často na myšlence sémantických rysů. Aktuální směr vývoje teorií i v těchto vědách však vede k tomu, že je pouhý rozklad významu na sémantické rysy považován za velmi zjednodušený přístup, a předpokládá se, že pojmové myšlení má mnohem složitější strukturu (viz např. Murphy, 2002).

Metodologie našeho výzkumu vychází ze studií Kena McRae a jeho kolegů (McRae — de Sa — Seidenberg, 1997; McRae et al., 2005) a z práce Davida P. Vinsona a Gabrielly Viglioccové (2008). Tito autoři respektují současný vývoj v dané oblasti a uvědomují si veškerá úskalí spjatá s rozkladem pojmových významů na sémantické rysy (ať už jde o verbální omezení metody, nebo o argument, že pojmová reprezentace se neskládá z prostého seznamu atributů; viz níže).

Cílem práce není získat z databáze sémantických rysů představu o kompletní mentální reprezentaci vybraných pojmů. Jsme si vědomy, že takové reprezentace mívají mnohem obsáhlejší, komplikovanější charakter (viz např. Murphy, 2002; McRae — de Sa — Seidenberg, 1997), a že námi získané reprezentace by měly proto být specifikovány jako „rysové“.

Jak vysvětlují McRae et al. (2005, s. 549), „[...] we do not believe that semantic knowledge is represented in the brain literally as a list of verbalizable features. [...] a participant's list of features represents a temporary abstraction that is constructed online for the purpose of producing feature names“.

Toto vysvětlení je také v souladu s výkladem, který ve své studii představují Barsalou et al. (1993, s. 8): „Rather than being a list of propositions stored in memory, a feature list reflects the sequential description of an experiential image. When people define a concept, they retrieve or construct a schematic image, focus attention on a subset of its perceptual symbols in a sequential manner, and describe the content of each focus with a linguistic description. On this view, feature lists do not exist in long-term memory as conceptual representations but are the result of a sequential on-line process that describes experiential images.“

Námi provedený sběr dat byl zaměřen na pojmy spjaté se základní slovní zásobou, které byly vybrány z rozličných sémantických kategorií. Zajímalo nás například, jak děti rozumí jednotlivým pojmům a na které vlastnosti těchto pojmů se zaměřují, jak děti (implicitně) vnímají vztahy mezi různými pojmy a jak tyto pojmy kategorizují.

Jak už bylo uvedeno výše, data jsou zpracovávána do obsáhlé baterie sémantických rysů. Předpokládáme, že z těchto dat a z jejich statistického zpracování si lze vytvořit představu o rysových reprezentacích vybraných pojmů u dětí, o vztazích mezi kategoriemi, pojmy a rysy. Výslednou datovou baterii bude možné využít jako nástroj pro další bádání v oblasti dětského jazyka.

Významný přínos naší práce vidíme ve vytvoření statisticky zpracované databáze sémantických rysů, která je zaměřena na české dětské probandy. Dosavadní nám známé sběry se totiž soustředily na dospělé anglické či české mluvčí (viz např. Nagy, 2014).

Jakou úlohu hrají rysy v kognitivním zpracování reálného světa do naší mysli, řeší McRae a jeho kolegové v návazných experimentech ve své studii (srov. McRae — de Sa — Seidenberg, 1997). Naším cílem je hypotézy o využití sémantických rysů v komputaci významu v rychlých sémantických úlohách otestovat u českých dětí, čímž ve vlastním návazném výzkumu využijeme materiál, který baterie sémantických rysů nabídne. Databáze vytvoří základ také pro tvorbu měřítka na testování hloubky slovní zásoby a pro samostatný experiment sledující hloubku slovníku, který bude s dětmi proveden v další fázi.

## 2. CÍLE PILOTNÍHO SBĚRU DAT

Pilotní sběr dat byl zaměřen především na testování metodologie, kterou jsme zvolily na základě výše zmíněných studií zpracovaných v anglickém prostředí. Po výběru vhodných pojmů k pilotnímu testování z celé připravené sady pro hlavní sběr bylo vytvořeno několik verzí pracovních archů, do kterých dětští probandi sepisovali své odpovědi. V původních zahraničních studiích tento sběr dat probíhal písemně, přičemž subjekt podle vzoru vypisoval jednotlivé sémantické rysy (tj. jednotlivé vlastnosti) pro daný pojem do sloupce pod sebe. S ohledem na dětské probandy bylo v pilotním sběru testováno, zda bude tato metoda funkční a zda nebude vhodnější dát probandům možnost psát odpovědi lineárně, přičemž si sami zvolí, zda budou psát celé věty, nebo jednotlivé informace oddělené čárkou. Byly proto testovány dvě písemné metody zápisů odpovědí (písemná horizontální metoda, kdy dítě mělo možnost odpovídat souvisle na řádcích, a písemná vertikální metoda, která předurčovala zaznamenávat jednotlivé vlastnosti zvlášť pod sebe ve sloupci). Rovněž byla testována ústní metoda — formou dialogu s osobou, která data sbírala.

Vhodnost jednotlivých metod sběru jsme posuzovaly podle několika kritérií. Porovnána byla celková kvalita odpovědí, počet získaných rysů a rovnoměrnost typologického rozložení rysů. Sledovaly jsme také, za jak dlouhou dobu jsou probandi schopni úkol vykonat a jak náročné či snadné pro ně je popsat vybrané pojmy. Nakonec pilotní sběr a následné zpracování získaných dat sloužilo k předběžnému ustanovení tzv. „kódovacích pravidel“, tedy k vytvoření pravidel, pomocí jakých jazykových kódů se budou zaznamenávat synonymní a rozdílné sémantické informace. Užší představení těchto pravidel je uvedeno níže.





### 3. METODOLOGIE

#### 3.1 PARTICIPANTI

Sběru dat se zúčastnili žáci třetích tříd několika běžných českých základních škol, věkové rozpětí probandů bylo tedy 8–9 let. Sběru se zúčastnilo celkem 39 dětí (viz tabulku 1).

varianta sběru	počet dětí	z toho dívek	z toho chlapců	z toho 8 let	z toho 9 let
vertikální	14	10	4	13	1
horizontální	19	7	12	13	6
ústní	6	3	3	4	2
<b>celkem</b>	<b>39</b>	<b>20</b>	<b>19</b>	<b>30</b>	<b>9</b>

**TABULKA 1:** Počet dětí, pohlaví, věk.

#### 3.2 MATERIÁL

V pilotním sběru dat bylo otestováno celkem 20 pojmů (substantiv; viz tabulku 2), které byly konfrontovány s výskytem ve Weslalexu (databáze slov vyskytujících se v učebnicích pro děti s ohledem na jednotlivé ročníky). Každé dítě popisovalo deset pojmů, odpovídalo na otázku „co to je, co to znamená?“, přičemž bylo upozorněno na to, že nemusí popisovat pouze vizuální vlastnosti pojmu, ale také jeho funkci, své chuťové, čichové, sluchové nebo hmatové vjemy, že může určit kategorii, do které pojem spadá, a uvést cokoli dalšího, co ho k danému pojmu napadne. Pro názornost byl úkol demonstrován na konkrétních vzorových příkladech. Pojmy v těchto sešitech byly pseudonáhodně namíchány skrze doplněk MS Excel, a to tak, aby v každém archu bylo deset pojmů a zároveň aby v každém byl maximálně jeden zástupce z vybraných sémantických kategorií. Dvě různé verze umožnily otestovat větší množství pojmů z různých sémantických kategorií. Testování ve dvou verzích má také pragmatické výhody — děti v lavici od sebe nemohou opisovat.

#### 3.3 PRŮBĚH SBĚRU DAT

Před sběrem dat byl rodičům odeslán informační dopis a dotazník zjišťující základní informace o dítěti (rodný jazyk, učené jazyky, věk) s žádostí o souhlas se zapojením dítěte do výzkumu. Písemné varianty sběru proběhly hromadně ve školních třídách ve vyučování, a to v průběhu jedné vyučovací hodiny. Probandům byly přečteny pokyny a na tabuli byl zaznamenán vzorový příklad. Po celou dobu práce zůstal příklad napsaný na tabuli. Ve zbytku hodiny děti samostatně vyplňovaly připravené archy. Ústní varianta sběru probíhala individuálně, dítěti byly přečteny pokyny a následně bylo vyzváno k ústnímu popisu pojmů, které mu předčítal vykonavatel sběru dat. Ze sezení byl pořizován audiozáznam.

pojem	ARF syn	SFI Weslalex	školní třída	počet slabik	počet fonémů	řešitel
bidýlko	365,92	3	5	3	7	1
břeh	59220,60	97	1 až 5	1	4	1
hranice	281472,79	20	2 až 5	3	7	1
chrám	29359,31	6	3, 4, 5	1	4	1
jinovatka	527,44	4	2, 3	4	9	1
lepidlo	3488,83	1	1	3	7	1
nestvůra	661,47	2	2, 5	3	8	1
paruka	2714,86	6	3, 4, 5	3	6	1
stádo	10510,82	1	3, 5	2	5	1
turista	117397,36	1	3	3	7	1
borovice	5494,04	23	1 až 5	4	8	2
cuketa	885,80	—	—	3	6	2
lavice	30982,31	66	1 až 5	3	6	2
racek	4412,89	4	1, 3, 4	2	5	2
rýč	956,72	12	1, 3, 4	1	3	2
svetr	6267,35	19	1 až 5	2	5	2
trolejbus	9044,78	1	4	3	9	2
varhany	10563,56	6	3, 4, 5	3	7	2
višeň	487,20	2	4	2	5	2
zub	44116,07	70	1 až 5	1	3	2

**TABULKA 2:** Vybrané pojmy.

### 3.4 KÓDOVACÍ METODA

Baterie sémantických rysů přináší náhled na rysovou (sémantickou) reprezentaci významů vybraných pojmů u žáků třetích tříd, tedy na znalosti významů daných pojmů u dětského probanda, na vztahy mezi sémantickými kategoriemi, pojmy, sémantickými rysy atd. Pro získání dat v komplexní a přehledné podobě, se kterou je možné dále pracovat, bylo potřeba odpovědi „kódovat“. Byla proto ustanovena určitá pravidla v souladu s pravidly předešlých studií, kterými jsme se inspirovaly (McRae — de Sa — Seidenberg, 1997; McRae et al., 2005; Vinson — Vigliocco, 2008). Základem bylo odpovědi rozdělit na části tak, aby každá část obsahovala pouze jednu sémantickou informaci, proto i informace jako „má čtyři kola“ byla rozdělena na dva sémantické rysy, tj. „má kola“ a „má čtyři kola“. Naopak určitá slovní spojení jsou natolik pevná a ustálená, že je potřeba je kódovat jako jeden rys, např. „je hudební nástroj“. Byl určen jednotný jazykový kód pro synonymní sémantické informace vyjádřené různými jazykovými výrazy, např. „píše se na tom“, „píšeme na tom“, „je ke psaní“, „psací“ apod. jsou u pojmu „lavice“ sjednoceny jako jeden sémantický rys, a to pod kódem „na psaní slouží“. Každá sémantická informace má tedy svůj vlastní kód, přičemž kódy se v celé baterii opakují pravidelně, ve stejných podobách, aby byla data soudržná.

Kódování zaručuje jednotný systém i v rámci různých sémantických rysů, tzn. že se používají ustálené kombinace verb a jmenných částí. Proto byl dodržován u jmen





singulár nominativu maskulina, u verb třetí osoba v přítomném čase a oznamovacím způsobu, např. vlastnosti pojmu („je zelené“, „je hezký“, „je strašidelná“ aj.) jsou zpravidla zaznamenávány jako *verbum být* ve třetí osobě singuláru indikativu přítomného času a jako adjektivum ve tvaru maskulina singuláru.

Je samozřejmě nemožné přesně určit, kdy je sémantická informace ještě synonymní a kdy už se jedná o odlišný rys. Problém vyvstává především mezi pojmy — pokud je zelená „cuketa“ a zelená je také „nestvůra“, nemusí jít o zcela totožnou barvu, nicméně těžko lze najít nějakou hranici a tyto informace kódovat jako dva samostatné rysy. Někdy naopak zdánlivě stejnou informaci u dvou různých pojmů dodatečně rozlišujeme, např. děti uvádějí „je dobrý“ nejenom v případech, kdy se jedná například o chování, ale také v případech, které kódujeme jako „je chutný“.

Kompletní odpovědi dětí byly z jistých hledisek zjednodušeny. Byly vyřazovány kvantifikátory typu „občas“, „někdy“, „může být“ aj. Informace, nakolik pojem skutečně vlastní daný rys, totiž vyplývá z produkční frekvence rysu (např. uvede-li deset dětí z patnácti, že je jablko červené, a jen tři děti zmíní, že je zelené, vyplývá z takových dat informace, že podle dětí je pro jablko typičtější červená barva).

Překódovány byly všechny sémantické informace, které dětští probandi zmínili, a všechny tyto informace jsou součástí databáze.

### 3.5 METODA SKÓROVÁNÍ ODPOVĚDÍ

Proband mohl za vyplnění archu získat maximální počet 30 bodů (3 body za pojem). Za žádnou nebo nesprávnou odpověď dítě získalo 0 bodů. Pokud odpověď byla uznána, získalo dítě 1 bod. Další bod byl přičten za uvedení diferenčního rysu či rysu, který výstižně charakterizuje pojem. Poslední bod byl přidělen v případě výrazně kvalitní definice, tedy při uvedení dalších diferenčních nebo důležitých rysů či většího množství (více než dvou) vypovídajících rysů.

## 4. VÝSLEDKY

Získaná data byla po kódování dále matematicky a statisticky zpracována. Byla zaznamenána průměrná doba vyplnění připravených archů a úspěšnost probandů v jednotlivých variantách. Statistické zpracování dat proběhlo ve dvou souběžných fázích.

V první fázi bylo hodnoceno, která ze tří metod je pro sběr dat nejvhodnější. V této fázi byly propočítány korelace hodnotících kritérií pro jednotlivé sběry (hodnotící kritéria viz výše) a také napříč všemi sběry. Byla provedena jednoduchá analýza rozptylu (ANOVA), která umožnila statisticky potvrdit homoskedasticitu dat napříč uvedenými metodami a díky které bylo možné porovnat průměry hodnotících kritérií pro jednotlivé metody. V rámci rozhodování o nejvhodnější metodě byla použita agregace hodnocení variant pomocí váženého průměru. Vzhledem k preferencím rozhodovatele byly varianty (metody sběru dat) ohodnoceny také pomocí Hurwitzova pravidla.

Druhá fáze pak byla zaměřena na dílčí zpracování statistik, kterými lze popsat vztahy mezi získanými daty (např. McRae et al., 2005). Byly provedeny výpočty uka-

zující podobu rysové reprezentace pojmů ve vztahu k probandům, informace o jednotlivých pojmech a jejich rysové reprezentaci a dále byly zjišťovány vzájemné vztahy mezi pojmy a jejich rysy.

## 4.1 VHODNOST JEDNOTLIVÝCH VARIANT SBĚRU

### 4.1.1 ŘAZENÍ DAT

Vzhledem k záměru vytvářet normy sémantických rysů byly statistické výpočty sledující vhodnost jednotlivých variant počítány z dat, která byla řazena podle pojmů, a to s výjimkou jednoduché analýzy rozptylu (podrobněji k tomu viz níže). Tabulky s daty pak obsahují informace o jednotlivých pojmech, tj. kvalitu odpovědí, kvantitu (počet rysů) a rozložení typů rysů v pojmu. V prvních dvou případech jsou v druhém rozměru tabulky zaznamenána data o probandech, v případě dat k typologickému rozložení nese druhý rozměr tabulky informace o typech rysů.

### 4.1.2 DÍLČÍ SROVNÁNÍ DAT

Byla pozorována průměrná doba, kterou probandi potřebovali k vyplnění sběrového archu, a také „úspěšnost“, s jakou daný arch vyplnili.

Průměrná doba vyplnění v případě vertikální písemné varianty odpovídala 26,5 minutám. V případě ústní varianty záznam odpovědí jednomu probandovi trval průměrně 3,36 minut. V případě horizontální varianty bohužel nebyla průměrná doba zaznamenána, poslední proband odevzdal sešit po 45 minutách.

„Úspěšnost“ v možných variantách sběru byla hodnocena dvojím způsobem, a to skrze hodnocení množství uznaných, či neuznaných položek a skrze hodnocení kvality. Jako rozhodující byla zvolena tato dvě kritéria. Stranou ponecháváme hodnocení úspěšnosti skrze množství uváděných rysů a skrze variabilitu rysové typologie, a to z toho důvodu, že při hodnocení úspěšnosti nelze jednoznačně rozhodnout, zda tato úspěšnost skutečně závisí na kvantitě uvedených rysů a vysoké variabilitě rysové typologie.

„Úspěšnost“ podle uznaných a neuznaných položek (tabulka 3) ukazuje, že nejvíce uznaných odpovědí z celkového množství možných odpovědí bylo dosaženo v ústní variantě (86,67 %), z písemných variant pak bylo dosaženo nejvyšší úspěšnosti ve variantě horizontální (77,37 %). Mezi písemnými variantami je pak v úspěšnosti typu uznané a neuznané položky rozdíl 15,23 %, mezi nejlepší písemnou (tj. horizontální) a ústní variantou činí rozdíl 9,2 %.

„Úspěšnost“ podle provedeného hodnocení kvality (tabulka 4) ukazuje, že nejlépe byly hodnoceny odpovědi probandů, kteří se účastnili ústní varianty sběru (57,22 %), nejlépe hodnocenou písemnou variantou je pak varianta horizontální (39,65 %). Rozdíl mezi ústní variantou a nejlépe hodnocenou písemnou variantou je 17,5 %, tento rozdíl je pak mezi písemnými variantami 5,84 %.

Z prezentovaných výsledků je patrné, že nejvyšší úspěšnosti na základě zvolených kritérií je dosaženo ústní variantou, z písemných variant je pak nejhodnější varianta horizontální.







varianta sběru	počet odpovědí	z toho „0“	z toho uznaných	úspěšnost (%)
vertikální	140	53	87	62,14285714
horizontální	190	43	147	77,36842105
ústní	60	8	52	86,66666667

**TABULKA 3:** Úspěšnost: uznané/neuznané.

varianta sběru	Σ body	z možných	úspěšnost (%)
vertikální	142	420	33,81
horizontální	226	570	39,65
ústní	103	180	57,22

**TABULKA 4:** Úspěšnost podle hodnocení — kvalita.

#### 4.1.3 POUŽITÉ STATISTICKÉ METODY PRO VYHODNOCENÍ KVALITY JEDNOTLIVÝCH TYPŮ SBĚRŮ

Následující řádky popisují statistické metody, které byly použity pro vyhodnocení kvality výše zmíněných tří typů sběrů.

##### **Korelace**

Metodou korelace byla zjišťována souvislost mezi jednotlivými kritérii hodnocení zvolených metod sběru, tj. souvislosti mezi kvalitou, kvantitou a typy rysů (srov. tabulky 5–7). Nejvyšší korelace byly zjištěny u metody vertikální, poměrně vysoké korelace lze pozorovat také u metody horizontální. Výrazně níže pak kritéria korelují u metody ústní, korelační koeficient je velmi nízký a nesevědí o závislosti náhodných veličin, tj. kritérií hodnocení. Domníváme se, že je to dáno nerovnoměrně rozrůzněnými typy rysů uváděnými ústní metodou (obdobná pozorování a potvrzení domněnky viz níže).

	kvalita	kvantita	typy rysů
kvalita	1	—	—
kvantita	0,892059220400532	1	—
typy rysů	0,8279245594239690	0,853825516958286	1

**TABULKA 5:** Korelace — vertikální metoda.

	kvalita	kvantita	typy rysů
kvalita	1	—	—
kvantita	0,74839944925841	1	—
typy rysů	0,733474303935229	0,834161057881922	1

**TABULKA 6:** Korelace — horizontální metoda.

	kvalita	kvantita	typy rysů
kvalita	1	—	—
kvantita	0,669636774743814	1	—
typy rysů	0,439629554492271	0,778264010754931	1

**TABULKA 7:** Korelace — ústní metoda.

### Jednoduchá analýza rozptylu

Abychom zjistily, zda mezi středními hodnotami zvolených kritérií jednotlivých metod existují signifikantní rozdíly, byla provedena jednoduchá analýza rozptylu (ANOVA). Tato metoda byla použita pro řazení dat s ohledem na probandy, a to z toho důvodu, že v případě řazení dat s ohledem na pojmy nebyly splněny podmínky pro provedení testu ANOVA — jedná se o rovnost rozptylů přes všechny metody, kterou zkoumáme Bartlettovým testem, a o normalitu pozorovaných dat. ANOVA byla popsáním způsobem propočtena pro hodnocení, počet rysů a typy rysů. U prvních dvou byly podmínky testu ANOVA splněny, rysovou typologií nelze touto metodou hodnotit.

Signifikantní rozdíly mezi středními hodnotami metod byly analýzou rozptylu potvrzeny — hodnocení (tabulka 8) na zvolené hladině významnosti  $\alpha = 0,05$ . Duncanovým testem (tabulka 9) pak bylo zjištěno, že tyto rozdíly jsou mezi metodami ústní—vertikální a ústní—horizontální. Mezi písemnými variantami žádné signifikantní rozdíly nejsou.

Test ANOVA provedený pro kvantitu (tabulka 10) čili počet rysů neprokázal žádné signifikantní rozdíly mezi metodami.

Z uvedených dat vyplývá, že na výsledné hodnocení, tj. kvalitu rysů (přesněji kvalitu odpovědí), bude mít zvolená metoda vliv, a to v případě volby mezi ústní a některou z písemných metod. Na počet uváděných rysů pak žádná z možných metod sběru rysů vliv mít nebude. V souvislosti s hodnocením úspěšnosti v jednotlivých metodách lze tato data interpretovat tak, že výrazně kvalitnější rysové produkce je při sběru dosaženo ústní metodou.

zdroj variability	SS	rozdíl	MS	F	hodnota P	F krit
mezi výběry	208,432138	2	104,216069	5,914486994	0,006011818	3,259446306
všechny výběry	634,3370927	36	17,6204748			
celkem	842,7692308	38				

**TABULKA 8:** ANOVA — hodnocení.



sRA (P3)	ústní/vertikální	12,65598991	<	20,356934	signifikanční výsledek
sRA (P2)	horizontální/vertikální	12,03893169	>	7,034012346	nesignifikanční výsledek
sRA (P2)	ústní/horizontální	12,03893169	<	15,92087893	signifikanční výsledek

**TABULKA 9:** Duncanův test — hodnocení.

zdroj variability	SS	roz- díl	MS	F	hodnota P	F krit
mezi výběry	436,2704839	2	218,135242	1,763699252	0,185898906	3,259446306
všechny výběry	4452,498747	36	123,6805207			
celkem	4888,769231	38				

**TABULKA 10:** ANOVA — počet rysů.

### Párové srovnání průměrů — vážený průměr

Metodou agregace dílčích hodnocení variant pomocí váženého průměru (tabulka 11) byla zjišťována vhodnost metody vzhledem ke sledovaným kritériím. Jednotlivým kritériím byly nejprve podle jejich předpokládané důležitosti uděleny různé váhy. Tyto váhy jsou normované, tudíž splňují podmínku, že jsou větší či rovny nule a jejich součet se rovná jedné. Vážený průměr pak byl propočítán dvakrát, nejprve byla nejvyšší důležitost (váha) udělena kvalitě a typům rysů, podruhé měla všechna kritéria stejné váhy. V prvním případě se jako nejvhodnější metoda vzhledem ke zvoleným kritériím ukazuje metoda písemná — horizontální, jejíž průměr činí 11,25; ve druhém případě, v němž měla kritéria váhy shodné (0,33), má nejvyšší průměr opět metoda horizontální.

Ze srovnání průměrů metodou váženého průměru vyplývá, že pro záměry sběru dat vykazuje nejlepší výsledky (s nejvyšším průměrem) s ohledem na výše zvolená kritéria písemná horizontální metoda.

### Variační koeficient

V rámci hodnocení jednotlivých variant sběru byl spočten také variační koeficient (tabulka 12), který v následných dílčích hodnoceních variant dle rozhodovacích kritérií umožnil zohlednit kromě průměru také směrodatnou odchylku (rozptyl) dat. Jinak řečeno pomohl podchytit situaci, kdy nějaké dítě bylo extrémně dobře hodnoceno podle nějakého kritéria či naopak. Takováto odlehlá pozorování by mohla do značné míry vychýlit průměr varianty (tedy dílčí hodnocení dle daného kritéria), a ta by tak ztratila výpovědní hodnotu. Pokud ovšem soubor dat takového odlehlé pozorování opravdu obsahuje, odrazí se to v rozptylu dat, a tedy i směrodatné odchylce.



různé váhy kritérií hodnocení				stejně váhy kritérií hodnocení			
	vertikální	horizontální	ústní		vertikální	horizontální	ústní
<b>kvalita</b>				<b>kvalita</b>			
váha		<b>0,4</b>		váha		<b>0,33</b>	
průměr	-0,7799	-0,3475	1,1274	průměr	-0,7799	-0,348	1,1274
<b>kvantita</b>				<b>kvantita</b>			
váha		<b>0,2</b>		váha		<b>0,33</b>	
průměr	-1,0582	0,1288	0,9294	průměr	-1,0582	0,1288	0,9294
<b>typy rysů</b>				<b>typy rysů</b>			
váha		<b>0,4</b>		váha		<b>0,33</b>	
průměr	-0,1438	1,0641	-0,9203	průměr	-0,1438	1,0641	-0,92
<b>vážený průměr</b>				<b>vážený průměr</b>			
	-0,5811	0,3124	0,2687		-0,6606	0,2818	0,3788

TABULKA 11: Vážený průměr.

<b>kvalita</b>		
vertikální	59,29990377 %	1,123510645
horizontální	40,59988861 %	-0,330902831
ústní	34,66355049 %	-0,792607814
<b>kvantita</b>		
vertikální	82,54545332 %	1,154563149
horizontální	39,42762445 %	-0,561855887
ústní	38,65261306 %	-0,592707262
<b>typy rysů</b>		
vertikální	52,26607745 %	1,138719887
horizontální	29,98970054 %	-0,735154288
ústní	33,93158453 %	-0,403565599
<b>průměr</b>		
vertikální	64,70381151 %	1,138931227
horizontální	36,67240453 %	-0,542637669
ústní	35,74924936 %	-0,596293558

TABULKA 12: Variační koeficient.

Variační koeficient je tvořen zlomkem, kde v čitateli se nachází hodnota směrodatné odchylky a ve jmenovateli střední hodnota. Pro kritérium s rostoucí preferencí (což je zde splněno pro všechna kritéria) potom vyžadujeme co největší střední hodnotu a co nejmenší směrodatnou odchylku, analogicky tedy hledáme co nejmenší hodnotu variačního koeficientu.

Z výsledků rozboru variability vyplývá, že nejvariabilnější data vykazuje metoda vertikální (63,76 %), nejméně variabilní data pak poskytuje metoda ústní. Jelikož naše požadavky na metodu sběru dat cílí na využití metody, která bude přinášet především stabilní výsledky (spolu s vysokou kvalitou odpovědi, rozrůzněnou rysovou typologií a přiměřeným počtem rysů), rozhodly jsme se volit mezi metodami s nízkou variabi-



litou dat, tj. mezi ústní metodou a písemnou horizontální metodou, jejichž rozdíl ve variabilitě dat činí jen 1,55 % (rozdíl mezi písemnou vertikální a mluvenou variantou je 28,02 %).

### **Hurwitzovo pravidlo**

Poslední metodou, kterou jsme uplatnily pro řešení rozhodovací situace, je tzv. Hurwitzovo pravidlo. Touto metodou lze „diskvalifikovat“ variantu, jež zcela nevyhovuje minimem svých dílčích hodnocení dle jednotlivých kritérií, zároveň je v ní však zohledněno, pokud varianta exceluje maximem svých hodnocení, tedy dosahuje maxima hodnocení lepšího než varianty ostatní. Toto pravidlo vychází z koeficientu optimismu  $\lambda$ , který leží v intervalu  $\langle 0, 1 \rangle$ , v našem případě tedy čím bude  $\lambda$  blíže jedné, tím více bude zvýhodněna metoda vynikající svým maximem, čím blíže bude k nule, tím více bude „diskvalifikována“ ona nevhodná metoda. Hurwitzovo pravidlo závisí na extrémních hodnotách min—max, které jsou pronásobeny  $\lambda$ , tedy vahou podle daného vzorce. Nejlepší výsledek je vyjádřen nejvyšší hodnotou kritéria optimismu. V hodnocení získaných dat bylo Hurwitzovo pravidlo propočítáno pro  $\lambda = 0,1$ .

Bez ohledu na udělené váhy bylo nejvyššího kritéria optimismu dosaženo v obou případech ve variantě písemné horizontální, která také podle jiných provedených statistických metod vhodně odpovídá zvoleným kritériím.

vertikální	-0,966727816
horizontální	-0,206378997
ústní	-0,715543917

**TABULKA 13:** Hurwitzovo pravidlo pro  $\lambda = 0,1$ .

## 4.2 ZPRACOVÁNÍ PRO NORMY SÉMANTICKÝCH RYSŮ

V následující části bude naznačeno, jakým způsobem mohou být data v baterii statisticky zpracována. S ohledem na malé množství testovaných subjektů pro jednotlivé metody jsou v tomto případě výsledky pouze ilustrační. Následující statistiky proto nejsou zásadní při určení, který z testovaných sběrů je nejvhodnější, ačkoli mohou k rozhodování v určité míře přispět.

Získaná data byla zpracována pomocí statistických metod, které jsou využívány při zpracování norem sémantických rysů (srov. McRae at al., 2005, s. 552–554). Statisticky byly zpracovány z matic, které uvádějí výskyt a produkční frekvenci rysů napříč pojmy — matice obsahují dvě proměnné, tj. sémantické rysy (řádky) a pojmy (sloupce), a byly zpracovány jednotlivě pro každý typ sběru. Dále byla vytvořena matice obsahující data pro všechny typy sběru.

Tyto statistiky poskytují především vlastní rozbor sesbíraných dat a přinášejí informace o: (a) rysové reprezentaci pojmů ve vztahu k probandům, (b) jednotlivých pojmech a jejich rysové reprezentaci a (c) vzájemných vztazích mezi pojmy a jejich rysy.

Na základě těchto výsledků je možné pozorovat, jakým způsobem fungují, z čeho se skládají a jak se prolínají tzv. rysové reprezentace pojmů, které by měly být součástí (počátečních) komputací slovních významů a měly by mít význam v sémantických úlohách, v nichž hraje roli rychlé vyhodnocení úkolu (McRae — de Sa — Seidenberg, 1997, s. 99).

#### 4.2.1 RYSOVÁ REPREZENTACE POJMŮ VE VZTAHU K PROBANDŮM

##### **Průměrný počet rysů**

U jednotlivých pojmů a kategorií je pozorováno, kolik probandi průměrně zmínili sémantických rysů v jednotlivých metodách sběru. To je vypočítáno jako aritmetický průměr ze sumy všech realizací (nikoli sémantických kódů) pro daný pojem vydělený počtem dětí, které pojem popisovaly.

Při větším množství zkoumaných pojmů v rámci kategorie lze vypočítat průměrný počet rysů pro celou kategorii na základě průměru rysů pro jednotlivé pojmy — suma průměrů pro jednotlivé pojmy dané kategorie je vydělena počtem pojmů v kategorii.

Z grafu níže (graf 1) vyplývá, že průměrně (napříč všemi typy sběrů) nejvíc rysů probandi uváděli u pojmů „zub“, „nestvůra“ a „bidýlko“, ke kterým uvedli i více než čtyři rysy. Na opačném konci se ocitly pojmy „varhany“ či „hranice“. Tato data by mohla ukazovat, které ze zvolených pojmů se nacházejí v centru dětské slovní zásoby (s omezením na zvolenou skupinu dětí) a které je třeba hledat spíše na její periferii. Získané hodnoty mohou poukazovat na to, v jaké míře probandi jednotlivé pojmy znají a co jsou o nich schopni vypovědět. Lze usuzovat, že počet uváděných rysů roste se znalostí daného pojmu — a znalost pojmu pak souvisí s tím, jak se vyskytuje ve světě a v jakých situacích se s ním dětští probandi setkávají.

S ohledem na cíl pilotního experimentu je hledána ta metoda, při které děti průměrně jmenují větší množství rysů — jsou otevřenější a ochotnější plnit svůj úkol. Průměrný počet uváděných rysů v horizontálním sběru i v ústním sběru byl 3,1 rysu na pojem, ve vertikálním sběru uváděly děti průměrně 2,8 rysu.

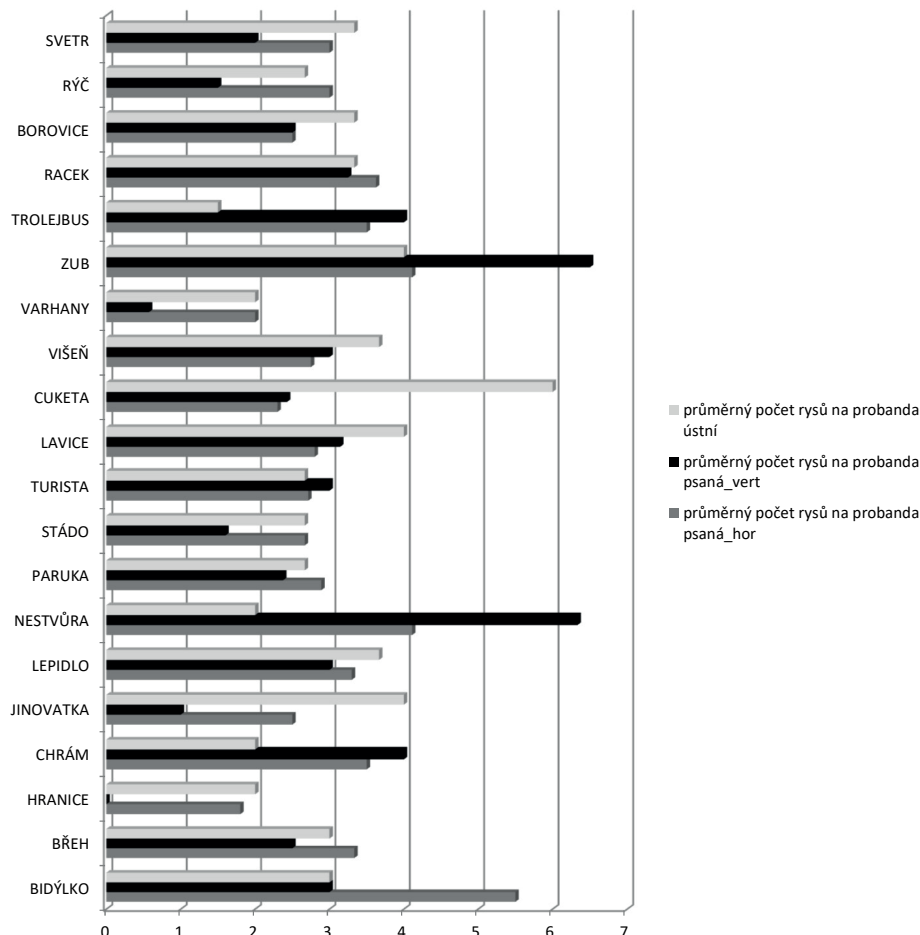
#### 4.2.2 POJMY A JEJICH RYSOVÉ REPREZENTACE

##### **Produkční frekvence rysů**

Základní hodnotou, kterou pozorujeme, je tzv. produkční frekvence rysu — ta udává, kolik dětí (z celkového počtu dětí, které popisovaly daný pojem) uvedlo daný rys pro pojem. Na základě produkce je možné pozorovat rozložení rysů v rámci jednotlivých pojmů (tj. získáme procentuální hodnotu, která znázorňuje poměr mezi produkcí daného rysu a celkovou produkcí rysů pro pojem) a zároveň lze určit tzv. rank rysů — tj. pořadí rysů v pojmu, které vychází z produkční frekvence rysů (nejnižší rank má rys, který byl zmiňován nejčastěji). Z toho vyplývá, že čím větším počtem probandů byl daný rys zmíněn, tím více je typický pro daný pojem. Zároveň lze pozorovat vztahy mezi jednotlivými rysy pojmu, od těch více typických (častěji zmiňovaných) až po ty méně typické, které byly zmíněny pouze jednou.







**GRAF 1:** Průměrný počet rysů.

Je zřejmé, že z většího vzorku probandů budou získána přesnější data a rys s hodnotou 1 lze potom jednoznačněji určit jako rys pouze okrajově svázaný s daným pojmem. V aktuálním sběru s pouze malým vzorkem probandů nabývá hodnoty 1 velké množství rysů a mnoho rysů, které lze pokládat za typické pro daný pojem a které jsou běžnou součástí charakteristiky pojmu, není zmíněno vůbec (což neznamená, že ony rysy si děti s pojmem vůbec nespojují).

Jako příklad uvádíme pojem „cuketa“ (tabulka 14), u kterého děti uváděly různorodé rysy, od taxonomických přes vizuální (percepční) až po encyklopedické, a jehož popis svědčí o jejich znalosti a zkušenosti s ním. Objevovaly se zde ale také „chybné“ rysy („cuketa je ovoce“, „je fialová“), které přínosně poukazují na to, že někteří probandi tento pojem zaměňují pravděpodobně s jinými pojmy a jejich znalost tohoto pojmu není (zatím) dostatečně přesná. U pojmu se objevují rysy s vysokou i nízkou produkční frekvencí.



<b>CUKETA – horizontální sběr</b> (11 rysů celkem)	<b>produkční frekvence</b> (x dětí z 10)	<b>rank</b> (pořadí podle produkční frekvence)	<b>produkční frekvence</b> (podíl v procentech)
zelenina je	5	2	21,73913
šišatý je	1	11	4,347826
zelený je	5	2	21,73913
dobrý je (chuť)	1	11	4,347826
k jídlu je	4	3	17,3913
opéká se	1	11	4,347826
řízek (z něj) dělá se	1	11	4,347826
smaží se	2	4	8,695652
v zahradě je	1	11	4,347826
zdravé jídlo je	1	11	4,347826
zelí (z něj) dělá se	1	11	4,347826
<b>CUKETA – vertikální sběr</b> (11 rysů celkem)	<b>produkční frekvence</b> (x dětí z 8)	<b>rank</b> (pořadí podle produkční frekvence)	<b>produkční frekvence</b> (podíl v procentech)
cuketa je	1	11	5,882353
ovoce je	1	11	5,882353
zelenina je	3	3	17,64706
velký je	1	11	5,882353
fialový je	1	11	5,882353
zelený je	3	3	17,64706
sladký je	1	11	5,882353
k jídlu je	3	3	17,64706
do buchtý dáva se	1	11	5,882353
obaluje se	1	11	5,882353
okurce je podobný	1	11	5,882353
<b>CUKETA – ústní sběr</b> (10 rysů celkem)	<b>produkční frekvence</b> (x dětí z 2)	<b>rank</b> (pořadí podle produkční frekvence)	<b>produkční frekvence</b> (podíl v procentech)
zelenina je	2	2	16,66667
puntíkováný je	1	10	8,333333
světle zelený je	1	10	8,333333
zelený je	2	2	16,66667
dobrý je	1	10	8,333333
krájí se	1	10	8,333333
na chalupě roste	1	10	8,333333
na kompostu roste	1	10	8,333333
plní se	1	10	8,333333
uvnitř světle zelený je	1	10	8,333333

**TABULKA 14:** Příklad – produkční frekvence rysů.



## Rozložení typů rysů v pojmu

S ohledem na zařazování rysů do odlišných typů podle mozkové aktivace bylo pozorováno také typologické rozložení uváděných rysů v pojmu. Toto typologické rozložení může být pozorováno jednak na počtu kódů abstrahovaných z odpovědí, tj. které sémantické rysy jakého typu se vůbec u pojmu objevily, jednak také přímo na odpovědích probandů. V druhém případě se určité realizace opakují, tj. jejich produkční frekvence je vyšší, narůstá proto i frekvence produkce typů rysů u pojmu, se kterou je operováno a ze které jsou výsledky vypočítány. Prakticky to znamená, že pokud u „jablka“ uvedli probandi 5× vizuální rys (týkající se barevnosti) „je červený“ a 2× „je zelený“, může být informace zpracována tak, že se u „jablka“ vyskytly dva vizuální rysy o barevnosti, můžeme ale také uvažovat tak, že u pojmu „jablko“ děti uvedly celkem 7 rysů o barevnosti. V prvním případě je tedy spočítáno typové rozložení s ohledem na daný pojem, v druhém případě s ohledem na probandy.

V rámci porovnání metod sběru lze ve výsledném grafu (znázorňujícím výpočet z kódů abstrahovaných z odpovědí) vidět, že v ústní metodě byly získávány především rysy encyklopedické (tj. nerecepční, naučené znalosti o vnějším světě) a taxonomické (přiřazení pojmu do nějaké sémantické skupiny, např. do nadřazené kategorie) a probandi méně často uváděli další typy rysů, zatímco nejrozmanitější rozložení uváděných rysů lze pozorovat u psaného, vertikálního sběru.

Zdá se, že data získaná v ústním sběru jsou do jisté míry odlišná od dat získaných písemně. V ústním sběru mohou děti více tíhnout k vyjmenovávání naučených znalostí, které o pojmu mají, a může jim připadat méně přirozené popisovat základní informace, jako je vzhled apod. Jinou roli zde hraje také osoba, která data sbírá — přímý kontakt s ní může dítě vést jiným směrem, než když odpovědi vypisuje (rozmlouvání s dospělou osobou může zastínit informaci z pokynů, podle které si má dítě představit, že vysvětluje například svému mladšímu sourozenci, co daný pojem znamená atd.). Nicméně je potřeba si i zde uvědomit, že u ústní realizace sběru mohl být zpracován pouze malý vzorek probandů.

V následující tabulce (tabulka 15) jsou vyznačeny nejnižší (kurzivou) a nejvyšší (tučným řezem písma) hodnoty. Stejná data jsou znázorněna také v grafu níže (graf 2).

### 4.2.3 VZÁJEMNÉ VZTAHY MEZI POJMY A JEJICH RYSY

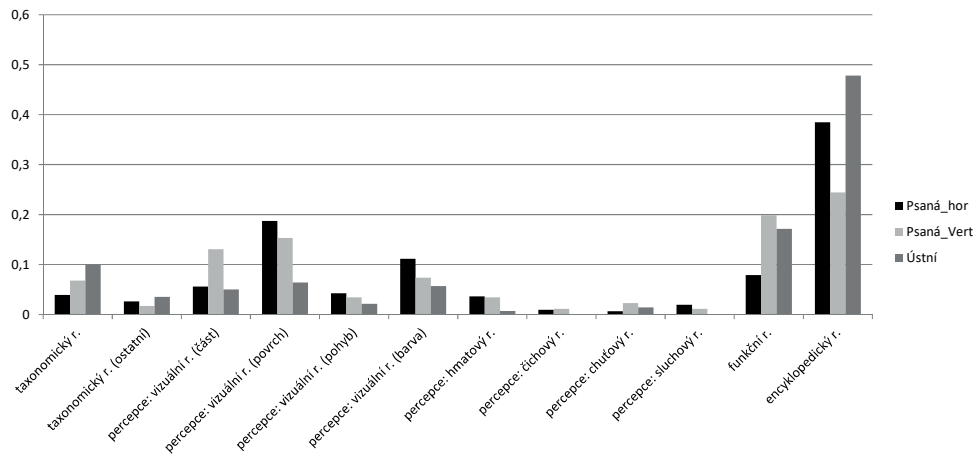
#### **Distinktivní rysy**

Tzv. distinktivní rysy jsou takové rysy, které odlišují daný pojem od pojmů ostatních. Daný rys náleží jednomu (max. dvěma) pojmům v baterii. Jsou to tedy rysy, které byly v rámci daného pojmu zmíněny alespoň jedním dítětem, zároveň se však nevyskytují ve více než dvou pojmech celkově. Byl spočítán také poměr mezi všemi rysy uvedenými k pojmu a těmi, které jsou distinktivní.

S předešlými výpočty je spřízněna tzv. distinktivnost neboli rozlišnost. Výsledkem je škála, která znázorňuje přehled rysů od vysoce distinktivních až po rysy značně sdílené mezi pojmy. Tyto údaje lze vypočítat jako inverzi čísla, které vyjadřuje, v kolika pojmech se daný rys vyskytuje. Pokud se rys vyskytuje pouze v jednom pojmu,

	psaná_hor	psaná_vert	ústní
taxonomický rys	3,95%	6,82%	10,00%
taxonomický rys (ostatní)	2,63%	1,70%	3,57%
percepce: vizuální rys (část)	5,59%	13,07%	5,00%
percepce: vizuální rys (povrch)	18,75%	15,34%	6,43%
percepce: vizuální rys (pohyb)	4,28%	3,41%	2,14%
percepce: vizuální rys (barva)	11,18%	7,39%	5,71%
percepce: hmatový rys	3,62%	3,41%	0,71%
percepce: čichový rys	0,99%	1,14%	0,00%
percepce: chuťový rys	0,66%	2,27%	1,43%
percepce: sluchový rys	1,97%	1,14%	0,00%
funkční rys	7,89%	19,89%	17,14%
encyklopedický rys	38,49%	24,43%	47,86%
	100,00%	100,00%	100,00%

TABULKA 15: Rozložení typů rysů.



GRAF 2: Rozložení typů rysů.

CUKETA	celkový počet rysů	počet distinktivních rysů v pojmu (výskyt max. ve dvou pojmech)	poměr distinktivních a všech rysů v pojmu	průměrná distinktivnost pojmu (aritmetický průměr z inverzních hodnot rysů pro pojem)
horizontální sběr	11	10	0,909091	0,893939
vertikální sběr	11	10	0,909091	0,74026
ústní sběr	10	10	1	0,95

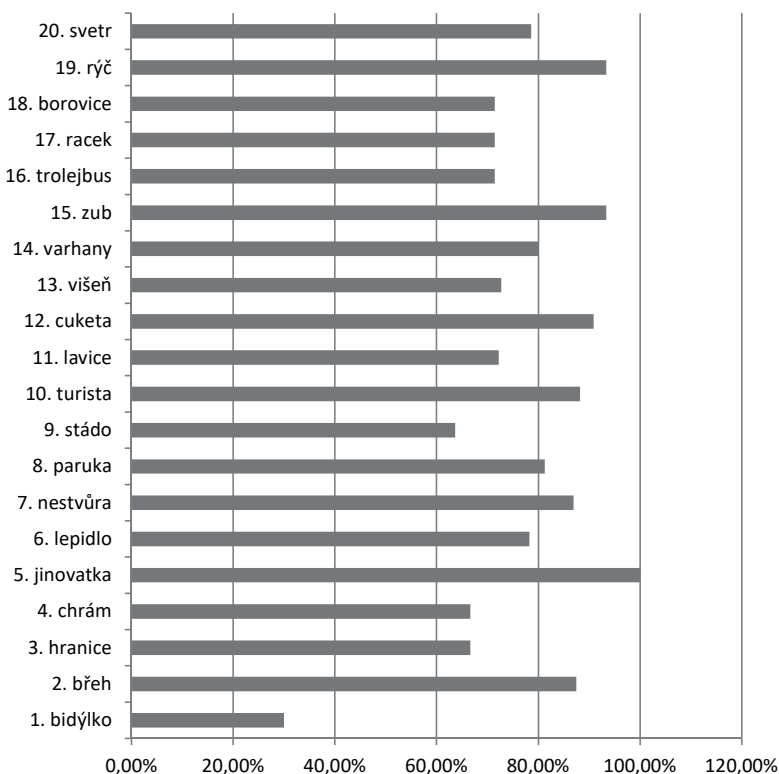
TABULKA 16: Příklad — distinktivnost rysů 1.



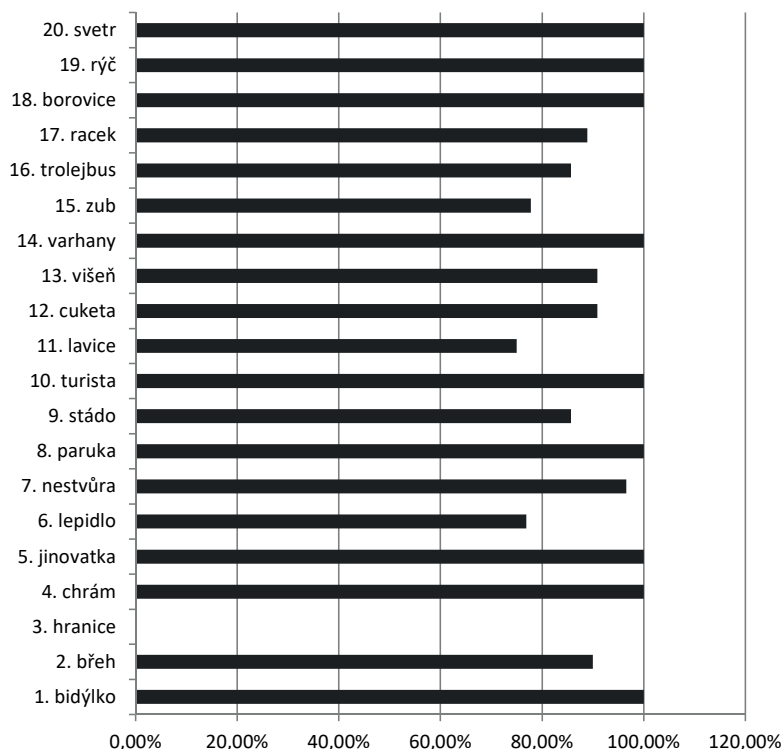
bude jeho hodnota 1, pokud ve dvou pojmech, bude jeho hodnota  $1/2$  (0,5) atd. Pak lze nalézt také střední hodnotu pro každý pojem, tj. je vypočítán aritmetický průměr z inverzních hodnot všech rysů pro daný pojem.

Následující grafy procentuálně vyjadřují počet distinktivních rysů (ze všech rysů uvedených pro daný pojem) v jednotlivých pojmech. Pro ústní sběr nebyl graf vytvořen, protože všechny rysy byly distinktivní. Protože se baterie sémantických rysů zaměřuje na vztahy mezi pojmy, je důležité, aby se tyto vztahy při sběru projevovaly a rysy se mezi pojmy prolínaly a opakovaly — jestliže se v horizontálním sběru objevuje menší procento distinktivních rysů, tj. takových, které se vyskytují maximálně ve dvou pojmech, zvyšuje se počet rysů, které se objevují v mnoha pojmech opakovaně, a proto je možné výsledek této statistiky interpretovat také tak, že se v horizontálním sběru nejvíce projeví vztahy a podobnosti mezi pojmy. Výsledky budou samozřejmě opět přesnější až po získání většího množství dat. Je potřeba respektovat skutečnost, že společnými rysy mohou být i takové, které nepřinášejí o pojmech žádnou relevantní informaci (viz např. níže problematiku rysu „je velký“).

V přesnějších měřeních (tj. v měřeních s větším vzorkem) mají distinktivní rysy významnou roli — odhalují, kterými rysy se pojem odlišuje od ostatních pojmů (od pojmů z kategorie).



**GRAF 3:** Distinktivnost rysů — horizontální varianta.



**GRAF 4:** Distinktivnost rysů — vertikální varianta.

### Podobnost mezi pojmy

Pro výpočet podobnosti mezi pojmy byl vypočítán kosinus. Čím bližší je hodnota číslu 1, tím jsou si dané pojmy podobnější. Jedná se o kosinus mezi dvěma vektory, tj. sloupci s produkční frekvencí pro jeden a následně pro druhý pojem.

Sledovat podobnost mezi pojmy, které spadají do značně odlišných sémantických kategorií, se zdá být irelevantní. Pojmy ze stejných sémantických kategorií si jsou sémanticky blízké, a proto jsou si pravděpodobně blízké také z hlediska sémantických rysů.

Ve vertikálním sběru jsou si nejbližší pojmy „cuketa“ a „višeň“ (s hodnotou 0,248807) a „cuketa“ s „nestvůrou“ (0,24004). V ústním sběru si byly podobné pojmy „višeň“ a „borovice“ (0,213201), „bidýlko“ a „chrám“ (0,166667) a „rýč“ s „varhany“ (0,158114). Pouze v horizontálním sběru si pojmy byly podobné ve vyšší míře, nejvíce to pak byly tyto čtyři dvojice: „chrám“ a „bidýlko“ s hodnotou (0,434122), „stádo“ a „chrám“ (0,544107), „nestvůra“ a „chrám“ (0,402939) a „nestvůra“ a „stádo“ (0,429886).

Jak je vidět, podobnosti jsou obecně velmi nízké a už intuitivně můžeme tušit, že „bidýlko“ a „chrám“ nemají zase tak moc společného. Zpravidla vybrané pojmy sblíží velmi obecné rysy jako například „je velký“ — takový rys je dětmi velmi často používán i v méně očekávaných případech. Navíc je jeho význam poměrně relativní — děti mají tendence uvádět tento rys u „chrámu“, „borovice“, ale například také u „paruky“ (pravděpodobně u každého pojmu používají jiná měřítka).





### Podmíněná pravděpodobnost

Podmíněná pravděpodobnost nám spolu s distinktivností udává informaci o výskytu daného rysu ve všech pojmech. Pokud je rys vysoce distinktivní a neobjevuje se v žádných jiných pojmech, tj. náleží pouze jednomu pojmu v baterii, pak bude jeho skóre pro distinktivnost i podmíněnou pravděpodobnost maximální, tj. rovno 1.

Podmíněná pravděpodobnost je vypočítána jako pravděpodobnost, že se rys vyskytne v daném pojmu, dělená pravděpodobností, že se rys vyskytne ve všech sledovaných pojmech. Proto se počítá s produkční frekvencí daného rysu v pojmu, která je dělena součtem všech produkčních frekvencí daného rysu ve všech pojmech.

$$P(C_j|F_i) = \frac{P(F_i|C_j)}{\sum P(F_i|C_j)} \text{ („C“ z angl. concept a „F“ z angl. feature)}$$

Jako příklad poslouží rys „je velký“, který se objevuje v mnoha pojmech (tabulka 17).

„je velký“ v horizontálním sběru	podmíněná pravděpodobnost — výskyt rysu (procenta)	„je velký“ ve vertikálním sběru	podmíněná pravděpodobnost — výskyt rysu (procenta)	„je velký“ v ústním sběru	podmíněná pravděpodobnost — výskyt rysu (procenta)
bidýlko	6,179111	bidýlko	0	bidýlko	0
břeh	2,265674	břeh	19,72738	břeh	0
hranice	7,552247	hranice	0	hranice	0
chrám	14,56505	chrám	0	chrám	0
jinovatka	0	jinovatka	0	jinovatka	0
lepidlo	0	lepidlo	16,43948	lepidlo	0
nestvůra	9,185166	nestvůra	10,38283	nestvůra	0
paruka	5,228479	paruka	0	paruka	0
stádo	21,2407	stádo	24,65923	stádo	0
turista	3,57738	turista	0	turista	0
lavice	2,427508	lavice	8,966991	lavice	0
cuketa	0	cuketa	11,60434	cuketa	0
višeň	0	višeň	0	višeň	0
varhany	7,552247	varhany	0	varhany	0
zub	0	zub	0	zub	0
trolejbus	4,855016	trolejbus	8,219742	trolejbus	0
racek	2,343801	racek	0	racek	0
borovice	10,19553	borovice	0	borovice	0
rýč	0	rýč	0	rýč	0
svetr	2,832093	svetr	0	svetr	0
	<b>100 %</b>		<b>100 %</b>		<b>0 %</b>

**TABULKA 17:** Příklad — podmíněná pravděpodobnost pro rys „je velký“.

## Rysové korelace

Rysové korelace hrají důležitou roli v organizaci sémantické paměti. Dá se předpokládat, že rysové reprezentace pojmů nemají jednoduchou strukturu, postavenou pouze na jednotlivých rysech, ale že mezi rysy se vytvářejí různé vztahy (především to pak platí u tzv. „biologických pojmů“; angl. *living things*). Vyhledávání vztahů mezi rysy je proto důležitou součástí zkoumání rysových reprezentací (McRae — de Sa — Seidenberg, 1997, s. 99).

Výpočet korelací mezi rysy přináší informaci o tom, které dva rysy se opakovaně vyskytují ve více pojmech. Získáváme tak páry rysů, které spolu korelují, to znamená, že pokud pojem „vlastní“ jeden z těchto rysů, „vlastní“ zpravidla i druhý z páru — jako jednoduchý příklad uveďme: vlastní-li pojem rys „létá“, bude mít pravděpodobně mezi rysy také „má křídla“. Pro výpočet je výchozí matice s pojmy a rysy s uvedenou produkční frekvencí (kolik probandů rys uvedlo). Aby se ve výsledcích neobjevovaly tzv. falešné korelace, byly do korelační analýzy zahrnuty pouze takové rysy, které se vyskytovaly ve třech a více pojmech.

Míra závislosti mezi rysy je vypočítána pomocí Pearsonova korelačního koeficientu (tj. parametrický nástroj korelační analýzy vhodný pro spojité veličiny, u kterých má být nalezen lineární vztah). Ten je vypočítán pro každou dvojici rysů a slouží jako míra vyjádření těsnosti vazby mezi náhodnými veličinami. Nabývá hodnot mezi  $-1$  a  $1$ , přičemž závislost je nejsilnější, když se koeficient přibližuje k těmto krajním hodnotám. Nezávislost veličin je pak dána  $0$ . Po vyřazení rysů, které se vyskytují v méně než třech pojmech, jsou provedeny výpočty a následně jsou odfiltrovány rysy, které mají korelační koeficient menší než  $0,8$  — pro identifikaci významnosti hodnot jednotlivých koeficientů je použita minimální míra sdílené variability (index determinace), která je rovna  $0,65$ , což odpovídá právě korelačnímu koeficientu o přibližné hodnotě  $0,8$ .

V pilotním výzkumu nebylo získáno dostatečné množství materiálu, proto byly nalezeny pouze dva korelační páry, jeden v případě horizontální varianty a jeden v případě vertikální varianty, přičemž v obou případech se spároval rys „je malý“ s rysem „je tvrdý“ (tyto rysy se oba zároveň vyskytují například u „chrámu“, u „lepídla“ či „lavice“).

## 5. ZÁVĚR

Cílem pilotního sběru dat bylo ověřit vhodnost jednotlivých metod sběru tak, aby mohla být pro hlavní sběr vybrána taková metoda, s jejíž pomocí nejlépe získáme požadovaná data. Vybraná metoda by měla splňovat požadavky na získání kvalitních popisů pojmů, přičemž tyto popisy by měly obsahovat různorodé informace (např. nepopisují pouze vizuální vlastnosti). Data získaná v tzv. pilotním sběru přinesla upřesňující informace o možnostech sběru, přičemž se z těchto dat na několika místech potvrzuje, že jako nejvhodnější metoda sběru se jeví tzv. písemný horizontální sběr (přináší velké množství rysů různého typu a zároveň se rysy jednotlivých pojmů prolínají a poukazují na vztahy těchto pojmů). Horizontální metoda sběru



umožňuje dětským probandům písemný lineární zápis odpovědi s možností vypisovat jednotlivé rysy oddělené čárkou či popisovat pojem v celých větách. Tato metoda je vhodná pro práci s dětskými probandy zřejmě proto, že striktně nepředepisuje, jak mají být odpovědi zaznamenány. Svobodnější způsob projevu může dětem usnadnit práci a nestresuje je nadbytečnými pravidly. I proto se dá očekávat, že touto metodou budou data získána bez větších komplikací.

Výsledky statistik, které popisují vztahy mezi rysovými reprezentacemi pojmů a mezi rysy, jsou pouze ilustrační. Statistiky ukazují, jakým způsobem lze s daty pracovat a jaké informace lze z dat získat, výsledky však nemohou být v tomto případě reprezentativní, protože pro pilotní sběr bylo zpracováno malé množství dat. Výpovědní hodnotu tyto statistiky získají až v následujícím hlavním sběru, který předpokládá účast přibližně 350 probandů a který bude popisovat větší množství pojmů z vybraných sémantických skupin, přičemž každý pojem bude popsán asi 20 probandy. Cílem bude také srovnání dat získaných od děvčat a od chlapců (srov. např. se studií zaměřující se na genderové rozdíly ve znalosti pojmů z vybraných sémantických kategorií u dětí; Barbarotto — Laiacona — Capitani, 2008), nabízí se rovněž možnost srovnat data s českými či anglickými dospělými mluvčími. Vytvořená baterie sémantických rysů bude zpřístupněna online, takže data mohou být využita v dalších výzkumech (viz také o našich vlastních návazných výzkumech výše). Může být také přínosným zdrojem informací o dětské řeči pro učitele a rodiče.

## LITERATURA:

- BARBAROTTO, Riccardo — LAIACONA, Marcella — CAPITANI, Erminio (2008): Does sex influence the age of acquisition of common names? A contrast of different semantic categories. *Cortex*, 44(9), s. 1161–1170.
- BARSALOU, Lawrence W. — YEH, Wenchi — LUKA, Barbara J. — OLSETH, Karen L. — MIX, Kelly S. — WU, Ling-Ling (1993): Concepts and meaning. In: Katharine Beals — Gina Cooke — David Kathman — Kyle E. McCullough — Sotaro Kita — David Testen (eds.), *Chicago Linguistic Society 29: Papers from the Parasession on Conceptual Representations*. Chicago, IL: The University of Chicago, Chicago Linguistic Society, s. 1–26.
- BLOOM, Paul (2000): *How Children Learn the Meanings of Words*. Cambridge, MA: MIT Press.
- ČERNÝ, Jiří — HOLEŠ, Jan (2004): *Sémiotika*. Praha: Portál.
- MCRAE, Ken — CREE, George S. — SEIDENBERG, Mark S. — McNORGAN, Chris (2005): Semantic feature production norms for a large set of living and nonliving things. *Behavior Research Methods, Instruments, and Computers*, 37(4), s. 547–559.
- MCRAE, Ken — DE SA, Virginia R. — SEIDENBERG, Mark S. (1997): On the nature and scope of featural representations of word meaning. *Journal of Experimental Psychology: General*, 126(2), s. 99–130.
- MURPHY, Gregory L. (2002): *The Big Book of Concepts*. Cambridge, MA — London: MIT Press.
- NAGY, Marek (2014): Produkce sémantických rysů u dospělých. In: Jana Kesselová — Maria Imrichová — Martin Ološtiak (eds.), *Registre jazyka a jazykovedy (I): Na počesť Daniely Slančovej*. Prešov: Filozofická fakulta Prešovskej univerzity v Prešove, s. 237–249.
- PRŮCHA, Jan (2011): *Dětská řeč a komunikace: Poznatky vývojové psycholingvistiky*. Praha: Grada.
- TANNENBAUM, Kendra R. — TORGESEN, Joseph K. — WAGNER, Richard K. (2006): Relationship

- between word knowledge and reading comprehension in third-grade children. *Scientific Studies of Reading*, 10(4), s. 381–398.
- VÁGNEROVÁ, Marie (2001): *Kognitivní a sociální psychologie žáka základní školy*. Praha: Karolinum.
- VINSON, David P. — VIGLIOTTO, Gabriella (2008): Semantic feature production norms for a large set of objects and events. *Behavior Research Methods*, 40(1), s. 183–190.



**Kristýna Konečná** | Katedra obecné lingvistiky FF UP  
<kristyn.konecna@gmail.com>

**Lucie Večeřová** | Katedra obecné lingvistiky FF UP  
<vecerova.luci@gmail.com>