

Korpusy rané češtiny

CoCzeFLA

Manuál pro přepisovatele

autoři: Anna Chromá, Filip Smolík, Klára Matiasovitsová

Tento manuál vznikl jako výtah z komplexního, pravidelně aktualizovaného manuálu pro přepis ve formátu CHAT [MacWhinney, B. (2000). The CHILDES Project: Tools for Analyzing Talk. 3rd Edition. Mahwah, NJ: Lawrence Erlbaum Associates].

- V nahrávkách dětské řeči lze očekávat poměrně vysoký výskyt nesrozumitelných nebo špatně srozumitelných úseků. Je potřeba nalézt kompromis mezi předčasným vzdáním snahy o porozumění a zbytečně obětovaným časem. Ani jedno není žádoucí. Neposlouchejte určitý úsek (orientačně) více než pětkrát (pokud je důvodem nesrozumitelnost).
- Název přepisu je přesně stejný jako název nahrávky či datapointu. Datapoint je tvořen několika nahrávkami pořízenými brzy po sobě.
- Vždy pracujte s **poslední verzí** manuálu. Při jeho čtení věnujte prosím důkladnou pozornost příkladům, které jsou nedílnou součástí.
- Přepisujte vždy v textovém editoru (doporučujeme Notepad++) a dodržujte formát souboru (**.txt**) a kódování znaků (**UTF-8**).

Obsah

1. Základní rámec přepisu: hlavička a ukončení	2
2. Typy řádků v přepisu a jejich zakončení	2
3. Ortografický přepis	3
3.1 Přepis v případě konkrétních typických variant	3
3.1.1 Redukce souhláskových skupin uvnitř a na začátku, proteze, změny kvantity	3
3.1.2 Redukce a další odchylky v sufixech	3
3.1.3 ý > ej / é > í/ý	3
3.1.4 Další vybrané jednotlivosti	4
4. Kódy na hlavním řádku	4
4.1 Neslova, fragmenty a nesrozumitelné úseky	4
4.2 Zakončení řádku: nedokončené výpovědi a překryvy	5
4.3 Opakované začátky a sebeopravy	6
4.4 Pauzy, prodloužení a dělení na výpovědi	7
4.4.1 Dělení na výpovědi	7
4.5 Neverbální jevy, komentáře a vysvětlivky	8
4.5.1 Tzv. nulový zápis	8
4.5.2 Začlenění komentářů do hlavního řádku	9
4.5.3 Využití řádků hlavičkového typu	9
4.5.4 Využití vedlejších řádků	9
4.5.5 Básničky a písničky	12
4.6 Morfologické inovace, idiosynkratická slova a chyby	12
4.7 Citoslovce	15
4.7.1 Zápis různých druhů citoslovcí	15
4.7.2 Zvláštní užití citoslovcí	15
4.8 Názvy, jména, citace a přímá řeč	16
5. Anonymizace a pseudonymizace	16
6. Hlavní rozdíly proti verzi 2.0	17

1. Základní rámec přepisu: hlavička a ukončení

V hlavičce přepisu se používají řádky uvozené znakem @. Uvádí se seznam osob na nahrávce, jazyk, datum nahrávky, atd. Hlavička je pouze jedna, a to na samém začátku přepisu (datapointu).

```
@Begin
@Languages:          ces
@Participants:       CHI Viktor Target_Child, MOT Mother
formát řádku Participants: ZKRATKA Jméno Role      jméno být nemusí, role se vybírají ze stanovené nabídky
@Transcriber:        TERBIN
```

Na konec **celého přepisu** (nikoli jednotlivých epizod) patří řádek označující konec: @End

V přepisu nikde **nesmí zůstat prázdné řádky**.

Za každou dvojtečkou řádkového úvodu následuje **přesně jeden tabulátor**.

2. Typy řádků v přepisu a jejich zakončení

Každá výpověď má svůj vlastní řádek, přičemž replika se může členit na více výpovědí. Každý řádek v přepisu začíná řídicím znakem, který indikuje, o jaký typ řádku se jedná. **Hlavní řádek** začíná hvězdičkou * a obsahuje ortografický přepis výpovědi. Všechny **vedlejší řádky** začínají znakem pro procento %, vážou se přímo k nejbližšímu předcházejícímu hl. řádku a dovysvětlují různé jeho aspekty.

Po úvodním znaku následuje třípísmenný kód označující mluvčího nebo typ řádku (v případě hlavních řádků velká písmena, v případě vedlejších malá písmena), dále dvojtečka, jeden tabulátor a vlastní obsah. **Každý řádek musí být poté zakončen a před koncovým znakem vždy musí být mezera** (viz znovu níže).

```
*CHI:      musím přinést polštářek .
%pho:      muším pineš poutáše .
```

Vedlejší řádky tedy dovysvětlují předcházející hl. řádek:

- **%pho** upřesňuje artikulační realizaci pronesené výpovědi. Tento řádek vždycky končí tečkou, bez ohledu na intonaci a také nesmí obsahovat žádné speciální kódy (jen písmena). Každý hlavní řádek s verbálním obsahem je vždy následován řádkem pho, ať se zápis na hlavním řádku a pho řádku liší či nikoliv
- **%err** blíže specifikuje chybu označenou v hlavním řádku hvězdičkou v hranatých závorkách (viz dále)
- **%com** je univerzální a umožňuje komentovat situaci – **obsahuje komentáře vztahující se pouze a přímo k dané výpovědi**
- další méně časté vedlejší řádky poznáte níže v manuálu

```
*CHI:      zejtra zvednul [*] .
%pho:      žejtja žvednul .
%err:      zvedne .
```

3. Ortografický přepis

Přepis je **na hlavním řádku** zaznamenáván ortograficky. To znamená, že pro zápis rekonstruuje ortografickou formu daného slova a neberete tak ohled na množství výslovnostních variant, a to jak **variant typických pro dospělou češtinu** (říká se “takle” > píšete *takhle*; říká se “kaničku” > píšete *tkaničku*; říká se “vod voka” > píšete *od oka*), tak **variant méně či více typických během osvojení** (říká se “motolka” > píšete *motorka*; říká se “kouajek” > píšete *korálek*). K záznamu artikulační realizace slouží %pho řádek, i ten je nicméně jen přibližný.

Odchytky od konvenční dospělé morfologie a syntaxe ale naopak ponecháváte. Stejně tak ponecháváte silně zkomolená slova. Podrobněji k těmto záležitostem viz níže.

Šva na %pho řádku můžete zaznamenat pomocí @ (p@ší), tento zápis se ale nesmí se vyskytnout na hlavním řádku (jde o jedinou výjimku z pravidla, že %pho řádek nesmí obsahovat jiné znaky kromě písmen).

Velká písmena píšete pouze u vlastních jmen, nikoli na začátku výpovědi.

3.1 Přepis v případě konkrétních typických variant

3.1.1 Redukce souhláskových skupin uvnitř a na začátku, proteze, změny kvantity na hlavní řádek **nezapisujete**, standardizujete ortograficky

- *MOT: takhle si zavážeš tkaničku .
- %pho: takle si zavážeš kaničku .
- *MOT: to jsem nevěděla .
- %pho: to sem nevěděla .
- *MOT: tam půjdu až za nějaký čas .
- %pho: tam pudu až za nějaký čas .
- *MOT: nevím .
- %pho: nevím .
- *MOT: s tím nic nenadělám takhle od oka .
- %pho: s tím nic nenadělám takle vod voka .

3.1.2 Redukce a další odchytky v sufixech na hlavní řádek **zapisujete**, nestandardizujete

- *MOT: s tím nemůžem takhle hejbat .
- %pho: s tím nemůžem takle hejbat .
- *MOT: s těma holčičkama už si nebudeš hrát ?
- %pho: s těma holčičkama už si nebudeš hrát .

3.1.3 ý > ej / é > í/ý

na hlavní řádek **zapisujete**, nestandardizujete

- *MOT: zejtra se vrátí mladej od sousedů .
- %pho: zejtra se vrátí mladej vod sousedů .

3.1.4 Další vybrané jednotlivosti

Kontrakce *jsi*

*MOT: přišels [: přišel jsi] pozdě .

%pho: přišels pozdě .

Demonstrativa *tohle, tohlencto* apod. s d-realizací (*todle, todlencto*)

*MOT: na tohlencto prosím tě nesahej .

%pho: na todlencto simtě nešahej .

Nespisovný kondicionál

(*a*)*bysem, (a)bysi, (a)bysme, (a)byste*

Prézens slovesa *být*

bez ohledu na výslovnost zapisujete na hl. řádek “*j*”: *jsem, jsi, jsme, jste, jsou*

neplatí pro tvar *seš* – vždy bez “*j*”

S *sebou*

u *s sebou* se předložka nikdy nevyslovuje, nicméně se vždy zapíše (podobně jako *jsem*)

Kroužek nad “*u*”

se píše jen tam, kam patří etymologicky, a tedy konvenčně. To neplatí pro žádné neologismy, idiosynkratické výrazy a ani pro citoslovce jako “*bú*” apod.

Zvuky (ne)souhlasu a prodlevy

<i>&eee</i>	zvuk zastavení po chybě a před opravou (něco tako “ <i>teda</i> ” nebo “ <i>vlastně</i> ”)
<i>hm</i>	souhlas 1
<i>mhm</i>	souhlas 2 (dvouslabičný)
<i>emem</i>	nesouhlas (dvouslabičný)
<i>nee</i>	dvouslabičné s rázem

4. Kódy na hlavním řádku

4.1 Neslova, fragmenty a nesrozumitelné úseky

Nesrozumitelné úseky značíte jako xxx. Pro označení fragmentů slov (chybných nebo přerušných začátků) použijete znak &, který připojíte bez mezery na začátek. Stejný znak se použije také pro označení hezitačních zvuků a dalších pazvuků, které nejsou fonologicky kódovatelné a zřejmě nepřidávají žádný význam. **Hezitační zvuk konzistentně zapisujete jako &em**. Kromě &em je obvyklé také hezitační **&to**, příp. i **&neto**. Podobnou povahu má i **&eee** (viz výše). Jejich významné prodlužování a zvláštní modulace je možné komentovat pomocí řádku %com.

Fragmenty s & je nutné vždy psát s malým písmenem, i když jde o fragment jména (“&jess” jako fragment z “*Jessinka*”).

Záměrný fragment Dospělí mluvčí někdy záměrně nakousnou slovo a potom čekají na doplnění od dítěte. I takové fragmenty značíme &, ale jejich záměrnost okomentujeme.

*MOT: a tohle je &če +...
%pho: a todle je če .
%com: čeká na doplnění "červená" .

*CHI: xxx .
%pho: xxx .
*CHI: bílý &em onp@c .
%pho: bílý em onp .
%com: onp = význam nejasný .
*CHI: &p prstík .
%pho: p p@stík .

Rozhodování mezi xxx a &

V případě, že jde o komplexnější výpověď, a pouze určitá její část má být zapsána jako neúplná nebo nesrozumitelná, je dobré se zabývat otázkou: "**Funguje daná výpověď syntakticky a sémanticky i bez nesrozumitelné části?**" Jde o to, že výpověď, která je jakkoliv neúplná (obsahuje xxx nebo končí znakem pro neúplnou výpověď), se standardně vyřazuje z analýz. To se ale netýká fragmentů se znakem &.

Pokud je tedy odpověď "**spíš ne**", pak zaznamenáte pomocí **xxx** (tím signalizujete, že tam **chybí podstatný syntaktický element, kterému nerozumíte**).

Pokud je odpověď "**spíše ano**", pak zaznamenáte pomocí **&** a přibližné hláskové rekonstrukce, např. **&s, &nel**, (tím signalizujete, že jde pravděpodobně **jen o fragment nějakého nedokončeného/spolknutého slova** nebo možná o zcela nejazykový zvuk).

Případně je také možné daný úsek/repliku rozčlenit vhodně na samostatné výpovědi, aby některá část zůstala jako srozumitelná a úplná.

4.2 Zakončení řádku: nedokončené výpovědi a překryvy

Každý řádek musí mít koncový znak. Před koncovým znakem je vždy mezera. Vedlejší řádky jsou VŽDY zakončeny tečkou. Nikdy nesmí být více koncových znaků na jednom řádku. Jakmile vložíte jeden koncový znak, musíte ukončit výpověď.

Na hlavních řádcích kromě tečky, otazníku a vykřičníku používáte i kódy pro nedokončené výpovědi: **+/**. znamená přerušeni jednoho mluvčího jiným mluvčím.

+< použijete pro překrývající se výpovědi: tento kód připojíte na začátek výpovědi, která začala dříve, než předcházející skončila. Za kódem následuje mezera. Kód pro překryv nemusí být vždy doprovázen kódem pro přerušeni, protože první mluvčí se nemusel vstoupením druhého mluvčího nechat ve své výpovědi přerušit.

+... použijete pro sebezrušení mluvčího, které není následované sebeopravou, tj. následuje pauza a potom zcela nová výpověď ať už téhož, nebo jiného mluvčího, popřípadě nová výpověď téhož mluvčího začne i bezprostředně bez pauzy (všimněte si rozdílu oproti kódu [//], viz níže).

*CHI: takhle &ši +/.
%pho: ťakle ťi .
*MOT: Adélko, na to se nesahá .
%pho: Adélko na to se nešahá .

*MOT: tak kam se posadí tahle panenka ?
%pho: tak kam se posadí tadle panenka .
*CHI: +< posadí 0se tady .
%pho: posadí tady .

*CHI: &tona my jsme +...
%pho: tona my ťme .
*CHI: mami, kde je ?
%pho: mami de je .
%com: vrací se k hledané ponožce .

*MOT: tak mu +...
%pho: tak mu .
*MOT: počkej !
%pho: počkej .
*MOT: ohnem mu ty nohy .
%pho: ohnem mu ty nohy .

4.3 Opakované začátky a sebeopravy

Další kódy se používají pro situaci, kdy mluvčí začne nějakou výpověď vícekrát (false starts), případně kdy se sám opraví. Jestliže se opakuje, použijeme kód <> [/], jestliže se opraví, pak použijete <> [//].

*MOT: tak <co mu teda> [//] jaký mu dáváš pleničky ?
%pho: tak co mu teda jaký mu dáváš pleničky .
*CHI: já mu dávám <ještě> [/] ještě jednu pemprsku .
%pho: já mu dávám eště eště jedu pep@sku .
*CHI: <až až až> [/] až tamtu poslední doděláme +...
%pho: až až až až tamtu posení doděuáme .

*CHI: <&tona my jsme vystupovali> [//] asi prvníma dveřma jsme vystupovali .
%pho: tona my ťme vyštopovauí asi p@níma dvema sme vykupovauí .

*MOT: a <chodíte> [//] jezdíte tam autobusem ?
%pho: a chodíte jezdíte tam autobusem ?

Jiné případy opakování, které nelze chápat jako opakované začátky (false starts), značíme jako [x počet_opakování]. Mezi "x" a číslem je mezera. V případě opakování fragmentů (false starts v případě, kdy ale není dořečeno celé slovo) toto ale neznačte, fragmenty rozepište.

*MOT: <počkej> [x 3] .
%pho: počkej počkej počkej .
*CHI: &kv &kv kvoká ?
%pho: kv kv kvoká .

Ne všechno, co se na povrchu jeví jako opakování, je skutečně opakování. Totéž slovo může být bezprostředně po sobě použito v referenci k odlišným konceptům, s odlišnou funkcí. Např. slovo “tohle” ve výpovědi “Tohle, tohle a tohle”. může referovat ke třem různým hračkám. Pokud jsme to schopni interpretovat, pak v takovém případě neznačíme opakování, nýbrž zapíšeme všechna slova.

4.4 Pauzy, prodloužení a dělení na výpovědi

K označení pauzy mezi slovy slouží znak (.), ale v principu je lepší začít při jakékoli (delší) pauze novou výpověď. Je také možné značit pauzy mezi slabikami. Jde o znak ^, který se bez mezer vloží mezi slabiky. Prodlužování jednotlivých slabik je možné zapisovat pomocí dvojtečky, opět bez mezer vložené na konec prodloužené slabiky.

Se značením prodlužování slabik i mezer mezi slabikami spíše šetřete a užívejte ho v markantních případech (významnost případu určujete s ohledem na typické vyjadřování dané osoby). V případě, že dítě často vkládá mezi slabiky pauzy (váhá nebo se zajímá), okomentujete (%com) to u konkrétní výpovědi nebo v obecném komentáři (@Comment) kdekoliv během přepisu (viz dále).

*CHI: my jsme jeli do lékárny pak (.) za tej^den .
%pho: my šme jeli do lékány pak (.) ža tej den .
*MOT: he:le@i, táta .
%pho: héle táta .

4.4.1 Dělení na výpovědi

Dělení na výpovědi se nemusí krýt s členěním dialogů na repliky (replika = co řekne jeden mluvčí, než ho vystřídá jiný mluvčí) a často se také nekryje, 1 replika se rozdělí na více výpovědí.

Dělíme podle klesavé i stoupavé koncové intonace a podle smyslu výpovědi, u velmi malých dětí často i po jednotlivých slovech.

Zároveň ale platí, že na jednom hlavním řádku jsou max 2 souřadně spojené hlavní věty ve slučovací poměru (bez spojky nebo s “a”).

Pokud se bezprostředně opakuje celá výpověď (byť by to bylo jen jedno slovo), pak značíme opakování [x 2] (viz výše). **Dělení na výpovědi je důležité, protože délka výpovědi je důležitým obecně využívaným měřítkem** stupně jazykového rozvoje, a to zejména zpočátku, dokud jsou výpovědi spíše krátké.

Pokud mluvčí opakuje po jiném mluvčím a potom promluvu dále rozvádí, rozvádí vlastní sdělení, snaží se přeformulovat (i vlastní) původní sdělení, např. vyjádřit se výstižněji, apod., rovněž rozdělujeme na samostatné výpovědi (často nám napoví i intonace).

MOT: popa@c, vid' .
%pho: popa vid' .
MOT: to jsou popeláři .
%pho: to sou popeláři .

MOT: koho tady vidíš, Sáři ?
%pho: koho tady vidíš sári .
MOT: kdo to je ?
%pho: kdo to je .

MOT: ano, pejsek je tam .
%pho: ano pejsek je tam .
MOT: haf@i .
%pho: haf .

Následující příklady jsou tedy zcela nepřijatelné, protože z automatické analýzy MLU (mean length of utterance) vyplyne tří- či pětislovná výpověď, což implikuje výpověď o třech či pěti syntaktických pozicích, což ale v tomto případě není:

chybně (pětislovná výpověď):

*CHI: houba, samá houba, samá houba .

%pho: houba samá houba samá houba .

správně (jednoslovná + dvouslovná výpověď):

*CHI: houba .

%pho: houba .

*CHI: <samá houba> [x 2] .

%pho: samá houba samá houba .

chybně (tříslovná výpověď):

*CHI: čí@z:ip miminko, čí@z:ip .

%pho: čí miminko čí .

správně (dvouslovná + jednoslovná výpověď)

*CHI: čí@z:ip miminko .

%pho: čí miminko .

*CHI: čí@z:ip .

%pho: čí .

4.5 Neverbální jevy, komentáře a vysvětlivky

4.5.1 Tzv. nulový zápis

V mnoha případech, kdy replika jednoho mluvčího neobsahuje žádná přepsatelná slova, přesto je ale potřeba ji zaznamenat, je možné univerzálně využít takzvaný nulový zápis. Při každém takovém použití nuly je nutné vždy připojit komentář (%com) a takový řádek nesmí obsahovat nic kromě nuly. **Tento zápis je univerzálně využitelný pro kašel, smích, u mladších dětí pro žvatlání a podobné vokalizace bez komunikační funkce, ale i pro složitější neverbální chování:**

*CHI: 0 .

%com: mumlá si pro sebe a MOT neposlouchá .

*CHI: 0 .
%com: vokalizuje zvuk žádosti a ukazuje na lžici .

4.5.2 Začlenění komentářů do hlavního řádku

Pokud je zvuk součástí výpovědi (je na řádku s přepsanými slovy), značte jej na hlavním řádku následovně (nikoli pomocí 0, viz výše):

&=smích neverbální hlasový projev; nekoinciduje s řečí (vložíme přímo na místě, kde se jev odehrál); podobně třeba **&=smrkání**

&=imit:auto takto můžete zapsat imitaci nejrůznějších zvuků (za dvojtečku bez mezery napíšeme jednoduše vlastními slovy, co se imituje)
imitace lze popsat i pouhým označením mluvčího, který daný zvuk produkoval:

*CHI: 0 .
%com: vokalizuje zvuk žádosti a ukazuje na polici .
*MOT: co **&=imit:CHI** ?
%pho: co .
%com: opakuje zvuk po CHI, vyjadřuje, že zvuku nerozumí .
*MOT: řekni, co chceš .
%pho: řekni co chceš .

<přepis> [?] zaznamenání nejistoty přepisu: slovo nebo více slov, kterými si nejste jisti, ale přece jen je chcete zaznamenat, dáte do špičatých závorek a připojíte k nim otazník

*MOT: a jakou **&=smích** barvu má tahle ?
%pho: a jakou barvu má tahle .
%com: z nahrávky není zřejmé, čemu se směje .
*CHI: **<žlutou> [?]** .
%pho: žuu .
*MOT: žlutou, ano .

4.5.3 Využití řádků hlavičkového typu

@New Episode signalizuje předěl na nahrávce, je bez dvojtečky a bez obsahu

@Situation: popis situace, objeví se hned v hlavičce, ale taky během přepisu, pokud dojde k předělu, např. vždy po “New Epizode”
Pokud se natáčelo během více dní (v rámci jednoho datapointu), přechod k další nahrávce označíme právě pomocí “New Epizode” a do @Situation uvedeme např. “o den později”

@Comment: cokoli dalšího, co je potřeba sdělit (od %com se liší tím, že se zde jedná o informace obecnější povahy, které se nevztahují jen k jedné konkrétní výpovědi)

4.5.4 Využití vedlejších řádků

%tim: časový údaj, kdy výpověď začala; referenčním bodem je 00:00 = začátek nahrávky; nepoužívá se často, ale může posloužit k upozornění na výraznější pauzy

- %add:** specifikuje adresáta dané výpovědi; **využijeme vždy, pokud se nahrávky účastní více než dvě osoby** (může být velmi důležité pro přehlednost)
- %err:** využíváme pro objasnění chyb (viz *dále*)
- %com:** komentář, který se týká přímo dané výpovědi, má univerzální využití

Preferované pořadí vedlejších řádků je: %pho; %tim; %add; %err; %com

- *MOT: já vám to tady takhle připravím, pane doktore, a můžete číst .
- %pho: já vám to tady takle připravím pane doktore a můžete číst .
- @Comment: MOT odejde z pokoje .
- *CHI: jednou ráno Geppetto chtělo mít Pinocchia [*] .
- %pho: jednou jáno Dzepeto chtělo mít Pinocchia .
- %err: chtělo = chtěl .
- %com: CHI "předčítá" podle knihy .

Řádek %com se využívá velmi často a velmi univerzálně. Jeho nejdůležitějším užitím je dovysvětlení intonace, resp. funkce výpovědi (která jinak bez nahrávky není zřejmá). Může to být důležité zejména u jednoslovných dětských výpovědí, které mohou dle intonace, doprovodných zvuků a gest mít funkci čistě popisnou, anebo žádací, a často také jen opakovací (dítě záměrně opakuje, učí se, zpřesňuje svou výslovnost). Komentuje se veškerý neverbální kontext, který dovysvětluje význam promluvy, tj. především to, kam mluvčí zrovna ukazuje a co během promluvy dělá (něco uchopí, gestikuluje, změní směr pohledu apod.), dále potom tón výpovědi. Pokud se činnost neváže k promluvě či pro její interpretaci není důležitá, komentovat ji nemusíme. Komentáře používejte hojně také pro dovysvětlení ukazovacích, popř. osobních či přivlastňovacích (on, jeho) zájmen, popř. zájmných příslovcí (tady, sem, tak, takhle).

- *CHI: &aj korálek .
- %pho: aj kouajek .
- %com: drží sáček, tónem a pokýváním hlavou žádá o otevření sáčku a vyndání korálků .
- *CHI: &te korálek .
- %pho: te koueuek .
- %com: komentuje vybraný korálek a ukazuje na něj .
- *FAT: a ještě tenhle dáš na tu šňůrku ?
- %pho: a eště tenle dáš na tu šňůrku .
- %com: ukazuje na jiný nenavlečený korálek ležící před CHI .
- *CHI: šňůrku .
- %pho: čupšu .
- %com: opakuje .
- *FAT: šňůrku .
- %pho: šňůrku .
- %com: opakuje a potvrzuje, nebo upřesňuje výslovnost .

Pokud něco neprobíhá úplně synchronizovaně s řečí, ale před nebo po ní, je lepší využít obecný řádek @Comment. U komentářů a situačních popisů dbejte na to, aby byl přepis společně s těmito komentáři co nejsrozumitelnější pro osobu, která nemá k dispozici nahrávku. Tedy s nimi opravdu nemusíte šetřit. Zároveň se soustřeďte na to, abyste ke komentování vybírali skutečnosti, které jsou relevantní pro komunikační výměnu mezi účastníky (např. kašel, pokud o něm někdo zrovna nemluví, není důležitý; pokud ale dítě zakašle a matka to poté komentuje, je důležité dětský kašel zaznamenat).

Legenda k následující ukázce z přepisu (pho řádky jsou záměrně vynechány):

- Červeně jsou vyznačeny komentáře, které jsou pro interpretaci promluv podstatné a jejich zachycení v přepisu je nezbytné. %com pomáhá vysvětlit konkrétní promluvu, ke které se váže (často vysvětluje obsažené deiktické výrazy) a @Comment pomáhá vysvětlit kontext promluv okolo (zde např. obsahuje popis toho, co CHI dělá v návaznosti na to, co mu MOT říká, tj. následuje její pokyny, byť CHI sám neříká nic).
- Zeleně jsou označeny komentáře, které v přepisu být vůbec nemusí.

*CHI: musím si to oloupat .
%com: o mrkvi .
*MOT: chceš to oloupat ?
%com: otevírá šuplík s příbory .
*CHI: asi krájet .
*MOT: krájet spíš to chceš, že jo .
*MOT: ale ono to je hrozně tvrdý .
%com: zavře šuplík a vrátí se k loupání mrkve .
*MOT: spíš zkus nakrájet bramboru .
*MOT: ta je taková měkčí .
*MOT: šlo by to ?
%com: podívá se na CHI .
*CHI: to bysem si dal .
%com: dává do pusy mrkev .
*MOT: dal by sis mrkev ?
*CHI: <krájí> [x 3] .
*MOT: hele@i, tady máš bramboru .
%com: položí před CHI prkýnko a na něj bramboru, pak otevře šuplík s příbory .
*MOT: tady máš nůž .
%com: položí nůž na prkýnko .
*CHI: tohleto si (.) nakrájím .
%com: bere si nůž do ruky; následně začne krájet bramboru .
*MOT: no, nakrájej bramboru .
*MOT: a dávej pozor na tu druhou ruku, jo ?
%com: setře utěrkou kuchyňskou linku vedle CHI .
*MOT: ať se do ní nekrojíš .
@Comment: CHI přestane krájet, pak začne krájet bramboru bez držení jen pravou rukou .
*MOT: víš ?
*MOT: takhle to drž, ale daleko od toho nože .
%com: ukáže, jak má chytit bramboru .
CHI: tak to chyť [] .
%err: reversal .
%com: drží nůž oběma rukama a snaží se nakrojit bramboru, MOT ho po očku sleduje .
*CHI: tak .
%com: přidrží si levou rukou bramboru .

4.5.5 Básničky a písničky

Pokud se na nahrávce objeví recitace básniček, zpívání nebo jen zlomky takových naučených textů, je potřeba je většinou vyřadit z analýzy. Takovou sekvenci vložíte do špičatých závorek a označíte ji komentářem jako **zpěv** nebo jako **básničku**. Jako **básničku** označujete i písňové texty, pokud nejsou zpívané, ale jen přeříkávané. V případech, kdy jsou takové úseky rozsáhlé, např. se na nahrávce zpívá úplně celá písnička, nemusíte ji přepisovat a vložte pouze vysvětlující komentář (případně s informací, v jakém časovém bodě zpívání/recitace začala; časování značíte pomocí %tim, viz výše). **Někdy je možné sekvenci normálně ponechat v přepisu bez závorek, pokud se text básničky/písničky jeví hodně aktualizovaný nebo nějak vědomě analyzovaný.**

*CHI: <bába leze do bezu, já tam za ní polezu> [=! básnička] .
%pho: bába leze do bezu já tam za ní polezu .

4.6 Morfologické inovace, idiosynkratická slova a chyby

Zejména u mladších dětí v jednoslovném období jsou velmi časté idiosynkratické formy označované kódem @c. V pozdějším věku tento typ idiosynkracie ustupuje a může naopak stoupat podíl morfologických inovací @n, popř. výskyt chyb. **Za hlavním řádkem obsahujícím slovo s jedním z těchto kódů vždy musí následovat řádek %com**, který vysvětlí význam daného slova, příp. obsahuje poznámku “význam nejasný”.

@c používejte tam, kde jde o zkomoleninu dospělé formy (manyky@c = mandarinky, díško@c = břicho), popř. slovo zcela idiosynkratického původu (fíšo@c = kostka), dále také u **rozšíření/zúžení/změny extenze konvenční formy** (čáp@c = jeřáb).

Občas se v nahrávce můžou vyskytnout **slova z jiných jazyků**, zejména z angličtiny, se kterou se děti v různé míře setkávají. Pro taková slova používejte kód @z:c, např. head@z:c, knees@z:c, bunny@z:c.

@n používejte tam, kde je jasně rozpoznatelná slovtvorná/tvarotvorná motivace (bábyfuk@n > bába + bublifuk, zachápal@n = pochopil > za + chápal, čokoládovář@n > čokoláda + (ov)ář, píšet@n = psát > píše, psu@n = píšu > psát).

Jako @n (a nikoli jako err!) značíme tedy zejména

1. použití sufixu z jiného paradigmatu (při zachování pádu/osoby a čísla, pokud se tato koncovka ve vlastním paradigmatu slova nevyskytuje):
 - na tomhle vozíkovi@n = vozíku
 - zpívu@n = zpívám (pokud se takový tvar vyskytuje v *Internetové jazykové příručce*, tak neznačit, např. *sypám* i *sypu*)
 - nakloňuješ@n = nakláníš
2. použití analytického futura namísto syntetického a naopak
 - bude@n jet@n = pojedě
 - nebudu@n dostat@n = nedostanu
 - pozávodí@n = bude závodit
3. nedodržení konvenční morfonologické alternace nebo naopak její nekonvenční užití:
 - pytl@n (= pytel > pytle)
 - schůdy@n (= schody > schůdek)
 - číselka@n (= čísílka > čísel)
 - psu@n (= píšu > psát)
 - píšet@n (= psát > píšu)

4. slovotvorné novotvary vzniklé přidáním/vynecháním/záměnou slovotvorného sufixu/prefixu:
 - čokoládovář@n, nabírák@n, nalejvák@n, čarodějnicový@n, šikulák@n
 - dovázali@n (zavázali); kolébavka@n (ukolébavka); zasnítíme@n (rozsvítíme); zapečený@n (upečený); zachápal@n (pochoopil); načárkovat@n (nakreslit čárky); pobombovaný@n; zaborala@n (nabourala); zaběhl@n (doběhl); ulepené@n (nalepené); výhledna@n (rozhledna); záhrada@n (ohrada); pjatá@n (napjatá)
5. Inovativní složeniny: bábyfuk@n, kolotočí@n, loďoletadlo@n, superschody@n

Slovotvorné novotvary mohou nabývat i konvenční podoby, ale s nekonvenčním významem!!!

- rybárna@n = místo, kde žijí ryby
- nočník@n = vlak, který jezdí v noci
- výše uvedené případy zapečený@n, zaběhl@n, ulepené@n

Jako @n (ani jako err!) naopak NEZACHYCUJEME jevy, které jsou frekventované i v dospělé češtině. Je důležité si uvědomit, že žádný jednotlivec nepoužívá všechny prostředky a není si třeba ani vědom, co je frekventované v komunitách, do kterých sám nepatří (časté zejména u teritoriální variace). Jde zejména o:

- obecně české koncovky: “o autách”, “s kamarádama”
- koncovku *-a* (životná) místo nulové koncovky (neživotné) ve 4sg: “skládačku a bagra”
- kolísání vzorů *hrad a les*: “do batoha”, “bez pilníka”, “bez suka”
- kolísání vzorů *žena a růže*: “uvážu ti mašlu”

Jako **chyby (err)** naopak označujeme zejména:

- záměny osoby, čísla, času u sloves (nedodržení shody)
- záměny pádu, čísla u jmen
- převrácené užití 1./2.osoby (jak u zájmen, např. “ty”, tak v časování) – do %err uvádíme “reversal”.

*MOT: chceš ?

CHI: chceš [] .

%pho: ceš .

%err: reversal .

V hlavním řádku na chybu upozorníte hvězdičkou v hranatých závorkách [*], kterou umístíte na konec výpovědi před koncový znak. Chybu pak zaznamenáte/okomentujete na řádku %err. Komentář může vyjadřovat nejistotu a uvádět i více variant. Pokud vůbec nelze definovat, v čem chyba spočívá, použijte zápis %err: **nejasné** .

CHI: my chodíme jenom normální cestu [] .

%pho: myd chodíme jenom nojmálný testu .

%err: **cestu = cestou** .

*CHI: já ti ukážu moje kladívko .

%pho: já ti ukážu moe kadívko .

*MOT: potom, jo ?

CHI: já ti ho ukážu do postele [] .

%pho: já ti ho ukážu do postele .

%err: **ukážu ti = přinesu ti ukázat** .

CHI: má mezi kočiček [] .
%pho: má mezi kocicek .
%err: nejasné .

Pokud chyba spočívá ve **vynechání slova**, napíšete ho už na hlavní řádek, přitom na začátek připojíte nulu (v tom případě **nepoužíváte hvězdičku**). Chybějící slovo jde nicméně často těžko takto jednoznačně rozřešit, proto je stanoven seznam takto kódovaných vynechaných kategorií, u nichž lze rekonstruovat konkrétní slovo: jde o **2. osobu pomocného slovesa být; zvrtná zájmena a další enklitika; předložky**. V některých případech je také možné doplnit jako chybějící pouze kategorii (zejm. *Objekt, Oprep*). Slovo, které je ve výpovědi navíc, se speciálně neznačí, pro označení poslouží hvězdička jako výše.

*CHI: koupu 0se 0ve vodě .
%pho: koupu odě .

CHI: zmrznul se [] .
%pho: sm@snul se .
%err: zmrznul .

Kdy zkomoleniny zaznamenat (@c) a kdy je standardizovat

V mnoha případech jsou dětské odchylky od dospělé formy jen mírné a je žádoucí je na hlavním řádku raději standardizovat na dospělou formu. Abyste to mohli udělat, tak:

1. musíte vědět, co daná forma v daném kontextu znamená (často totiž můžete jako @c označit i formy, které jsou pro dítě zjevně slovem, ale ani rodiče třeba netuší, co mají znamenat)
2. musí být zachovaná gramatická část (koncovka/sufix: *-dlo*, resp. *-o* u “yvuladlo”)

dále přihlížejte k následujícím faktorům:

3. počet slabik
4. pořadí slabik
5. zachování vokálů
6. pokud jsou konsonantické odchylky, měli byste si je dokázat odůvodnit (chybějící nebo jinak tvořený konsonant = zjednodušení artikulace; je dobré se odrazit od vytvoření seznamu zkomolenin daného dítěte pro danou nahrávku a podívat se, které odchylky dělá pravidelně – pravděpodobně to bude stačit jednou u prvního přepisu)

Pokud nejsou kritéria ve větším počtu splněna, ponecháme na hlavním řádku realizaci dítěte označenou jako @c. Kritéria mají částečně sestupnou důležitost (1. a 2. jsou nejdůležitější). Viz příklady níže: např. tuť > kluk bylo standardizováno, protože dané dítě substituovalo *t/t' > k* pravidelně.

Pokud s dětskou formou operuje i rodič, opravuje dítě, nebo naopak přistupuje na jeho tvar, pak tuto formu vždy zachovejte s kódem @c i na hlavním řádku!

*CHI: to není holčička .
%pho: to něni oučiča .
*SIT: a kdo je to ?
%pho: a kdoe to .
*CHI: to je kluk .
%pho: toe tuť .

*SIT: co je tohle ?
%pho: co je todle .
*CHI: &e yvuladlo@c .
%pho: e yvuladlo .
%com: umyvadlo .

4.7 Citoslovce

Citoslovce je třeba důsledně rozlišovat od odpovídajících neverbálních jevů: **&=smích vs chichi@i** nebo **&=povzdech vs uh@i**. Chichi@i nebo uh@i se v přepisu velmi pravděpodobně vůbec neobjeví.

Dětské varianty citoslovcí označujte jako dětské slovo a opomínejte tak na hlavním řádku jejich citoslovečnou platnost (např. kukukuku@c = kykyryký@i), kterou případně můžete zachytit v komentáři.

Stejně jako u jiných slov, pokud není význam citoslovce úplně jasný, okomentujte kontext v řádku %com (např. u citoslovce na@i uveďte, zda jde o citoslovce vyjadřující podávání; příp. uveďte, jaká činnost či gesto jeho užití doprovází).

Při nejistotě určování, co je a co není citoslovce, se řiďte podle SSJČ. Pozor na @i u citoslovcí, jako je **aha@i**, **hele@i**, **jé@i**, **fuj@i** (často se na jejich označení zapomíná), rovněž **ahoj@i**, **nazdar@i** a **čau@i** jsou citoslovce.

4.7.1 Zápis různých druhů citoslovcí

1. Citoslovce (často tzv. zvukomalebná) s paralelní stavbou (**čáry_máry@i**, **hola_hej@i**) a podobná citoslovce, kde se složky neopakují doslova (**dupy_dup@i**, **cáky_cák@i**), jsou spojena podtržítkem. Stejně tak další ustálená citoslovečná spojení, konkrétně: **ach_je@i**, **a_je@i**, **ty_je@i**, **ty_brd@i**, **pane_je@i**, **no_teda@i**, **heleď_se@i**, **ty_bláho@i**.
2. Citoslovce s multiplikací slabik podtržítka nikdy nemají.
 - slabiky o více než dvou hláskách (**brum@i**, **haf@i**, **píp@i**, **chro@i**) nebo víceslabičné (**houp@i**, **hafy@i**, **drnky@i**) se při multiplikaci značí jako opakování (brum@i [x 4])
 - slabiky o stavbě konsonant+vokál (**je@i**, **pa@i**, **pí@i**) se při multiplikaci píšou bez mezery za sebou, ale maximálně třikrát (**jejeje@i**, **jejeje@i**, **papa@i**, **pípípí@i**, **pípípíp@i**); konkrétní realizaci zachycujeme na fonetickém řádku.

4.7.2 Zvláštní užití citoslovcí

Další speciální kódy využíváme pro zvláštní užívání citoslovcí v raných stádiích řeči. Pro užití citoslovce jako predikátu použijeme kód **@z:ip** (ve spojení s (u)dělat), pro citoslovce v nominální pozici **@z:in**, v atributivní pozici **@z:ia**. Pokud je citoslovce součástí verbonominálního predikátu, neznačíme jej tedy jako @z:ip, ale @z:ia nebo @z:in ("Brambory jsou blé@z:ia", "To je <bum@z:in> [x 2]"). Tyto funkce kódujeme pouze v jednoznačných kontextech, nikdy je tedy nekódujeme v jednoslovných výpovědích (toto kódování se častěji vyskytne u dospělých).

*MOT: uděláme ham@z:ip ?
%pho: uděláme ham .

*MOT: podívej, to je <brm@z:in> [x 2] !
%pho: podívej to je brmbrm .
%com: brmbrm = auto .

*MOT: tady je oheň, to je <au@z:ia> [x 2] .
%pho: tady je oheň to je auau .

4.8 Názvy, jména, citace a přímá řeč

Jako uvozovky vždy používejte dvojité anglické uvozovky nahoře: "" (Unicode pro počáteční uvozovky je 201C a pro koncové uvozovky 201D).

1. názvy písní a pohádek či knih, stejně jako předčítaný text a jiné citace se vkládají do uvozovek; v případě názvů se píše velké písmeno na začátku
*CHI: já chci "Hajný lesa pán" .
*MOT: světový bestseller "Víš co má v plínce myš" ?
*CHI: a on zpívá "Hajdom_hajdom_tydlidom@i" .
2. jména postav, míst apod. – všechny složky velkými písmeny, bez uvozovek
*MOT: a potom **Červená Karkulka** potkala vlka .
*MOT: v **Country Lifu** .
 - jména obsahující čísla: číslice není možné v CHATu použít a název se tedy zapíše slovy
*CHI: můj oblíbenej je **Fiat Pět Set** .
3. názvy písmen: podle konkrétní realizace buď "M" nebo "Em" (bez uvozovek, ale s velkým písmenem)
*CHI: a tohle je M ?
*MOT: jo, to je velké Em .
 - Zkratky je nutné psát bez teček (tečka je pouze koncový znak): ET (= E.T. mimozemšťan, "ítý"), EC@c (zkomolené označení letadla RC)

Výzvu "řekni" chápejte v přepisu jako přímou řeč a zapisujte s uvozovkami. Při bezprostředním sousedství uvozovek s dalšími kódy je nutné po a před uvozovkami vložit mezeru (" <haf@i> [x 2] "). Viz také příklad s &=imit:CHI výše.

*MOT: řekni "žlutá" .
%pho: řekni žlutá .

*MOT: řekni " <haf@i> [x 2] " .
%pho: řekni haf haf .

*MOT: co "didi@c" ?
%pho: co didi .
%com: nerozumí, co chce CHI říct a opakuje po něm .

5. Anonymizace a pseudonymizace

Jméno hlavního dítěte nahraďte pseudonymem, nejlépe ihned v průběhu přepisu. Pro každý pseudonym vytvoříte i set příslušných hypokoristik, které následně v celém korpusu používáte konzistentně (např. Báro = Sáro; Barunko = Sárinko). Jméno je nutné vhodným způsobem nahradit i na řádku %pho. Jména ostatních lidí ponecháváte. Příjmení, adresy apod. se na hlavním a %pho řádku nahradí kódem zzz a v komentáři se doplní, jaký typ informace je vynechán. Pokud si nejste jistí, jak s daným údajem naložit, označte ho kódem @an, abyste se k němu mohli snadno vrátit později.

*CHI: já se jmenuju Adam zzz .
%pho: já se menuu Adam zzz .
%com: vynechané příjmení .

6. Hlavní rozdíly proti verzi 2.0

VŠE JE NUTNO OPRAVIT V PŘEPISECH CHROMAT UDĚLANÝCH PŘED VYDÁNÍM VERZE 2.1

- změna používání podtržitek u citoslovcí (paralelní vs multiplikovaná)
- zrušení podtržitek u názvů a jmen
- zavedení uvozovek u citací/přímé řeči
- zavedení &=imit:CHI
- podrobné info o anotaci @n, @c a chyb
- odstranění kódování **Bezpríznačového agramatického dětského tvaru**

(Pokud v rané řečové fázi dítěte u určité formy slova, která naznačuje nějaký konvenční tvar (mami - 5sg ; sype - 3sg), není z kontextu jasné, že jde funkčně opravdu o daný tvar (forma se vyskytuje v kontextech, které nenaznačují vokativ/3sg), můžeme tuto formu označit jako "**mami@c**" / "**sype@c**" a do komentáře uvést kromě významu i poznámku "bezpríznačový agramatický dětský tvar". Toto označení je často vázané pouze na kratší období vývoje dítěte, pokud dítě začne používat i jiné tvary daného slova a kontexty ukazují, že daný tvar používá převážně funkčně správně, slovo přestáváme značit jako dětské slovo - bezpríznačový agramatický dětský tvar (a pokud je použito nesprávně, značíme "error").