

Korpusy rané češtiny

CoCzeFLA | Corpora of Czech as the First Language in Acquisition

www.coczefla.ff.cuni.cz

Manuál pro přepisovatele

autoři: Anna Chromá, Filip Smolík, Klára Matiasovitsová

Tento manuál vznikl na základě 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]. <https://talkbank.org/manuals/CHAT.pdf>

- 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, a ne třeba překryvy).
- Název přepisu je přesně stejný jako název nahrávky či data pointu. Data point je tvořen několika nahrávkami pořizovanými 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).



FILOZOFICKÁ FAKULTA
Univerzita Karlova



MINISTERSTVO ŠKOLSTVÍ,
MLÁDEŽE A TĚLOVÝCHOVY



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 slova, 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	8
4.5.3 Využití řádků hlavičkového typu	9
4.5.4 Využití vedlejších řádků	9
4.5.5 Básničky a písničky	11
4.6 Morfologické inovace, idiosynkratická slova a chyby	11
4.7 Citoslovce	15
4.7.1 Zápis různých druhů citoslovcí	15
4.7.2 Zvláštní užití citoslovcí	15
4.8 Názvy, jména, citace, čtení (zápisy v uvozovkách)	16
5. Anonymizace a pseudonymizace	17
6. Rozdíly mezi verzemi	18
6.1 Co je nového ve verzi 3.0 (proti 2.0)	18
6.2 Co je nového ve verzi 3.1 (proti 3.0)	18

1. Základní rámec přepisu: hlavička a ukončení

V hlavičce přepisu se používají řádky uvozené znakem @. Uvádí se seznam osob na nahrávce a informace o nich (věk vyplňuje koordinátor projektu), jazyk a přepisovatel. Hlavička je pouze jedna, a to na samém začátku přepisu (data pointu).

```
@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
@ID:                 ces|ChroMat|CHI|1;08.18|male|||Target_Child|||
@ID:                 ces|ChroMat|MOT||female|||Mother|||
@Transcriber:        TERBIN
```

V přepisu nikde nesmí zůstat prázdné řádky.

Na konec celého přepisu (nikoli jednotlivých epizod, které jsou značené pomocí @New Episode, viz níže) patří řádek označující konec:

```
@End
```

2. Typy řádků v přepisu a jejich značení

Každá výpověď má svůj vlastní řádek, přičemž replika (= celý úsek před vystřídáním mluvčích) 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 (ke konkrétním způsobům zakončení viz níže). Před a po interpunkci (čárky, středníky, uvozovky) na kterémkoli řádku je také vždy mezera.

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

Vedlejší řádky dovysvětlují předcházející hl. řádek:

- %pho upřesňuje artikulační realizaci pronesené výpovědi. Tento řádek vždycky končí tečkou (bez ohledu na intonaci) a nesmí obsahovat žádné speciální kódy ani interpunkci, 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.
- %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:      já neunesu .
%pho:      já neuřešu .
%com:      myslí matčin batoh .
```

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 osvojování (říká se "motolka" > píšete *motorka*; říká se "kouajek" > píšete *korálek*). K záznamu artikulační realizace slouží %pho řádek, i ten je nicméně jen přibližný.

Šva (neutrální středový vokál výslovností podobný anglickému neurčitému členu) na %pho řádku můžete zaznamenat pomocí @. Šva může např. v některých případech nahrazovat správnou výslovnost konsonantů (p@ší = prší, v@k = vlk) nebo se objeví jako hodně redukovaný vokál (k@máu = komár). Tento zápis se ale nesmí vyskytnout na hlavním řádku (jde o jediný speciální kód, který se používá na %pho řádku).

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

Odchytky od konvenční dospělé morfologie a syntaxe na hlavním řádku zachycujete. Stejně tak ponecháváte silně zkomolená slova. Podrobněji k těmto záležitostem viz níže.

3.1 Přepis v případě konkrétních typických variant

3.1.1 Redukce souhláskových skupin uvnitř a na začátku slova, 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 nějakej čas .

%pho: tam puđu až za ňákej 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*)

na hl. řádek zapisujete "h":

*MOT: na tohlencto prosím tě nesahej .

%pho: na todlencto simtě nešahej .

Nepisovný kondicionál

na hl. řádek zapisujete (*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 *j* u *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

&eee zvuk zastavení po chybě a před opravou (něco jako "teda" nebo "vlastně")

hm souhlas 1

mhm souhlas 2 (dvouslabičný)

emem nesouhlas (dvouslabičný)

nee dvouslabičný nesouhlas 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 přerušovaných 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. 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").

*CHI: xxx .

%pho: xxx .

*CHI: bílý &em onp@c .

%pho: bílý em onp .

%com: onp = význam nejasný .

*CHI: &p prstík .
%pho: p p@stík .

Záměrný fragment Dospělí mluví 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 tohle je če .
%com: čeká na doplnění "červená" .

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 výpovědi s fragmenty se znakem &.

Pokud je tedy odpověď "spíš ne", pak zaznamenáte pomocí xxx (tím signalizujete, že ve výpovědi chybí podstatný syntaktický element, kterému nerozumíte, a necháte výpověď vyřadit z analýz).

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).

Je také důležité se zamyslet nad tím, zda není 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. 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.

+... (s třemi tečkami, nikoli trojtečkou!) použijete pro sebezpřerušeni 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).

Vedlejší řádky jsou VŽDY zakončeny tečkou.

*CHI: takhle &ši +/.
%pho: takle ši .
*MOT: Adélko , na to se nesaá .
%pho: Adélko na to se nešahá .

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

*CHI: &tona my jsme +...
 %pho: tona my šme .
 *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, použijete kód <> [/], jestliže se opraví, pak použijete <> [//].

*MOT: tak <co mu teda> [/] jaký mu dávaš pleničky ?
 %pho: tak co mu teda jaký mu dávaš 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 šme vyštupovaui asi p@níma dvema sme vykupovauí .

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

Případy opakování s ukončenou intonací, které nelze chápat jako opakované začátky (false starts), značíte jako [x 3]. Mezi x a číslem je mezera. V případě opakování fragmentů toto ale neznačíte, fragmenty rozepíšete (viz &kv &kv níže).

*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 uvnitř výpovědi slouží znak (.). 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 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 zajíká), 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, jedna replika se rozdělí na více výpovědí. Dělíte podle klesavé i stoupavé koncové intonace a podle smyslu výpovědí, u velmi malých dětí často i po jednotlivých slovech.

Zároveň ale platí, že na jednom hlavním řádku mohou být max dvě souřadně spojené hlavní věty ve slučovacím poměru (bez spojky nebo s "a"), a to jen v případě, že spolu významově úzce souvisí. Pokud je lze rozdělit, pak by na jednom řádku měla optimálně být jen jedna hlavní věta (pokud následující navazují ve slučovacím poměru). Jinak je jedna výpověď tvořena celkem více hlavních vět v jiném než slučovacím poměru nebo celkem jedné hlavní věty a jejich závislých vět.

Pokud se bezprostředně opakuje celá výpověď (byť by to bylo jen jedno slovo), pak značíme opakování [x 2] (viz výše). U malých dětí i rodičů s nimi komunikujících hrají velkou roli různé eliptické výpovědi (jedno-, dvouslovné) a opakování. 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). V příkladech níže jsou pro přehlednost vynechané pho řádky.

~~*MOT: popa@c , vid' , to jsou popeláři .~~
*MOT: popa@c , vid' .
*MOT: to jsou popeláři .

~~*MOT: koho tady vidíš , Sári , kdo to je ?~~
*MOT: koho tady vidíš , Sári ?
*MOT: kdo to je ?

~~*MOT: ano , pejsek je tam , haf@i .~~
*MOT: ano , pejsek je tam .
*MOT: haf@i .

*MOT: " Je Marshall hasičský, nebo záchranářský pes " ?
 %com: čte .
 *CHI: požární .
~~*MOT: požární, takže je hasičský, nebo záchranářský ?~~
 *MOT: požární .
 %com: opakuje po CHI .
 *MOT: takže je hasičský , nebo záchranářský ?

Přeškrtnuté verze přepisů jsou tedy zcela nepřípustné, 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:

~~*CHI: houba, samá houba, samá houba.~~
 *CHI: houba .
 *CHI: <samá houba> [x 2] .
~~*CHI: čí@i miminko, čí@i.~~
 *CHI: čí@i miminko .
 *CHI: čí@i .

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 (v případě víceslovného pojmenování spojujeme podtržítkem, např. &=zvuk_žádosti)
 &=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: O .
 %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 začátek přepisu nové nahrávky, 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 data pointu), přechod k další nahrávce označíte právě pomocí "New Epizode" a do @Situation uvedete např. "o den později"

@Comment: cokoli dalšího, co je potřeba sdělit (od %com se liší tím, že se zde jedná o komentáře, které se nevztahují k jedné konkrétní výpovědi, ale k něčemu, co se děje mezi výpověďmi či na nějakém širším úseku).

4.5.4 Využití vedlejších řádků

%add: specifikuje adresáta dané výpovědi; využijete vždy, pokud se nahrávky účastní více než dvě osoby (může být velmi důležité pro přehlednost)
 u CHI výpovědí určujete podle potřeby/pocitu
 u dospěláckých výpovědí určujete povinně, pokud adresátem není CHI (CHI je základní adresát a není třeba ho určovat)

%err: využíváte pro objasnění chyb (viz dále)

%com: komentář, který se týká přímo dané výpovědi, má univerzální využití

Preferované pořadí vedlejších řádků je: %pho; %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 můžou 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 také 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 nemusíte. Komentáře používejte hojně také pro dovysvětlení ukazovacích, popř. osobních či přivlastňovacích (on, jeho) zájmen, popř. zájmenných 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 (viz výše).

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: **otevívá 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 chyt [] .
 %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 přeřikávané. 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 a skončila). Někdy je možné sekvenci 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. Někdy také děti používají existující slova, ale v jiných/posunutých významech, což kódujeme jako @z:m. Všechny tyto případy však mohou být vytvořeny i dospělými

mluvčími (např. lumpací@n kalhotky, knížtička@n). 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 (fišo@c = kostka).

@z:m používejte tam, kde jde o konvenční, slovotvorně nemotivovanou formu použitou s rozšířeným, zúženým nebo jinak změněným významem (čáp@z:m = jeřáb).

@n používejte tam, kde je jasně rozpoznatelná slovotvorná/tvarotvorná motivace. Jako @n (a nikoli jako err!) značte 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 = pojed
 - nebudu@n dostat@n = nedostanu
 - pozavodí@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); zasvítíme@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, lodoletadlo@n, superschody@n

Slovotvorné novotvary přitom mohou nabývat i konvenční podoby, ale s nekonvenčním významem, např.:

- 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!) NEZACHYCUJETE jevy, které jsou frekventované i v dospělé češtině. 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 označujete zejména:

- záměny osoby, čísla, času či jmenného rodu u sloves
- záměny pádu, čísla u jmen
- záměny předložek a spojek
- 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 .

- odchylky na syntaktické nebo diskurzívní úrovni, které často nelze ve výpovědi přesně lokalizovat (do %err píšeme „nejasné“, viz níže):

CHI: ty hamoune , kvůli tomu jsi zničila tyhle &fila na mě fialky [] ?

%pho: ty hajoune kuli tomu si zničila tyle fila na mně filalky .

%err: nejasné .

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 = přinesu 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, eventuálně spojky. 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 .

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

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; viz např. níže bylo tuť > kluk standardizováno, protože dané dítě substituovalo *t/ť* > *k* pravidelně.)

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).

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 jeho 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 tohle .

*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_jo@i*, *ach_bože@i*, *a_je@i*, *ty_je@i*, *pane_je@i*, *ty_brdo@i*, *ty_bláho@i*, *no_teda@i*, *heled_se@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 (*houpy@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*, *pípípí@i*); konkrétní realizaci zachycujeme na fonetickém řádku.

4.7.2 Zvláštní užití citoslovcí

Další speciální kódy využíváte pro zvláštní užívání citoslovcí v raných stádiích řeči. Pro užití citoslovce jako predikátu použijete 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, neznačíte 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 kódujete pouze v jednoznačných kontextech, nikdy je tedy nekódujete v jednoslovných výpovědích.

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 .

%pho: tady je oheň .

MOT: to je <au@z:ia*> [x 2] .

%pho: to je auau .

4.8 Názvy, jména, citace, čtení (zápisy v uvozovkách)

Do uvozovek se vkládají

1. názvy písní a pohádek či knih s velkým písmenem na začátku:

*CHI: já chci " Hajný lesa pán " .

*MOT: světový bestseller " Víš , co má v plínce myš " ?

*CHI: a on zpívá " Hajdom@i hajdom@i tydlidom@i " .

2. předčítání: text přepisujete a členíte standardně, pouze přímou řeč uvnitř čtených úseků neznačíte. Každému čtenému řádku musíte přiřadit řádek %com: čte . Komentář může být v případě potřeby i obširnější, ale vždy musí obsahovat slovo „čte“. Pokud byste vyhodnotili předčítaný úsek jako velmi dlouhý a vůbec by na něj nenavazovala konverzace, konzultujte případ s některým z koordinátorů.

*CHI: ještě .

%pho: eště .

%com: ukáže , kde chce číst .

*MOT: " spánek dopřeje mozku odpočinek " .

%pho: spánek dopřeje mozku odpočinek .

%com: čte .

*MOT: " <když se> [/] některé části mozku <se> [?] při spánku vypnou , avšak ty části , ve kterých se tvoří sny , jsou aktivní " .

%pho: když se některé části mozku sp při spánku vypnou avšak ty části ve kterých se tvoří sny jsou aktivní .

%com: čte .

*MOT: " někteří lidé mají mozek , který odpočívá méně " .

%pho: někteří lidé mají mozek který odpočívá méně .

%com: čte .

3. přímá řeč nebo citace (viz také příklad s &=imit:CHI výše)

*MOT: řekni " žlutá " .

%pho: řekni žlutá .

*MOT: říkáš " pejsek " ?

%pho: říkáš pejsek .

%com: nerozumí .

*MOT: co " didi@c " ?

%pho: co didi .

%com: didi = význam nejasný ; MOT nerozumí , co chce CHI říct , a opakuje po něm .

Jako uvozovky vždy používejte dvojité anglické uvozovky nahoře: " " (Unicode pro počáteční uvozovky je 201C a pro koncové uvozovky 201D; klávesová zkratka je Alt+0147 a Alt+0148), a to na všech řádcích (včetně %com). Po úvodních a před koncovými uvozovkami vždy vkládáte mezeru.

Naopak do uvozovek nevkládáte

1. jména postav, míst apod. (všechny složky píšete velkými písmeny)

*MOT: a potom Červená Karkulka potkala vlka .

*MOT: v Country Lifu .

číslice není možné v CHATu použít a název s nimi se tedy zapíše slovy:

*CHI: můj oblíbenej je Fiat Pět Set .

2. názvy písmen: podle konkrétní realizace buď "M", nebo "Em" (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).

5. Anonymizace a pseudonymizace

Jméno hlavního dítěte nahrazujte pseudonymem. Pro každý pseudonym vytvoříte i set příslušných hypokoristik, které následně v celém korpusu používáte konzistentně (např. Kláro = Sáro; Klárinko = 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.

*CHI: já se jmenuju Adam zzz .

%pho: já se menuu Adam zzz .

%com: vynechané příjmení .

6. Rozdíly mezi verzemi

VŠE JE NUTNO OPRAVIT V PŘEPÍSECH CHROMAT UDĚLANÝCH PŘED VYDÁNÍM VERZE 3.0

6.1 Co je nového ve verzi 3.0 (proti 2.0)

- 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 Bezpříznakového agramatického dětského tvaru

(Pokud v rané řečové fázi dítěte u určité formy slova, která naznačuje nějaký konvenční tvar (mami – vokativ, 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 "bezpříznakový 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 – bezpříznakový agramatický dětský tvar (a pokud je použito nesprávně, značíme "error").

6.2 Co je nového ve verzi 3.1 (proti 3.0)

- mezery před všemi interpunkčními znaménky na všech řádcích (nově čárky, středníky a uvozovky)
- značení předčítání
- značení @z:f a @z:m
- zrušení vedlejšího řádku tim
- zpřísnění vedl. řádku add
- ach_bože@i