

9. SEZNAM PŘÍLOH

Příloha 1: Pravidla přepisu sond pro korpus banátské češtiny BANÁT	70
Příloha 2: Zdrojový kód pro vytvoření vertikály ze souboru TRS a vytvořená vertikála	79
Příloha 3: Jednotlivé příklady nalezených jevů z korpusů	82

Příloha 1: Pravidla přepisu sond pro korpus banátské češtiny BANÁT

Ve velké většině případů vychází z *Pravidel nahrávání a přepisu sond pro korpus ORAL2013*. Změny oproti těmto pravidlům jsou pro přehlednost zvýrazněna.
Datum poslední změny 24. května 2014.

Veškerý popis k anonymizaci přepisu byl zatím odstraněn a bude dodán později, proto zatím přepisujte veškerá jména tak, jak jsou vyřčena (s velkým písmenem).

9.1. Přepis

Přepisujeme na počítači v přepisovacím programu Transcriber (details viz manuál Přepisování pomocí programu Transcriber). Pro přepis volíme transkripci blízkou folkloristické (nikoliv tedy fonetické), tj. co nejbližše běžnému záznamu psanému, ale se speciálními úpravami pro účely počítačového zpracování podle úzu zavedeného v Českém národním korpusu.

Obecné zásady přepisu

1. **Přepisujeme věrně vše, co slyšíme, nic nevynecháváme.** Zachycujeme i všechna přechytnutí, často se opakující přívětné výrazy jako *no* nebo *jo* a hezitační zvuky (viz bod 10. a 11. v oddílu Vlastní transkripce) (*já nevím mmm . no tak sem tam normálně šel . no jo ale mmm . nikdo tam už nebyl . jo*).
2. Promluvy mluvčích v přepisovacím programu Transcriber dělíme do tzv. **segmentů**. Je to většinou úsek tvořící přirozený syntagmatický, sémantický nebo prozodický celek (velmi přibližně odpovídající jednoduché větě či souvětí) v propojeném zvuku a přepisu, tj. každý segment má svou část ve zvukovém záznamu a v přepisu. Délka segmentu nesmí přesáhnout 20 slov. Např. *ale já sem to už udělal* nebo *povídej co si neska dělal* či *Pepo . já nevím co ...*: Simultánní úseky, kdy mluví dva mluvčí současně, nám Transcriber umožňuje zapsat zvláštním způsobem, viz manuál Přepisování pomocí programu Transcriber.
3. Používáme pauzovou interpunkci. To znamená, že opouštíme běžnou interpunkci a hranice vět (ani tam, kde by snad byly zřetelné) se **nevyznačují** ani velkým písmenem na začátku, ani čárkou nebo tečkou na konci, **ale odpovídají jednomu segmentu v přepisu**. Velká písmena jsou vyhrazena jen pro vlastní jména. Všude tam, kde se uprostřed promluvy jednoho mluvčího nebo tam, kde se mluvčí střídají, tj. na hranicích promluv, objeví **pauza**, vyznačíme ji **jednou** nebo **dvěma tečkami**. Jednu tečku použijeme v případě **krátké pauzy**, takové, kterou již zaznamenáme jako drobnou, často velice kratičkou mezeru mezi slovy, ale ne ještě jako zaváhání nebo odmlčení (*šli sme nahoru . no a tam už byli připravený . no* nebo *řek mi že . o tom fakt nic neví*). Dvě tečky používáme v případě **střední pauzy**, tedy o něco výraznější, delší, než je krátká, např. malého zaváhání (*vona mi to .. uvařila výbornou polívku*). Rozdíl mezi krátkou a střední pauzou nelze vymezit dobou jejich trvání, posuzujeme je ve vztahu k tomu, jak rychle mluvčí mluví. Před tečkou/tečkami i za ní/nimi děláme mezeru. Mějte na paměti, že se pauzy nemusí krýt s místem běžné interpunkce (tj. např. s tečkami a čárkami v tradiční interpunkci). **Dlouhou pauzu** vnímanou třeba jako odmlčení zapisujeme poznámkou v kulatých závorkách (*odmlčení*). Více o jejím zapsání viz bod 10. C).

Tečka nebo dvě tečky tedy znamenají pauzu, nikdy nic jiného. Čárky nepoužíváme vůbec.

4
já myslím že . to tak nějak bude ..
co říkáš ?
1
já nevím

Pokud se pauza vyskytuje na hranici promluv dvou mluvčích, zapíšeme ji na konci první z promluv.

1
tak já nevím co s tím ..
3
no jo no

Nezapomínejte psát tečky vyznačující pauzy i na konci segmentů v průběhu promluvy jednoho mluvčího, samozřejmě pokud jsou tam pauzy realizované, např.

2
já sem mu řek že . nevim .
že se mi do toho fakt nechce ..
proč bych to zas měl dělat zrovna já
a ne třeba Pavel .
to už fakt nevim

4. V případě otázek zjišťovacích a doplňovacích a dalších vět a větných fragmentů s tázací intonací zapíšeme na jejich konci otazník odsazený od předchozího slova mezerou (*vidíš to ? cos to přines ? že jo ? vid' ?*). V případě věty nebo větného ekvivalentu s výraznou zvolací intonací zapíšeme na jejich konci vykřičník opět odsazený od předchozího slova mezerou (*nech toho ! Pepo ! fuj !*).

5. Zvláštní rysy syntaktické stránky mluvených projevů (překnutí, přerušování a změny větné perspektivy, přiřazování vět a větných úseků apod.) zachycujeme všechny tak, jak se v projevu objevují za sebou. Pokud se mezi nimi vyskytnou pauzy, všechny je zachytíme (*d'álo se to z týdle mouky . no . z pšeničný .. pak už začínaj prasata křičet*). Pokud jsou řazeny za sebou bezpauzově, neoddělujeme je žádnými znaménky (*no zas je potřeba tym tem prasatum honem něco zrichtovat no a tam eště zas děti procejtnou . tak honem se vaří snídaně*). Pokud se slova opakují za sebou bez pauzy, také je nijak neoddělujeme (*byl tam ten ten ten*). Pokud se opakují s pauzou nebo pauzami, zaznamenáme je (*byl tam nákej . nákej .. člověk*).

6. **Neukončené věty** označujeme pouze na konci promluvy jednoho mluvčího, tj. tam, kde byl mluvčí přerušen jiným mluvčím, ale již na svou promluvu dále nenavázal. Vyznačujeme je **třemi tečkami s dvojtečkou odsazenými mezerou**, tedy ...: Nepoužíváme trojtečku, ale tři tečky za sebou.

0
já sem si pak začal říkat že snad nikam ...:
2
no o tom vůbec nezačínaj
0
no dobře

7. **Přerušené výpovědi** (případ, kdy je mluvčí přerušen, tedy přestane mluvit, mluví druhý mluvčí, a když přestane mluvit, první mluvčí dokončí svůj projev, tedy dořekne, co začal říkat před přerušením) v dialogu se značí **třemi tečkami** na začátku i na konci přerušení. Tři tečky jsou od textu **odsazené na začátku i na konci mezerou**. Opět nepoužíváme trojtečku, ale tři tečky za sebou.

4
potom ud'ali ten kolchoz ...
5
v třicátym vosmym to bylo
4
... tak teprvá to . tu vesnici začali stahovat dohromady

Podobně označujeme případ, kdy byl mluvčí přerušen a jeho nedořečenou výpověď dořekl jiný mluvčí.

0
a vypadal že se z toho snad ...
1
... zjeví . to chápu

Případ, kdy byl první mluvčí přerušen a jeho nedořečenou výpověď dořekl druhý mluvčí, ale první se nedal a dořekl to také, zapisujeme takto:

0
a vypadal že se z toho snad ...
1
... zjeví . to chápu
0
... zjeví jo

Netýká se to případu, kdy první mluvčí mluví a pak současně s ním začne mluvit druhý mluvčí. Zápis by pak vypadal takto:

4
potom ud'ali ten kolchoz

4 + 5
1: tak teprva to . tu vesnici začali stahovat dohromady
2: v třicátym vosmym to bylo

8. Přímou řeč nijak zvlášť nevyznačujeme, tj. nepoužíváme uvozovky. Často totiž není jasné, kde má přímá řeč své hranice.
9. Apostrof se neuzívá, spojovníky u složenin se píší těsně (*česko-polské pohraničí*). Vyskytne-li se částice „li“, píše se s odsazením (*bude -li*). Nikdy se nepoužívají dlouhé pomlčky (–).
10. Pokud je nutné zapsat do textu poznámku nebo vysvětlivku k nějaké mimořádné situaci, užívají se kulaté závorky (). Vysvětlí se tak např. náhlé přerušování toku mluvy nějakým vnějším zásahem nebo ruchy na pozadí, např. televize (pokud v sondě zaznívá hovor z televize, nepřepisujeme ho). Používají se ale jen v těch případech, je-li takové vysvětlení nezbytně nutné pro porozumění zapisovanému textu (tedy jen výjimečně a co nejstručněji). Standardizovaně je používáme v následujících případech:

A) (*smích*) Přerušování promluvy smíchem značíme jednotně touto poznámkou v kulatých závorkách.

1
(smích)
3
ty jo no jo (smích) .. to seš blázen
2
no jo (smích)
1
co sem to ud'ál ?
1 + 3
1: (smích)
2: (smích)

Pokud se směje najednou víc mluvčích než dva, není možné je v programu Transcriber vyjmenovat, a proto se poznámka píše do samostatného segmentu bez označení mluvčího. Detaily viz manuál Přepisování pomocí programu Transcriber.

(Žádný mluvčí)
(smích)

(se smíchem) Pokud mluvčí pronese slovo nebo více slov se smíchem, запиšte před něj nebo před ně tuto poznámku.

3
a von tam normálně (se smíchem) spad

Pokud se mluvčí nesměje, ale vysloví např. něco jako *no haha . ty si vtipnej*, přepisujeme to takto doslovně.

B) (*mluví dítě*) Pokud se stane, že do hovoru vstoupí dítě, zapisujeme jeho projev do kulatých závorek jako poznámku uvedenou slovy *mluví dítě* do zvláštního segmentu bez čísla mluvčího (označení v obdélníčku zůstane *Žádný mluvčí*).

1
pak sme byli kupovat tu novou postel

(*Žádný mluvčí*)
(*mluví dítě mamí já mám hlad*)

1
ted' nic nedostaneš

V přepisu pro banátský korpus se tato značka většinou nepoužívá.

C) (*odmlčení*) Pokud se v rozhovoru objeví dlouhá a výrazná pauza, zapisujeme ji jako poznámku v kulatých závorkách. Nepřidělujeme jí číslo žádného mluvčího, píšeme ji do samostatného segmentu bez označení mluvčího. Detaily viz manuál Přepisování pomocí programu Transcriber.

(*Žádný mluvčí*)
(*odmlčení*)

D) (*zpěv*) Poznámku (*zpěv*) zapisujeme k číslu mluvčího, který zpívá, nebo do samostatného segmentu bez označení mluvčího, pokud zpívá mluvčích více. Slova písně nezapisujeme.

E) (*cizojazyčný projev*) Pokud mluvčí pronese větu nebo delší projev v cizím jazyce, nepřepisujeme ho, nýbrž zapíšeme k číslu mluvčího poznámku v kulatých závorkách. Pokud jde jen o jedno slovo nebo spojení slov, např. frazém, přepisujeme je pravopisem daného jazyka (*von je takovej big boss*).

11. Úseky, které se při přepisu nepodaří rozluštit, se vyznačují sérií tří krátkých pomlček následujících za sebou bez mezer, od okolního textu však mezerou odsazených. Při opětovném poslechu v časovém odstupu se takové místo někdy rozluštit podaří.

3
já sem prostě --- úplně sám
2
já sem ---
0
já sem --- .. ale ted' nevím

12. [...] Poznámky k anonymizaci jmen budou doplněny později.

13. Velká písmena jsou vyhrazena pouze pro vlastní jména. Názvy jedinečných produktů časem přecházejí v názvy obecné, a píšeme je tedy s malým písmenem (*chlebičky s nivou*). Velké písmeno u nich ponecháme pouze v případě, že se o nich mluví jako o jedinečném produktu (*vyrábí se tam sýr Niva*).

Vlastní transkripce

1. Co nejvíce se držíme běžného standardního zápisu, pouze tam, kde je více možností výslovnosti, zachycujeme ji tak, aby ze zápisu bylo možno rekonstruovat, co přesně zaznělo. Zachycujeme různé varianty spisovné výslovnosti i výslovnost, která je s českou výslovnostní normou v rozporu. Různými možnostmi výslovnosti se ale nemyslí nezřetelná nebo nedbalá výslovnost (kromě případů zjednodušení souhláskových skupin výslovně uvedených v bodu 5.), nýbrž pravidelné, systémové varianty.
2. Ty **pravidelné jevy, kde se i spisovná forma mluvená od formy psané** (ať spisovné, či nespisovné) **odlišuje** (tj. tam, kde je výslovnost vždy jedna, přestože je odlišná od spisovného zápisu), se zapisují

spisovnou formou zápisu (rozdíl *i-y*; *dě, tě, ně; bě, pě, mě*; znělé a neznělé souhlásky *dub, sbírat, jabko* atd.).

3. Pokud jde o souhláskové skupiny, spisovnou grafiku respektujeme zvláště tam, kde běžná i spisovná výslovnost předpokládá výslovnost zjednodušenou (*kamennej, každodenní, činněj, rozzlobit, rozsvítit*). Do standardní podoby sjednocujeme i zápis těch souhláskových skupin, jejichž konkrétní zvuková realizace závisí na míře pečlivosti při vyslovování, na tempu apod., kdy tedy vlastně nejde o jazykový jev (zjednodušování, které zachycujeme, najdete v seznamu pod bodem 5.). Neregistrujeme ani běžné splývání souhláskových skupin nebo splývavou artikulaci koncových souhlásek na mezislovních předělech (*pod čepicí, i když je vysloveno jako počepicí*).
4. Příznakové rysy běžné mluvy, a to včetně rysů regionálních, se však zachycují (píšeme tedy např. *volno, dôle, vzdádu, kamen, zrouna, k pánoj, ulicej, perkýnko, vo našom, moravské mět, větr, zešit, zebnat, zešit, kameň, lokeť, kost*). Pečlivě zachycujeme všechny podoby slovesa *být – su, sem, seš, sme, zme, ste, sou, só* i v negaci moravské *něsu, něsem, něseš, něsme, něste, něsou, něsó* (a podobně moravskou předponu třetího stupně adjektiv a příslovci *nélepši*).
5. V těch případech, **kdy se svými ustálenými podobami běžná mluva od výslovnosti spisovné odlišuje, spisovný zápis nerespektujeme a tuto odlišnou výslovnost zaznamenáváme**. Píšeme tedy např. *sem* (= jsem), *pudu* (= půjdu), *von výde* (= vyjde), *já si to vezmu* (= vezmu), *dyť, kerej, práznej, muskej, pocem, von řek*. Tam, kde ustálená zjednodušená výslovnost byla nevýrazná (tj. kdy zjednodušení souhláskové skupiny bylo nezřetelné), dáváme v zápise přednost podobě spisovné. V zápisech mohou tedy být i dublety. Někdy jde o dublety regionální (*jedenáct–jedenást, žencká–ženská, počta–pošta, velikánckej–velikánskej*).

K nejčastějším variantám, na které je třeba dát velký pozor, patří

- *sem-jsem, eště-ještě, dyž-když, dyby-kdyby*
- *čtyři-štyři-štyry-čtyry, čtvrt-štvrt, čtvrtek-štvrtek*
- *jestli-esli-jesi, neska-dneska, nákej-nějakej, d'áli-dělali*

Další uvádí následující přehled:

- *jméno-méno-meno, jmenovat-menovat, jmeniny-meniny*
- *jdu-du, jdeš-deš, jde-de, jdeme-deme, jdete-dete, jdou-dou-dó*
- *půjdu-pudu, půjdeš-pudeš, půjde-pude, půjdeme-pudeme-pudem, půjdete-pudete, půjdou-pudou-pudó*
- *přijdu-přidu, přijdeš-přideš, přijde-přide... přijdou-přidou-přidó, imperativ přijď-přid'...*
- *půjčím-pučím-pučím, půjčíš-pučíš-pučíš, půjčí-pučí-pučí... půjčí-pučijou-pučijó-pučej, imperativ – půjč-puč, půjčte-pučte...*
- *vezmu-vezmu, vezmeš-vezmeš... vezmou-vezmou-vezmó*
- *chci-chcu, ... chtějí-chtě-chtěj*
- *myslím-mysim-mysím, pojd' sem-pocem, vyjde-výde, mohl-moh, otevřeš-votevřeš-vodevřeš*
- *hanba-hamba, tkanička-kanička, žencká, muský, člověk-čověk-čovek-ček,*
- *prázdněj-práznej, lepší-lepčí, menší-menčí*
- *šestnáct-šesnáct-šesnác, sedm-sedum-sedn, osm-osum-osn, osmdesát-osumdesát-osndesát*
- *který-kerej-kerý-keré, kterou-keró-kerou..., některý-někerej-někery-někeré..., takový-takovej-takové, nějaký-nějakej-nějaké-nejakej-nejaké-ňákej, tenhle-tendle-tenle,*
- *jestli-jesli-jesi-ješi-estli-esli-esi-eši, protože-protoe-proe-prože-prtože-prže-pže-poče-prče, však-šak, vždyť-dyť, však-šak*
- *odsud-vodsud-vodsud'-vodsad', vždycky-dycky-ždycky, vlastně-vlasně, kvůli-kůli*
- skupina vz- na začátku slova: *vzpomenu-zpomenu, vzbudit-zbudit, vz máhat se-zmáhat se, vztekly-zteklý, říct-říc,*

Slova *dcera, srdce, ctnost* píšeme takto, jejich výslovnost není variantní, i když je odlišná od spisovného zápisu.

Případ, kdy je sloveso nahrazeno zájmenem *to*, příp. *tento* a je k němu připojena předpona, píšeme vždy dohromady jako jedno slovo (*já sem to neto . nechťel* nebo *ten pes se tam vyto . no vytento*).

6. Zachycujeme i ty případy, kdy může v jednom slově (ve stejném hláskovém okolí) dojít k různé realizaci spodoby znělosti. V takovém případě zachytíme to, co skutečně zaznělo. Jde o realizace souhláskové skupiny *sh* (*schoda/zhoda, naschledanou/nazhledanou, schodit/zhodit*), spojení s předložkou *s, se* (*s máslem/z máslem, s ledem/z ledem, se sestrou/ze sestrou*), nářeční asimilace jako *řůj*

(*tvůj*), *křítko* (*kvítko*), *zme* (*jsme*), *gvůli* (*kvůli*), imperativ 1. os. pl. – *kubme/kupme*, *zavězme/zavěsme*. Předložku *přes* píšeme vždy jako *přes*. Případy, ve kterých vždy dojde ke stejné spodobě a v kterých se tedy nevyskytují varianty, zapisujeme ve spisovné podobě (*dub*, *jabko*).

Zachycujeme i zvláštnosti asimilační artikulace – případy, kdy se *s*, *z* před *č* asimiluje na *š*, (event. *ž*) (*piščitý*, *hešči*, *hežči*, *roščilený*, *rožčilený*, *roščesal*) nebo nářeční realizaci skupiny *šť* jako *šč* (*ešče*, *ščestí*, *sidlišči*, *pracovišče*).

7. Odlišná kvantita se registruje, a to jak kvantita fonologická (*domů/domu*, regionální podoby *rohlik*, *klucí*, *pánove*, *pět kluku*), tak emfatická, emocionální (*bóže*, a to i v případech jako *bylo to dlouhý*). Neoznačuje se však prodlužování slabik na konci promluvového úseku a prodlužování dané váháním (např. chybně: *já sem jim to tam dal na polici .. a voni to tam nenašli* – správně má být: *na polici*).

Moravské realizace komparativů a superlativů s dlouhým *ě* zapisujeme s *j* nebo souhláskou s háčkem a dlouhým *é* (*novjější*, *pěknější*, *vlasatější*).

Na kvantitu dejte pozor především v těchto případech:

- krácení v koncovech i v základu slova (*myslím*, *vim*, *viš*, *tvuj*, *muj*, *svuj*, *vubec*, *pani*, *řikal*, *mit*, *jit*, *Pepík*)

- krátké koncovky adjektiv – přestože pak splývají s příslovci (*žensky*, *mužsky*, *obrovsky*, *přátelsky*, *slovensky*, *lidsky*)

- krátký středový vokál jednoslabičných infinitivů a jejich odvozenin (*byt*, *dat*, *brat*, *psat*)

- krátkost u některých jednoslabičných a dvojslabičných substantiv (*mak*, *hrach*, *žaba*, *skala*)

- zdložené vokály (*brál*, *řikál*, *volál*, *stóji*, *dólu*, *móře*, *pívo*)

- moravské *ú* na místě českého *ou* se zapisuje vždy s čárkou, nikdy s kroužkem (*mnú*, *sebú*, *kúsek*, *nesú*, *kupují*)

8. Cizí slova, původně cizí slova, která se běžně nezapisují počeštěným pravopisem, a cizí vlastní jména (osobní i zeměpisná) se píše původním pravopisem, např. zeměpisné názvy (*New York*), firmy (*Oriflame*), výrobky (*discman*) a jiné (*hardware*, *e-mail*). Slova, kde je běžný počeštěný zápis, zapisujeme počeštěně (*volejbal*, *fotbal*). Pokud mluvíci pronese větu nebo delší projev v cizím jazyce, nepřepisujeme ho, nýbrž zapíšeme k číslu mluvíčího poznámku v kulatých závorkách (*cizojazyčný projev*). Pokud jde jen o jedno slovo nebo spojení slov, např. frazém, přepisujeme je pravopisem daného jazyka (*von je takovej big boss*).

9. Nedořečené slovo označíme hvězdičkou za slovem (bez mezery) (*to bylo vše**). Za hvězdičkou následuje mezera, v případě pauzy následuje mezera a tečka (*to bylo vše* . převčirem*) nebo v případě střední pauzy mezera a dvě tečky (*to bylo vše* .. převčirem*). Podobně označujeme případ, kdy není vysloven začátek slova, např. při opravě koncovky: *takže dala do klobouky *ku jedničky a dvojky*.

Pomocí hvězdiček značíme i případy s příklonným *s*, to je spolu s hvězdičkou vždy odsazeno od textu mezerou (*dělala *s tam něco ? *s tam byl, jo ?*). Případy příklonného *s* typu *sis*, *ses*, *a/bys*, *kdybys* píšeme dohromady.

10. Interjekce zapisujeme přesně tak, jak zazněly, co do počtu i vyznačení pauz (*no no no . no . no*). Skupiny interjekcí přepisujeme rozděleně, tj. *ty jo*, *no jo*, *že jo*, *no jo no*.

11. Responzní zvuky rozlišujeme přitakací a odporovací, zapisujeme je uskupením písmen, a to *hmm* pro přitakací zvuk a *emem* pro odporovací, nesouhlasný. Na rozdíl od nich hezitační zvuky doprovázející mluvený projev označujeme seskupením vždy tří písmen, a to *mmm* pro zvuky spíše souhláskové a *eee* pro zvuky spíše vokalicke. Pokud jsou v projevu odděleny pauzou, použijeme pro oddělení jednu nebo dvě tečky.

12. Zkratky zapisujeme právě tak, jak byly vysloveny, tedy ne DVD, ale *dé vé dé* nebo *dý ví dý*, případně *dývídíčko*. Při zápisu se držíme toho, zda byla zkratka vyslovena dohromady jako jedno slovo (*aids*), pomocí pojmenování hlásek (*ú es á*) nebo byly jednotlivé hlásky vysloveny zvlášť (*v š e*). Podobně přepisujeme (*iks rokú*). Zkratky vždy přepisujeme malými písmeny.

Specifika korpusu BANÁT

1. Označování mluvčích

- Mluvčí už mají předem přidělený kód, který používáme dále v nahrávce.
- Promluvu českého mluvčího (jeho zkratka začíná na CZ) neupravujeme ani nepřipisujeme nic navíc, pouze pečlivě vyznačíme hranice segmentů. Jeho přepis bude doplněn později a zčásti vypuštěn tam, kde je to z kontextu možné.
- V případě překrývajícího se segmentu banátského a českého mluvčího přepisujeme důsledně projevy obou mluvčích.

2. Speciální značky

- **Banatismy** („cizí slova“) – značíme znakem křížku těsně před slovem, tj. bez mezery - #slovo (křížek napíšete pomocí klávesové zkratky alt+x). Jedná se o slova, nejčastěji příslovce a jiné nesklonné slovní druhy, která mají cizí původ (a jsou užívána i v okolních zemích), např.: #makar by přišel, chodí od #numera k #numeru, že #barem to mám, tady se mi #baš nevejdou. Dále sem řadíme zeměpisné názvy, např.: Bígr, Gerník, (Svatá) Helena, Rovensko, Eibentál, Berzáska, Temešvár. Stejnou značkou značíme i použitá rumunská slova, která přepíšeme tak, jak je slyšíme (rumunský výraz bude doplněn v závorce později). V neupravených prepisech jsou tak mohou být značeny i tvary, které patří do další kategorie.
- („běžné a české“) **Hovorové tvary**– značíme zavináčem, @slovo (alt+v), jedná se o slovní tvar, který není běžný ani v beletrii a který nebyl zmíněn výše. Značku použijte, pokud máte pocit, že jde o netypický slovní tvar, který znáte, pokud vás jeho tvar zarazí. Označení bude využito v dalším zpracování. Např: (ve) vnitřku, vobarvit, zasejc
- **Banátské „hovorové tvary“, charakteristické znaky mluvčích** – značíme ampersandem, &slovo (alt+c). Jsou to buď tvary používané často v Banátu, které ale nepatří do běžné zásoby obecné češtiny (např. *ičko, stavínko, vodzamknout*), jsou typické pro Banát či české nářečí (např. koncovka genitivu *deset &rokuch, &nikerakš jsem ho neviděla*) anebo jsou to tvary nestandardně vyslovované, které takto označujeme kvůli jednodušší budoucí lemmatizaci (&mušet místo *mušet*, &ně místo *mně*).
- Pokud si nejste jisti, do které kategorie daný tvar zařadit, raději ho označte, i když konkrétní kategorie nemusí být zřejmá. Všechny značky budou případně opraveny a sjednoceny při kontrole, protože je zřejmé, že kategorie nemají příliš jasné hranice.

3. Nesrozumitelné či špatně srozumitelné úseky

Pokud si v přepisu nejste něčím jisti, nebo dokonce celému úseku nerozumíte, tak **nehádejte**. Jedná se v některých případech o náročný přepis, slyšíte cizí slova či narážky na realitu, kterým nerozumíte a nedávají vám smysl a je pak velmi jednoduché z toho „udělat normální větu“ (která je pak velmi sugestivní pro kontrolora). Takový úsek v případě velké nesrozumitelnosti nepřepisujte vůbec a vyznačte ho třemi pomlčkami (---) anebo na jeho konec napíšete značku pro „nesrozumitelné“ – otazník v závorce (?), při kontrole to bude opraveno či zkontrolováno pečlivěji.

4. Anonymizace obsahu (pouze pro české mluvčí)

XA(počet vteřin) – anonymizace z důvodu citlivého obsahu hovoru

XM(počet vteřin) – anonymizace, monolog českých mluvčích, banátský mluvčí přítomen (ale následuje potom např. dlouhé odmlčení)

XN(počet vteřin) – anonymizace, hovor českých mluvčích, když je banátských mluvčích nepřítomný

Seznam použitého označování:

- . krátká pauza v promluvě jednoho mluvčího nebo na hranici promluv
- .. střední pauza v promluvě jednoho mluvčího nebo na hranici promluv (zaváhání)
- ? tázací intonace
- ! výrazně zvolací intonace
- ...: označení neukončené věty na konci promluvy jednoho mluvčího
- ... označení přerušené výpovědi
- úsek, který se nepodařilo rozluštit
- (?) úsek, který je špatně srozumitelný a přepisující si s ním není zcela jist

* nedořečené slovo

hmm přitakací responzní zvuk

emem nesouhlasný responzní zvuk

mmm označení souhláskových hezitačních zvuků

eee označení samohláskových hezitačních zvuků

(smích) přepis smíchu

(se smíchem) slovo nebo slova pronesená se smíchem

(mluví dítě) přepis promluvy dítěte

(odmlčení) dlouhá a výrazná pauza

(zpěv) zaznamenání zpěvu (slova písně nepřepisujeme)

(cizojazyčný projev) věta nebo delší projev v cizím jazyce

() komentář k situaci

#banatismus

&banátský tvar

@hovorový výraz

XA(počet vteřin) – anonymizace z důvodu obsahu hovoru

XM(počet vteřin) – anonymizace, monolog českých mluvčích, banátský mluvčí přítomen

XN(počet vteřin) – anonymizace, hovor českých mluvčích, když je banátských mluvčích nepřítomný

Příklady použití znamének:

- krátká pauza uprostřed promluvy
4
to sou teda věci . tohle
- střední pauza uprostřed promluvy
1
no jo .. a co je eště novýho ?
- střední pauza na hranici promluv
2
tak něco řekni ..
1
a co ?
- ukončení promluvy střední pauzou, za tím v závorkách smích
3
to měl taky Honza .. (smích)
- nesrozumitelný úsek a smích na konci promluvy
5
vod někoho vopiš --- (smích)
- nesrozumitelný úsek, po něm krátká pauza a pokračování promluvy
1
povídá vo tom zas --- . tajdle máš štrúdl
- nesrozumitelný úsek uprostřed promluvy bez pauz
2
a von tak prostě jako --- celej večer jo
- pouze smích
2
(smích)
- pouze nesrozumitelná promluva
1

- přerušení jednoho mluvčího druhým

1

voni nesnídaj voni teda nesnídaj ...

2

musíš kupovat těch rohlíků dvacet

1

... je teda pravda že tím pádem .

že voni nesnídaj

- postupné přerušení promluv dvou mluvčích (ne současné mluvení! Zde zachyceno tak, jak části promluv zazní postupně za sebou. Zápis souběžného mluvení viz manuál Přepisování pomocí programu Transcriber.)

1

no asi jo . aby prostě jim ...

2

to je divný ...

1

... to došlo

2

... tohleto

- nedořečené slovo

1

já to ne*

- nedořečené slovo a krátká pauza

1

já to ne* . já prostě nevím

- nedořečené slovo současně s nedořečenou větou

1

já jim přine* ...:

2

nic nenos . už to maj

Příloha 2: Zdrojový kód pro vytvoření vertikály ze souboru TRS a vytvořená vertikála

```
#!/usr/bin/perl
use warnings;
use locale;
my $dir = "files"; #složka s trs
my @rep=<$dir/*trs>;
foreach $file (@rep) {
    my $name_input = substr($file,6,-4); #nazev sondy
    my $output = substr($file,0,-4).".txt"; #exportni soubor
    my %mluvci = (); #mluvci v sonde
    my $prekryv = "0"; #prekryv (0/1)
    my %spk = (); #aktualni mluvci
    open (I,"$file") or die "Nelze otevrit soubor";
    open (O,">$output") or die "Nelze uložit soubor";
    #tisk zacatku souboru
    print O "<doc file=\"$name_input\">\n";
    while ($row = <I>) {
        chomp $row;
        #mluvci
        if ($row =~ /Speaker id=\"(.{4})\" name=\"(..)-(..)(..)\"/) {
            $mluvci{$1}="<sp id=\"$2\" pohlavi=\"$3\" vek=\"$4\">\n";
        }
        #zmena mluvčího
        if ($row =~ /<Turn/ ) {
            #ziskani aktualnich mluvci
            if ($row =~ /speaker=\"(spk.) ( )?(spk.)?\"/) {
                $spk{1}=$1;
                if($2){
                    $prekryv = "1";
                    $spk{2}=$3;
                }
                else {
                    $prekryv = "0";
                    print O $mluvci{$spk{1}};
                }
            }
            #zadny mluvci
            else {
                print O "<sp id=\"no_speaker\">\n";
                $spk{1}=0;
                $spk{2}=0;
            }
        }
        #kdyz neni prekryv
        if ($prekryv == "0") {
            #tisk segmentu do vertikaly
            if($mluvci{$spk{1}} =~ /C/ && $row !~ /^</){
                print O "[".$row."] \n";
            }
            else {
                slova($row);
            }
        }
        #překryv
        else {
            #tisk aktualniho mluvciho
            if ($row =~ /<Who nb=\"(.)\"/) {
                print O $mluvci{$spk{1}};
                $n = $1;
            }
            #tisk segmentu do vertikaly
            else {
                if($mluvci{$spk{$n}} =~ /C/ && $row !~ /^</){
                    print O "[".$row."] \n";
                }
            }
        }
    }
}
```

```

        else {
            slova($row);
        }
    }
}
}#konec while
print O "</doc>\n";
close (I);
close (O);
}#konec foreach
my @rep_txt=<$dir/*txt>; #seznam vytvorených vertikál pro každý soubor(txt)
open (O,">vertikala_banat_2014") or die "Nelze uložit soubor";
foreach $file (@rep_txt) {
    open (I,"$file") or die "Nelze otevřít soubor";
    my $cspk =0; #aktualni mluvci
    my $lspk =0; #posledni mluvci
    while ($row = <I>) {
        #if oznaceni mluvciho
        if ($row =~ /<sp id="(.)"/){
            $lspk = $cspk;
            $cspk = $1;
            #if predchozi mluvci stejny
            if ($lspk ne $cspk) {
                print O $row;
            }
        }
        else {
            #oprava
            $row =~ s/[#@][ii]čko/&ičko/g;
            $row =~ s/@//g;
            print O $row;
        }
    }#konec while
    #konec cteni zdroje
    close (I);
}#konec foreach
#zavření vertikály
close (O);
print "done";
sub slova { #SUBPROGRAM segment > co slovo to radek
    if ($_[0] !~ /</) {
        $row = zavorky($row);
        $row = $_[0]." "; #mezera na konec
        $row =~ tr/ +/\n/; #nahrazení mezer koncem radku
        $row =~ s/_/ /g; #zpet mezery v zavorkách
        $row =~ s/&\;/&/g; #nahrazení entity znakem &
        print O $row; #tisk segmentu
    }
} #konec sub
sub zavorky { #SUBPROGRAM slova v zavorkách - " " > "_"
    my $n=0;
    my $z=0;
    my $text;
    while ($n < length($_[0])){
        $ch = substr($_[0],$n,+1);
        if ($ch eq "("){$z=1;}
        if ($ch eq ")"){ $z=0;}
        if ($z ==1 && $ch eq " "){$ch = "_";}
        $text .= $ch;
        $n++;
    }
    return $text;
}
}

```

<doc file="11A001N">
<sp id="no_speaker">
(odmlčení)
<sp id="01" pohlavi="F" vek="79">
tak
bude
spát
tam
v
tom
stavení
.
ju
?
<sp id="C1" pohlavi="F" vek="21">
[ne . my máme půjčený byt od Knappů]
<sp id="01" pohlavi="F" vek="79">
máte
tady
.
spát
?
<sp id="C1" pohlavi="F" vek="21">
[tady .]
<sp id="01" pohlavi="F" vek="79">
tale
jedno
stavení
tam
hodně
spí
víte
Češi
..
aha
.
no
tady
budete
spát
tam
dyž
máte
&mašinu
?

B	.. no do pečení no má vůni . bába taky	si	dává . [si dáva nebo do pečení] [si dáva ?] ne ne do mmm do p	3
B	suším si nějakou trávu na podestýlku .. pojď ddyžtak sedneme	si	tady dovnitř nebo lepší venku ? [tady je lepší vzduch] na luřě je	3
O	samozřejmě . no . no já nevím já sem dycky	si	myslela že dokud jako teda bude tatínek tak by* .	3
O	mlčení) co to je za zvíře ? (ovce se zelenou vlnou) takle voní	si	voznačují ty ovce jestli to sou samice nebo samci ..	3
O	tu firmu . která . je si schopná spočítat ..	si	z návštěvnosti stránek .. kolik . kolik prostě . maj	3
O	jak sem ho teda dostal . že sem mu tam	si	dovolil zaparkovat .. (smích) a úplně prstě všichni z toho	3
O	.. aha . jako že je vysokej a eště myslim	si	že na ně i působí že je světlej . no	3
O	takový to zdanění . tak to . to to zdaňovat	si	nebudou no hmm a ty sou větší než ty jejich	3
B	eee viš . dala] [sem ti je do ruky] [(smích)] ty seš furt uvázaná .	tě	bolí krk nebo copá ? [já ...] [... sem taková zimomřivá někdo to	1
O	mlčení) třeba (odmlčení) jinak mě nic nenapadá (odmlčení)	tě	asi napadnu já (odmlčení) proč by si mě napadával zase	1
B	von mně povídá . a nebudeš se zlobit dyž já	tě	budu volat --- .. a já povídám nebudu . a	3
B	přinesli kafe ..] jo (smích) no copá se vobtěžujete nenoste nic ..	ti	to pomůžu .. [jo (?) tys* tys* mi to] každá jednu [děkuju .] [chceš	1
B	meje po zemi .. no to je takový já ..	ti	to ukážu .. s tym se meje nemusim se vohybat	1
B	to šití už nejsou holt hadry neprodávají se (odmlčení) ale	ti	to už mrká ňák to světylko [hmm .. to nic] eee ? [to nic to nevím	1
B	projde jako tuřým druhým dětom ona na to uměla (odmlčení)	ti	řikám my jsme měli vel napřed to tady nebylo tak	1
B	babička pude .. eee já sem jejich babička . sem	ti	to řikala . já sem vychovala jeden mě* jedenáct rokůch	1
O	by mu to jen utíkalo/no . to je hrozný	ti	řeknu no/fo je vopravdu příšerný no/vim že von	1
O	jich ptala Báro . mně to je tak blbý .	ti	dávat ty peníze takhle jenom .fo by stálo .	1
O	jako asi .ten př* ten překladač to pozná/a	ti	vypíše . ale prostě taji sem někde udělal ňakej .	1
O	ňaky vyše těch . těch eee třeba stupně uspořeno .	ti	to dali/(odmlčení) ti to vyměnili ňak jako trošku min	1
O	zase něco dáš / no co dyť to de votevřít normálně	ti	dám votevřít normálně --- / a že to . / dyť tady	1
O	. / máte program do kolika . do kolikátého ? / (odmlčení)	ti	nevím ani . / je to aji na příští týden ?	1
O	vztahu k . eee k natu a tak no .	ti	budou rozkazovat co máš dělat jo ? / to ne to	1
O	pak sme vykročily ./ale vona je i tak najednou	ti	ted' mluví .tak jako ro* trochu rozřeseně/hmm/jak	1
B	.. stačí to stačí .. pozor na ubrus ---] dědku hele .. dědku já už	ti	mám chuť plesknout .. jo jo ty si přilej eště	3
B	(odmlčení) viš jak mys* eee my sme byli možná to	ti	može vokázat Anka teta Cílka řekni [hmm] depa bydlela Cecilie	3
B	k ní viš . do Prahy . tak já tady	ti	dám pení* ten co si po ňom poslala ten dopis	3
B	ne my . my ti dáme redžim ale ten prý	ti	dá do hrobu jo jo jo . neni zdravěj tak	3
B	mají dávat z kapsy zpátky . [hmm] že to brali	ti	vod lidíh já nevím . [hmm] a holt asi .	3
B	si řekla . tu vychvaluješ ale ta ne ta není	ti	chválíhodná ne ne dej jí pán bůh ...	3
B	ale dyž někery kerý zašelřený tak by ještě	ti	vydrželo ale dyž je ten benzin taky u vás tak	3
O	vona tam má malou předsíňku/hmm .. hmm/vona je	ti	taková . taková najednou jak sem tam se sta*/(se smichem)	3
O	změnit není to jako definitivní .že by prostě .	ti	řekla tak ve čtyřeceti se vdáš/---/a budeš mít	3
O	že vám budeme kupovat dyvidýčko . tak sme koupili už	ti	dar / (odmlčení) / takže už ho mam schovanej půl roku / (odmlč	3
O	zase . to je takovej lék / ktorej j* jako .	ti	* umožní vodkašlat a přitom ti kašel hlumí . / hmm	3
O	chodit / na to . ale von by to tam vyštval	ti	jo ? . / no neska sme tam . / no .	3
O	tu . / eee . rozmarynu . a já poám to	ti	vaří . / eee Magda tedkon jako no . / že to	3
O	jednou se dostaneš do problému . / a k tomu to	ti	no / chci právě říci / ten Josef / NP / to nevím jak	3

Všechny příklady jsou přejaty z uložené konkordance, kromě zvýraznění klíčového slova (KWIC) nedošlo k žádným dalším úpravám. Za pravopisné chyby proto nenese autor této práce odpovědnost. Dále je ponecháván pouze originální kontextu, nikoliv pouze vybraná věta ani není vzdálený kontext doplňován. V případě dat z korpusu BANÁT2014 nejsou vysvětlována cizí slova.

použité popisky:

B = data z korpusu BANÁT2014

O = data z korpusu ORAL2013

s = zájmeno *svůj* možné

z = *nic* ve smyslu vůbec

1 = první pozice ve větě

3 = pozice ve větě