

Bērnu runas transkribēšana un ar to saistītās problēmas

Ilze Auziņa, Guna Rābante-Bušā, Roberts Darģis

LU Matemātikas un informātikas institūts

Bērnu runas korpusa izveide

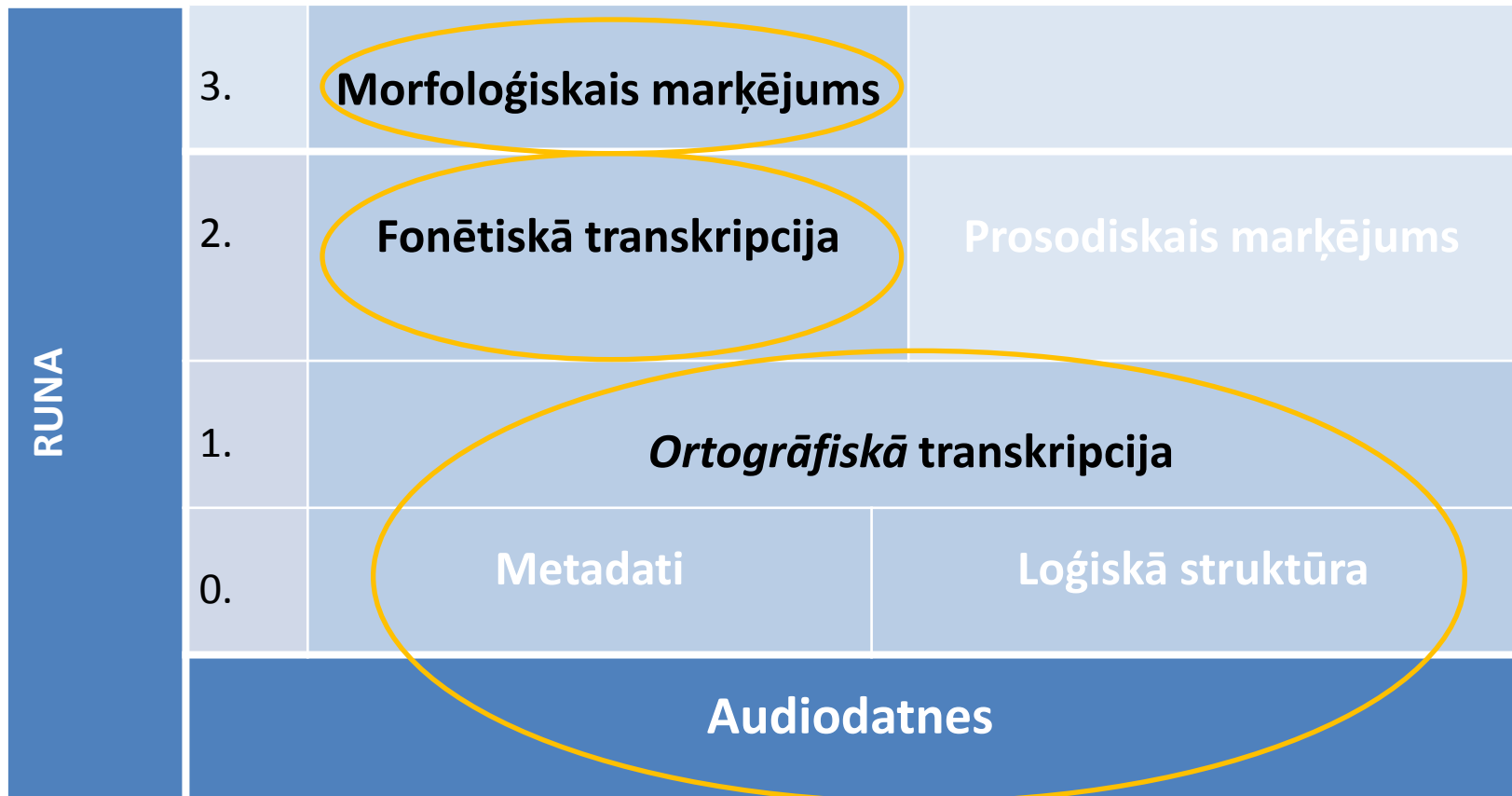
- Projekts «Latviešu valodas monolingvāla un bilingvāla apguve: rīki, teorijas un lietojums»
 - Projekta ilgums: 03.2015.–04.2017.
 - 4 savstarpēji saistītas sadaļas
 - Korpusu izstrāde sadaļā «Morfoloģiski marķēta specializēta bērnu runas korpusa izstrāde»
- Mērķis:
 - Izveidot divus monolingvāli latviski runājošu bērnu runas korpusus
 - Izveidot divus bilingvāli – latviski un krieviski – runājošu bērnu runas korpusus
 - Datus morfoloģiski un daļēji fonētiski marķēt

Runas korpuss

Korpuss – reprezentatīvs rakstīta teksta un / vai transkribētas runas kopums elektroniskā formā, ko izmanto datorizētā valodas analīzē un aprakstīšanā.

Runas korpuss – runas ierakstu kopums, kas pieejams mašīnlasāmā formā, ir marķēts un dokumentēts.

Runas datu marķēšanas līmeņi



Bērnu runas datu iegūšana

- Bērnu runa tiek ierakstīta katru nedēļu 16 mēnešu ilgā periodā
- Katra ieraksta ilgums:
 - Monolingvāli bērni – 30 minūtes
 - Bilingvāli bērni – 30 minūtes katrā valodā
- Sasniedzamais apjoms: 192 stundas
- Bērnu vecums: 18-36 mēneši
- Ieraksta ierīces: diktofons vai mobilais telefons
- Ieraksta vide: bērnu dzīvesvieta
- Par katru ierakstu tiek saglabāta papildu informācija, piem., ieraksta vieta, nodarbe u. tml.

Rezultāti

(2015. gada marts – augusts)

- Izstrādāta metodika audiodatu ieskaņošanai un aprakstīšanai (metadati)
- Izstrādāti ortogrāfiskās transkripcijas principi bērnu runas datu pierakstīšanai mašīnlasāmā formā (CHAT un latviešu valodas ortogrāfiskās transkripcijas adaptēšana)
- Izstrādāta vietne datu pievienošanai un aprakstīšanai
- Ir sākta 4 bērnu runas datu ierakstīšana un audioierakstu atšifrēšana ortogrāfiskajā transkripcijā

Vietne runas datu pievienošanai un raksturošanai (1)

Pievienot ierakstu

Saglabātie ieraksti

Beigt darbu

Jauns ieraksts

vieta

Jauna vieta

vietas nosaukums

apraksts

fotoattēls

Browse...

No file selected.

nodarbe

datums

ieraksts

Browse...

No file selected.

Vietne runas datu pievienošanai un raksturošanai (2)

[Pievienot ierakstu](#)[Saglabātie ieraksti](#)[Beigt darbu](#)

Saglabātie ieraksti

datums	nodarbe	vieta	ieraksts	darbības
2015.05.05	Bērns spēlējas ar koka puzzle (liek lāču ģimenes locekļu apģērba komplektus), ar lupatu lelli, pavārtās pa paklāju, spēlējas ar gumijas dinozauriņiem, šķirsta [...] iatu.	Dzīvojamā istaba Istabā ir divi logi, starp kuriem stāv dīvāns, pretī dīvānam - pie pretējās sienas - ir kamīns un podiņu krāsns, blakus kamīnam atrodas drēbju skapis; perpendikulāri dīvānam pie vienas sienas stāv grāmatu skapis, pie pretējās sienas ir liels ovāls galds ar diviem datoriem un diviem krēsliem. Lielāko daļu istabas grīdas klāj paklājs. Istabā atrodas lielākā daļa bērna spēļmantu.	Lejupielādēt	labot
2015.05.14	Lauma ēd brokastis - mocarellas sieru ar tomātu un baziliku, pēc tam baltmaizi, tad banānu, beigās vēl arī šokolādi. [...]	Virtuve Virtuves vienā galā ir logs, gar vienu sienu novietota elektriskā plīts, blakus ir izlietne, pie tās stāv veļasmašīna; gar pretējo sienu novietoti grīdas skapīši, galds, pie tā ir bērnu barošanas krēsls, blakus stāv ledusskapis. Pie abām garākajām sienām piekārti vairāki sienas skapīši. Loga pretējā sienā ir durvju aila uz dzīvojamo istabu. Pie loga stāv apavu plaukts, uz kura iekārtota gulvieta kaķim.	Lejupielādēt	labot
2015.04.29	Lauma spēlējas ar pogām (ber plastmasas pudelē un laukā), šķirsta grāmatas, spēlējas ar rotaļu traukiem. [...]	Dzīvojamā istaba Istabā ir divi logi, starp kuriem stāv dīvāns, pretī dīvānam - pie pretējās sienas - ir kamīns un podiņu krāsns, blakus kamīnam atrodas drēbju skapis; perpendikulāri dīvānam pie vienas sienas stāv grāmatu skapis, pie pretējās sienas ir liels ovāls galds ar diviem datoriem un diviem krēsliem. Lielāko daļu istabas grīdas klāj paklājs. Istabā atrodas lielākā daļa bērna spēļmantu.	Lejupielādēt	labot
2015.05.19	Lauma spēlējas ar puzzle, pogām, gumijas dinozauriem, kopā lasām pasaku par pogu.	Dzīvojamā istaba Istabā ir divi logi, starp kuriem stāv dīvāns, pretī dīvānam - pie pretējās sienas - ir kamīns un podiņu krāsns, blakus kamīnam atrodas drēbju skapis; perpendikulāri dīvānam pie vienas sienas stāv grāmatu skapis, pie pretējās sienas ir liels ovāls galds ar diviem datoriem un diviem krēsliem. Lielāko daļu istabas grīdas klāj paklājs. Istabā atrodas lielākā daļa bērna spēļmantu.	Lejupielādēt	labot

Ortogrāfiskās transkripcijas principi (1)

- Tiek pierakstīts viss teiktais
 - Vārdu atkārtojumi, pārteikšanās u.tml.
 - Vai pierakstīt arī bērna radītās neartikulētās skaņas?
 - Kā tās atspoguļot?
- Pieturzīmes netiek izmantotas, izņēmums – jautājuma zīme, izsaukuma zīme un punkts
- Bērna izrunātajam vārdam, ja tas neatbilst latviešu valodas ortogrāfijas normām, tā «skaidrojums» tiek likts kvadrātiekvās
 - Piemēram, *aiģīte* [vardīte], *buba* [bumba]
 - Ne vienmēr ir iespējams precīzi atveidot bērnu izrunāto vārdu
 - Lai precīzāk atveidotu bērna runu, tiek izmantoti papildu apzīmējumi latviešu valodas skaņām: *ō* (garais patskanis o) un *uo* (divskanis o)

Ortogrāfiskās transkripcijas principi (2)

Pauzes

- vienkāršas pauzes

(.) (0.35)	Pauze (pārtraukums runas plūsmā) tiek apzīmēta ar punktu, kas likts apaļajās iekavās; pauzēm, kas ilgst vismaz 0,3 s, norādīts laiks.
---------------	---

- aizpildītās pauzes

(āā), (ēē), (khm), (mm), (em), (am) u.c.	Aizpildītas pauzes tiek pierakstītas, izmantojot atbilstošu burtu savirknējumus.
--	--

Ieelpa un izelpa

(h.)	izelpa
(.h)	ieelpa

Ortogrāfiskās transkripcijas principi (3)

Neverbālas skaņas tiek speciāli marķētas.

Neverbāli elementi - saistīti gan ar pašu runātāju, piem., smieklī, klepus, ieelpa, gan ar runātāju nesaistīti neverbāli elementi.

raudas	<r/>; <r> <i>teksts</i> (laiks) </r>
čalošana	 (laiks)
troksnis	<tr/>; <tr> <i>teksts</i> (laiks) </tr>
fizioloģisks troksnis	<ftr/>; <ftr> <i>teksts</i> (laiks) </ftr>

Transkripcijas piemērs

MO_1: **daguns** [deguns]

MO_1_mamma: nē tas ir knābis

MO_1: **gnābs** [knābis]

MO_1_mamma: k pirmais burtiņš ja? knābis

MO_1: (.h) **jā ā uķ**

MO_1_mamma: (.) @ @ @

MO_1: (.) **ķī**

MO_1_mamma: (.) nu putniņam tagad sāpēja jāsamiļo

MO_1_mamma: miļo putniņ

MO_1_mamma: vai ne? (.) kā miļojās?

MO_1_mamma: @ (.) nemīļosi putniņu(.)

MO_1: **viņš laudās** [raudās]

MO_1_mamma: mhm



Problēmas

- Nesaprotama bērna runa
- Neverbālās komunikācijas atspoguļošana ortogrāfiskajā transkripcijā
- Pēc iespējas precīzāka bērnu artikulēto skaņu un vārdu atveide transkripcijā