

## Otrās baltu valodas korpusa sintaktiska anotēšana

*Inga Znotiņa*

*Rīgas Stradiņa universitāte, Valodu centrs, Latvija*

**Ievads.** Lai veicinātu baltu valodu apguves izpēti, tiek veidots otrās baltu valodas apguvēju korpus – elektronisks tekstu vākums, kurā ir iekļauti valodas apguves procesa laikā tapuši teksti. Tekstu autori ir Latvijas un Lietuvas augstākās izglītības iestādēs studējošie: latvieši, kas apgūst lietuvišu valodu, un lietuvieši, kas apgūst latviešu valodu.

Minētais korpus ir anotēts. Tajā ir anotētas vārdu pamatformas, vārdšķiras, kļūdas, kā arī sintaktiskie teikumu veidi. Sintaktiskās anotēšanas pazīmju kopa ir sastādīta tā, lai atbilstu abām korpusā pārstāvētajām baltu valodām – latviešu un lietuviešu valodai – un būtu izmantojama arī gadījumos, kad apguvēju rakstītais teksts ir grūti saprotams vai kļūdaini, aprūtinot teikuma tipa noteikšanu.

**Darba mērķis, materiāls un metodes.** Šī darba mērķis ir raksturot un pamatot otrās baltu valodas korpusa anotēšanai izveidoto sintaktiskās anotēšanas pazīmju kopu. Tās izveidē izmantotas esošo baltu valodu korpusu anotēšanā izmantotās sintaktisko pazīmju kopas, kā arī akadēmiskas publikācijas par baltu valodu sintakses jautājumiem.

**Rezultāti.** Korpusu anotējot sintaktiski, tajā iekļautie teksti ir sadalīti izteikumos un izsacījumos, un izteikumi ir iedalīti vairākās pamatgrupās pēc tā, kādam teikuma modelim tie atbilst.

Sintaktiskajā anotējumā izmantota šāda klasifikācija:

VNT – vienkāršs nepaplašināts teikums. Šajā grupā ir tādi izteikumi, kas atbilst vienkārša teikuma minimālajam teikuma modelim – t. sk. tad, ja daļa no gramatiskā centra ir reducēta;

VPT – vienkāršs paplašināts teikums. Šajā grupā ir tādi izteikumi, kas atbilst vienkārša teikuma paplašinātajam teikuma modelim – t. sk. tad, ja daļa no gramatiskā centra vai viss gramatiskais centrs ir reducēts;

SPT – salikts pakārtots teikums. Šajā grupā ietilpst izteikumi, kuros ir vairākas daļas, starp kurām ir pakārtojuma attiecības – t. sk. tad, ja vienā vai vairākās daļās gramatiskais centrs vai daļa no tā ir reducēta;

SST – salikts sakārtots teikums. Šajā grupā ietilpst izteikumi, kuros ir vairākas daļas, starp kurām ir sakārtojuma attiecības – t. sk. tad, ja vienā vai vairākās daļās gramatiskais centrs vai daļa no tā ir reducēta;

SJT – salikts jaukts teikums. Šajā grupā ir tādi izteikumi, kuros ir vairākas daļas, starp kurām ir gan pakārtojuma, gan sakārtojuma attiecības – t. sk. tad, ja vienā vai vairākās daļās gramatiskais centrs vai daļa no tā ir reducēta;

IS – izsacījums. Gadījumā, ja teksta vienībā nav saskatāmas teikuma modeļa pazīmes, tā tiek ieskaitīta šajā grupā;

NT – neskaidri veidots izteikums. Šajā grupā tiek ieskaitītas tās teksta vienības, kuras autors acīmredzami ir iecerējis veidot kā izteikumus, taču izteikta neatbilstība mērķvalodas normām neļauj noteikt teikuma veidu.

Parcelācijas gadījumā viena teikuma struktūra tiek dalīta vairākos izteikumos, līdz ar to acīmredzamas parcelācijas gadījumā parcelētie izteikumi tiek anotēti kā viens teikums, nenodalot šos izteikumus atsevišķi. Gadījumos, kad rodas šaubas, vai izteikums ir parcelāts vai atsevišķa teikuma modeļa daļēja realizācija, tas tiek uzskatīts par atsevišķa teikuma modeļa daļēju realizāciju.

**Secinājumi.** Līdz šim valodas apguvēju korpusos lietotajās pazīmju kopās ne vienmēr ir skaidrs, kas ir darāms redukcijas vai parcelācijas gadījumos. Otrās baltu valodas apguvēju korpusā izmantotā klasifikācija šādus gadījumus paredz, kā arī definē anotētāja rīcību gadījumos, kad nav saskatāmas teikuma modeļa pazīmes vai izteikums neatbilst mērķa valodas sintakses normām. Izveidotā pazīmju kopa var tikt izmantota ne tikai valodas apguvēju korpusu, bet arī citu veidu korpusu anotēšanai.