

Latviešu valodas FrameNet izveide

Gunta Nešpore-Bērzkalne, Baiba Saulīte, Laura Rituma
LU Matemātikas un informātikas institūts

“Daudzslāņu valodas resursu kopa teksta semantiskai analīzei un sintēzei latviešu valodā”, proj. nr.
1.1.1.1/16/A/219

“No abstraktās nozīmes reprezentācijas līdz dabiskam teikumam un saistītam tekstam”, proj. nr.
Nr. 1.1.1.2/VIAA/1/16/188

FrameNet

Freimu semantika – leksiskās nozīmes teorija, pēc kuras vārda nozīme var tikt saprasta tikai semantiskā situācijā (freimā), kurā iesaistīti noteikti situācijas dalībnieki un apstākļi.

Self_motion

Definition:

The **Self_mover**, a living being, moves under its own direction along a **Path**. Alternatively or in addition to **Path**, an **Area**, **Direction**, **Source**, or **Goal** for the movement may be mentioned.
She WALKED along the road for a while.

Many of the lexical units in this frame can also describe the motion of vehicles (e.g., as external arguments). We treat these as belonging in this frame.

The cars SCOOTED slowly towards the intersection.

Self_motion most prototypically involves individuals moving under their own power by means of their bodies. Many words also specify the manner of motion (*swim, walk*). This frame is impossible, necessary, or unspecified. Lexical units that involve separate vehicles are associated with FEs that are not appropriate for the more general case of motion, so they are placed

FEs:

Core:

Area [Area]
Semantic Type: Location

Area is used for expressions which describe a general area in which motion takes place when the motion is understood to be irregular and not to consist of a path. It is also used for a path meaning with a non-motion target, since these should be annotated with the **Place** FE.

The mouse **SCURRIED about**.

Stop **RUNNING around the room** and sit down!

Direction [dir]
Excludes: Area

The direction that the **Self_mover** heads in during the motion.

You should **WALK south** about a block .

Goal [Goal]
Semantic Type: Goal
Excludes: Area

Goal is used for any expression which tells where the **Self_mover** ends up as a result of the motion.

The children **SKIPPED into the park**.

Some particles imply the existence of a **Goal** which is understood in the context of utterance.

The principal **WALKED over** and sat down.

A dog **RAN up** and licked our hands.

WebAnno

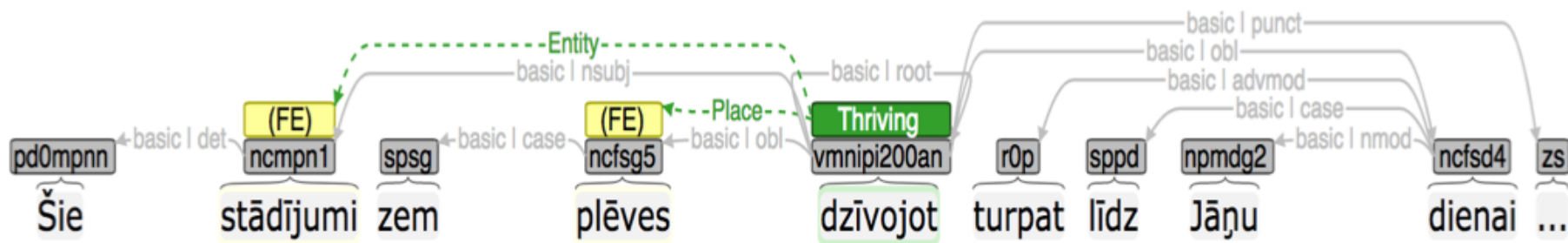
- Freimi tiek marķēti, izmantojot platformu *WebAnno*, kas ļauj semantisko informāciju kombinēt ar sintakses koku.
- Latviešu valodas verbi tiek sastatīti ar atbilstošajiem angļu valodas *FrameNet* freimiem un marķēti, pēc iespējas ievērojot *FrameNet* nozīmes un semantisko lomu definīcijas.
- Freimi tiek marķēti konkordances skatā, jo, redzot vienkopus dažādus verba lietojuma piemērus, var ātrāk un precīzāk piemeklēt piemērotāko *FrameNet* freimu.

WebAnno

- 1 Viņa dzīvoklis bija aptīrīts, un tagad viņš **(FE)** **(FE)** **Residence** dzīvoja pie **(FE)** **(FE)** vecākiem šķūnī.
Time-Resident *Co_resident* *Location*
- 2 Šie stādījumi zem plēves **(FE)** **(FE)** **Thriving** dzīvojt turpat līdz Jāņu dienai...
Entity *Place*
- 3 Kā izklūtu no situācijas, kā **(FE)** **Manner_of_life** dzīvotu tālāk?
Manner
- 4 Un parasti jau nākas atzīt – lai kā, bet es **(FE)** **Dead_or_alive** dzīvotu tālāk.
Protagonist
- 5 Vajadzēja **Manner_of_life** dzīvot **(FE)** ārkārtīgi taupīgi, tāpēc gadījās arī tā, ka mums, bērniem,
Manner
- 6 Ja **Residence** dzīvotu **(FE)** **(FE)** Amerikā, viņa būtu karsējmeitene un visi zēni beigtos nost viņa
Resident *Location*
- 7 Un tad **(FE)** tu nevari ne aizmigt, ne **Subsisting** dzīvot bez **(FE)** .
Entity *Support*

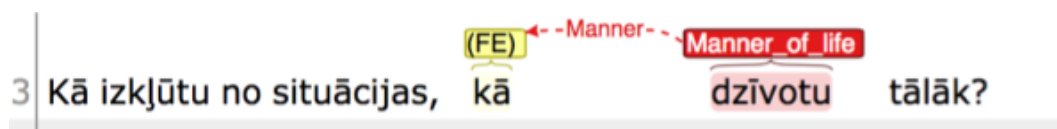
Kādus datus marķējam

- Tiek marķēti biežākie latviešu valodas verbi un no šiem verbiem atvasinātie lietvārdi (tiek izmantots TOP 2000 verbu saraksts, kas iegūts no LVK2013, dati tiks precizēti pēc LVK2018 publicēšanas)
- Par pamatu semantiskajam marķējumam tiek izmantoti sintakses koki: no sintaktiski marķētajiem datiem tiek atlasīti teikumi, kuros parādās freimu marķēšanai izvēlētais verbs



Kādu informāciju marķējam

- Tiek marķētas centrālās semantiskās lomas (freimu elementi) + LAIKS un VIETA
- Tiek marķēts tikai tas, kas parādās tekstā (atšķirībā no angļu FrameNet, kur marķētas arī tekstā neminētās semantiskās lomas)



- Ir iespēja marķēt vārdu savienojumus kā vienu vienību, piekārtojot freimu visam vārdu savienojumam, nevis atsevišķam verbam:
 - *iet bojā* 'mirt'
 - *nākt klajā* 'tikt publicētam'