



TARTU ÜLIKOOL

Pronominaalsete viitesuhete analüüs asendussõnade suhtes käsitsi märgendatud korpuses

Linda Freienthal

04.06.2018



Pronominaalne viitesuhe

Linda on üliõpilane. **Ta** elab Tartus.





Analüüsi vajalikkus ja eesmärk

- Töö eesmärk on uurida pronoomenite *mina-meie, sina-teie, tema-nemad, kes, mis* ja *see-need* viitesuhteid ning leida tugevaid seoseid ja seaduspärasusi, mida saab automaatse lahendaja koostamisel ära kasutada.
- Oluline abivahend paljudele keeletehnoloogilistele vahenditele.
- Eestis puudub laialt kasutatav asendussõnade automaatne lahendaja.

Anafooride (asendussõnade) lahendamismeetodeid

- Reeglipõhine (eelistav ja välistav) vs masinõpe
 - Teadmistel põhinevad vs teadmistevaesed
- (10) John peitis Billi võtmed ära. *Ta* oli purjus. (Lappin 2005: 7)
- (11) John peitis Billi võtmed ära. *Ta* mängis talle vingerpussi. (Lappin 2005: 7)



Eestis seni tehtud töö

- Tiina Puolakainen (2015) „Anaphora resolution experiment with CG rules“.
- Pilleriin Mutso (2008) „Knowledge-poor Anaphora Resolution System for Estonian“.

Asendussõnade suhtes käsitsi märgendatud korpus

- Ajalehetekstid
- ca 107000 tekstisõna
- Märgendatud projekti „Sihipärane süntaks korpuse jaoks“ raames
- Korpuses on märgendatud pronoomenid *mina-meie, sina-teie, tema-nemad, kes, mis, see-need*

Pronoomen	Mitu korda esines korpuses	Viitealuste arv
<i>see-need</i>	1489	1220
<i>tema-nemad</i>	922	958
<i>mis</i>	716	582
<i>kes</i>	337	299
<i>sina-teie</i>	135	84

Korpuse väljund

"<Hansapanga>"

"Hansa_pank" L0 S prop sg gen @NN> #8->9

"<tehinguid>"

"tehing" Lid S com pl part @OBJ #9->7 {Viitealus}

"<,>"

"," Z Com #10->10

"<mis>"

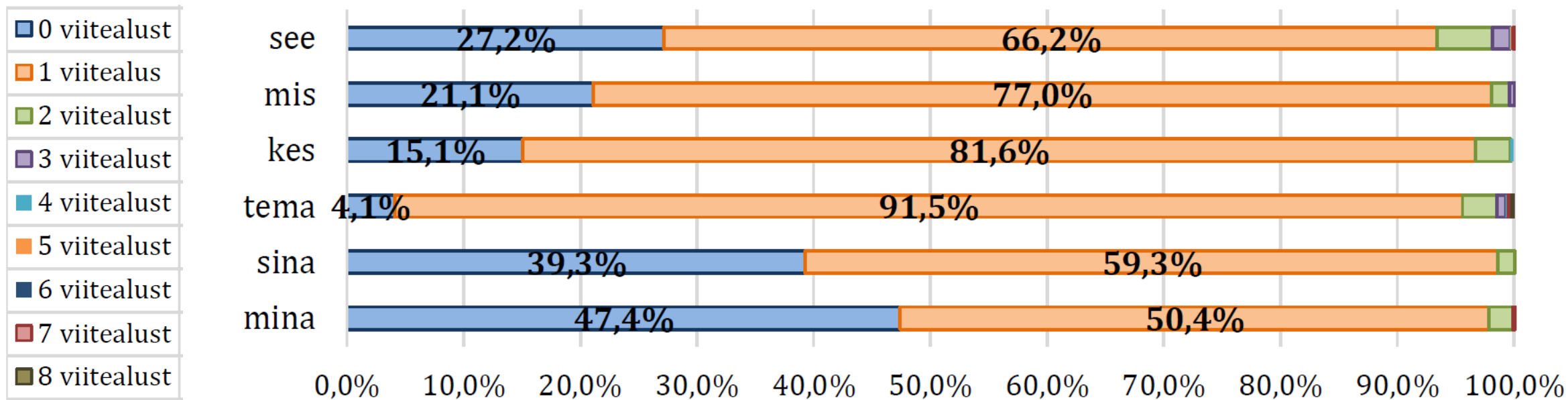
"mis" L0 P inter rel pl nom @SUBJ #11->12 {Pronoomen} {Coref:2.9}

"<toimusid>"

"toimu" Lsid V main indic impf ps3 pl ps af @FMV #12->9

Töö tulemused (1)

- Asendussõnade viitealuste arv.



Joonis. Erinevate viitealuste arvu osakaalud pronoomenite kaupa.

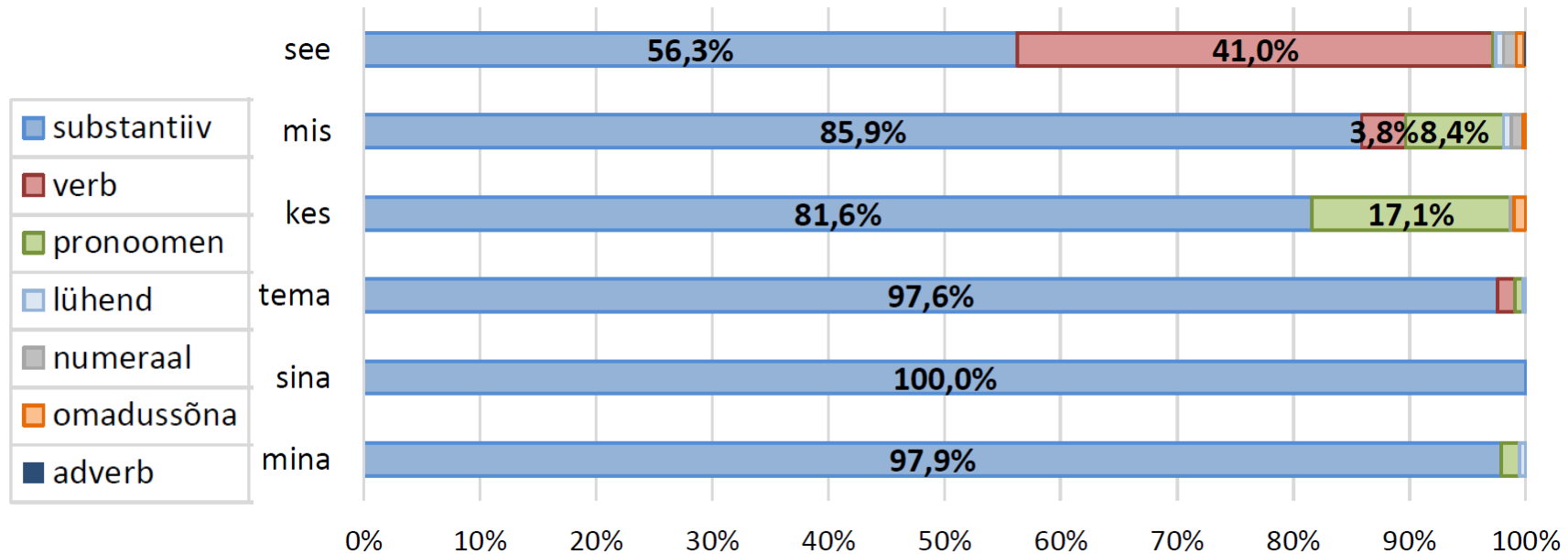
Mitme viitealusega asendussõna näide

- Jüri, Mari ja mina väitsime: „**Meie** pole seda teinud!“





Töö tulemused (3)

- Viitealuste sõnaliigiline jaotus



Joonis. Viitealuste sõnaliigiline jaotus asendussõnade kaupa

Viitealuste näiteid (1)

- Soovitus neile, **kellel** on kalduvus koduteel kaduma minna.

- Maapind liigub Liibanoni ranniku lähedal, **mis** tähendab...


Töö tulemused (4)

- Substantiivsete kandidaatide korral:
 - Asendussõnade *mina-meie* ja *sina-teie* puhul võib eelistada pärisnimesid (78,1% ja 81%).
 - Asendussõnade *mis* ja *see-need* puhul tuleb eelistada üldnimesid (98,1% ja 95,2%).
- Pronoomenitest kandidaatide korral:
 - Asendussõna *mis* puhul võib eelistada näitavaid asesõnu (85,7%).
- Verbidest kandidaatide korral (*mis*, *tema-nemad* ja *see-need*):
 - Peab eelistama põhiverbe.
 - Peab eelistama kindlat kõneviisi või *da*-infinitiivi.

Viitealuste näiteid (2)

- Tundkem rõõmu **selle** üle, mida ta suutis meile pärandada.



- See, **mida** kunstiteraapias kujutatakse, on metafoor.





Töö tulemused (5)

- Asendussõna ja viitealuse arvuline ühildumine.
 - Asendussõnade *tema-nemad* ja *mina-meie* puhul võib eelistada samas arvus kandidaate (81,7% ja 82%).
 - Asendussõnade *kes* ja *mis* puhul võib eelistada või peab eelistama samas arvus kandidaate (90,6% ja 91,2%).
 - *Teie* viitab suure tõenäosusega ühele ainsuses viitealusele (viisaka pöördumisvormina).
- Asendussõna ja viitealuse süntaktiline ühildumine.
 - Tugevaid seoseid ei leidunud.
- Asendussõna ja viitealuse käändeline ühildumine.
 - Enimlevinud on nimetav kääne. Seda võib eelistada asendussõnade *mina-meie* ja *sina-teie* puhul (78,8% ja 83,3%).

Töö tulemused (6)

- Viitealuse kaugus asendussõnast sõnedes.
 - Asendussõnade *kes* ja *mis* puhul tuleb eelistada kandidaate, mis on oma asendussõnast 2-7 sõnet eespool.
 - Asendussõnade *kes* ja *mis* puhul võib eelistada kandidaate, mis on oma asendussõnast kaks sõnet eespool (84% ja 79%, relatiivlaused).
 - Asendussõna *tema-nemad* puhul võib eelistada kandidaate, mis on vahemikus 1-27 sõnet asendussõnast eespool (83%).
- Viitealuse kaugus asendussõnast lausetes.
 - Asendussõnadele *kes* ja *mis* tuleb otsida viitealuseid samast lausest.
 - Asendussõnale *see-need* samas või sellele eelnevas lauses (93,6%).
 - Asendussõnale *tema-nemad* samas või kuni kaks lauset eespool (90,1%).

Relatiivlause

- Taanis kohtus ta **kahe** tudengiga, kes kutsusid ta väikesele peole ja pakkusid öömaja.





TARTU ÜLIKOOL



TARTU ÜLIKOOL

TARTU ÜLIKOOL

UNIVERSITY OF TARTU