Lauseliikmed andmepuus

Uuritud lauseid üle kümne tuhande, varem sobivatest lausetest kinnitatud reegleid üle kuuesaja. Uuritud lausetest reeglitele vastavusi 3177 – sellise hulga pilguga haaramiseks kuluvad abivahendid marjaks ära. Üheks võimaluseks on paigutada reeglid ning nendele vastavused andmepuusse.

Järgnevalt töö kirjeldus näite põhjal.

Näitlauseks:

(1) *Loomulikult ei tohi tunnistada, et sul endal ka need olemas on.*

Parseri väljund:

"<s>"

"<Loomulikult>"

"loomulikult" L0 D cap @ADVL #1->3

"<ei>"

"ei" L0 V aux neg cap @NEG #2->3

"<tohi>"

"tohti" L0 V main indic pres ps neg cap <FinV> <Intr> @FMV #3->3

"<tunnistada>"

"tunnista" Lda V main inf cap <NGP-P> @IMV #4->4

"<,>"

"," Z Com CLB #5->5

"<et>"

"et" L0 J crd cap @J #6->12

"et" L0 J sub cap @J #6->12

"<sul>"

"sina" Ll P pers ps2 sg ad cap @NN> @ADVL #7->7

"<endal>"

"ise" Ll P pos det refl sg ad cap @<NN #8->8

"<ka>"

"ka" L0 D cap @ADVL #9->12

"<need>"

"see" Ld P dem pl nom cap @SUBJ #10->12

"<olemas>"

"ole" Lmas V main sup ps in cap <Intr> @ADVL #11->12

"<on>"

"ole" L0 V main indic pres ps3 pl ps af cap <FinV> <Intr> @FMV #12->12

"<.>"

"." Z Fst #13->13

"</s>"

Pärast eemaldamisi jäi reeglisse alles järgnevus

'@NEG', '@FMV', '@IMV'

Vaadatakse ainult esimest osalauset osalause piirini (CLB), milleks siin on koma. Siinsest neljast sõnast arvestatakse reegli kontrollimisel kolme. Sõna „loomulikult“ ei mõjuta reeglitele vastavust, sest ta on sellises loetelus koos veel paarikümne ballastsõnaga nagu nt. „arvatavasti“, „tõenäoliselt“, „niisiis“.

’@NEG’-iga algavaid reegleid ja reeglitele vastavaid lauseid on vaadeldud tekstis vähim, siitkaudu saab puu ehitamise etapid lihtsaimini läbi teha. Uuritud tekstis oli eitusega algavaid lauseid kuus:

1. *Loomulikult ei (@NEG) tohi (@FMV) tunnistada (@IMV) , et sul endal ka need olemas on .* ['@NEG', '@FMV', '@IMV']
2. *Seega ei (@NEG) saa (@FMV) me (@SUB) oma läkitust (@OBJ) kodeerida (@IMV) nii , et see oleks mõistetav .* ['@NEG', '@FMV', '@SUBJ', '@OBJ', '@IMV']
3. *Siis ei (@NEG) lausu (@FMV) ma (@SUBJ) enam ühtki sõna (@OBJ) ega mõtle enam millelegi , kõik vajub enneaimamatusse õndsusemerre .* ['@NEG', '@FMV', '@SUBJ', '@OBJ']
4. *Ei (@NEG) ole (@FMV) enam millest (@ADVL) rääkida (@SUBJ) , ta tahab maale saada , ehk ta seda küll ei ütle .* ['@NEG', '@FMV', '@ADVL', '@SUBJ']
5. Ei (@NEG) ole (@FMV) jahu (@SUBJ) põskedel (@ADVL) ja huuled on loomulikult värsked . ['@NEG', '@FMV', '@SUBJ', '@ADVL']
6. Siiski ei (@NEG) ole (@FMV) ma (@SUBJ) teie peale (@ADVL) väga tige , et te mu üles ajasite . ['@NEG', '@FMV', '@SUBJ', '@ADVL']

Et kaks viimast reeglit ['@NEG', '@FMV', '@SUBJ', '@ADVL'] on ühesugused, saab nad kokku grupeerida.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| 1 | '@NEG' | '@FMV' | '@IMV' |  |  |
| 1 | '@NEG' | '@FMV' | '@SUBJ' | '@OBJ' | '@IMV' |
| 1 | '@NEG' | '@FMV' | '@SUBJ' | '@OBJ' |  |
| 1 | '@NEG' | '@FMV' | '@ADVL' | '@SUBJ' |  |
| 2 | '@NEG' | '@FMV' | '@SUBJ' | '@ADVL' |  |

Küllalt paljud algusmärgendite järjestused reeglites aga kattuvad, seetõttu andmepuuna saab nad lühemalt kirja panna nii, et ei pea samu märgendeid kordama. Samuti on võimalik iga märgendi juurde märkida, mitme analüüsitud lause juures see reeglis vastaval kohal ette tuli. Andmetest paistab välja, et kõigis @NEG-iga alanud lausetes on teiseks liikmeks @FMV. Järgneva liikmena on levinuim @SUBJ. Levinumate järgnevuste mugavamaks leidmiseks, sorteeriti see haru puus ülespoole.

Öeldistäitega algavaid lauseid leiti kirjandustekstidest 12. Sellistele lausetele vastavatest reeglitest koostatud puu on juba mõnevõrra põhjalikum.

Levinuimaks järgnevuseks nagu paistab

'@PRD', '@SUBJ', '@FMV'

Ning näitlause selle järgnevuse kohta:

1. Ja saatanlik (@PRD) nagu ta (@SUBJ) oli (@FMV), ei suutnud ta jätta puusi diskreetselt nõksutamata

Sageduselt järgmine sõnajärjestruktuur on @FCV-ga algavad laused.

Näiteks:

1. *Olin* (@FCV) *tihti* (@ADVL) *tundnud* (@IMV), *et see suvi oli mind välja valinud, kutsunud täitma prohvetlikku ülesannet.*

Sihitisega algavate lausete puu:

Nt

1. Nad (@OBJ) tunneb (@FMV) siin (@ADVL) kohe ära

'@OBJ', '@FMV', '@ADVL'

@FMV-öeldisega algavate lausete reeglite puu

näitena

1. Jäin (@FMV) seejärel (@ADVL) kohe (@ADVL) magama

'@FMV', '@ADVL', '@ADVL'

Edasi kasvab puu juba nõnda suureks, et tuleb mahutamise tarbeks lihtsamalt ja väiksemalt joonistada.

Määrusega algavaid lauseid leiti tekstist 892

'@ADVL' 892

'@FMV' 500

'@SUBJ' 207

'@ADVL' 110

'@PRD' 5

'@ADVL' 2

'@OBJ' 2

'@OBJ' 22

'@ADVL' 19

'@ADVL' 7

'@ADVL' 2

'@IMV' 2

'@IMV' 2

'@PRD' 10

'@IMV' 1

'@ADVL' 142

'@ADVL' 32

'@ADVL' 19

'@SUBJ' 6

'@SUBJ' 13

'@SUBJ' 30

'@OBJ' 2

'@OBJ' 12

'@IMV' 2

'@OBJ' 42

'@ADVL' 6

'@ADVL' 6

'@SUBJ' 4

'@PRD' 17

'@SUBJ' 3

' P ' 16

'@OBJ' 16

'D cap @ADVL' 15

'@SUBJ' 15

' S ' 4

'@OBJ' 4

'@SUBJ' 128

'@FMV' 103

'@ADVL' 53

'@ADVL' 7

'@ADVL' 7

'@OBJ' 6

'@IMV' 1

'@OBJ' 7

'@ADVL' 7

'@IMV' 1

'@NEG' 17

'@FMV' 17

'@OBJ' 9

'@ADVL' 8

'@OBJ' 2

'@ADVL' 1

'@OBJ' 1

'@IMV' 1

'@SUBJ' 1

'@IMV' 3

'@NEG' 2

'@FMV' 2

'@FMV' 1

'@FCV' 2

'@ADVL' 1

'@IMV' 1

'@OBJ' 1

'@IMV' 1

'@ADVL' 2

'@ADVL' 2

'@FMV' 1

'@NEG' 1

'@FMV' 1

'@OBJ' 1

'@NEG' 1

'@FMV' 1

'@ADVL' 122

'@FMV' 82

'@SUBJ' 39

'@ADVL' 16

'@ADVL' 20

'@ADVL' 7

'@SUBJ' 3

'@OBJ' 1

'@SUBJ' 1

'@ADVL' 22

'@FMV' 16

'@ADVL' 9

'@SUBJ' 4

'@SUBJ' 6

'@ADVL' 4

'@OBJ' 2

'@ADVL' 2

'@OBJ' 1

'@ADVL' 1

'@SUBJ' 1

'@ADVL' 4

'@FMV' 4

'@IMV' 1

'@SUBJ' 1

'@FMV' 1

'@FCV' 5

'@IMV' 4

'@OBJ' 4

'@SUBJ' 1

'@ADVL' 1

'@IMV' 1

'@IMV' 1

'@SUBJ' 5

'@OBJ' 3

'@FMV' 2

'@NEG' 1

'@FMV' 1

'@FMV' 2

'@ADVL' 2

'@ADVL' 2

'@OBJ' 1

'@NEG' 5

'@FMV' 5

'@OBJ' 3

'@FMV' 3

'@FCV' 73

'@SUBJ' 36

'@ADVL' 14

'@IMV' 3

'@IMV' 12

'@ADVL' 4

'@OBJ' 5

'@IMV' 2

'@IMV' 24

'@SUBJ' 6

'@OBJ' 4

'@ADVL' 1

'@OBJ' 1

'@IMV' 1

'@ADVL' 1

'@ADVL' 11

'@IMV' 7

'@ADVL' 3

'@IMV' 3

'@OBJ' 1

'@ADVL' 1

'@IMV' 1

'@OBJ' 2

'@IMV' 1

'@OBJ' 1

'@ADVL' 1

'@ADVL' 1

'@ADVL' 1

'@ADVL' 1

'@IMV' 1

'@SUBJ' 1

'@NEG' 37

'@FMV' 37

'@ADVL' 12

'@SUBJ' 4

'@SUBJ' 9

'@ADVL' 5

'@OBJ' 2

'@ADVL' 1

'@OBJ' 1

'@OBJ' 4

' P ' 5

'ad' 1

'@ADVL' 1

'@IMV' 1

'@SUBJ' 1

'@OBJ' 1

'@ADVL' 1

'@SUBJ' 1

'@IMV' 14

'@OBJ' 14

'@FMV' 7

'@SUBJ' 3

'@ADVL' 1

'@ADVL' 1

'@NEG' 3

'@FMV' 3

'@ADVL' 2

'@ADVL' 1

'@ADVL' 1

'@FMV' 1

'@FMV' 1

'@SUBJ' 1

'@FCV' 1

'@SUBJ' 1

'@IMV' 1

'@SUBJ' 1

'@NEG' 1

'@FMV' 1

'Z Com' 4

'@SUBJ' 4

'@FMV' 4

'@ADVL' 4

näitena

Esimest korda (@ADVL) satub (@FMV) ta (@SUBJ) silme ette

'@ADVL', '@FMV', '@SUBJ'

Ja suurim, aluse ehk subjektiga algavate lausereeglite puu:

'@SUBJ' 1642

'@FMV' 1310

'@ADVL' 677

'@ADVL' 35

'@ADVL' 19

'@ADVL' 11

'@OBJ' 6

'@PRD' 2

'@PRD' 10

'@ADVL' 1

'@OBJ' 6

'@ADVL' 6

'@OBJ' 21

'@ADVL' 21

'@OBJ' 2

'@PRD' 9

'@ADVL' 9

'@PRD' 141

'@ADVL' 11

' S ' 115

'D cap @ADVL' 27

'@OBJ' 1

'@OBJ' 88

'@ADVL' 65

'@ADVL' 3

'@ADVL' 3

'@ADVL' 2

'@IMV' 1

' S ' 8

'D cap @ADVL' 6

'@OBJ' 1

' P ' 4

'@IMV' 3

' P ' 56

'D cap @ADVL' 21

'@OBJ' 2

'@IMV' 8

'@PRD' 2

'Z Fst' 1

'@FCV' 132

'@IMV' 64

'@ADVL' 28

'@ADVL' 11

'@ADVL' 5

'@OBJ' 2

' S ' 9

' P ' 2

'@ADVL' 48

'@IMV' 37

'@OBJ' 7

'@ADVL' 6

'@ADVL' 5

'@IMV' 3

'@ADVL' 2

'@ADVL' 1

'@ADVL' 1

'@IMV' 1

'@IMV' 1

'@IMV' 1

'@ADVL' 1

'@OBJ' 5

'@ADVL' 3

'@IMV' 3

'@IMV' 2

'@OBJ' 10

'@ADVL' 5

'@IMV' 4

'@OBJ' 1

'@ADVL' 1

'@IMV' 1

'@IMV' 4

'@ADVL' 3

'D cap @ADVL' 1

'@IMV' 1

' S ' 7

'D cap @ADVL' 3

'@IMV' 3

'@IMV' 2

' P ' 3

'@IMV' 2

'D cap @ADVL' 1

'@IMV' 1

'@NEG' 129

'@FMV' 119

'@OBJ' 19

'@ADVL' 11

'@ADVL' 13

'@ADVL' 12

'@ADVL' 5

'@OBJ' 1

'@PRD' 1

' P ' 12

'@OBJ' 3

' S ' 8

'@OBJ' 2

'@IMV' 3

'@FCV' 10

'@IMV' 4

'@ADVL' 4

'@IMV' 2

'@OBJ' 2

'@IMV' 2

'@OBJ' 2

'@ADVL' 2

'@IMV' 2

'@ADVL' 46

'@FMV' 17

'@ADVL' 10

'@OBJ' 5

'@ADVL' 2

'@PRD' 2

'@ADVL' 17

'@FMV' 11

'@ADVL' 2

'@PRD' 1

'@ADVL' 4

'@FMV' 3

'@ADVL' 1

'@FMV' 1

'@FCV' 1

'@ADVL' 1

'@ADVL' 1

'@IMV' 1

'@NEG' 1

'@FMV' 1

'@NEG' 5

'@FMV' 5

'@ADVL' 3

'@FCV' 2

'@ADVL' 2

'@IMV' 2

'@IMV' 2

'@OBJ' 2

'@FMV' 1

'@NEG' 1

'@FMV' 1

'@SUBJ' 1

'@FMV' 1

'@PRD' 1

'@OBJ' 14

'@FMV' 11

'@ADVL' 2

'@ADVL' 2

'@ADVL' 3

'@FMV' 3

'indic pres' 6

' P ' 3

'@OBJ' 3

' S ' 3

'@OBJ' 3

'indic impf' 3

' S ' 2

'@OBJ' 2

' P ' 1

'@OBJ' 1

'@PRD' 2

'@NEG' 2

'@FMV' 2

'@IMV' 1

Näitena

Mu mõrsja (@SUBJ) istub (@FMV) salongi laua ääres (@ADVL)

'@SUBJ', '@FMV', '@ADVL'

Kogu leitud reeglite puud korraga silma ette tekitada pole mõistlik – seda nagunii ei haara. Aga puudeosa lõpetuseks on välja toodud levinuimate järgnevuste puu – siin on nähtavaks jäetud märgendid, mis uuritud tekstis kasutusel vähemalt neljakümnes lauses.

Levinuim lausekuju siis @SUBJ, @FMV, @ADVL ehk näitena lauses

Ta (@SUBJ) ärkab (@FMV) hommikul (@ADVL).

Esineb kokku 677 korda 3177-st reeglile vastanud lausest ehk reeglile vastanutest ligikaudu igal viiendal juhul.