

TEKSTIDE LEKSIKAAISE SEOSE MÕOTMISEST

=====

J. Tuldava

Artiklis vaadeldakse tekstide leksikaalse seose mõt-
mise üldisi põhimõtteid ja esitatakse eksperimendi tulemu-
sed 20 eesti ilukirjandusteksti (valimi) võrdlemise kohta
korrelatsioon- ja faktoranalüüsi abil.

Üldised põhimõtted. Tekstide leksikaalse seose (läheduse) mõõtmine võib osutada vajalikuks individuaalsete või funktsionaalsete stiilide uurimisel, sagedussõnastike võrdlemisel ja miinimum-sõnastike koostamisel, tekstide auto-
maatklassifitseerimisel jm. Seepärast on küsimus leksikaalse seose adekvaatsetest mõõtudest juba pikemat aega olnud nii keelestatistikute, kui ka informaatikaga tegelevate teadlaste huvi keskpunktis (vt. näit. Вахабов, 1970; Дядюк, 1973; Маршакова, 1974).

Leksikaalse seose kvantitatiivsel uurimisel on mitu eri aspekti. Alljärgnevalt vaatleme neist tähtsamaid.

1. Tekstide sõnavara võib võrrelda ja leksikaalset seost mõõta nii sõnastiku kui teksti tasandil. Esimesel juhul kõrvutatakse ainult sõnastikke, s.o. vaadeldavates tekstides esinevaid sõnu (või sõnavorme), arvestamata nende sagedusi tekstis. Teisel juhul arvestatakse ka sagedusandmeid, s.t. eri sõnade osakaalu ühes või teises tekstis.

2. Leksikaalset seost (lähedust resp. kaugust) võib mõista kahte moodi. Esimesel juhul peetakse suurimaks läheduseks tekstide sõnavara identsus, s.o. sõnavara täielikku kokkulangemist sõnastiku või teksti tasandil. Vastavad leksikaalse seose mõõdud konstrueeritakse, lähtudes identsuse põhimõttest. Kõige sagedamini kasutatakse sel juhul spetsiaalseid leksikaalse läheduse indekseid (näit. Rosengren, 1971; Тулдава, 1974) või korrelatsioonanalüüsi (astakkorrelatsiooni- või lineaarkorrelatsioonikordajaid kasutavad näit. R. Frumkina, vt. Фрумкина, 1964; A. Kolgušin, vt. Колгушин, 1970; S. Kļaviņa, vt. Кļавиня, 1976; sõltuvuskordajaid on kasutanud allakirjutanu, vt. Tuldava, 1974).

Teisel juhul lähtutakse leksikaalse seose mõõtmisel hoopis teisest seisukohast, nimelt tekstide võrdlemisel ei nõtuta maksimaalse mõõdnuna sõnastike identsust, vaid sõnade allsageduste optimaalset jaotumust sel puhul, kui võrreldavad tekstid eelnevalt ühendada ja seejärel juhuslikul teel uuesti kaheks jagada (lähemalt vt. Тумдава, 1972). See meetlus lubab kindlaks määrata tekstide läheduse vastavalt sõnastike h o m o g e e n s u s e (mitte identsuse) astmele.

3. Sõnastike ja sagedussõnastike võrdlemisel on olulise tähtsusega küsimus sellest, m i d a j a m i l m ä ä r a l võrrelda. Võib kõrvutada ja vastavad lähedusindeksid arvutada kogu sõnastiku ulatuses, kuid võib ka teha valiku, võttes vaatluse alla ainult osa sõnastikust, näit. teatud sõnaklassi kuuluvad sõnad (ainult nimisõnad või omadussõnad, nn. struktuurisõnad või täistähenduslikud sõnad), eri pikkusega sõnad (näit. ühesilbilised sõnad), kõrg- või madalsagedusega sõnad jne. Võrreldavate sõnade valik oleneb töö eesmärgist ning teoreetilistest kaalutlustest selle kohta, milliseid sõnu pidada antud ülesande lahendamisel olulisteks. Informaatika-alastes töödes võetakse tavaliselt aluseks nn. võtmesõnad (deskriptorid), mis võivad olla nimisõnad või ka nimisõnaühendid, kusjuures tekstide lähedust või kaugust määravateks, nn. "diagnoosivateks" sõnadeks peetakse sageli harva esinevaid deskriptoreid (vt. Маршакова, 1974). Stilostatistilistes uurimustes on võrreldud tekste nimisõnade esinemuse järgi (Yule, 1944), omadussõnade alusel (Шайкевич, 1968), kõrgsagedustsooni kuuluvate sõnade järgi (Каджануни, 1970; Дарчук, 1975), kesksooni sõnade järgi (Krallmann, 1966) jne.

4. Tekste võib võrrelda t ä i e s u l a t u s e s või o s a d e (valimite) p õ h j a l. Ilukirjanduslike tekstide puhul võib vaadelda eraldi autori- ja tegelaskõnet. Uurida võib nähtusi teosesiseselt veelgi diferentseeritumalt, näit. võrreldes eri tegelaste kõnes esinevat sõnavara (vt. Muller, 1968), autorikõne eri liike ühes ja samas teoses - jutustust, kirjeldust, sisemonoloogi jm. (vt. Шкловер, 1974).

5. Tekstide sõnavara otsese võrdlemise kõrval võib kõrvutavat analüüsi teostada mingi e t a l o n i vahendu-

sel. Etaloniks võib olla allkeele põhisõnavara, paljude võrreldavate tekstide sõnavara summa või ühisosa, nn. "nullstiili" (Пефеденос, 1972) põhimõttel konstrueeritud sõnastik jpm.

6. Leksikaalse seose mõttmisel võib piirduda ainult suhtelise läheduse või kauguse hinnanguga või nõuda ka statistilise olulisuse määramist vastavalt matemaatilise statistika ja tõenäosusteooria reeglitele.

Õeldule võib lisada, et tekstide leksikaalse seose mõõtmine võimaldab vastavalt eksperimendi iseloomule hinnata korruga või eraldi tekstide temaatilist, žanrilist, semantilist jm. seost. Tavalise leksikaalse seose mõõtmise kõrval mõõdetakse tänapäeval ka spetsiaalselt "semantilist lähedust", näit. kontent-analüüsi vahendusel saadud semantiliste üksuste ehk nn. kategooriate põhjal (Laffal, 1969) või lausete "semantilise seotuse" näitajate abil (Кроподоцько, 1973). Kui on võimalik esitada teksti sisu mingisuguste täpselt piiritletud semantiliste kategooriate abil, siis saab tekste võrrelda ja nende omavahelist seost (lähedust või kaugust) mõõta põhimõtteliselt sama meetodika järgi, mille esitame allpool leksikaalse seose mõõtmiseks.

Materjal ja meetod. Käesolevas töös vaatleme üht teatavat võimalust tekstide leksikaalse seose mõõtmiseks ja tekstide rühmitamiseks. Vaatlusmaterjaliks on 20 valimit à 5000 sõnet eesti tänapäeva ilukirjandusteoste autorikõnests (sama materjal, mida kasutati eesti keele sagedussõnastiku koostamisel).[✳] Kvantitatiivset vaatlust

[✳] Vaadeldavad teosed ja nende lühendid on järgmised:

1. A. Beekman, Kartulikuljused (1968) - AB
2. V. Gross, Pinginaabrid (1965) - VG
3. A. Hint, Tuuline rand IV (1966) - AH
4. H. Kiik, Tondiõomaja (1970) - HK
5. J. Kross, Kolme katku vahel I (1970) - JK
6. P. Kuusberg, Südasuvel (1966) - PK
7. L. Promet, Primavera (1971) - LP
8. V. Saar, Ukuaru (1969) - VS
9. H. Sergio, Põgenike laev (1966) - HS
10. R. Sirge, Kolmekesi lauas (1970) - RS
11. M. Traat, Tants aurukatla ümber (1971) - MT
12. E. Vetemaa, Väike romaaniaaamat (1968) - EV
13. A. Kaal, Saaremaa laastud II (1970) - AK
14. T. Kallas, Puiesteede kummaline valgus (1968) - TK
15. J. Peegel, Lühikesed lood (1970) - JP
16. J. Tuulik, Vana loss, Abruha lood (1972) - JT

täiendame võimaluse korral kvalitatiivse analüüsiga.

Lähtudes põhimõttest, et tekstide leksikaalset lähedust määrab võimalikult suurem arv ühiseid ja lähedase sagedusega sõnu, on loomulik kasutada tekstide sõnavara võrdlemisel korrelatsioonanalüüsil põhinevaid kaugusmõõte. Valime antud juhul Pearsoni-Bravais' korrelatsioonikordaja r. Kuid korrelatsioonanalüüs olgu vaid vaheetapiks veelgi sügavama analüüsi teostamiseks, nimelt faktoranalüüsi kasutamiseks, mis võimaldab tekstide vahelisi seoseid kompaktselt ja seejuures paindlikult esile tuua (üksikasjalikumalt faktoranalüüsi põhimõtete kohta ja rakendamise kohta keeleteaduses ja stilistikas vt. Тулдова, 1976). Faktoranalüüsi kasutamine objektide (antud juhul tekstide) rühmitamisel on eriti otstarbekohane sel juhul, kui rühmade kohta on juba ette teada, et tegemist on nn. umbmääraste hulkadega (inglise k. "fuzzy sets", vt. Zadeh, 1965; vene keeles kasutatakse väljendeid "размытые множества" või "нечеткие множества", vt. Пиотровский и др., 1977, 12). Sel juhul ei teki järgalt determineeritud objektide rühmitusi, vaid uurijal avaneb võimalus rühmi käsitleda osaliselt kattuvatena, ebaselgeid piirjooni omavatena. Ka tekstide rühmitamine keeleliste või stilistiliste tunnuste alusel ei anna tavaliselt ühemõtteliselt kindlaid lahendusi, vaid selline rühmitamine eeldab eri variantide ja tõenäosusfaktori arvestamist.

Tänu elektronarvuti kasutamise võimalusele on korrelatsioonanalüüsi ja faktoranalüüsi rakendamine tänapäeval hõlpsasti läbiviidav, kuigi ka siin on (vähemalt esialgu) teatavad piirid võrreldavate objektide või tunnuste arvu määramisel.* Teostasime korrelatsioonanalüüsi ja faktoranalüüsi

17. A. Valton, Luikede soo, Karussell (1971) - AV
18. M. Unt, Kuu nagu kustuv päike (1968) - MU
19. E. Niit, J. Kross, Muld ja marmor (1968) - K/N
20. J. Smuul, Jaapani meri, detsember (1965) - JS

* Kõik käesolevas uurimuses tehtud arvutustööd (eksperimentideks vajalike jaotussõnastike - sagedusloendite koostamine, korrelatsioon- ja faktoranalüüs) teostati TRÜ arvutuskeskuses väljatöötatud programme alusel arvutil "Minsk-32".

si eespool nimetatud 20 ilukirjandusproosa teksti (valimil) põhjal kolme eri nimistu alusel: 50 sagedamat sõnavormi, 40 väljavalitud täistähenduslikku sõna (lekseemi) ja 40 välja- valitud omadussõna (lekseemi). Eksperimendi lähteandmed esitatakse tabelites 1, 4 ja 6 (tabelis 1 on kasutatud järg- misi lühendeid: k - konjunktsioon, d₁ - iseseisev määrsõna, d₂ - rühmäärsõna, d₃ - abimäärsõna, pp - postpositsioon, pl - plural, nom - nimetav kään, gen - omastav kään).

Tekstide võrdlemine "struktuurisõnade" alusel. Esimene nimistu-50 sagedamat sõnavormi - hõl- mab peamiselt nn. struktuurisõnu, s.o. abisõnu, asesõnu ja üldise tähendusega määrsõnu (vt. tabel 1). Need on kõrgsa- gedustsooni kuuluvad sõnad (sagedus $F \geq 165$), mis tekstist katavad kokku ligikaudu 25 %. Paljud uurijad peavad sellist nimistut oluliseks individuaalstiilide diferentseerimisel, arvestades seda, et nimistusse kuuluvad struktuurisõnad ei ole otseselt mõjutatud teemast, moodustavad indiviidi sõna- vara "vähem teadliku" osa ja seepärast toovad esile teksti "varjatud stiilimustri" (Eliasson, 1972, 57). Tähtis on ka asjaolu, et kõrgsagedustsooni kuuluvate sõnade sagedusand- med on statistiliselt kõige usaldusväärsemad.

Korrelatsioonanalüüs (tekstide vahel) näitab, et 50 sagedama sõnavormi põhjal arvatatud korrelatsioonikordajate väärtused on üldiselt kõrged (kriitiline Pearsoni-Bravais' korrelatsioonikordaja r väärtus 20 objekti ja 5 %-lise olu- lisusnivoo korral on 0,45). See tähendab kõigepealt seda, et allkeele piirides on kõrgsagedustsooni sõnade kasutamine küllaltki ühtlane, kuigi mõnede sõnade suhtes võib konsta- teerida suhteliselt suurt hajuvust (näit. sõnavormid on, oli, ning, kuid, küll). Võrreldavate autorite (tekstide) hulgas on 50 sagedama sõnavormi sageduse põhjal kõige lähemas seos- ses V. Gross ja A. Valton ($r = 0,94$) ning A. Kaal ja M. Unt ($r = 0,94$). Võrreldes konkreetset sagedusandmeid näit. A. Kaalu ja M. Undi puhul 10 sagedama sõnavormi ulatuses võime- nentida arvnaõtajate suurt lähedust (vrd. tabel 1):

A. Kaal: 185 - 71 - 76 - 65 - 84 - 56 - 40 - 40 - 37 -
- 26;

M. Unt: 186 - 69 - 71 - 66 - 70 - 51 - 37 - 37 - 37 -
- 18.

Arvud tähistavad järjekorras sõnavormide ja, ta, on, ei, et, oli, kui, see, oma, nagu esinemissagedust vastavates valimites. Sagedused on lähedased ka suure osa ülejäänud sõnavormide puhul, kuid võib siiski täheldada üksikuid olulisi erinevusi. Näiteks esineb küll A. Kaalu valimis 17 korda, M. Undi valimis ainult 1 kord. Sidesõna kuid sagedused on vastavalt 19 ja 4, aga sagedused 21 ja 31.

Leksikaalse korrelatsiooni hinnangut võib modifitseerida sel teel, et arvutatakse nn. määratuse ehk determinatsiooni kordaja r^2 . Kordaja r^2 väärtust võib esitada protsentides ja sisuliselt interpreteerida, nimelt mõeldab see "sõnakasutuste sarnasust, mis on sõltumatu juhusest" (Herdan, 1966, 160). Nii näiteks on $r = 0,94$ korral $r^2 = 0,94^2 = 0,88$. Sel juhul võib arvestada, et mõõdetud leksikaalne lähedus on 88 % ulatuses tingitud "valikust" (G. Herdani mõttes), s.t. mingitest tegelikult ühimestest stiiliomadustest võrreldavate autorite puhul.

Korrelatsioonanalüüsi tulemusi üldistab faktoranalüüs. 20 teksti võrdlemine 50 sagedama sõnavormi põhjal toob esile kaks põhilist faktorit, mis kokkuvõttes kirjeldavad 87,1 % kõigi tunnuste varieeruvusest (vt. tabel 2). Arvutades faktorite individuaalväärtused (vt. tabel 3), võime lähemalt määratleda ka faktorite sisulist külge. Esimese faktori puhul on täheldatav sõnavormide ta, oli, polnud tugev mõju; teisele faktorile on iseloomulik sõnavormide on, pole, osalt ka mis, nagu dominantne seisund. Seega on põhiliseks erinevuseks faktorite vahel verbi ajavormide kasutamine, mida antud juhul kajastavad olema vormid. Oleviku- või minevikuvormide eelistamine autorikõnes toob endaga kaasa ka mõned muud erinevused jutustamis- või kirjeldamislaadis. Tundub, et olevikuvormide kasutamisel tekib nihe lihtsama struktuuriga tegelaskõne poole (sellele viitab tegelaskõnele iseloomulike sõnavormide mis ja küll sagedus).

Esimese faktori ("minevik") mõju on suur selliste autorite juures nagu A. Hint, T. Kallas, A. Beekman (faktor-

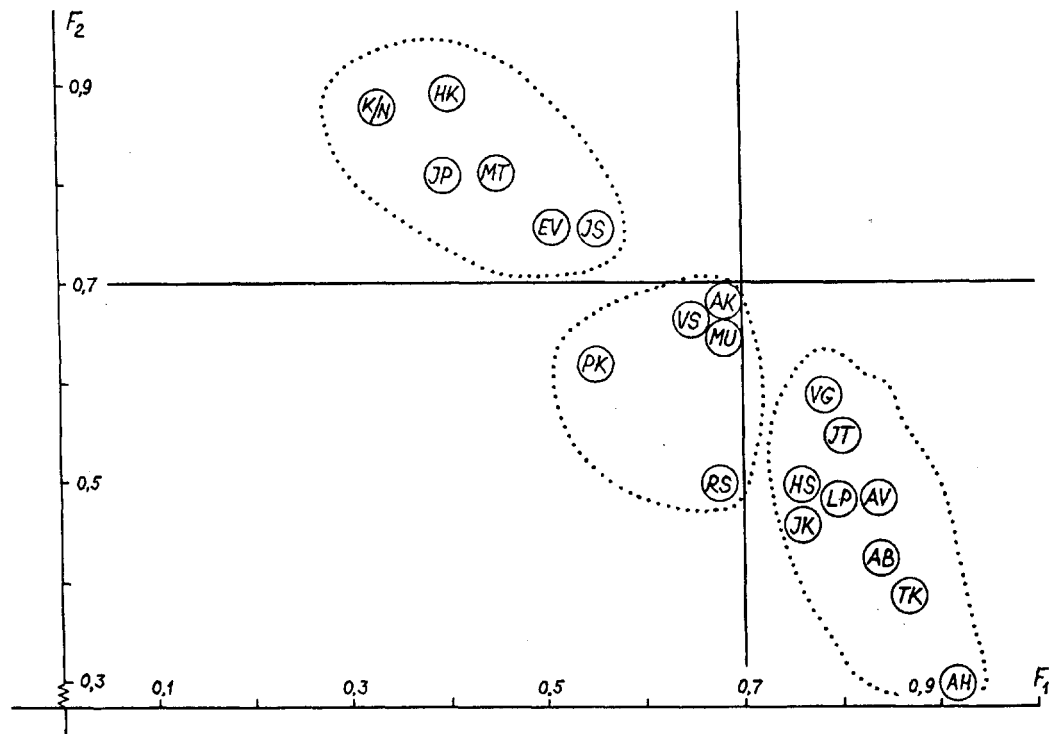
kaalud üle 0,8; vt. tabel 2). Teine faktor ("olevik") domineerib H. Kiigel, J. Peeglil, M. Traadil jt. Huvitav on täheldada, et J. Krossi ajaloolises romaanis "Kolme katku vahel" määrab stiili esimene faktor, kuna aga E. Niidu - J. Krossi reisikirjelduses on ülekaalus teine faktor.

Kahe faktori koosmõju võib kujutada graafikul, mis toob eriti selgelt esile tekstide omavahelised kaugused ja rühmitused (vt. joon. 4). Võib nentida kolme suure rühma olemasolu. Esimene rühm, mida iseloomustavad suured faktorkaalud esimese faktori puhul (üle 0,7) ja väikesed faktorkaalud teise faktori puhul (alla 0,7), hõlmab 9 teksti (A. Hint, T. Kallas, H. Sergo jt.). Teise rühma kuuluvad 6 teksti; siin domineerib teine faktor, kuna aga esimene faktor on olulise mõjuta (J. Smuul, J. Peegel jt.). Kolmas rühm moodustub nendest tekstidest, mille puhul mõlema faktori mõju on ühtlaselt mõõdukas (faktorkaalud 0,5 - 0,7). Sii kuuluvad A. Kaal, M. Unt, V. Saar, P. Kuusberg, R. Sirge.

Tekstide võrdlemine täistähenduslike sõnade alusel.

Leksikaalse seose mõõtmise teine faas seisnes kesksagedustsooni sõnade võrdlemises vaadeldava 20 teksti lõikes, kusjuures vaatluse alla võeti rühm t ä i s t ä h e n d u s - l i k k e s õ n u (l e k s e e m e). Valik toimus põhimõttel vältida temaatilisi või "terminoloogilisi" sõnu (näit. spetsiifilisi maa- või linnaelu termineid) ning koondata tähelepanu "üldinimlikele" mõistetele, nagu aistingud, teadmised, tundmused, tegevus. 40 väljavalitud täistähendusliku sõna hulgas on 8 omadussõna, 16 tegusõna ja 16 nimisõna (vt. tabel 4).

Võrreldes eelmise eksperimendiga on korrelatsioonid tekstide vahel täistähenduslike sõnade sageduste põhjal märksa nõrgemad. Suurim korrelatsioonikordaja väärtus esineb autoritepaaril V. Gross - H. Sergo ($r = 0,82$ ja $r^2 = 0,67$ ehk 67 %). Kõige rohkem erinevad teineteisest V. Gross ja T. Kallas ($r = 0,16$). Korrelatsioonanalüüsist selgub, et vaadeldavate sõnade kasutamise poolest erineb T. Kallas üldse kõige rohkem teistest autoritest - seda kinnitab asjaolu, et 9 juhul on korrelatsioonikordaja väärtused (T. Kalda ja teiste autorite vahel) statistiliselt ebaolulised. Edasine analüüs näitab (vrd. tabel 4), et T. Kaldal esinevad suhte-

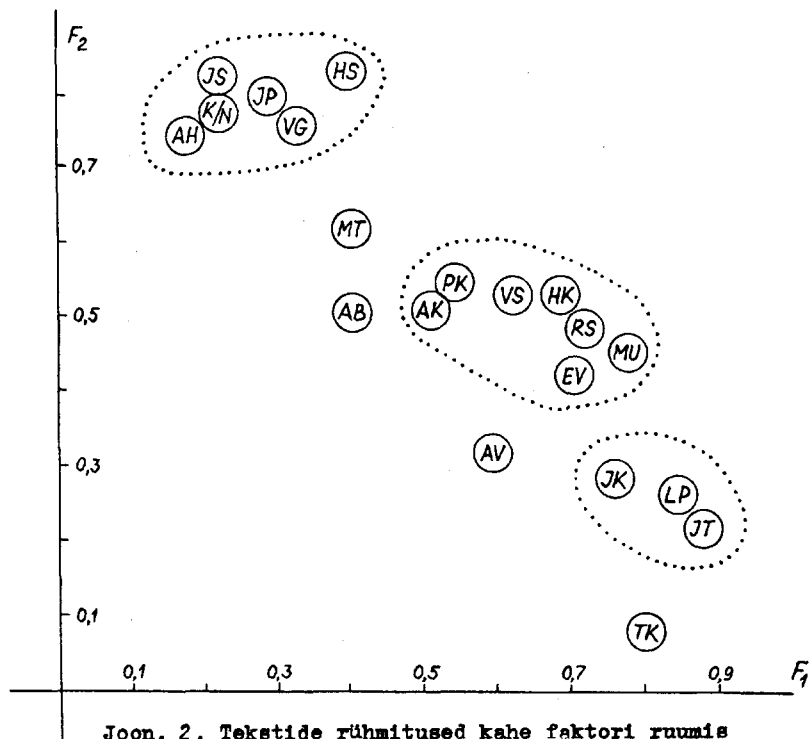


Joon. 1. Tekstide rühmitused kahe faktori ruumis
50 sagedama sõnavormi põhjal

liselt rohkem kui teistel autoritel sellised sõnapaarid nagu mõtleva ja tundma, mõte ja tunne, rääkima ja küsima, kuna aga keskmisest näitajast oluliselt vähem on sõnu mees, naine, töö, teadma. T. Kalda teost "Puiesteede kummaline valgus" iseloomustavad peategelase meelisklused ja meenutused, mida seaduspäraselt kajastab sõnavara valik.

Faktoranalüüsi andmeil võib eristada kahte enam-vähem ühesuguse tähtsusega faktorit, mis kirjeldavad tunnuste varieeruvust vastavalt 34,5 ja 32,3 % (vt. tabel 5). Esimese faktori mõju on tugev J. Tuuliku, L. Prometi, J. Krossi, M. Undi jt. teostes (faktorikaalud alates 0,7-st). Faktorite individuaalväärtuste järgi otsustades on esimese faktori puhul dominantsõnadeks (võtmesõnadeks) silm, nägema, rääkima, tundma, mõningal määral ka mõtleva ja tahtma - seega peamiselt inimese meeltegevust, mõtteid ja tundmusi väljendavad sõnad. Teise faktori puhul on dominantsõnadeks elu, vana, teadma, mees, naine. Tekstide paiknemist kahe faktori "ruumis" näitlikustab joonis 2, millest nähtub, et esimesele faktorile ("aistingud, mõtted, tunded") seisab eriti lähedal T. Kallas. Moodustuvad üksikud rühmad, näiteks esimese faktori mõjupiirkonnast: J. Tuulik - L. Promet - J. Kross. Teise faktori ("elukogemused, teadmised") alla koonduvad A. Hint, H. Sergo, J. Smuul, J. Peegel jt. (huvitava kombel - peamiselt saartelt pärit mehed). Faktorite suhtes neutraalse rühma moodustavad V. Saar, H. Kiik, R. Sirge jt.

Tekstide võrdlemine omadussõnade alusel. Tekstide läheduse mõõtmise kolmas eksperiment hõlmas 40 o m a d u s - s õ n a (vt. tabel 6). Väljavaliitud sõnade (lekseemide) hulgas esinevad mitmed antonüümid, nagu hea - halb, ilus - ine-



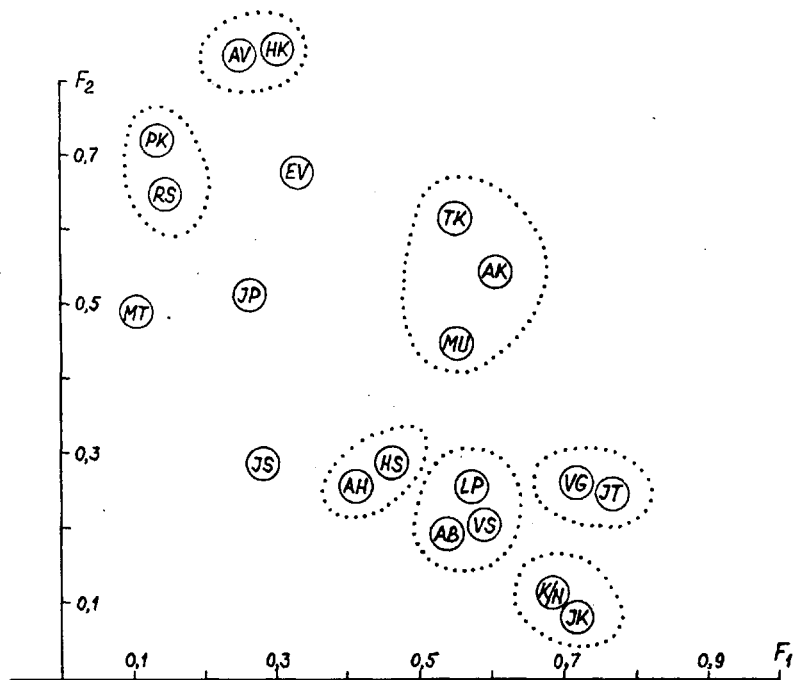
Joon. 2. Tekstide rühmitused kahe faktori ruumis
40 täistähendusliku sõna sageduste põhjal

tu, noor - vana jt. Peale selle on esindatud tähtsamad värve tähistavad omadussõnad. Paljude uurijate arvates (vt. näit. Шапкеевич, 1968; Лятина, 1969) peegeldavad omadussõnade valik ja sagedused eriti ilmekalt autoristiilide omapära. Omadussõnad on suhteliselt nõrgalt seotud lause süntaktilise struktuuriga ja nende kasutamine sõltub suurel määral autori individuaalsest maitsest.

Korrelatsioonanalüüsi põhjal selguvad tugevamad korrelatsioonid: H. Kiik - A. Valton ($r = 0,86$) ja A. Hint - H. Sergio ($r = 0,81$). Vaadeldes omadussõnade sageduste jaotumust tekstides, võime konstateerida, et näit. H. Kiige ja A. Valtoni lähedus on tingitud sellest, et mõlemad autorid kasutavad vaadeldavais valimites eriti palju sõna hea ja ühtlaselt väikeste sagedustega selliseid sõnu nagu raske, lihtne, õige, puhas, vaba, tühi, pikk, lühike jne. Mõlemas valimis puuduvad peaaegu täiesti värve tähistavad sõnad.

Faktoranalüüs aitab selgitada mõningaid varjatud seoseid ja täpsustada tekstide rühmitust omadussõnade kasutamise alusel (vt. tabel 7). Selguvad kolm enam-vähem võrdse tähtsusega faktorit (osatähtsus 23-24 %). Esimese faktori dominantsõnadeks osutuvad raske, täis, suur, väike, pikk, hall, valge. Need on peamiselt suurust ja värve tähistavad omadussõnad. Autoritest on esimese faktori mõjupiirkonnas V. Gross, J. Tuulik, J. Kross ja E. Niit - J. Kross. Teisele faktorile on iseloomulik sõnade hea ja uus aktiivsus ning samal ajal värve tähistavate sõnade puudumine või vähesus. Faktorkaalude järgi otsustades on teisele faktorile eriti lähedased H. Kiik, A. Valton ja P. Kuusberg, vähemal määral ka E. Vetemaa ja R. Sirge. Vahest kõige markantsemalt väljendub faktor A. Valtoni tekstis, arvestades dominantsõnade suurt osatähtsust, näit. hea esineb A. Valtoni valimis 21 korda (allkeele keskmine - 9), uus 13 korda (keskmine - 5). A. Valtonil on "pluss-sõnaks" ka ilus sagedusega 10 (keskmine - 3).

Kahe esimese faktori põhjal tehtud skemaatiline kujutus (joon. 3), näitab tekstide hajutatust tasapinnal, kuid see-



Joon. 3. Tekstide rühmitused kahe peamise faktori ruumis
40 omadussõna sageduste põhjal

juures moodustuvad väiksemad, tihedalt seotud rühmad. Nii näiteks selgub, et J. Krossi "Kolme katku vahel" ning E. Niidu ja J. Krossi "Muld ja marmor" on stiililt väga lähedased omadussõnade kasutamise aspektist vaadates. Suurt omavahelist lähedust võib täheldada ka naiskirjanike A. Beekmani, L. Prometi ja V. Saare puhul. Veel esinevad rühmitused: V. Gross - J. Tuulik, A. Hint - H. Sergo. Kõiki nimetatud autoreid ühendab esimene faktor. Teise faktori mõjupiirkonnas esinevad rühmad A. Valton - H. Kiik ja P. Kuusberg - R. Sirge. Neutraalses tsoonis on kolmik: A. Kaal - T. Kallas - M. Unt.

Faktoranalüüs annab ka kolmanda faktori, mida iseloomustavad dominantsõnad vana, suur ja osalt ka must. Autoritest kuuluvad siia rühma A. Hint, H. Sergo, J. Smuul ja J. Peegel (jällegi mehed saartelt). Vaadeldes näiteks sõnade vana ja suur sagedusi nende autorite tekstides võrreldes allkeele keskmisega (5000-sõnelise valimi korral):

| | AH | HS | JS | JP | Keskmine |
|------|----|----|----|----|----------|
| vana | 18 | 11 | 11 | 20 | 8 |
| suur | 18 | 24 | 18 | 28 | 14 |

Nimetatud autoritel (välja arvatud J. Peegel) esinevad "miinus-sõnadena" hea ja ilus, s.t. sagedusega oluliselt alla keskmine.

Mõningad järeldused. Tekstide võrdlemine ülalkirjeldatud leksikaalse seose mõõtmise eksperimendi alusel näitab, et tekstid võivad olla lähedased ühe teatava sõnavara rühma alusel, kuid kaugemad mõne teise rühma põhjal. Tekstide leksikaalse seose määramisel on seepärast loomulik arvestada erinevaid sõnavara kihte. Antud juhul sooritasime eksperimendi kolme sõnavararühma alusel, mis kuuluvad kõrg- või kesksagedustsooni. Kui võrrelda kolme katse tulemusi omavahel, siis selgub, et ainult I katse (50 sagedama sõnavormi alusel) ja II katse (40 sageda täistähendusliku sõna alusel) tulemused on teineteisele mõningal määral lähedased (katsete omavahelise korrelatsiooni tugevus $r = 0,28$). III katse (40 omadussõna põhjal) tulemustel pole aga kokkuvõttes olulist korrelatsiooni esimese kahe katse tulemustega (kuiigi üksikud tekstid jäävad lähedasteks nii II kui ka III katse tulemuste põhjal, näit. AH, HS, JS ja JP). Mõnel juhul

võib nentida suhtelist lähedust kahe teksti vahel kõigi kolme sõnavararühma alusel (AK ja NU, JP ja JS, AH ja HS). Siinjuures tuleb meenutada, et sõnakasutuse võrdlemisel ei arvestatud temaatilisi sõnu, mis tähendab, et käesoleva eksperimendi tulemused viitavad sügavamatele, varjatud stilistilistele omadustele, mille poolest autorid on üksteisele lähedased või erinevad üksteisest.

Võib teha üldise järelduse, et kõik kolm katset näitavad mingeid erinevaid aspekte stiilide võrdlemisel leksikaalsete tunnuste alusel. Stilostatistilisest seisukohast peab eriti rõhutama faktoranalüüsi vahendusel saadud kompaksete ja sisuliselt tõlgendatavate tunnuste - f a k t o r i t e tähtsust teoste sõnavara hindamisel. Leksikaalse seose kompleksel mõõtmisel on otstarbekohane võtta arvesse kõigi kolme katse põhjal saadud faktoreid (kokku 7). Märkides iga teksti (autori) puhul faktori olulisuse pluss-märgiga (juhul kui faktorikaalu väärtus on vähemalt 0,7, mis vastab ligikaudu kahe standardhälbe suurusele)* ning ebaolulisuse nulliga (vt. tabel 8), võime eristada küllaltki põhjendatult mitmesuguseid stiiliomastreid (struktuure). Selgub, et täpselt ühesuguseid struktuure (oluliste faktorite kombinatsioonide põhjal) esineb vähe, mis lõppkokkuvõttes viitab individuaalsete stiilide mitmekesisusele. Faktoranalüüsi kasutamise peamine mõte seisneb selles, et tekstide võrdlemisel ja rühmitamisel on võimalik leksikaalse seose puhtkvantitatiivse mõõtmise kõrval lähemalt määratleda ka tekstidevahelise seose sisulist külge.

Stiilide võrdlemine kvantitatiivsete näitajate põhjal ei tarvitse piirduda sõnavarastatistiliste andmete kõrvutamisega. Ülal kirjeldatud meetodikat on võimalik üle kanda ka mitmesuguste munde stiilinähtuste (stiiliomaduste ja -tunnuste, kujundite jne.) vaatlusele kirjandusliku stilistika valdkonnas. Kvantitatiivse uurimise tulemused võivad olla kvalitatiivse analüüsi objektiivseks aluseks.

* Faktorikaalu olulisuse määramise kohta vt. Хорван, 1972, 448-450. Antud juhul lähtusime keskmistest korrelatsioonikordajate väärtustest $\bar{r} = 0,25 \div 0,30$ ja objektide arvust $n = 20$.

T a b e l 1

50 sagedama sõnavormi jaotussõnastik 20 valimi lõikes
(r - sõnavormi astak F järgi; F - sõnavormi tegelik esinemissagedus
koondvalimis; U - kasutuskordaja, vt. lk. 107)

| r | Sõna- vorm | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 | F | U |
|---|---------------------------|-----|-----|-----|-----|-----|-----|-----|-----|-----|----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|------|------|
| | | AB | VG | AH | HK | JK | PK | LP | VS | HS | RS | MT | EV | AK | TK | JP | JT | AV | MU | K/N | JS | | |
| | 1 ja (k) | 137 | 129 | 141 | 174 | 283 | 118 | 171 | 171 | 155 | 80 | 168 | 131 | 185 | 169 | 121 | 147 | 177 | 186 | 181 | 197 | 3221 | 3138 |
| | 2 ta (nom) | 98 | 78 | 112 | 45 | 75 | 95 | 76 | 123 | 25 | 68 | 123 | 78 | 71 | 127 | 37 | 87 | 138 | 69 | 27 | 50 | 1602 | 1491 |
| | 3 on | 23 | 54 | 21 | 134 | 22 | 86 | 29 | 76 | 15 | 28 | 163 | 118 | 76 | 41 | 119 | 50 | 79 | 71 | 144 | 90 | 1439 | 1187 |
| | 4 ei | 36 | 80 | 56 | 71 | 24 | 118 | 97 | 108 | 74 | 80 | 78 | 80 | 65 | 49 | 71 | 64 | 76 | 66 | 32 | 50 | 1375 | 1297 |
| | 5 et (k) | 53 | 47 | 54 | 67 | 44 | 114 | 100 | 95 | 51 | 56 | 45 | 71 | 84 | 64 | 47 | 60 | 78 | 70 | 31 | 33 | 1264 | 1186 |
| | 6 oli | 43 | 66 | 122 | 7 | 84 | 28 | 69 | 27 | 91 | 29 | 21 | 52 | 56 | 107 | 32 | 74 | 112 | 51 | 11 | 34 | 1116 | 926 |
| | 7 kui (k) | 51 | 54 | 52 | 32 | 56 | 36 | 62 | 64 | 35 | 64 | 55 | 41 | 40 | 30 | 77 | 56 | 44 | 37 | 57 | 52 | 995 | 951 |
| | 8 see | 11 | 39 | 32 | 33 | 23 | 11 | 42 | 48 | 22 | 32 | 42 | 58 | 40 | 58 | 55 | 33 | 46 | 37 | 27 | 24 | 713 | 669 |
| | 9 oma | 24 | 29 | 56 | 35 | 30 | 33 | 21 | 14 | 25 | 24 | 25 | 20 | 37 | 19 | 29 | 23 | 38 | 37 | 24 | 43 | 592 | 566 |
| | 10 nagu | 23 | 16 | 29 | 35 | 25 | 17 | 20 | 28 | 25 | 37 | 43 | 37 | 26 | 33 | 28 | 45 | 16 | 18 | 23 | 18 | 542 | 515 |
| | 11 mis | 21 | 30 | 30 | 30 | 15 | 21 | 23 | 33 | 24 | 29 | 19 | 26 | 22 | 24 | 27 | 26 | 24 | 25 | 36 | 32 | 517 | 509 |
| | 12 ka | 9 | 23 | 48 | 23 | 15 | 29 | 11 | 20 | 22 | 35 | 23 | 24 | 49 | 16 | 17 | 10 | 16 | 17 | 19 | 8 | 434 | 387 |
| | 13 siis (d ₁) | 13 | 18 | 27 | 21 | 16 | 14 | 18 | 30 | 20 | 15 | 17 | 16 | 27 | 25 | 25 | 30 | 31 | 29 | 17 | 13 | 422 | 407 |
| | 14 ma | 1 | - | - | - | - | 61 | 48 | 31 | - | 15 | - | 69 | 19 | 16 | 22 | 47 | - | 61 | - | 26 | 416 | 305 |
| | 15 ning (k) | 26 | 18 | 28 | 1 | 30 | 3 | 8 | 6 | 21 | 17 | 3 | 22 | 16 | 12 | 67 | 26 | 46 | 9 | 18 | 16 | 393 | 315 |

Tabel 1 (järg)

| r | Sõna- vorm | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 | F | U |
|----|------------------|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|-----|----|-----|-----|
| | | AB | VG | AH | HK | JK | PK | LP | VS | NS | RS | MT | EV | AK | TK | JP | JT | AV | MU | K/N | JS | | |
| 16 | aga (d_2) | 4 | 23 | 25 | 13 | 22 | 10 | 17 | 19 | 15 | 20 | 11 | 21 | 31 | 26 | 25 | 27 | 12 | 16 | 34 | 13 | 384 | 359 |
| 17 | seda | 6 | 15 | 23 | 19 | 12 | 17 | 29 | 25 | 6 | 14 | 20 | 28 | 33 | 22 | 12 | 18 | 24 | 24 | 9 | 22 | 378 | 354 |
| 18 | aga (k) | 2 | 7 | 24 | 25 | 18 | 19 | 7 | 11 | 21 | 13 | 16 | 19 | 21 | 16 | 22 | 28 | 15 | 31 | 13 | 12 | 340 | 313 |
| 19 | nii (d_1) | 8 | 19 | 9 | 20 | 3 | 15 | 15 | 28 | 12 | 16 | 11 | 18 | 19 | 23 | 21 | 22 | 20 | 27 | 9 | 14 | 329 | 308 |
| 20 | või (k) | 6 | 21 | 11 | 19 | 13 | 21 | 12 | 26 | 12 | 37 | 19 | 7 | 15 | 12 | 28 | 13 | 9 | 8 | 11 | 29 | 329 | 292 |
| 21 | kes | 16 | 22 | 19 | 10 | 17 | 21 | 9 | 8 | 19 | 17 | 22 | 9 | 12 | 6 | 11 | 11 | 16 | 11 | 13 | 28 | 297 | 279 |
| 22 | veel (d_1) | 7 | 11 | 24 | 20 | 13 | 15 | 15 | 14 | 20 | 4 | 15 | 17 | 22 | 16 | 8 | 11 | 13 | 23 | 14 | 10 | 292 | 276 |
| 23 | ja (d_2) | 1 | 8 | 14 | 13 | 25 | 3 | 17 | 16 | 7 | 1 | 13 | 18 | 27 | 20 | 9 | 9 | 21 | 10 | 18 | 22 | 272 | 225 |
| 24 | nað | 15 | 14 | 6 | 13 | 15 | 24 | 8 | 17 | 9 | 5 | 33 | 3 | 9 | 14 | 9 | 13 | 9 | 18 | 9 | 14 | 257 | 228 |
| 25 | pole | 1 | 9 | 5 | 15 | 2 | 24 | 6 | 22 | 6 | 4 | 33 | 20 | 12 | 5 | 13 | 15 | 12 | 12 | 20 | 13 | 249 | 205 |
| 26 | kõik | 6 | 8 | 5 | 12 | 10 | 15 | 9 | 12 | 7 | 13 | 12 | 18 | 13 | 19 | 13 | 17 | 16 | 20 | 8 | 13 | 246 | 234 |
| 27 | polnud | 21 | 14 | 33 | 3 | 14 | 6 | 9 | 9 | 22 | 7 | 1 | 9 | 7 | 16 | 4 | 21 | 22 | 14 | 1 | 5 | 238 | 194 |
| 28 | selle | 9 | 11 | 9 | 8 | 13 | 5 | 10 | 11 | 18 | 7 | 8 | 14 | 11 | 14 | 19 | 12 | 19 | 14 | 12 | 13 | 237 | 227 |
| 29 | juba (d_1) | 6 | 13 | 14 | 13 | 11 | 12 | 7 | 17 | 15 | 11 | 9 | 12 | 11 | 10 | 12 | 11 | 8 | 16 | 13 | 10 | 231 | 225 |
| 30 | midagi | 4 | 12 | 5 | 5 | 7 | 21 | 18 | 11 | 7 | 6 | 23 | 20 | 10 | 12 | 10 | 9 | 9 | 23 | 7 | 5 | 224 | 187 |
| 31 | teda | 13 | 10 | 14 | 1 | 12 | 27 | 9 | 17 | 7 | 18 | 10 | 10 | 5 | 9 | - | 10 | 17 | 22 | 5 | 7 | 223 | 185 |
| 32 | nüüd | 13 | 11 | 11 | 22 | 4 | 15 | 8 | 16 | 8 | 10 | 10 | 13 | 13 | 6 | 9 | 16 | 13 | 11 | 7 | 3 | 219 | 204 |
| 33 | ainult (d_2) | 3 | 13 | 20 | 10 | 12 | 10 | 9 | 10 | 16 | 8 | 18 | 15 | 2 | 9 | 12 | 4 | 7 | 15 | 12 | 13 | 218 | 196 |
| 34 | oleks | 5 | 22 | 26 | 16 | 17 | 11 | 3 | 5 | 5 | 9 | 8 | 11 | 13 | 5 | 14 | 20 | 8 | 6 | 8 | 1 | 213 | 176 |

T a b e l 1 (järg)

| r | Sõna- vorm | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 | P | U |
|---------|----------------------------|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|-----|----|-----|-----|
| | | AB | VG | AH | HK | JK | PK | LP | VS | HS | RS | MT | EV | AK | TK | JP | JT | AV | MU | K/N | JS | | |
| | 35 ära (d ₃) | 7 | 7 | 9 | 11 | 10 | 10 | 18 | 14 | 5 | 12 | 21 | 8 | 14 | 3 | 5 | 12 | 12 | 15 | 6 | 7 | 206 | 187 |
| | 36 tema (gen) | 3 | 13 | 9 | 14 | 9 | 11 | 12 | 19 | 10 | 13 | 7 | 8 | 5 | 11 | 2 | 9 | 10 | 8 | 19 | 6 | 198 | 182 |
| | 37 olid (pl) | 7 | 7 | 23 | - | 20 | 5 | 13 | 3 | 17 | 5 | 1 | 3 | 14 | 15 | 10 | 13 | 19 | 9 | 3 | 6 | 193 | 155 |
| | 38 kuid (k) | 6 | 13 | 8 | 5 | 5 | 22 | 7 | 12 | 10 | 8 | 9 | 23 | 19 | 5 | 3 | - | 23 | 4 | 2 | 8 | 192 | 156 |
| | 39 kas (d ₂) | 10 | 10 | 8 | 18 | 4 | 15 | 14 | 10 | 8 | 13 | 5 | 9 | 18 | 4 | 7 | 6 | 4 | 16 | 6 | 4 | 189 | 168 |
| | 40 ta (gen) | 8 | 5 | 23 | - | 26 | 12 | 16 | 9 | - | 19 | 7 | 7 | 6 | 9 | 6 | 8 | 7 | 7 | 4 | 10 | 189 | 154 |
| | 41 tal | 10 | 9 | 17 | 3 | 8 | 4 | 6 | 10 | 6 | 6 | 20 | 12 | 10 | 21 | 5 | 8 | 19 | 6 | 2 | 4 | 186 | 154 |
| | 42 küll (d ₂) | 3 | 7 | 5 | 4 | 9 | 10 | 3 | 12 | 15 | 9 | 15 | 19 | 17 | 9 | 15 | 4 | 10 | 1 | 13 | 5 | 185 | 153 |
| | 43 üle (pp) | 10 | 4 | 4 | 10 | 16 | 8 | 5 | 11 | 9 | 10 | 7 | 5 | 5 | 7 | 10 | 8 | 6 | 16 | 19 | 10 | 180 | 165 |
| - 143 - | 44 välja (d ₃) | 11 | 11 | 9 | 11 | 8 | 10 | 11 | 3 | 8 | 8 | 11 | 9 | 10 | 6 | 8 | 5 | 17 | 12 | 8 | 2 | 178 | 168 |
| | 45 mitte | 4 | 14 | 4 | 3 | 7 | 13 | 6 | 14 | 7 | 14 | 4 | 9 | 6 | 14 | 11 | 6 | 9 | 9 | 10 | 13 | 177 | 164 |
| | 46 mida | 3 | 13 | 4 | 7 | 4 | 15 | 11 | 13 | 3 | 11 | 13 | 9 | 3 | 1 | 6 | 8 | 21 | 6 | 6 | 17 | 174 | 144 |
| | 47 ise | 4 | 6 | 21 | 9 | 5 | 9 | 2 | 8 | 10 | 13 | 10 | 1 | 9 | 6 | 12 | 9 | 13 | 16 | 4 | 2 | 169 | 140 |
| | 48 kus | 8 | 13 | 4 | 6 | 6 | 7 | 12 | 13 | 7 | 10 | 12 | 2 | 16 | 1 | 11 | 2 | 5 | 11 | 6 | 14 | 166 | 148 |
| | 49 pärast (pp) | 12 | 16 | 17 | 4 | 4 | 6 | 11 | 7 | 10 | 9 | 6 | 8 | 6 | 9 | 10 | 7 | 4 | 7 | 8 | 5 | 166 | 153 |
| | 50 enam | 3 | 6 | 11 | 9 | 2 | 10 | 11 | 12 | 17 | 3 | 11 | 11 | 10 | 5 | 9 | 6 | 7 | 9 | 8 | 5 | 165 | 152 |

Tabel 2

20 autoritekti faktormatriksi
50 sagedama sõnavormi põhjal
(tähtsusega on märgitud olulised faktorikaalud: $F \geq 0,7$)

| Jrk. nr. | Autor | Faktorikaalud | | Kommunaliteet (k^2) |
|---------------|-----------------------|-------------------|-------------------|----------------------------|
| | | f_1 | f_2 | |
| 1. | A. Beekman | 0,84 ^m | 0,45 | 0,90 |
| 2. | V. Gross | 0,77 ^m | 0,58 | 0,93 |
| 3. | A. Hint | 0,91 ^m | 0,25 | 0,89 |
| 4. | H. Kiik | 0,40 | 0,89 ^m | 0,95 |
| 5. | J. Kross | 0,76 ^m | 0,47 | 0,79 |
| 6. | P. Kuusberg | 0,55 | 0,64 | 0,71 |
| 7. | L. Promet | 0,79 ^m | 0,52 | 0,89 |
| 8. | V. Saar | 0,67 | 0,68 | 0,91 |
| 9. | H. Bergo | 0,74 ^m | 0,50 | 0,80 |
| 10. | R. Sirge | 0,68 | 0,51 | 0,73 |
| 11. | M. Traat | 0,44 | 0,82 ^m | 0,86 |
| 12. | H. Vetemaa | 0,52 | 0,76 ^m | 0,86 |
| 13. | A. Kaal | 0,69 | 0,68 | 0,94 |
| 14. | T. Kallas | 0,87 ^m | 0,41 | 0,92 |
| 15. | J. Peegel | 0,39 | 0,81 ^m | 0,81 |
| 16. | J. Tuulik | 0,79 ^m | 0,55 | 0,93 |
| 17. | A. Valton | 0,81 ^m | 0,50 | 0,91 |
| 18. | M. Unt | 0,67 | 0,68 | 0,91 |
| 19. | H. Niit - J. Kross | 0,32 | 0,88 ^m | 0,87 |
| 20. | J. Saaul | 0,56 | 0,76 ^m | 0,90 |
| Faktori panus | | 9,24 | 8,17 | - |
| (%) | | 46,2 | 40,9 | 87,1 |

Tabel 3

Individaalsed faktorite väärtused 50 sagedama sõnavormi põhjal

| Sõnavorm | f_1 | f_2 | Sõnavorm | f_1 | f_2 |
|----------------|------------------|------------------|------------------|-------|-------|
| ja (k) | 3,5 ^m | 3,9 ^m | kõik | -0,4 | -0,2 |
| ta (nom) | 3,6 ^m | -0,6 | polnud | 0,4 | -1,0 |
| on | -2,2 | 5,0 ^m | selle | -0,3 | -0,3 |
| ei | 1,0 ^m | 1,2 ^m | juba (d_1) | -0,4 | -0,2 |
| et (k) | 1,3 ^m | 0,7 | midagi | -0,4 | -0,2 |
| oli | 3,6 ^m | -2,1 | teda | -0,1 | -0,6 |
| kui (k) | 0,7 | 0,5 | nüüd | -0,5 | -0,1 |
| see | 0,3 | 0,4 | ainult (d_1) | -0,3 | -0,3 |
| oma | 0,2 | 0,1 | oleks | -0,5 | -0,2 |
| nagu (k) | 0,0 | 0,3 | ära (d_2) | -0,5 | -0,2 |
| mis | -0,1 | 0,3 | tema (gen) | -0,5 | -0,2 |
| ka | 0,0 | -0,1 | olid (pl) | -0,4 | -0,3 |
| siis (d_1) | 0,1 | -0,1 | kuid (k) | -0,1 | -0,7 |
| ma | -0,2 | 0,1 | kas | -0,1 | -0,6 |
| ning (k) | 0,2 | -0,5 | ta (gen) | 0,0 | -0,8 |
| aga (d_2) | -0,1 | -0,1 | tal | -0,6 | -0,2 |
| seda | -0,1 | -0,1 | küll | -0,6 | -0,2 |
| aga (k) | -0,4 | 0,1 | üle (pp) | -0,4 | -0,3 |
| nii (d_2) | -0,2 | 0,0 | välja (d_2) | -0,4 | -0,4 |
| või (k) | -0,5 | 0,1 | mitte | -0,5 | -0,2 |
| kes | -0,2 | -0,2 | nida | -0,5 | -0,2 |
| veel (d_1) | -0,3 | -0,1 | ise | -0,3 | -0,4 |
| ja (d_2) | -0,4 | -0,1 | kus | -0,5 | -0,2 |
| nad | -0,4 | -0,1 | pärast (pp) | -0,2 | -0,5 |
| pole | -0,8 | 0,3 | enam | -0,5 | -0,3 |

40 täistähendusliku sõna sagedused 20 valimi lõikes
(F - esinemissagedus koondvalimis)

| Jrk. | Sõna | AB | VG | AH | HK | JK | FK | LP | VS | HS | RS | MT | EV | AK | TK | JP | JT | AV | MU | K/N | JS | F |
|------|-----------|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|-----|-----|-----|
| nr. | (lekseem) | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 | |
| 1. | suur | 7 | 15 | 18 | 8 | 11 | 6 | 11 | 5 | 24 | 5 | 7 | 12 | 16 | 10 | 28 | 12 | 16 | 20 | 21 | 18 | 270 |
| 2. | vana | 3 | 5 | 18 | 3 | 4 | 9 | 6 | 2 | 11 | 9 | 10 | 7 | 8 | 2 | 20 | 5 | 5 | 7 | 10 | 11 | 155 |
| 3. | hea | 2 | 7 | 3 | 17 | 4 | 10 | 6 | 6 | 5 | 9 | 5 | 12 | 11 | 16 | 15 | 8 | 21 | 12 | 5 | 6 | 180 |
| 4. | ilus | 1 | 1 | 1 | 1 | 2 | - | 5 | 2 | 1 | 1 | 5 | 2 | 5 | 8 | 4 | 5 | 10 | 5 | 3 | 1 | 63 |
| 5. | soe | 2 | 2 | 1 | 1 | 1 | - | 3 | - | - | 4 | 2 | - | 1 | 4 | 5 | 5 | - | 7 | - | 2 | 40 |
| 6. | rõõmus | 1 | 1 | - | 2 | 3 | 2 | - | - | - | - | 3 | - | 2 | 2 | - | 3 | 4 | - | 1 | 2 | 26 |
| 7. | imelik | 1 | 2 | - | - | 3 | 5 | 3 | 4 | - | 1 | 3 | 7 | - | 4 | - | 7 | - | 1 | - | - | 41 |
| 8. | hall | 2 | 6 | 8 | 1 | 10 | - | 1 | 4 | - | - | 2 | - | 2 | 7 | 3 | 10 | 1 | 6 | 5 | 2 | 70 |
| 9. | nõgema | 14 | 10 | 10 | 21 | 11 | 15 | 16 | 14 | 14 | 14 | 9 | 10 | 10 | 14 | 17 | 25 | 8 | 19 | 6 | 11 | 268 |
| 10. | vaatama | 21 | 7 | 2 | 26 | 18 | 5 | 22 | 11 | 14 | 13 | 16 | 17 | 9 | 16 | 16 | 35 | 19 | 19 | 3 | 8 | 297 |
| 11. | kuulma | 5 | 9 | 3 | 4 | 1 | 8 | 4 | 7 | 5 | 5 | 3 | 3 | 5 | 2 | 2 | 9 | - | 1 | 3 | 7 | 86 |
| 12. | kuulama | 4 | - | 1 | 1 | 2 | 2 | 2 | 1 | 1 | 2 | 4 | 6 | - | 1 | 1 | 4 | 4 | 4 | - | 10 | 50 |
| 13. | rääkima | 5 | 3 | 6 | 6 | 7 | 9 | 6 | 8 | 10 | 7 | 8 | 15 | 17 | 4 | 27 | 9 | 17 | 3 | 4 | 180 | |
| 14. | küsima | - | 3 | 2 | 2 | 4 | 5 | 8 | 2 | 7 | 3 | - | 1 | 3 | 10 | 3 | 2 | - | 7 | - | 2 | 64 |
| 15. | käima | 6 | 6 | 8 | 11 | 3 | 11 | 5 | 10 | 13 | 11 | 15 | 1 | 11 | 2 | 11 | 15 | 10 | 11 | 3 | 8 | 171 |
| 16. | sõitma | 2 | 5 | 2 | 6 | 4 | 7 | 4 | 6 | 6 | 4 | 8 | 1 | 6 | 6 | 6 | 3 | - | 7 | 7 | 6 | 96 |
| 17. | teadma | 5 | 15 | 18 | 12 | 13 | 14 | 17 | 20 | 20 | 12 | 14 | 12 | 9 | 7 | 21 | 14 | 22 | 17 | 10 | 13 | 285 |
| 18. | mõtleva | 6 | 7 | 6 | 10 | 4 | 13 | 17 | 6 | 4 | 2 | 15 | 8 | 6 | 17 | 3 | 10 | 21 | 11 | - | 8 | 174 |
| 19. | tundma | 12 | 10 | 10 | 11 | 13 | 16 | 20 | 7 | 9 | 17 | 4 | 19 | 3 | 12 | 6 | 18 | 10 | 9 | 3 | 9 | 218 |

T a b e l 4 (järg)

| | AB | VG | AH | HK | JK | PK | LP | VS | BS | RS | MT | EV | AK | TK | JP | JT | AV | MU | K/N | JA | F |
|--------------|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|-----|----|-----|
| 20. tahtma | 12 | 7 | 14 | 9 | 3 | 14 | 17 | 15 | 9 | 10 | 5 | 11 | 12 | 5 | 8 | 12 | 10 | 15 | 2 | 5 | 195 |
| 21. kartma | 1 | 1 | - | 1 | 1 | 5 | 2 | 2 | 1 | 1 | 3 | 2 | 4 | 1 | 3 | 4 | 2 | 6 | - | 1 | 41 |
| 22. lootma | 3 | 2 | - | 3 | - | - | 2 | 3 | 1 | 1 | 2 | 2 | 3 | - | - | 1 | 3 | - | - | - | 26 |
| 23. naerma | 4 | 1 | 4 | 4 | 3 | 3 | 3 | 2 | 1 | 1 | - | 2 | 2 | 2 | 1 | 3 | 2 | 2 | 1 | 1 | 42 |
| 24. meeldima | - | 3 | - | - | - | 3 | 1 | 1 | - | - | 2 | 6 | 3 | 6 | 2 | 2 | 4 | 5 | 3 | - | 41 |
| 25. kevad | - | 3 | 5 | 1 | - | - | 3 | 6 | 1 | 1 | 1 | - | 2 | - | 1 | - | - | 3 | - | - | 27 |
| 26. õhtu | 2 | 6 | 3 | 4 | 1 | 4 | 5 | 5 | 13 | 4 | 4 | 2 | 1 | 7 | 3 | 8 | 7 | 12 | - | 3 | 94 |
| 27. mõtte | 4 | 4 | 5 | 11 | 6 | 6 | 4 | 7 | 3 | 4 | 6 | 4 | 5 | 6 | 7 | 3 | 10 | 6 | 2 | 5 | 108 |
| 28. tunne | 4 | 3 | - | 1 | 1 | 1 | 3 | - | - | 7 | 2 | 5 | 1 | 5 | 1 | 3 | 13 | 3 | 2 | 2 | 57 |
| 29. elu | 2 | 15 | 12 | 10 | 1 | 2 | 4 | 3 | 7 | 1 | 8 | 1 | 2 | 7 | 13 | 6 | 7 | 5 | 4 | 12 | 122 |
| 30. surm | 1 | - | 6 | - | - | - | 2 | 2 | 4 | 1 | 1 | 2 | - | - | - | 3 | 2 | - | 2 | 3 | 29 |
| 31. rõõm | - | 2 | 3 | 1 | - | 2 | 8 | 1 | - | 2 | - | 2 | 1 | 2 | 4 | 3 | 3 | 2 | - | 1 | 37 |
| 32. mure | 3 | - | 8 | 2 | - | 2 | 1 | 2 | - | - | 1 | - | 3 | 1 | 9 | - | - | - | - | 1 | 33 |
| 33. töö | - | 13 | 6 | 12 | - | 3 | 2 | 6 | 8 | 4 | 9 | 4 | 5 | - | 12 | 3 | 6 | 1 | 2 | 9 | 105 |
| 34. sõna | 3 | 10 | 5 | 9 | 2 | 12 | 3 | 6 | 11 | 5 | 13 | 6 | 2 | 2 | 4 | 7 | 5 | 3 | 1 | 15 | 124 |
| 35. mees | 52 | 25 | 12 | 31 | 10 | 18 | 9 | 16 | 36 | 18 | 27 | 16 | 12 | 9 | 20 | 17 | 6 | 14 | 11 | 14 | 373 |
| 36. naine | 39 | 17 | 13 | 5 | 1 | 8 | 8 | 6 | 14 | 10 | 1 | 9 | 2 | 1 | 4 | 7 | 13 | 14 | 6 | 7 | 185 |
| 37. süda | 3 | - | 8 | 1 | 1 | 2 | 3 | 2 | 3 | 7 | 4 | 5 | 2 | 1 | 2 | 7 | 2 | 2 | 1 | 2 | 58 |
| 38. silm | 15 | 18 | 7 | 18 | 24 | 7 | 18 | 13 | 13 | 17 | 5 | 8 | 10 | 12 | 7 | 26 | 10 | 19 | 8 | 5 | 260 |
| 39. päike | 3 | 1 | - | 3 | - | - | 7 | 5 | 7 | 3 | 6 | 9 | 2 | - | 8 | 8 | - | 7 | 5 | 1 | 75 |
| 40. ilm | - | 2 | 1 | - | - | - | 1 | 1 | 6 | - | 1 | - | 1 | 1 | 4 | 2 | 2 | 1 | - | 8 | 31 |

Tabel 5

20 autoritekti faktormatriksi 40 täistähendusliku sõna põhjal

| A u t o r | Faktorikaalud | | Kommunali- teet (h ²) |
|---------------------------|--------------------|--------------------|--------------------------------------|
| | F ₁ | F ₂ | |
| 1. A. Beekman | 0,40 | 0,51 | 0,43 |
| 2. V. Gross | 0,36 | 0,77 ^{xx} | 0,71 |
| 3. A. Hint | 0,17 | 0,75 ^{xx} | 0,58 |
| 4. H. Kiik | 0,68 | 0,53 | 0,74 |
| 5. J. Kross | 0,76 ^{xx} | 0,29 | 0,66 |
| 6. P. Kuusberg | 0,56 | 0,55 | 0,61 |
| 7. L. Promet | 0,85 ^{xx} | 0,26 | 0,79 |
| 8. V. Saar | 0,62 | 0,53 | 0,67 |
| 9. H. Sergio | 0,37 | 0,86 ^{xx} | 0,87 |
| 10. R. Sirge | 0,71 ^{xx} | 0,49 | 0,75 |
| 11. M. Traat | 0,39 | 0,62 | 0,54 |
| 12. E. Vetemaa | 0,70 ^{xx} | 0,42 | 0,66 |
| 13. A. Kaal | 0,56 | 0,54 | 0,61 |
| 14. T. Kallas | 0,80 ^{xx} | 0,08 | 0,65 |
| 15. J. Peegel | 0,29 | 0,79 ^{xx} | 0,71 |
| 16. J. Tuulik | 0,87 ^{xx} | 0,23 | 0,81 |
| 17. A. Valton | 0,59 | 0,33 | 0,46 |
| 18. M. Unt | 0,77 ^{xx} | 0,46 | 0,79 |
| 19. E. Niit - J. Kross | 0,22 | 0,75 ^{xx} | 0,61 |
| 20. J. Smuul | 0,22 | 0,79 ^{xx} | 0,67 |
| Faktori panus | 6,89 | 6,45 | - |
| (%) | 34,5 | 32,3 | 66,8 |

T a b e l 6

40 omadussõna sagedused 20 valimi lõikes
(F - absoluutsagedus koondvalimis)

| Jrk. nr. | Sõna (lekseem) | AB | VG | AH | HK | JK | FK | LP | VS | ES | RS | MT | EV | AK | TK | JP | JT | AV | MU | K/N | JS | F | |
|-------------|-------------------|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|-----|----|-----|----|
| | | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 | | |
| 1. | hea | 2 | 7 | 3 | 17 | 4 | 10 | 6 | 6 | 5 | 9 | 5 | 12 | 11 | 16 | 15 | 8 | 21 | 12 | 5 | 6 | 180 | |
| 2. | halb | - | 1 | 1 | - | - | 1 | - | - | - | 1 | 2 | 3 | 2 | 1 | 2 | - | 2 | 2 | - | 3 | 21 | |
| 3. | ilus | 1 | 1 | 1 | 1 | 2 | - | 5 | 2 | 1 | 1 | 5 | 2 | 5 | 8 | 4 | 5 | 10 | 5 | 3 | 1 | 63 | |
| 4. | inetu | - | - | - | 1 | - | - | - | - | - | - | - | 4 | - | 1 | - | 1 | - | - | - | - | 7 | |
| 5. | tugev | - | 3 | 1 | - | 1 | 2 | - | - | - | 3 | 2 | 2 | 2 | 4 | 3 | 4 | 1 | 3 | 1 | 2 | 34 | |
| 6. | nõrk | - | 2 | 1 | 3 | - | - | - | - | - | - | - | 1 | - | - | - | 1 | 1 | 1 | - | 2 | 12 | |
| 7. | vana | 3 | 5 | 18 | 3 | 4 | 9 | 6 | 2 | 11 | 9 | 10 | 7 | 8 | 2 | 20 | 5 | 5 | 7 | 10 | 11 | 155 | |
| 8. | noor | 4 | 3 | 9 | 4 | 3 | 2 | 4 | 1 | 6 | 2 | 6 | 4 | 1 | 6 | 3 | 2 | 6 | - | 2 | 7 | 75 | |
| 9. | uus | 3 | 2 | 6 | 6 | 2 | 4 | 1 | 5 | 6 | 5 | 8 | 4 | 6 | 6 | 8 | 2 | 13 | 6 | 4 | 6 | 103 | |
| 10. | tore | - | 2 | - | - | 5 | 1 | 1 | - | - | 2 | 1 | 2 | 1 | 4 | - | 1 | 1 | 1 | 1 | 2 | 25 | |
| 11. | hirmus | - | 3 | - | 1 | 1 | - | 1 | - | - | 2 | - | - | - | - | 1 | 4 | - | 5 | - | 1 | 19 | |
| 12. | rõõmus | 1 | 1 | - | 2 | 3 | 2 | - | - | - | - | 3 | - | 2 | 2 | - | 3 | 4 | - | 1 | 2 | 26 | |
| 13. | õnnelik | - | 1 | 1 | - | - | 1 | 3 | 1 | 1 | 1 | - | - | 1 | - | 2 | 1 | 1 | 1 | - | 1 | 16 | |
| 14. | raske | 3 | 2 | 3 | 3 | 5 | 4 | 2 | 5 | 2 | - | - | 1 | 6 | 1 | 1 | 9 | 2 | 3 | 1 | 5 | 58 | |
| 15. | kurb | - | - | 1 | - | - | 1 | - | 1 | - | - | 3 | - | - | 2 | - | - | - | - | - | - | 4 | 12 |
| 16. | kerge | 2 | 1 | 1 | 1 | 2 | 3 | - | 1 | - | 3 | 1 | 2 | 2 | 2 | 2 | 6 | - | 2 | - | 4 | 35 | |
| 17. | lihtne | - | 1 | 1 | 2 | 1 | 1 | - | 1 | - | - | 1 | 1 | - | - | 1 | 2 | 2 | 3 | - | 7 | 24 | |
| 18. | imelik | 1 | 2 | - | - | 3 | 5 | 3 | 4 | - | 1 | 3 | 7 | - | 4 | - | 7 | - | 1 | - | - | 41 | |
| 19. | õige | - | 3 | 4 | 3 | 1 | - | 2 | 7 | 5 | 1 | 3 | 2 | - | 1 | 4 | 3 | 3 | 1 | 4 | 7 | 54 | |
| 20. | puhas | 1 | 4 | 3 | 1 | 1 | 1 | 1 | 4 | - | - | 3 | 1 | 4 | - | 11 | 1 | 2 | 1 | 1 | 3 | 43 | |

Tabel 6 (järg)

| | AB | VG | AH | HK | JK | PK | LP | VS | HS | ES | MT | EV | AK | TK | JP | JT | AV | MU | K/N | JS | F |
|---------------------|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|-----|----|-----|
| 21. vaba | 1 | 5 | 2 | 1 | - | 3 | 1 | - | - | 2 | 2 | 1 | 1 | 2 | 4 | - | 1 | 1 | 2 | - | 29 |
| 22. selge | - | 1 | 2 | 3 | 2 | 1 | 3 | 2 | 4 | 1 | 2 | 1 | 3 | 1 | 2 | 5 | 5 | 3 | 5 | 1 | 47 |
| 23. täis | 2 | 7 | 5 | 4 | 4 | 1 | 8 | 6 | 6 | 3 | 6 | 4 | 4 | 4 | 5 | 4 | - | 3 | 9 | 2 | 87 |
| 24. tühi | - | - | 5 | 2 | 4 | - | 1 | 4 | 1 | 3 | 3 | 3 | 1 | 6 | 3 | 2 | 2 | 1 | 2 | 1 | 44 |
| 25. suur | 7 | 15 | 18 | 8 | 11 | 6 | 11 | 5 | 24 | 5 | 7 | 12 | 16 | 10 | 28 | 12 | 16 | 20 | 21 | 18 | 270 |
| 26. väike | 1 | 8 | 6 | 1 | 7 | 1 | 6 | 9 | 3 | 4 | 1 | 5 | 6 | 6 | - | 7 | 3 | 5 | 8 | 7 | 94 |
| 27. pisike | - | - | - | - | - | - | 1 | 3 | 3 | - | - | - | 3 | 6 | 8 | - | 1 | - | - | - | 25 |
| 28. pikk | 4 | 9 | 4 | 3 | 3 | 2 | 4 | 6 | 4 | 4 | 2 | 2 | 7 | 10 | 4 | 6 | 2 | 3 | 13 | 2 | 94 |
| 29. lühike | 1 | 3 | 3 | 1 | 4 | 1 | 1 | 1 | 1 | 3 | - | 3 | - | 2 | 1 | - | 1 | 1 | - | 2 | 29 |
| 30. külm | - | 1 | - | 1 | - | 3 | 4 | - | - | - | 1 | - | 1 | 2 | 2 | 4 | 1 | 7 | 1 | 7 | 35 |
| 31. soe | 2 | 2 | 1 | 1 | 1 | - | 3 | - | - | 4 | 2 | - | 1 | 4 | 5 | 5 | - | 7 | - | 2 | 40 |
| 32. hall | 2 | 6 | 8 | 1 | 10 | - | 1 | 4 | - | - | 2 | - | 2 | 7 | 3 | 10 | 1 | 6 | 5 | 2 | 70 |
| 33. valge (värv) | 6 | 2 | 1 | 3 | 2 | 1 | 2 | 4 | 1 | - | 1 | - | 6 | 5 | 2 | 4 | - | 6 | 10 | 1 | 57 |
| 34. must | 1 | 2 | 1 | - | 3 | - | 6 | - | 6 | - | 3 | 2 | 1 | 2 | 1 | 3 | - | 4 | 10 | 6 | 51 |
| 35. punane | - | 3 | - | - | 5 | 2 | 2 | 2 | 4 | 2 | 1 | 2 | 3 | 1 | 2 | 1 | 1 | 3 | 7 | 3 | 44 |
| 36. sinine | 2 | 2 | - | - | 6 | 1 | 3 | 2 | 2 | 1 | - | 3 | - | 3 | 1 | 3 | - | 2 | 1 | 2 | 34 |
| 37. kollane | 4 | - | - | 1 | 1 | - | - | 1 | - | - | 1 | 1 | 1 | 2 | 5 | 4 | - | 4 | 2 | - | 27 |
| 38. roheline | - | 2 | - | 1 | 1 | - | 4 | 2 | - | - | - | - | 2 | 3 | - | 1 | - | - | 1 | 2 | 19 |
| 39. hele | 2 | 2 | 1 | - | 3 | - | 1 | 1 | 1 | - | 1 | 3 | - | 2 | - | 3 | - | - | 4 | 4 | 28 |
| 40. tume | 2 | 1 | - | - | 2 | - | 1 | 4 | - | 2 | - | - | - | 2 | - | 2 | - | - | 1 | - | 17 |

Tabel 7

20 autoritekti faktometriks
40 esimese põhjal

| Autor | Faktorikaalud | | | Kommuna- litget (h ²) |
|---------------------------|-------------------|-------------------|-------------------|---|
| | F ₁ | F ₂ | F ₃ | |
| 1. A. Beekman | 0,53 | 0,18 | 0,37 | 0,45 |
| 2. V. Gross | 0,71 [±] | 0,27 | 0,44 | 0,77 |
| 3. A. Hint | 0,36 | 0,26 | 0,78 [±] | 0,80 |
| 4. H. Kiik | 0,30 | 0,84 [±] | 0,15 | 0,81 |
| 5. J. Kross | 0,71 [±] | 0,08 | 0,28 | 0,59 |
| 6. P. Kuusberg | 0,14 | 0,71 [±] | 0,38 | 0,68 |
| 7. L. Promet | 0,57 | 0,27 | 0,48 | 0,63 |
| 8. V. Saar | 0,57 | 0,26 | 0,07 | 0,60 |
| 9. H. Sergo | 0,45 | 0,28 | 0,78 [±] | 0,89 |
| 10. R. Sirge | 0,16 | 0,66 [±] | 0,41 | 0,63 |
| 11. M. Traat | 0,10 | 0,49 | 0,62 | 0,63 |
| 12. E. Vetemaa | 0,34 | 0,69 [±] | 0,38 | 0,73 |
| 13. A. Kaal | 0,60 | 0,54 | 0,44 | 0,85 |
| 14. T. Kallas | 0,57 | 0,62 | 0,01 | 0,71 |
| 15. J. Peegel | 0,27 | 0,52 | 0,68 [±] | 0,80 |
| 16. J. Tuulik | 0,73 [±] | 0,26 | 0,18 | 0,63 |
| 17. A. Valton | 0,26 | 0,82 [±] | 0,28 | 0,82 |
| 18. M. Unt | 0,54 | 0,45 | 0,46 | 0,70 |
| 19. E. Niit - J. Kross | 0,69 [±] | 0,10 | 0,56 | 0,80 |
| 20. J. Smuul | 0,29 | 0,27 | 0,74 [±] | 0,70 |
| Faktori panus | 4,74 | 4,73 | 4,56 | - |
| (%) | 23,7 | 23,7 | 22,8 | 70,2 |

Tabel 8

Oluliste faktorite jaotumine tekstide lõikes
(leksikaalse seose määramise eksperimentide põhjal)

| Tekst (Autor) | Faktorid | | | | | | |
|---------------------------|----------|------------|--------------|------------|----------|------------|--------------|
| | I (I) | II (II) | III (III) | IV (IV) | V (V) | VI (VI) | VII (VII) |
| 1. A. Beekman | + | 0 | 0 | 0 | 0 | 0 | 0 |
| 2. V. Gross | + | 0 | 0 | + | + | 0 | 0 |
| 3. H. Hint | + | 0 | 0 | + | 0 | 0 | + |
| 4. H. Kiik | + | 0 | 0 | 0 | 0 | + | 0 |
| 5. J. Kross | 0 | + | + | 0 | + | 0 | 0 |
| 6. P. Kuusberg | 0 | 0 | 0 | 0 | 0 | + | 0 |
| 7. L. Promet | + | 0 | + | 0 | 0 | 0 | 0 |
| 8. V. Saar | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 9. H. Sergo | + | 0 | 0 | + | 0 | 0 | + |
| 10. R. Sirge | 0 | 0 | + | 0 | 0 | + | 0 |
| 11. M. Traat | 0 | + | 0 | 0 | 0 | 0 | 0 |
| 12. E. Vetemaa | 0 | + | + | 0 | 0 | + | 0 |
| 13. A. Kaal | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 14. T. Kallas | + | 0 | + | 0 | 0 | 0 | 0 |
| 15. J. Peegel | 0 | + | 0 | + | 0 | 0 | + |
| 16. J. Tuulik | + | 0 | + | 0 | + | 0 | 0 |
| 17. A. Valton | + | 0 | 0 | 0 | 0 | + | 0 |
| 18. M. Unt | 0 | 0 | + | 0 | 0 | 0 | 0 |
| 19. E. Niit - J. Kross | 0 | + | 0 | + | + | 0 | 0 |
| 20. J. Smuul | 0 | + | 0 | + | 0 | 0 | + |
| Ekspereiment | 1. | | 2. | | 3. | | |

Viidatud kirjandus

- Eliasson, S. Vokabulärstudier i läroplanstexter 1935-1969. Stockholm, Språkförlaget Skriptor, 1972.
- Herdan, G. The Advanced Theory of Language as Choice and Chance. Berlin - Heidelberg - New York, 1966.
- Krallmann, D. Statistische Methoden in der stilistischen Textanalyse. Diss., Bonn, 1966.
- Laffal, J. Total or Selected Content Analysis. - International Conference on Computational Linguistics. Preprint No. 24. Stockholm, 1969.
- Muller, Ch. Initiation à la statistique linguistique. Paris, Librairie Larousse, 1968.
- Rosengren, I. The Quantitative Concept of Language and its Relation to the Structure of Frequency Dictionaries. - "Etudes linguistiques appliquées", 1971, No. 1.
- Tuldava, J. Korrelatsioonanalüüs keeleteaduses (2). - Linguistica, V. Tartu, TRU, 1974.
- Yule, G.U. The Statistical Study of Literary Vocabulary. Cambridge University Press, 1944.
- Zadeh, L.A. Fuzzy Sets. - "Information and Control", vol. 8, 1965, No. 3.
- Вахабов В. К. Некоторые результаты статистических исследований дескрипторного языка. - "Научно-техническая информация. Серия 2. Информационные процессы и системы." М., 1970, № 4, 27-31.
- Дарчук Н. П. Индивидуальное и общее в лексической системе авторского стиля (на материале современной украинской художественной прозы). АКД. Киев, 1975.
- Дядык Ю. В. Лексическое расстояние между текстами и его применение к классифицированию научно-технической информации. - "Научно-техническая информация. Серия 2. Информационные процессы и системы." М., 1973, № 6, 9-11.
- Каджазунчи Д. К. Статистический анализ индивидуального стиля. АКД. М., 1970.
- Клявина С. Корреляция между текстами разных функциональных стилей (по использованию словарного состава и частей речи). - Linguistica, VIII. Учен. зап. ТГУ, вып. 401. Tartu, 1976, 33-48.
- Колгушин А. Н. Лингвистика в военном деле. М., 1970.
- Ляпина А. М. Цветовая гамма прилагательных в романе М. А. Шолохова "Поднятая целина". - Актуальные проблемы лексикологии. Доклады 2-ой лингвистической конференции. Новосибирск. Изд. НГУ, 1969, 249-261.
- Маршакова И. В. Классификация документов на основе лексики (по ключевым словам документов). - "Научно-техническая информация. Серия 2. Информационные процессы и системы". М., 1974, № 5, 3-10.
- Перебийнос В. И. О моделировании нулевого стиля. - Linguistica, IV. Tartu, 1972, 96-107.
- Пиотровский Р. Г., Бектаев К. Е., Пиотровская А. А. Математическая лингвистика. М., "Высшая школа", 1977.

- Скороходько Э. Ф. Анализ текстов с применением речевых семантических сетей. - Математическая лингвистика, 1. Киев, Изд. Киевского университета, 1973, 97-103.
- Тулдава Ю. Статистический метод сравнения лексического состава двух текстов. - Linguistica, IV. Tartu, 1972, 199-220.
- Тулдава Ю. Об измерении лексической связи текстов на уровне словаря. - Вопросы статистической стилистики. Киев, "Наукова думка", 1974, 35-42.
- Тулдава Ю. Опыт квантитативного анализа художественного стиля. - Studia metrica et poetica, 1. Учен. зап. ТГУ, вып. 396. Тарту, 1976, 122-141.
- Фрумкина Р. М. Статистические методы изучения лексики. М., "Наука", 1964.
- Харман Г. Современный факторный анализ. М., "Статистика", 1972.
- Шайкевич А. Я. Опыт статистического выделения функциональных стилей. - "Вопросы языкознания", 1968, № 1, 64-76.
- Шкловер Э. Г. О взаимодействии авторской речи и речи персонажей в романах Э. Штрिटтматтера. - Исследования по общему и немецкому языкознанию. Тула, 1974, 73-81.

ОБ ИЗМЕНЕНИИ ЛЕКСИЧЕСКОЙ СВЯЗИ ТЕКСТОВ

Ю. Тулдава

Р е з ю м е

В статье рассматриваются результаты сравнения лексического состава 20 текстов современной эстонской художественной прозы. Анализу подвергаются выборки по 5000 словоупотреблений из авторской речи произведений (выборки состоят из 5 подвыборок по 1000 словоупотреблений из разных мест произведений). Эксперимент охватывает три этапа: тексты сравниваются на основе трех списков слов - 1) 50 наиболее частотных словоформ (в основном местоимения, союзы, наречия; см. табл. 1); 2) 40 избранных полнозначных слов-лексем (16 существительных, 16 глаголов, 8 прилагательных); 3) 40 избранных прилагательных-лексем. В списке полнозначных слов представлены слова неспецифичные, "общечеловеческие", такие как человек, мужчина, женщина, слово, работа, жизнь, смерть, чувство, чувствовать, думать, знать, видеть, смотреть, слышать, слушать, говорить, большой, старий, хороший и т.д. (см. табл. 4). В списке прилагательных представлены антонимичные пары, такие как хороший - плохой, сильный - слабый, радостный - печальный, длинный - короткий, холодный - теплый и др., названия основных цветов и ряд прилагательных общего значения, например, правильный, чистый, свободный, ясный, пустой и др. (см. табл. 6). Методами сравнения и измерения связи (близости) текстов, а также группировки текстов служили корреляционный анализ по Пирсону-Браве и факторный анализ (метод "главных факторов"). Результаты экспериментов приводятся в факторных матрицах и в разных схемах, изображающих взаимное расположение текстов по данным корреляционного анализа и в пространстве выявленных факторов. Результаты количественного анализа комментируются и интерпретируются содержательно. Констатируется, что тексты могут быть близкими по одной группе слов, но далекими по другой, хотя в отдельных случаях наблюдается близость текстов на основе разных групп слов. Наконец, тексты сравниваются на основе всей совокупности выявленных факторов лексической связи (табл. 8). Вполне совпадающих факторных структур не удалось обнаружить, но в отдельных случаях тексты близки между собой на основе двух или трех факторов. Сравнение лексического состава текстов методами корреляционного и факторного анализа позволяет выявить некоторые внутренние, скрытые для простого наблюдения закономерности индивидуальных стилей.

ON THE MEASURING OF LEXICAL CORRELATION OF TEXTS

J. Tuldava

S u m m a r y

The article presents the results of an experiment where the vocabularies of various texts are compared, the texts are grouped and their mutual lexical nearness resp. distance is calculated with the help of correlation and factor analysis. The samples consisting of 5,000 entries each were taken from 20 texts of contemporary Estonian prose fiction (non-conversational material only). The texts were compared on the basis of three various lists of words: 1) the 50 most frequent word forms (mainly pronouns, conjunctions and adverbs; see Table 1); 2) 40 selected words (lexemes) of full meaning (16 nouns, 16 verbs, 8 adjectives); 3) 40 selected adjectives (lexemes). The list of words of full meaning included non-specific words of general human interest, such as man, woman, word, work, life, death, feeling, feel, think, know, see, hear, listen, talk, great, old, good etc. (see Table 4). The list of adjectives consisted of antonymic pairs (good - bad, strong - weak, merry - sad, long - short, cold - warm etc.), the names of the principal colours, and some frequent adjectives of general meaning, e.g. right, clean, clear, free, empty (see Table 6). The results of the experiment are presented in tables and schemes and the quantitative observation is complemented by qualitative analysis. It is stated that the texts may be in good lexical correlation on the basis of one group of words but rather distant on the basis of another group of words. However, in some cases the correlation between a number of texts remains constant throughout various groups of words. In the last stage of the experiment the texts were compared on the basis of all the factors which were established in the course of the factor analysis. No fully coinciding structures were found but in a few cases there was coincidence in two or three significant factors (Table 8). The comparison of the vocabularies of texts with the help of correlation and factor analysis reveals some intrinsic phenomena and regularities of individual styles hidden from simple observation.