

## Puhutun kielen ymmärrettävyydestä

*Kokeellinen tutkimus Yleisradion puheohjelmien ymmärrettävyydestä sekä ymmärrettävyyden ja ohjelman aiheuttaman mielenkiinnon välisestä suhteesta.*

### *Tutkimuksen tarkoitus*

Radion esiintyjille, esitelmäsijoille, opettajille ym. puhutun sanan käyttäjille on luonnollisesti tärkeitä tietää, mitkä tekijät vaikuttavat sanoman ymmärrettävyyteen. Erilaisissa puheoppaissa annetaan ohjeita »selvien» puheiden pitämiseksi, ja radion käyttöön on olemassa erilaisia esiintymisohjeita »korvalle kirjoittamista» varten. Ohjeet ovat kuitenkin yleensä kokemukseen perustuvia, eikä tietävästi ole kovinkaan monta kokeellista tutkimusta, joissa sanoman eri ominaisuuksia muuttamalla olisi selvitetty yksityisten tekijöiden vaikutusta puhutun sanoman ymmärrettävyyteen.

Tässä selostetun tutkimuksen tarkoituksena on tarkastella puhutun sanoman *rakenneominaisuuksien* vaikutusta sanoman ymmärrettävyyteen. Rakenneominaisuuksilla tarkoitetaan tässä sellaisia ominaisuuksia kuin *sanojen ja virkkeiden pituutta, sanaluokkien välisiä suhteita* jne. Sen sijaan ei kiinnitetä huomiota sanoman *sisältöön*, joka tietysti myös vaikuttaa ymmärrettävyyteen. Useinhan sanoman vaikeutta tai helpoutta arvostellaan *pelkästään* sisällön, aiheen ja vaikeuden perusteella.

Jos puhutun kielen ymmärrettävyyttä on kokeellisesti tutkittu varsin vähän, niin sitä enemmän on tutkittu kirjoitetun kielen ymmärrettävyyttä eli luettavuutta. Niinpä tässäkin tutkimuksessa käytetään apuna luettavuuden tutkimuksen menetelmiä.

### *Luettavuus ja »kuunneltavuus»*

Kirjoitetun kielen *luettavuutta* on tutkittu 1920-luvulta lähtien pääasiassa Yhdysvalloissa. Luettavuus tarkoittaa tällöin seuraavia kolmea seikkaa<sup>1</sup>:

- 1) selvä käsiala tai painojälki
- 2) kirjoitustyylin helpous
- 3) sisällön aiheuttama mielenkiinto.

Tutkimus on pääasiallisesti keskittynyt kohtaan 2 eli kielen ymmärrettävyyteen. Tutkimusten mukaan englanninkielisen tekstin ymmärrettävyyteen vaikuttavat

<sup>1</sup> WIIO, OSMO A: Ymmärretäänkö sanomasi? Weilin & Göös 1968, s. 95.

mm. sanojen yleisyys ja pituus, lauseiden pituus, määreiden lukumäärä ja sanaluokkien väliset suhteet. Tunnetuin luettavuuden tutkija on itävaltalais-amerikkalainen RUDOLF FLESCHE, joka on vuodesta 1943 lähtien julkaissut useita ns. luettavuuskaavoja. Kyseessä ovat regressiokaavat, joihin luettavuuteen vaikuttavista tekijöistä valitaan tärkeimmät: Fleschilla tekijöinä ovat esim. sanan ja lauseen pituus. Kaavan avulla voidaan ennustaa tutkittavan tekstin todennäköinen vaikeustaso.

Vastaavaa puhutun kielen ymmärrettävyyttä, »kuunneltavuutta», on tutkittu huomattavasti vähemmän, eikä näissä tutkimuksissa ole yleensä havaittu samalla tavalla selviä korrelaatioita kielen rakenneominaisuuksien ja ymmärrettävyyden välillä. GEORGE KLARE luettelee kirjassaan *The Measurement of Readability*<sup>2</sup> kahdeksan tutkimusta, joista vain kahdessa oli selvä korrelaatio kuunnellun tekstin ymmärrettävyyden ja tekstin rakenneominaisuuksien välillä.

Toinen kirjoittajista on väitöskirjassaan<sup>3</sup> tutkinut suomenkielisen kirjoitetun tekstin luettavuutta. Tutkimuksessa ilmeni samantapaisia riippuvuussuhteita kirjoitetun kielen rakenneominaisuuksien ja ymmärrettävyyden välillä kuin aikaisemmissa englannin kieltä koskevissa tutkimuksissa. Erillisenä työnä kirjoittajat tutkivat puhutun kielen ymmärrettävyyttä samalla menetelmällä, jota oli käytetty väitöskirjassa.

### *Ymmärrettävyyden mittaaminen*

Olellisinta luotettavien tulosten saamiseksi tämäntapaisessa tutkimuksessa on *ymmärrettävyyden* mittaamenetelmä. Tarkoitukseen on käytetty kahta päämenetelmää: arviontia ja ymmärrettävyysoetta. Arvioinnin suorittavat yleensä asiantuntijahenkilöt kuten opettajat, ammattitoimittajat jne. Ymmärrettävyysokeita puolestaan tehdään erilaisille koehenkilöryhmille, jotka lukevat tai kuulevat sanoman. Eri menetelmin mitataan, kuinka sanoma on ymmärretty.

Tässä tutkimuksessa käytetään ensisijaisesti jälkimmäistä menetelmää, ja näin saatuja tuloksia verrataan arviointimenetelmällä saatuihin. Ymmärrettävyyden mittaamiseen käytetään amerikkalaisen WILSON L. TAYLORIN<sup>4</sup> kehittämää ns. *täydennysmenetelmää*. Menetelmän mukaan tutkittavasta sanomasta poistetaan osia, jotka koehenkilöiden tulee arvata. Oikein arvattujen osien lukumäärästä saadaan numeerinen arvo sanoman ymmärrettävyydelle eli ns. täydennysarvo. On varsin luonnollista, että helpon sanoman poissaolevat osat on helpompi arvata kuin vaikean. Taylor on osoittanut, että tulkintatuloksen luotettavuus on paras, jos hyväksytään vain täsmälleen oikein arvatut osat; siis synonyymeja ei hyväksytä.

<sup>2</sup> KLARE, G. R.: *The Measurement of Readability*, Iowa State University Press, 1963, s. 148—53.

<sup>3</sup> WILIO, OSMO A.: *Readability, Comprehension and Readership*, Weilin & Göös. (Ilmestynyt joulukuussa 1968).

<sup>4</sup> TAYLOR, W. L.: *Cloze Procedure: A New Tool for Measuring Readability*. *Journalism Quarterly* 30, 1953, s. 415—33. Sama: *Recent Developments in the Use of Cloze Procedure*. *Journalism Quarterly* 33, 1956, 42—48.

SOLA POOL, I. DE: *Trends in Content Analysis*. University of Illinois, (1958), s. 78—88.

Täydennysmenetelmä valittiin lähinnä sen vuoksi, että tulokset olisivat vertailukelpoisia toisen kirjoittajan väitöskirjan tulosten kanssa. Sitä paitsi on käynyt ilmi, että täydennysmenetelmä saattaa olla luotettavampi sanoman ymmärrettävyyden mitta kuin esimerkiksi ns. vaihtoehtokysely, jota on myös käytetty. MILTON DICKENS ja FREDERICK WILLIAMS<sup>5</sup> ovat osoittaneet, että täydennysmenetelmällä saadut mittaustulokset puhutun sanoman ymmärrettävyydestä ovat riippumattomampia sanoman laadusta ja tilastollisin menetelmin arvosteltuina myös luotettavampia kuin vaihtoehtokysely.

Täydennysmenetelmä on osoittautunut käyttökelpoiseksi paitsi luetun myös kuullun tekstin ymmärrettävyyttä määritettäessä. Taylor (1956 s. 47) antoi koehenkilöiden täyttää kahdesta samaa aihetta käsittelevästä uutistekstistä puuttuvat sanat (joka 14. sana oli korvattu lyhyellä tauolla) ja havaitsi, että »helposti ymmärrettäväksi» kirjoitettu teksti sai selvästi suuremman täydennysarvon kuin »vaikealla kielellä» kirjoitettu »huono» uutisteksti.

### *Esikoe*

Täydennysmenetelmän soveltuvuutta suomenkielisen kuullun tekstin yhteydessä on kokeiltu v. 1965 Yhteiskunnallisen Korkeakoulun (nyk. Tampereen Yliopiston) Lehdistö- ja tiedotusopin laitoksessa. Metodinen esitutkimus toimeenpantiin tämän kirjoittajien johdolla tiedotustutkimuksen praktikumkurssin yhteydessä. Tavoitteena oli selvittää, luokitteleeiko täydennysmenetelmä *a priori* vaikeiksi tai helpoiksi katsotut erilaiset kuullut tekstit oletusten mukaiseen järjestykseen, kuinka ymmärrettäväksi täydennysmenetelmä luokittelee saman tekstin luettuna ja kuultuna, parantaako kuullun tekstin toistaminen sen täydennysarvoa, onko poistettujen sanojen tiheydellä merkitystä kuullun tekstin yhteydessä sekä päädytäänkö täydennysarvon avulla samaan vaikeusjärjestykseen kuin pelkällä koehenkilöiden suorittamalla tekstien vaikeusasteen yleisarvioinnilla.

Koehenkilöinä olivat harjoituskurssin osanottajat, jotka oli sattumanvaraisesti jaettu kahteen 12 hengen ryhmään. Esikoe oli suunniteltu siten, että ryhmien väliset erot ilmaisivat vastauksia ongelmanasettelussa esitettyihin kysymyksiin: toinen ryhmä (koeryhmä) täydensi luetusta tekstistä puuttuvat sanat ja toinen ryhmä (vertailuryhmä) taas saman tekstin kuullusta versiosta poistetut sanat. Koeryhmä kuuli »helpon», »korvalle kirjoitetun» uutistekstin ja vertailuryhmä kuuli saman uutistekstin »vaikeana», »silmälle kirjoitettuna» versiona jne.

Kokeessa käytetyt uutis- ja esitelmätekstit oli otettu valmiista radio-ohjelmista. Muut tekstit oli tätä tutkimusta varten puhuttu ääninauhaan mahdollisimman

<sup>5</sup> DICKENS, M. ja WILLIAMS, F.: An Experimental Application of »Cloze» Procedure and Attitude Measures to Listening Comprehension. Speech Monographs, 31, 1964, s. 103—108.

WILLIAMS, F.: An Experimental Application of the Semantic Differential and »Cloze» Procedure as Measurement Techniques in Listening Comprehension. Julkaisematon väitöskirja, University of Southern California, 1962.

selkeällä ja »neutraalilla» äänellä. Jokaisesta tekstistä poistettiin 20 sanaa. Poistettavat sanat — joka 11. sana — leikattiin ääninauhasta ja korvattiin 5 sekunnin pituisella kohinatauolla, jonka aikana kuuluttajan ääni ilmoitti poistetun sanan järjestysnumeron tekstissä. Näin meneteltäessä puheen luonnollinen rytmi tietysti särkyi, mutta toisaalta pidettiin tärkeänä, että koehenkilöille jäisi kunkin aukon kohdalla aikaa kirjoittaa muistiin puuttuva sana.

Kokeen tulokset voidaan tiivistää seuraavasti:

1) Täydennysmenetelmä luokittelee kuullut tekstit samaan vaikeusjärjestykseen kuin radiotoimittajien *apriorinen* käsitys samojen tekstien vaikeudesta. Yleisradion uutistoimituksen STT:n sähkömateriaalin pohjalta laatima »korvalle kirjoitettu» uutisteksti sai 25 yksikköä suuremman täydennysarvon kuin sama uutisteksti käsittelemättömänä. Helposti seurattavana pidetyn esitelmöitsijän (Eero Saarenheimon, joka puhui havainnollisesti ja lyhyitä lauseita,) tekstin täydennysarvo oli 16 yksikköä suurempi kuin vaikeasti seurattavana pidetyn esitelmöitsijän (Pentti Renvallin, joka käytti abstraktisia käsitteitä ja pitkiä lauseita) tekstin täydennysarvo.

2) Täydennysmenetelmä luokittelee kuullut tekstit samaan vaikeusjärjestykseen kuin luetut tekstit. Kuitenkin kerran kuullun tekstin täydennysarvo oli toistakymmentä yksikköä pienempi kuin (saman) luetun tekstin täydennysarvo. Kun kuultu teksti toistettiin ja puuttuvia sanoja sai siten mieltiä (tai arvata) kahteen kertaan, täydennysarvo suureni n. 10 yksikköä eli suunnilleen samaksi kuin luetun tekstin täydennysarvo.

3) Kun kuullusta tekstistä poistetaan joka 11. sana, tekstin täydennysarvo on 8 yksikköä pienempi kuin tapauksessa, jossa samasta tekstistä on poistettu vain joka 21. sana. Ero selittynee pääasiassa siitä, että jälkimmäisessä tapauksessa koehenkilöille jää enemmän aikaa keskittyä vastausten muistiin merkitsemiseen — eikä niinkään siitä, että hyvin harvassa olevat aukot tekisivät tekstikokonaisuuden selkeämmäksi.

4) Täydennysmenetelmän avulla saatu täydennysarvo ja koehenkilöiden yleisarvioinnit tekstien mielenkiintoisuudesta ja ymmärrettävyydestä tuntuvat korreloivan positiivisesti: mitä suurempi on tekstin täydennysarvo, sitä mielenkiintoisempaan ja helpommin ymmärrettävänä sitä on pidetty. Arvoituksellisenä poikkeuksena tästä yleissäännöstä oli »helppo» uutisteksti, joka sai hyvin suuren täydennysarvon, mutta joka kuitenkin arvioitiin yhdeksi »huonoimminkin ymmärrettävistä» teksteistä, jopa heikommaksi kuin suhteellisen pienen täydennysarvon saanut »huono» uutisteksti.

Esikoe osoitti, että täydennysmenetelmä on hyvin käyttökelpoinen suomenkielisen kuullun tekstin ymmärrettävyyttä mitattaessa. Täydennysarvo näytti mittaavan pääasiassa samaa kuin kuulijoiden yleisarviointi tekstin »ymmärrettävyydestä». Esiin tullut räikeä poikkeustapaus kuitenkin osoittaa, että subjektiivisella ymmärrettävyyden arviolla ja mitatulla ymmärrettävyydellä on myös joitakin eroja. Metodisiksi ongelmiksi esikokeen jälkeen jäivät edelleen tekstistä poistettavien sanojen tarkoituksenmukaisin tiheys sekä poistettavien sanojen tilalle jätettyjen auk-

kojen kesto. Näiden ongelmien selvittäminen kokeilla olisi edellyttänyt varsin laajoja koejärjestelyjä. Siksi ne päätettiin jatkotutkimuksessa ratkaista harkinnanvaraisesti.

### *Jatkotutkimus*

Esikokeen tulosten rohkaisemina kirjoittajat toimeenpanivat v. 1967—1968 laajemman jatkotutkimuksen, jonka tavoitteeksi asetettiin toisaalta koko radion puheohjelmiston kattava yksityisten ohjelmien ymmärrettävyyden määrittäminen ja toisaalta puhutun kielen ymmärrettävyyteen vaikuttavien kielellisten tekijöiden erittely. Lisäksi päätettiin systemaattisesti verrata toisiinsa puheohjelmien valmistamisesta vastaavien toimittajien käsityksiä ohjelmien ymmärrettävyydestä ja samoista ohjelmista koehenkilöiden avulla saatuja »mitattuja ymmärrettävyydsarvoja».

### *Menetelmä*

Tutkittavaksi valittiin n. 2 minuutin pituinen näyte lähes kaikista puheohjelmista, jotka sisältyivät Yleisradion radio-ohjelmaan lauantaina ja sunnuntaina 17.—18. 6. 1967. Usein toistuvista ohjelmista, esim. uutislähetyksistä ja ajankohtaisohjelmien haastatteluista, otettiin mukaan kuitenkin vain muutama sattumanvaraisesti poimittu tapaus. Tutkimusaineistoon valituista 35 ohjelmanäytteestä 14 oli poimittu ajankohtaislähetyksistä (haastatteluja, lausuntoja ja mm. »Pääministerin haastattelutunti»), 8 edusti esitelmiä (mukana myös sanomalehtikatsaus ja »Maamiehen tietolaari»), 7 haastatteluja (erilaisista selostuksista ja viihdeohjelmasta), 2 uutislähetystä ja 1 urheilukilpailuja, kuunnelmia, levypakinoita ja iltahartauksia.

Lisäksi tutkimusaineistoon otettiin mukaan 5 tekstiä radio-ohjelmiston ulkopuolelta. Nämä oli aikaisemmin tutkittu täydennysmenetelmällä luettuina teksteinä. Nyt ne puhuttiin tutkimusta varten ääninauhaan, jotta kuullun ja luetun tekstin välistä ymmärrettävyyttä voitaisiin vertailla. Koko tutkimusaineisto käsitti siis 40 tekstinäytettä.

Tällä kertaa teksteistä päätettiin poistaa joka 8. sana ja yhdestä tekstistä poistettiin aina 15 sanaa. Näin ollen tekstinäytteet saatiin riittävän lyhyiksi, 1—2 minuutin pituisiksi. Se mikä tällöin menetettiin mittauksen tarkkuudessa (15 poistoa on tilastollisesti varsin vähän), voitettiin siinä, että mukaan saatiin enemmän tekstejä. Aukkojen huomattava tiheys (joka esikokeen valossa johtaa yhden kuuleman perusteella hyvin pieniin ja epätarkkoihin täydennysarvoihin) puolestaan korvautui sillä, että kukin teksti päätettiin kuunteluttaa koehenkilöillä kolme kertaa.

Ensimmäinen kuuntelu oli tarkoitettu yleiskuvan saamiseksi tekstistä (vastaa kirjoitetun tekstin ylimalkaista lukua), ja siinä kukin poistettu sana oli korvattu tarkalleen poiston pituisella »kohinapalalla». 5 sekunnin pituisista vakiotaouista luovuttiin, koska niiden katsottiin liiaksi rikkovan puheen rytmiä ja siten ratkai-

sevasti häiritsevän puuttuvien sanojen kohtuullista »arvaamista». Samalla luonnollisesti annettiin tärkeä vihje puuttuvasta sanasta: sen pituus.

Vaihtelevan pituisiin »kohinapätkiin» ei tällä kertaa yhdistetty poiston järjestysnumeroa; sen sijaan itse kuuntelutilaisuudessa kokeen johtaja osoitti karttakepillä taululle piirretyistä numeroista sitä, jota vastaava poisto kulloinkin oli käsillä.

Ensimmäisen kuuntelukerran jälkeen koehenkilöt arvioivat myös tekstin mielenkiintoisuutta ja ymmärrettävyyttä 5-portaisella asteikolla.

Teksti kuunneltiin välittömästi toisen ja kolmannen kerran. Näillä kerroilla tekstiä ei kuunneltu alusta loppuun yhtäjaksoisesti, vaan lauseiden välissä pidettiin 5 sekunnin taukoja puuttuvien sanojen muistiin merkitsemistä varten. Tällaisia taukoja oli keskimäärin 10 kpl tekstiä kohti, ja ne oli sijoitettu tekstin ajatuksenkulun kannalta »luonteviin» paikkoihin.

Yhden tekstin kuuntelu kolmeen kertaan kesti koetilaisuudessa 4—8 minuuttia lukijan puhenopeudesta riippuen. Koko aineiston kuuntelu jaettiin neljään yhden viikon kuluessa pidettyyn koetilaisuuteen, joista kussakin käsiteltiin 10 tekstiä. Koetilaisuudet kestivät 1 1/2—2 tuntia, ja niiden kuluessa pidettiin yksi pitempi väliaika (5 tekstin jälkeen). Koe suoritettiin luentosalissa, johon oli asennettu hyvät äänentoistolaitteet.

### *Koehenkilöt*

Koehenkilöinä toimi 54 Yleisradion ääni- ja kuvatarkkailijakurssin osanottajaa. Nämä ovat pohjakoulutukseltaan ylioppilaita, ja heidän ikänsä vaihtelee 20—25. vuoteen. Koehenkilöjoukko edustaa tavallista opiskelija-aineistoa, mutta ei tietenkään ns. suurta yleisöä. Homogeeniseen — vaikkakin koko kuuntelijakunnan kannalta epäedustavaan — koehenkilöjoukkoon päädyttiin paitsi siksi, että tällainen oli käytännössä helposti saatavissa tutkimuksen tarkoituksiin, myös siksi, että tutkimuksessa pyrittiin nimenomaan selvittämään eri tekstien välisiä eroja ymmärrettävyyden kannalta eikä eri yleisöryhmien välisiä eroja.

### *Tulokset*

Tulokset käsiteltiin siten, että laskettiin jokaisen koehenkilön jokaisen tekstin täsmälleen oikein arvattujen sanojen lukumäärä. Kaikkien koehenkilöiden vastauksista laskettiin kullekin 40 tekstille oikeiden vastausten keskiarvo. Saadut keskiarvot vaihtelivat välillä 1,6—9,1. Kokonaiskeskiarvo oli 5,0 eli viisi oikein arvattua sanaa 15:stä.

Vaikein teksti oli erään kultaseppämestarin haastattelu ja helpoimmaksi osoittautui iltahartaus. Täydennysarvoksi muunnettuna keskimääräinen »ymmärtämisaste» oli 33 % ja parhaimmillaan 61 %. Samantapaisia arvoja on saatu vastaavissa kirjoitetun kielen täydennyskokeissa. Eri tekstien väliset erot olivat varsin selviä, joten menetelmä oli erotteliva.

Verrattaessa kokeellisesti saatuja ymmärrettävyydsarvoja koehenkilöiden itsensä antamiin yleisarviointeihin tekstien ymmärrettävyydestä ja kiinnostavuudesta osoittautui, että kumpikaan subjektiivisista arviomitoista ei ole merkittävässä riippuvuussuhteessa täydennysarvon kanssa. Täydennysarvon ja ymmärrettävyydsarvoinnin välinen korrelaatio on vain —.05. (Kahdella ilmiöllä ei ole keskinäistä riippuvaisuussuhdetta, ja korrelaation arvo on lähellä nollaa; suurta riippuvaisuussuhdetta ilmoittavat arvot, jotka ovat lähellä + tai — 1:tä.) Täydennysarvon ja mielenkiintoisuusarviointin välinen korrelaatio on .11. Keskenään nämä subjektiiviset arvioinnit korreloivat selvästi ( $r = .63$ ). Kuulijoiden omat arvioinnit tekstien ymmärrettävyydestä ja mielenkiintoisuudesta tuntuvat siten heijastavan eräänlaista yleisvaikutelmaa, joka ei riipu tekstien sanatarkasta ymmärtämisestä. Arviomitta heijastanee tekstin kuulemisen helppoutta ja miellyttävyyttä pikemminkin kuin todellista ymmärrettävyyttä.

Kokeellisesti saadun ymmärtämisen astetta osoittavan tuloksen vertaamiseksi tekstin rakenneominaisuuksiin valittiin viisi muuttujaa, joiden voi olettaa aikaisempien tutkimusten perusteella vaikuttavan kuullun tekstin ymmärrettävyyteen. Muuttujat olivat:

1. LS = virkkeiden keskimääräinen pituus sanoina. Lausejako suoritettiin kuulusta tekstistä puhtaaksikirjoitettuun kopioon kieliopillisten ja ajatuksellisten kokonaisuuksien perusteella.

2. MS = määresuhde, jolla tarkoitetaan adjektiivien ja adverbien lukumäärän suhdetta verbien ja substantiivien lukumäärään. Lasketaan kaavasta:

$$100 \times \frac{\text{Adjektiivit} + \text{adverbit}}{\text{Substantiivit} + \text{verbit}}$$

3. PS = pitkät sanat, joilla perusmuodossa on 4 tai useampia tavuja.

4. KT = sellaisten sanojen lukumäärä, jotka esiintyvät kirjassa Karvonen—Takala—Röman—Ylimentalo: Opettajan sanastokirja (Kasvatustieteiden Edistämisseura 1966). Luettelossa on 600 suomen kielen yleisintä sanaa yli 8000 kansakoululaisen aineista.

5. SM = sanaa minuutissa (keskimäärin).

Muuttujien keskiarvot ja hajonnat laskettuna koko aineistosta olivat:

Muuttuja	Keskiarvo	Hajonta
Täydennysarvo	5,0	1,9
LS	7,6	2,2
MS	34,5	13,8
PS	20,4	7,4
KT	48,6	10,6
SM	96,0	12,4

Täydennysarvon kanssa muuttujat korreloivat seuraavasti (Pearsonin tulomomenttikerroin):

	LS	MS	PS	KT	SM
Täydennysarvo	.03	— .19	.00	— .13	— .01

Merkitseviä, odotusten mukaisia riippuvuussuhteita ei tullut esiin. Määresuhteen korrelaatio —.19 täydennysarvon kanssa on *melkein merkitsevä*. Sama korrelaatio-suhde on tullut esiin kirjoitettuja tekstejä tutkittaessa, mutta huomattavasti merkitsevämpänä (Wiio, 1968). Siis mitä enemmän määreitä verbeillä ja substantiiveilla on eli mitä enemmän adjektiiveja ja adverbeja tekstissä on substantiiveihin ja verbeihin verrattuna, sitä vaikeampaa teksti on.

Toinen edes hiukan merkitsevä korrelaatio on täydennysarvon ja KT:n eli opettajan sanastokirjan sanojen välillä (— .13). Negatiivinen korrelaatio tässä tapauksessa on kuitenkin hypoteesin vastainen ja päinvastainen kuin kirjoitetun tekstin tutkimuksissa selvinnyt.

Korrelaatioiden pienenudesta huolimatta kokeiltiin regressioanalyysia, jolloin tarkoituksena oli saada viisi muuttujaa »paremmuusjärjestykseen» sen perusteella, miten niiden avulla voidaan ennustaa täydennysarvoa.

Järjestykseksi tuli:

1. Määresuhde
2. Sanoja minuutissa
3. Virkkeiden pituus
4. KT-sanat
5. Pitkät sanat

Regressiokaavan yhteiskorrelaatio ei kuitenkaan ole merkitsevä, joten järjestys on vain viitteellinen. Käyttökelpoista regressiokaavaa ei tällä aineistolla saatu: täydennysarvolla mitattavaa ymmärrettävyyttä ei pystytty ennustamaan käytetyillä muuttujilla.

Aikaisemmasta luetun kielen tutkimuksesta tähän mukaan otettujen viiden tekstin kirjoitetun kielen ja kuullun kielen täydennysarvojen välinen korrelaatio ei ollut merkitsevä. Luetut ja kuullut sanomat ymmärrettiin eri tavalla.

#### *Vertailututkimus kansalaisopiston oppilaille*

Ymmärrettävyys on luonnollisesti riippuvainen koehenkilön koulunkäynnistä ja kokemuspieristä. Koetulosten yleistettävyyden selville saamiseksi valittiin kontrolliryhmä eri puolilta maata kotoisin olevista ja eri ikäryhmiä edustavista kansalaisopistolaisista, jotka olivat valtakunnallisten kurssien osanottajia. Kokeeseen osallistui 19 henkeä. Näille esitettiin seitsemän sattumanvaraisesti valittua pääkokeissa käytettyä tekstiä. Koejärjestely oli muuten sama kuin edellä esitetystä kokeesta.



Kokeiden täydennysarvojen kokonaiskeskiarvo kansalaisopiston oppilailla oli 3,1 ja samojen tekstien kokonaiskeskiarvo opiskelijoilla 6,3. Kansalaisopistolaiset ymmärsivät siis koetekstit paljon huonommin kuin opiskelijat. Tästä huolimatta tekstien väliset ymmärrettävyyserot olivat molemmilla ryhmillä saman suuntaisia: järjestyskorrelaatio opiskelijoiden täydennysarvon ja opistolaisten täydennysarvon välillä on peräti .96. Ryhmien välinen ero johtuu siis siitä, että opistolaiset ymmärsivät *kaikki* tekstit huonommin, mutta *samat* tekstit olivat kummallekin ryhmälle vaikeita tai helppoja.

Myös kansalaisopistolaiset arvioivat kuulemiensa tekstien ymmärrettävyyttä ja mielenkiintoisuutta. Vaikutelma-arviointien ja täydennysarvon väliset riippuvuussuhteet olivat samanlaiset kuin opiskelija-aineistolla. Täydennysarvo ei korreloi merkittävästi kummankaan subjektiivisen arvioinnin kanssa, mutta arvioinnit ymmärrettävyydestä ja mielenkiinnosta korreloivat merkittävästi keskenään ( $\rho = .88$ ).

#### *Asiantuntijoiden arvioinnit tekstien ymmärrettävyydestä ja mielenkiintoisuudesta*

Tutkimusaineisto oli koottu radio-ohjelmista, joten tutkijoiden mielestä oli aiheellista verrata saatuja tuloksia myös kokeneiden radiotoimittajien arviointeihin siitä, kuinka *ymmärrettäviä* ja *mielenkiintoisia* samat tekstit heidän mielestään olivat. Kaikkien kokeissa käytettyjen tekstien arvioimiseen osallistui viisi Yleisradion puheohjelmien osastolla pitkään työskennellyttä toimittajaa. Toimittajat kuuntelivat kaikkien 40 koetekstin alkuperäiset nauhat, joista ei ollut poistettu sanoja.

*Tulokset.* Yhdenkään toimittajan arvioinnit tekstien *ymmärrettävyydestä* eivät korreloi merkittävästi pääkokeessa saatujen täydennysarvojen eivätkä pääkokeen koehenkilöiden esittämien *ymmärrettävyyсарviointien* kanssa. Sen sijaan jokaisen toimittajan arvioinnit tekstien *mielenkiintoisuudesta* korreloivat merkittävästi koehenkilöiden esittämien mielenkiintoisuusarviointien kanssa ( $r = .48 - .64$ ). Voidaan siis päätellä, että ammattimiesten arvioinnit tekstien mielenkiintoisuudesta pitävät melko tarkasti yhtä ainakin opiskelijaryhmän mielenkiintoisuusarviointien kanssa. Sen sijaan tekstien ymmärrettävyydestä toimittajien käsitys on erilainen kuin kokeelliset tulokset sekä täydennysarvolla että pelkällä arvioinnilla mitattuna.

#### *Tulosten tarkastelua*

Opiskelijaryhmällä saatujen koetulosten varmennus kansalaisopiston oppilailla suoritettulla vertailututkimuksella osoittaa, että tuloksilla on yleispätevyyttä. Täydennysmenetelmässä poistettujen sanojen lukumäärä — 15 — on kuitenkin varsin vähäinen, joten poistojen vaikeusaste ei pääse tasoittumaan saman koetekstin sisällä. Taylor suosittaa 50 sanan poistamista vakavan tuloksen saamiseksi. Näin ollen tässä kokeessa täydennysarvolla saatu ymmärrettävyyden mitta ei ole erityisen luotettava. Useampien poistojen käyttäminen puhutun sanoman tutkimisessa tekee kuitenkin koetilanteen vaikeaksi ja pitkäksi.

Täydennysarvon ymmärrettävyyden vaikutelman vertailu molemmilla koeryhmillä osoittaa, että sanoman »todellinen ymmärtäminen» ja »ymmärtämisen vaikutelma» eivät merkitse samaa. Aikaisemmin ylioppilailta suoritetun *kirjoitetun* testin mukaan sen sijaan täydennysarvo ja ymmärtämisen vaikutelma korreloivat ( $r = .56$ ) merkittävästi<sup>6</sup>. Puhutun sanoman ymmärrettävyyden vaikutelma muotoutuu siis toisella tavalla kuin kirjoitetun sanoman, mitä myös viiden vertailutekstin (luetun ja kuullun version) ymmärrettävyyden pieni korrelaatio osoitti.

Eron voinee selittää siten, että puhuttu kieli välittää enemmän *tunnelmaa* kuin kirjoitettu kieli. Kirjoitettu kieli on *denotatiivista*, siinä sanoilla on hallitsevana niiden varsinainen merkitys: mitä sana tarkoittaa. Puhuttu kieli on *konnotatiivista*, sanoilla on lisänä sivumerkitys: mikä on sanan tunnesisältö. Puhutussa sanomassa puhuja vakuuttaa ääneensävyllään ja persoonallisuudellaan. Kuulija voi olla varma, että hän on sanoman ymmärtänyt, mutta tarkempi tutkimus osoittaa monien yksityiskohtien jääneen ymmärtämättä.

Ymmärrettävyys- ja mielenkiintoarvioiden vertailu tukee edellä esitettyä tulkintaa puhutun ja kirjoitetun kielen ymmärrettävyyden erosta. Opiskelijaryhmillä oli ymmärrettävyys- ja mielenkiintoarvioiden keskinäinen korrelaatio .63, mutta kansanopistoryhmillä .88. Aikaisemmissa *kirjallisissa* kokeissa on näiden arviointien korrelaatioksi saatu ylioppilasryhmillä .14<sup>7</sup>. Tämän mukaan siis kirjoitetussa sanomassa osataan varsin hyvin erottaa toisistaan sanoman aiheuttama mielenkiinto ja sanoman ymmärrettävyys. Puhutussa sanomassa tämä erottelu on vaikeampaa ja tulee sitä vaikeammaksi, mitä vähemmän muodollista koulutusta asianomaisella on.

Englantilainen J. M. TRENAMAN<sup>8</sup> on tutkinut BBC:n televisio-ohjelmien ymmärrettävyyttä ja havainnut, että paljon koulutusta saaneilla ohjelmien ymmärrettävyyteen vaikuttivat kielelliset tekijät, mutta vähän lukeneilla ohjelman dramatisoinnin aste sekä sanoman yhdistyminen esiintyjän henkilöllisyyteen. Samantapaisen tuloksen olivat jo ennen Trenamania saaneet A. H. NIAS ja H. KAY<sup>9</sup> yhden ohjelman vaikutusta tutkiessaan. P. F. VERNON<sup>10</sup> tutki BBC:n opetusohjelmia käyttämällä suurta määrää sekä kielellisiä ominaisuuksia että sisältöä mittaavia muutujia. Hänen mukaansa ymmärrettävyyteen liittyvät ohjelman mielenkiinto- ja sisältötekijät selittivät vaihtelusta 40 %, mutta kielelliset tekijät vain 15 %. Sekä hänen että Trenamanin tutkimustulosten mukaan ohjelman käsitteellisyys taso (abstraktinen-konkreettinen) oli tärkein ymmärrettävyyteen vaikuttava tekijä.

Eri menetelmällä saadut mittaustuloksemme tukevat edellä selostettuja englantilaisia havaintoja. Radion kautta välitetyn puhutun sanoman ymmärrettävyyteen

<sup>6</sup> WIIO, OSO A.: Luettavuus. Julkaisematon liseniaatintyö, Tampereen yliopisto, 1966, s. 57.

<sup>7</sup> Kuten edellä, s. 57.

<sup>8</sup> TRENAMAN, J. M.: *Communication and Comprehension*. Longmans, 1967, s. 110.

<sup>9</sup> NIAS, A. H. W. ja KAY, H.: *Immediate memory of a broadcast feature programme*. *British Journal of Educational Psychology*, 24, marraskuu 1954.

<sup>10</sup> VERNON, P. F.: *An Investigation into the Intelligibility of Educational Broadcast*, BBC, 1950.

tuntuvat kielelliset tekijät vaikuttavan vähän, joskin vaikutus on saman suuntainen kuin kirjoitetun sanoman kohdalla. Ohjelman aiheuttama mielenkiinto sekä ohjelman »näytelmällisyys» ovat ratkaisevia ainakin vähän koulutusta saaneilla kuulijoilla. Tarkempaa tutkimusta kuitenkin vaatisi hyvin vaikean tekstin vaikutus mielenkiintoon.

Tutkimuksesta voi tehdä erään ohjelmapoliittisen päätelmän: ammattitoimittajien käsitys ohjelman ymmärrettävyydestä ei aina vastaa todellisuutta, vaikka toimittajat ilmeisesti osaavatkin hyvin arvioida ohjelman aiheuttaman mielenkiinnon. Tutkimuksemme mukaan opiskelijaryhmällä ja kansalaisopistoryhmällä on hyvin samantapaiset käsitykset ohjelmien ymmärrettävyydestä ja mielenkiintoisuudesta, mutta toimittajien käsitys *ymmärrettävyydestä* eroaa kummankin muun ryhmän käsityksistä. Koetuloksista ei voi päätellä, että toimittajat olisivat yli- tai aliarvioineet kuulijat, sillä toimittajat ja kuulijat pitävät *eri* ohjelmia vaikeina tai helppoina.

Samantapaisia tuloksia on saatu niissä amerikkalaisissa tutkimuksissa, joissa on verrattu toisiinsa asiantuntijoiden, suuren yleisön ja ammattitoimittajien käsityksiä eräistä tieteeseen liittyvistä asioista. Esimerkiksi mielisairauksiin suhtautuivat asiantuntijat ja yleisö suunnilleen samoin, mutta molempien käsitykset erosivat toimittajien arvioinneista.<sup>11</sup> Samoin tiedemiesten, tiedetoimittajien ja yleisön arvioinnit poikkesivat ammattitoimittajien arvioinneista arvosteltaessa eri tieteenalojen saavutuksia. Vaikuttaa siltä, että toimittajien koulutus ja ammattikokemus painottavat niin suuressa määrin sanoman *mielenkiintoisuutta*, että tämä arviointi värittää myös käsitystä sanoman ymmärrettävyydestä. USA:ssa tehtyjen tutkimusten mukaan jopa käsitys mielenkiintoisuudestakin oli vinoutunut painottamaan poikkeuksellisuutta ja epätavallisuutta. Kärjistetysti voisi sanoa, että toimittajat tekevät mielenkiintoisia ohjelmia, joita seurataan, mutta joita ei todellisuudessa ymmärretä. Viestintä on pääasia, ei tieto, jota viestinnällä pitäisi välittää.

<sup>11</sup> TANNENBAUM, P. H.: Communication of Science Information. Science, 140/3567, 1963, s. 580—83.

Kirjoittajat haluavat lausua erityiset kiitokset Suomen Kulttuurirahastolle, jonka apuraha on tehnyt tämän tutkimuksen mahdolliseksi, sekä Oy Yleisradio Ab:lle koetilaisuuksien ja tutkimusaineiston järjestämisestä.

## Listenability

An experimental study of the comprehension and interest of Finnish radio programs

by OSMO A. WIIO—KAARLE NORDENSTRENG

The *readability* of written messages has been the subject of many studies but little is known of the comprehension or *listenability* of spoken messages. The purpose of this study was to find out to what extent the structural characteristics of spoken language correlate with the comprehension of messages. The research method used was WILSON L. TAYLOR's 'cloze procedure', in which parts of the messages are deleted and the subjects try to guess the missing parts. DICKENS AND WILLIAMS (1964) have shown that the cloze procedure is a useful and reliable method of measuring the comprehension of spoken language. The most suitable technique for the study was chosen after a preliminary test with 24 university students.

Thirty-five excerpts of about two minutes each were copied from most speech programs which were broadcast on the 17th and 18th of June, 1967, through the network of the Finnish Broadcasting Corporation. Thus 35 samples were obtained; in addition, there were 5 control samples.

Every eighth word was deleted, the total number of the deletions in each sample being 15 words, and substituted with a »white noise» of equal length. The subjects were to guess the missing words; in addition, they had to fill a special »connotative test», in which they were asked to evaluate the relative *difficulty* of each test passage according to a five-step scale. A similar scale was also used to test the *interest* aroused by the program.

*Subjects.* The subjects were 54 students participating in a sound and picture controller course arranged by the FBC. A homogeneous group was selected because the main purpose was to study the comprehension of the types of messages regardless of the educational or occupational level of the subjects. This aspect was, however, included in the study as a separate experiment with 19 persons of lower education.

*Tests.* The tests were administered to the group of students in four 90-minute sessions. The passages were played three times, and 10 pauses of 5 seconds were inserted in the passages to give the subjects time to write down their answers. The experiment with 19 persons took place in one session, and it included seven passages picked at random from among the 40 test passages.

*Results.* The cloze scores, i.e. the number of the right guesses for each test and person, and the means of the individual cloze scores for each test were computed. The means varied between 11 % and 61 % for the students and between 3 % and 42 % for the control group.

Comparison of the comprehension estimates given by the cloze score with the judged comprehension score showed that the correlation was not significant for the students or for the control group. Similarly, the correlation between the cloze score and the interest judgement was not significant for either group. On the other hand, the correlations between the two judgements (comprehension and interest scales) were significant for both groups,  $r = .63$  for the students and  $\rho = .88$  for the control group.

The cloze scores were then correlated with five message variables, which were:

1. Number of sentences per 100 words.
2. Modification ratio ( $100 \times \text{adjectives} + \text{adverbs} / \text{nouns} + \text{verbs}$ ).
3. Long words, i.e. the number of the words with a root of 4 or more syllables per 100 words.
4. Easy words from a list of the 600 most common words used by Finnish school children.
5. Words per minute.

The following correlations were found between the cloze score means and the message variables (students):

Sentences	.03
Modification ratio	— .19
Long words	.00
Easy words	— .13
Words / minute	— .01

No significant correlations were found; the correlation between the cloze score and the modification ratio was *nearly significant*. A similar but more significant correlation was found by one of the writers in the analysis of written messages. This means that the more adjectives and adverbs the spoken message contains the more difficult it is.

A regression analysis gave the variables the following order of relative importance: modification ratio, words per minute, sentence length, easy words, and long words.

The formula does not, however, reach the level of significance, and listenability can not be predicted with the variables available.

The students and the control group comprehended the corresponding test passages in a very similar way; the control group reached a lower level of comprehension but the rank order correlation between the two cloze scores was .96.

*Expert judgement.* Five experienced program editors of the broadcasting corporation also judged the 40 test passages through a connotative test. The original tapes without deletions were played to them. The correlations between their judgements and the cloze scores and comprehension judgements of the two test groups are not significant. On the other hand, the correlations between the *interest* judgements of the editors and those of the test groups are significant ( $r = .48-.64$ ). This shows that the program editors judge the interest aroused by the program quite well but can only poorly predict its comprehension by the listeners.

*Conclusions.* In an earlier study (with a similar test and group), which was concerned with written messages, one of the writers found a correlation of .56 between the judged comprehension and the cloze score. In the present study the same correlation was not significant. It seems that there is a difference between

spoken and written messages: a speaker may create a feeling of understanding without his message being really understood. In writing this is more difficult: written language is more denotative and less connotative than spoken language.

In the present study, the correlation between interest and comprehension was significant, while in the previous tests with written messages it was not so. It seems that it is more difficult to separate the two things — interest and comprehension — in a spoken message than in a written message, and it becomes even more difficult when persons with lower education are concerned. The results support those obtained by TRENAMAN through a study of the comprehension of the television programs of the BBC. He found out that as for persons with a high level of education the linguistic elements of the message are important for comprehension while for persons with a lower level of education the content factors are of greater importance. For persons with low education the degree of dramatisation and concreteness are the most important elements of comprehension.

The results also suggest that the program editors seem to value the *interest* aspect of the message so much that they misjudge its comprehension.