

SANANTUNNISTUSTESTIN PERIAATTEET

Sanantunnistustestin avulla voidaan arvioida testattavien sanaston laajuutta. Testi perustuu tietoon sanojen yleisyydestä ja sen vaikutuksesta sanojen oppimiseen: Mitä yleisempi sana on, sitä useammin kielenoppija kohtaa sanan ja sitä paremmin hän oppii kyseisen sanan. Tähän perustuen voidaan henkilön sanaston laajuutta arvioida sen mukaan, kuinka paljon ja kuinka harvinaista sanastoa hän tuntee (Nation 2001; Milton 2009). Suurin osa aihepiirin tutkimuksesta on tehty englannin ja muutamien muiden suurten kielten aineistosta, mutta myös suomen kielellä tehdyt tutkimukset osoittavat, että sanaston yleisyyspohjainen arviointi toimii (Honko 2013; Tervola 2019).

Sanaston laajuuden on useissa tutkimuksissa osoitettu olevan vahvin yksittäinen yleisen kielenkäyttötaidon indikaattori. Etenkin lukemisessa sanaston laajuudella on suuri merkitys. Siksi sanaston laajuutta mittaavia testejä käytetään paljon juuri korkeakouluohjelmien pääsykokeissa (ks. esim. Lam 2006; Harrington & Michael 2009).

Tämän sanatestin sanat on poimittu Finnish Internet Parsebankin korpuksesta (FIP), josta on saatu myös sanojen yleisyystiedot. Korpukseseen kerätään kaikki Internetin suomenkieliset tekstit, joten korpus ei rajoitu minkään tietyn aihepiiriin tai rekisteriin. Tätä testiä varten korpuksen sanalistasta on poimittu kaikki yleiskieliset sanat – puhekieliset, slangisanat, murteelliset sekä vanhahtavat sanat on jätetty pois. Myös pronominit ja konjunktiot on jätetty pois. Testissä on siis lähinnä substantiiveja, verbejä, adjektiiveja ja joitakin partikkeleita. Testin kaikilla tasoilla on sekä konkreettista sanastoa (avata [taso 1], hius [2], saari [3], terävä [4] ja vaaka [5]) että abstraktimpaa yleisakateemista sanastoa (tarkoittaa [1], muodostaa [2], hallinta [3], oikeuttaa [4] ja edustus [5]). Testi sopii siis juuri mittaamaan korkeakoulukontekstissa tarpeellisen sanaston tuntemusta.

Testiä voi käyttää kuka hyvänsä, sillä se testaa suomen kielen yleisimpien sanojen tuntemusta. Testin tekeminen kuitenkin vaatii kirjoitetun kielen tuntemusta ja sanan perusmuotojen hahmottamista, joten se ei sovellu henkilöille, joilla on hyvin vähän

suomenkielistä kirjoitus- ja lukukokemusta. Testi on suunnattu erityisesti korkeakouluopintoja harkitseville aikuisille kielenkäyttäjille.

TESTIN TULOKSEN TULKINTA

Testin tasoista muodostuu yhdessä 6600 suomen kielen yleisimmän sanan tuntemuksen mittari. Kyseinen sanasto kattaa noin 75 % korpuksesta, eli keskimäärin joka neljäs tekstissä vastaantuleva sana on vielä harvinaisempi.

1. taso: 600 yleisintä sanaa, tekstipeitto noin 50 %
2. taso: sanat yleisyysväliltä 600.–1200., tekstipeitto noin 60 %
3. taso 1200.–2000., tekstipeitto noin 65 %
4. taso 2000.– 4000., tekstipeitto noin 70 %
5. taso 4000.–6600., tekstipeitto noin 75

Tätä sanatestiä ei ole vielä testattu oikeilla käyttäjillä, joten tulkintaohjeet perustuvat muihin sanaston laajuudesta tehtyihin tutkimuksiin. Yleisesti sanaston koon perustasona pidetään 2000 yleisimmän sanan tuntemusta, joka on vähimmäisvaatimus akateemisen opiskelun sujumiselle (Nation 2001: 15). Samansuuntainen on Miltonin (2009: 190) arvio, jonka mukaan Eurooppalaisen viitekehysten keskitasolla sanaston tulisi kattaa noin 3000 sanaa. Tähän perustuen voidaan arvioida, että taso 3, joka mittaa 2000 yleisimmän sanan tuntemusta, on hyvä lähtökohta, kun ryhdytään pohtimaan sanaston laajuuden ja sitä kautta kielitaidon riittävyyttä korkeakouluopintoihin. **Jos tasolla 3 henkilö saa pääsääntöisesti yli 90 %:n tuloksia, voidaan katsoa, että hänellä ei pitäisi olla kielitaidon puolesta estettä päästä aloittamaan korkeakouluopintoja suomen kielellä.** Tälläkin tasolla korkeakouluopiskelija tarvitsee silti paljon kielellistä tukea opiskelussaan. Alemmat tasot 1 ja 2 ovat apuna määriteltäessä, kuinka kaukana perustason saavuttaminen on, eli tuleeko sanaston suppeus ilmi vasta 3. tasolla vai jo alemmilla, yleisemmän sanaston tasoilla. Jos taas 3. tason suorittaa yli 90-prosenttisesti, neljättä ja viidettä tasoa kannattaa kokeilla sillä ajatuksella, kuinka vaikeaa suomenkielinen opiskelu mahdollisesti olisi. Mitä paremmat

tulokset saavuttaa 4. ja 5. tasoilla, sitä vaivattomampaa opiskelu suomen kielellä korkeakoulussa todennäköisesti on.

Ohjaustilanteessa voidaan käydä läpi opiskelijan kielellistä historiaa (esimerkiksi suomen kielen käyttökokemus vuosissa, kielikurssit ja suoritettut tutkinnot, suomen kielen käyttötottumukset vapaa-ajalla) ja havainnoida keskustelun sujumista ohjaustilanteessa, esimerkiksi onko ohjaajan tarvetta muokata puhettaan ymmärtämisen varmistamiseksi vai onnistuuko keskustelu helpottamattomalla puheella. Sanantunnistustesti tuo kielitaitokeskusteluun hieman syvemmän tason. Se kertoo eräänlaista hiljaista tietoa kielitaidosta, joka ei tule näkyviin keskustelussa.

Testin on laatinut Tampereen yliopiston Suomen kielen ja kulttuurin yliopisto-opettaja Maija Tervola. Parannusehdotukset ja muun palautteen saa mielellään lähettää osoitteeseen maija.tervola@tuni.fi.

Lähteet:

FIP = Finnish Internet Parsebank: Lexical search <https://bionlp.utu.fi/finnish-internet-parsebank.html>. Sanojen yleisyystiedot tarkistettu 05–12/2020.

Harrington, Michael & Carey, Michael 2009: The on-line Yes/No test as a placement tool. *System* 37, s. 614–626.

Honko, Mari 2013: Alakouluikäisten leksikaalinen tieto ja taito. Toisen sukupolven suomi ja S1-verrokki. *Acta Universitatis Tamperensis* 1865. Tampereen yliopisto.

Lam, Yvonne 2006: Yes/no tests for foreign language placement at the post-secondary level. – *Canadian Journal of Applied Linguistics / Revue canadienne de linguistique appliquée* 13(2), 54–72.

Milton, James 2009: *Measuring Second Language Vocabulary Acquisition*. Multilingual Matters.

Nation, I. S. P. 2001: *Learning Vocabulary in Another Language*. Cambridge University Press.

Tervola, Maija 2019: Maahanmuuttajalääkärien suomen kielen taito ja kielitaitotarpeet. Erityisalan kielitaidon näkökulma lääkärin työhön. Tampereen yliopiston väitöskirjat 92. Tampereen yliopisto.