

Kieliteknologian korkeakouluopetus:  
Tarve, tarjonta ja ehdotukset toimenpiteiksi

1. syyskuuta 1999

# Sisältö

|   |           |
|---|-----------|
| <b>1 Johdanto</b>   | <b>2</b>  |
| 1.1 Kieliteknologian rajaus . . . . .                                   | 2         |
| 1.2 Kieliteknologia muissa maissa . . . . .                             | 5         |
| <b>2 Kieliteknologiaan liittyvän opetuksen tarjonta Suomessa</b>        | <b>7</b>  |
| <b>3 Kieliteknologian tilanteen ja mahdollisuuksien arviointia</b>      | <b>9</b>  |
| <b>4 Kieliteknologian koulutuksen tarve</b>                             | <b>11</b> |
| 4.1 Puheteknologian opetus . . . . .                                    | 12        |
| 4.2 Käyttöliittymien kieliteknologian opetus . . . . .                  | 13        |
| 4.3 Dokumenttien hallinnan ja tiedon haun opetus . . . . .              | 14        |
| 4.4 Kielenopettamisen ja -oppimisen teknologia . . . . .                | 15        |
| 4.5 Kääntämisen apuvälineiden opetus . . . . .                          | 16        |
| 4.6 Kieliteknologian menetelmien ja teorian opetus . . . . .            | 16        |
| 4.7 Työmarkkinatilanne . . . . .  | 17        |
| <b>5 Ehdotukset toimenpiteiksi</b>                                      | <b>19</b> |
| 5.1 Kieliteknologian ammattilaisten koulutuksen järjestäminen . . . . . | 20        |
| 5.2 Kieliteknologiaa hyödyntävien ammattilaisten koulutus . . . . .     | 20        |
| 5.3 Kieliteknologiaa lähellä olevien alojen koulutus . . . . .          | 20        |
| 5.4 Kieliteknologian opetuksen ohjaus ja koordinointi . . . . .         | 21        |
| 5.5 Kurssien yhteisen hyödyntämisen periaatteista . . . . .             | 21        |
| 5.6 Tarvittavia panostuksia . . . . .                                   | 22        |
| 5.7 Kustannusarvio . . . . .  | 23        |

# Luku 1

## Johdanto

### 1.1 Kieliteknologian rajaus

Kieliteknologia on ihmisten kieleen kohdistuvaa tai sitä hyväksikäyttävää teknologiaa mukaan lukien teorialat, joihin se perustuu. Kieliteknologia on tutkimuksen, korkeakouluopetuksen ja teollisen toiminnan kohteena verraten uusi, laaja-alainen ja rajoiltaan epätarkka alue. Siitä ja sen erinäisistä osa-alueista on käytetty ja käytetään edelleen erilaisia muitakin nimiä, kuten *tietokonelingvistiikka* (engl. *computational linguistics*, ruotsiksi *datalingvistik* tai *datorlingvistik*), luonnollisen kielen käsittely (engl. *natural language processing*, *NLP*) tai *matemaattinen kielitiede* ja usein osia siitä on sisällytetty esimerkiksi tekoälyn tutkimukseen. Euroopan yhteisö on tutkimusohjelmissaan käyttänyt alasta esim. englanninkielisiä nimityksiä *Language Research and Engineering (LRE)*, *Language Engineering*, ja käyttää nykyään nimitystä *Human Language Technology*. Kielitieteellinen tietojenkäsittely (linguistic computing) on myös läheisessä suhteessa kieliteknologiaan.

Tekesin rahoittama esiselvityshanke on julkaissut vuonna 1998 raportin *Kieliteknologia Suomessa*, toim. Manne Miettinen, CSC Raportti R02/98, joka selvittää kieliteknologian nykytilaa ja mahdollisuuksia Suomessa sekä tutkimuksen että yritystoiminnan kannalta. Tämä raportti on varsin hyödyllinen lähde arvioitaessa minkälaisia asioita kieliteknologiaan kuuluu.

Kieliteknologian tehtävää voi hahmotella luonnollisten kielten automaattisen käsittelyn teorian ja menetelmien (tietokonelingvistiikan) sovelluskohteiden kautta. Tällöin voidaan nimetä esim. seuraavia aihepiirejä, joista monet tunnetaan myös muiden oppiaineiden tai niiden haarojen osina tai sovelluskohteina (näitä on mainittu suluissa):

- luonnollisen kielen (kuten englannin tai suomen) käyttö tietokoneen käyttöliittymissä, mikä usein edellyttää kykyä tunnistaa ja jäsentää ihmisen tuottamaa (kirjoitettua tai puhuttua) kieltä ja/tai kykyä muuntaa tietokoneen vastaukset luonnolliseksi kieleksi (tietojenkäsittely, tekoäly, kognitiotiede, diskurssin tutkimus),

- dokumenttien ja tekstimuotoisen tiedon hallinta, tiedonhaku vapaamuotoisista (myös monenkielisistä) teksteistä, tekstien indeksointi, tekstin tiivistelmän automaattinen tuottaminen, mikä usein edellyttää sanojen taipumisen ja pääpiirteittäisen lauserakenteen tunnistamista (informaatiotutkimus, rakenteisten dokumenttien tutkimus, hahmontunnistus, tietokantojen tutkimus, tiedonhaku algoritmit, neuraalilaskenta, tekstilingvistiikka),
- erilaiset kirjoittajan apuvälineinä toimivat tietokoneohjelmat kuten oikeinkirjoituksen ja oikeakielisyyden tarkistus, tavutus (leksikografia, kielenhuolto, automaattien teoria),
- välineet kielenkääntämisen tueksi sekä automaattiset kielenkääntöohjelmat ja tietokoneohjelmien sopeuttaminen eri kielten ja kulttuurien mukaisiksi (käännöstiede, terminologian tutkimus),
- puhesynteesi ja automaattinen puheentunnistus (puheen käsittely, digitaalinen signaalinkäsittely, fonetiikka, akustiikka, hahmontunnistus, neuraalilaskenta),
- tietokoneavusteinen (tai teknologia-avusteinen) kielenoppiminen ja -opettaminen (soveltava kielitiede, kielten didaktiikka, multimedia) sekä
- sanakirjojen ja oppimateriaalien laatiminen ja toimittaminen, tekstien luettavuuden parantaminen (leksikografia, kasvatustiede, käytettävyytystutkimus).

Kaikki kieleen liittyvät sovellukset eivät kuitenkaan ole keskeistä kieliteknologiaa. Esimerkiksi tekstinkäsittelyssä käyttäjä kirjoittaa kieltä, jota ohjelma myös käsittelee ja tallettaa, ilman että tekstinkäsittelyohjelmaa pitäisi tämän vuoksi pitää kieliteknologisena sovelluksena. Esimerkiksi oikeakielisyyden tarkistustoiminto, joka *mallintaa kieliopin ja sanakirjojen tietoja*, on vasta varsinaista kieliteknologiaa. Myös kielenopiskeluohjelma voi olla olennaisilta osiltaan kuten ääni- ja videonauhuri sekä piirtoheitin, ilman varsinaista toimivaa tietoa tai mallia kielen rakenteesta. Kieliteknologiseksi tällainen sovellus muuttuu, kun ohjelmaan liitetään tietoa kielen toiminnasta ja rakenteesta, jolloin ohjelma voi neuvoa oppilasta oikeaan suuntaan.

Termillä **tietokonelingvistiikka** on usein korostettu kieliteknologian tutkimusluonnetta,<sup>1</sup> jolloin haasteena on kuvata ihmisen (yleensä kirjoitetussa muodossa olevan) kielen rakennetta ja toimintaa tietojenkäsittelytieteen ja matematiikan malleilla ja menetelmillä. Tavoite on silloin lähellä yleisen kielitieteen tavoitteita, mutta menetelmät ja teoriat ovat toisenlaisia, koska niiltä vaaditaan tietokoneen edellyttämää eksplisiittisyyttä ja konkreettisuutta. Erilaiset tällaisen tutkimuksen piirissä syntyneet *kielenkäsittelyn mallit, kuten sananmuotojen, lauserakenteiden ja merkitysten jäsentämisen menetelmät muodostavat edelleen kieliteknologian ytimen*.

<sup>1</sup>Helsingin yliopiston humanistisen tiedekunnan oppiaine ”tietokonelingvistiikka” muutettiin asetuksella 1.8.1999 alkaen ”kieliteknologia”-nimiseksi oppiaineeksi.

**Ihmisen ja tietokoneen vuorovaikutus** on olennainen osa tietotekniikkaa ja siinä kieli-tekniologia on mukana monella tasolla: sekä ihmisen vuorosanojen tulkitsemisessa, että tietojärjestelmän vastausten muuntamisessa ihmisten kielelle. Sekä puheen että tekstin muotoisen kielen analysoinnin laatu nousee tärkeäksi rajoittavaksi tekijäksi monissa järjestelmissä. Koneen ja ihmisen onnistunut dialogi ja järjestelmien hyvä käytettävyys edellyttää monipuolista kielen ja dialogin toiminnan ymmärrystä.

**Puhetekniologia** on noussut erääksi informaatiotekniikan keskeisistä alueista ja on myös esimerkki aidosti monitieteisestä alasta. Tämän selvityksen yhteydessä voidaan puhetekniologiaa pitää eräänä kielitekniologian tärkeimmistä alueista, kun puhetta tarkastellaan kielenä. Puhetekniologian osia, jotka liittyvät läheisesti esimerkiksi tietoliikennetekniikkaan (kuten puheenkoodaus ja siistaus) jää kuitenkin kielitekniologian ulkopuolelle.

Organisaatioiden tiedon hallinnasta ja ihmisten välisestä viestinnästä merkittävä osa tapahtuu digitaalisten dokumenttien välityksellä. Suurin osa dokumentteihin sisältyvästä tiedosta on tallennettu luonnollisten kielten avulla, useimmiten tekstinä, mutta äänen osuus on kasvussa. **Tiedonhaku** tai haluttua tietoa sisältävien dokumenttien etsintä on tärkeä osa monia normaaleja tehtäviä. Tietomäärien kasvaessa **automaattinen tekstin tiivistäminen, dokumenttien luokittelu, indeksointi, suodatus ja reititys**<sup>2</sup> tulevat yhä keskeisemmiksi ja tarpeellisemmiksi. Näiden tehtävien ratkaisemiseksi on käytetty toisaalta kielitekniologisia menetelmiä, toisaalta (varsinkin englannin kielen kohdalla) puhtaasti tilastollisia menetelmiä ja usein molemmanlaisia menetelmiä yhdessä. Informaatiotutkimuksen ja kielitekniologian suhteen tässä esiintyy samanlaista yhteistä aluetta kuin puhetekniologiassakin tietoliikennetekniikan ja kielitekniologian välillä.

Kielitekniologia liittyy **rakenteisiin dokumentteihin ja multimediaan** usealla tavalla. Rakenteiset dokumentit voivat sisältää tekstiä, kuvaa, ääntä tai linkkejä toisiin dokumentteihin. Niiltä osin, kun rakenteinen dokumentti sisältää kieltä, voidaan kielitekniologian menetelmillä tuottaa esimerkiksi linkkejä, metatietoa dokumentin luokitteluksi tai löytymisen helpottamiseksi. Toisaalta myös rakenteisen dokumentin keinoin voidaan kielitekniologian menetelmillä löydettyä kielellistä rakennetta (esim. sanojen perusmuotoja, avainsanoja tai termejä) helposti merkitä myöhempää hyödyntämistä varten.

**Kielenoppimisen ja -opettamisen tekniologia** on hyvin monitieteinen teema, joka koskettaa lähitulevaisuudessa mahdollisesti hyvin suuria ihmismääriä, mm. koska Euroopan unioni on poliittisestikin sitoutunut monikielisyyteen. Oppimateriaalin valmistamisessa, aktiivisen kielenkäytön harjoitteluvälineissä (elektronisissa työkirjoissa, kielenoppimispeleissä, ym.) voidaan enenevässä määrin hyödyntää kielitekniologiaa.

Tietokoneen ja kielitekniologian mukanaolon vuoksi itse luonnolliseen kieleen ja siihen, miten sitä käytämme, kohdistuu uutta mielenkiintoa. Esim. terminologiset työkalut auttavat ihmistä kirjoittamaan yksiselitteisempää tekstiä ja dialogin tutkimuksen menetelmät auttaisivat tekemään tietokoneohjelmien viestejä ihmiselle helpommin ymmärrettäviksi.

<sup>2</sup>Reititys (engl. routing) tarkoittaa tässä dokumenttien ohjailua sisällön perusteella sopiville vastaanottajille.

Yksinkertaistetut, ns. kontrolloidut (luonnolliset) kielet ansaitisivat itsessään lisää tutkimusta, koska niiden avulla voidaan joissakin tapauksissa päästä toimivampaan koneen ja ihmisen väliseen kommunikaatioon ja mm. monikielisiin sovelluksiin.

## 1.2 Kieliteknologia muissa maissa

Pohjoismaissa kieliteknologia (tietokone-lingvistiikka) esiintyy omana oppiaineenaan mm. Tukholman yliopistossa (datorlingvistik), Uppsalan yliopistossa (språkteknologiprogrammet, mutta jatkokoulutus nimellä datorlingvistik), Göteborgin yliopistossa (datalogvistik), Bergenin yliopistoissa (datalogvistik) ja Syddansk universitetissa (sproglig informatik). Monessa korkeakoulussa on yleisen kielitieteen tai muun aineen yhteydessä (ilmeisesti resursseiltaan vaatimattomampi) kieliteknologian koulutusohjelma kuten Oslon yliopistossa (språk, logikk og informasjon, språkinformatik), Linköpingin yliopistossa tietotekniikan yhteydessä (språkinformatik). Oppiaine (datalogvistik) ja vastaava oppituoli löytyy myös Norges teknisk-naturvitenskapelige universitetista (NTNU) Trondheimista. Puheteknologiaa opetetaan useissa pohjoismaisissa teknillisissä korkeakouluissa (esim. Kungliga Tekniska Högskolan, NTNU) ja se liittyy usein tietoliikennetekniikkaan tai digitaaliseen signaalinkäsittelyyn (näitä ei ole sisällytetty seuraavaan arvioon).

Vuoden 1998 syksyllä suoritetun (epävirallisen) kyselyn perusteella saatiin arvio eri pohjoismaiden kieliteknologian alueen korkeakouluopetuksen laajuudesta, joka on tiivistetty ohaiseen kuvassa 1.1 olevaan taulukkoon:

|                              | Norja | Ruotsi | Suomi | Tanska |
|------------------------------|-------|--------|-------|--------|
| sisäänotto kieliteknologiaan | 9     | 50     | 5     | 28     |

Kuva 1.1: Arvio yliopistotasaisen perustutkintoon johtavan kieliteknologian koulutuksen laajuudesta Pohjoismaissa

Saksassa kieliteknologian opetusta on toisaalta omalla nimellään (Computerlingvistik), mutta suurin osa opetuksesta sijainnee tietotekniikan tai tekoälyn osana. Saksassa on panostettu voimakkaasti kieliteknologian tutkimusohjelmiin. Näistä merkittävin on Verbmobil-niminen suurhanke, joka on useampivaiheinen ja yhteensä lähes 1000 henkilötyövuoden laajuinen, jossa suuret teollisuusyritykset ovat olleet myös mukana.

Belgiassa toimii Lernout & Hauspie -niminen kieliteknologian alalla huomattavan kokoinen yritys, jonka liikevaihto vuonna 1998 oli noin 211 miljoonaa USD, ja jolla oli noin 1700 työntekijää. Yritys kehittää kaikenlaista kieli- ja puheteknologiaa, mutta toistaiseksi huomattava osa toiminnasta on tavanomaisempaa, esim. kielenkääntämisen palvelua. Belgian valtio on tukenut L&H:ta ja muita Flanders Language Valleyn yrityksiä ja tutkimuslaitoksia merkittäväillä summilla.

Monet telealan yritykset, erityisesti laitevalmistajat (esim. Nokia, Ericsson, Siemens) mutta myös operaattorit (esim. Sonera, Telia) ovat jo usean vuoden ajan kehittäneet erityisesti matkapuhelinviestintään liittyvää puheteknologiaa. Tämä työ on kohdistunut varsinkin puheenkoodaukseen ja -tunnistukseen. Näiden yritysten tuotekehitysyksiköissä työskentelee tällä hetkellä useita satoja puheteknologiaan erikoistuneita suunnittelijoita.

Euroopan unioni on panostanut melko runsaasti lähes kahdenkymmenen vuoden aikana erilaisiin kieliteknologian tutkimus- ja kehitysohjelmiin. Nyt käynnistyneeseen 5. puiteohjelmaan sisältyy informaatioyhteiskunnan teknologia -niminen osa, johon sisältyy median sisällön ja työkalujen osa, jonka näkyvänä komponenttina on ”Human Language Technology” eli kieliteknologia.

## Luku 2

# Kieliteknologiaan liittyvän opetuksen tarjonta Suomessa

Suomen korkeakoululaitoksessa on tarjolla melko runsaasti kieliteknologiaan läheisesti liittyvää tai kieliteknologian ammattilaiseksi valmistuville hyödyllistä opetusta tietotekniikan, informaatiotutkimuksen, multimedian, yleisen ja soveltavan kielitieteen, kääntämisen, fonetiikan, viestinnän, digitaalisen signaalinkäsittelyn ym. alueilla. Nykytilanteessa on kuitenkin varsinaisten kieliteknologian ammattilaisten koulutukseen suunnattuna opetuksen tarjontaa kovin harvoille, lähinnä Helsingin yliopiston (HY) yleisen kielitieteen laitoksella, jossa voi suorittaa alan syventävät opinnot ja jatko-opintojakin. Puheteknologian opetusta Teknillisessä korkeakoulussa (TKK) ja Tampereen teknillisessä korkeakoulussa (TTKK) voi myös pitää alalle suuntautuneina. Suomen yliopistoissa tarjotaan kuitenkin hyvin laajalti kieliteknologiaan läheisesti liittyvää opetusta, jota voidaan osittain hyödyntää kieliteknologeja koulutettaessa.

- **Kieliteknologian** (ent. tietokone-lingvistiikan) arvosanaopetusta on vain Helsingin yliopiston humanistisessa tiedekunnassa.
- **Tietojenkäsittelyn** opetusta on useissa yliopistoissa tarjolla *tietojenkäsittelytieteen*, tai *-opin* sekä *tietotekniikan* nimisenä eriasteisena, perus- ja aineopintojen sekä syventävien opintojen tasoisena. Laajalti on tarjolla opetusta myös kieliteknologian kannalta tärkeissä aiheissa, kuten *multimediassa*, *käyttöliittymissä*, *rakenteisissa dokumenteissa*, *neuraalilaskennassa* ja *hahmontunnistuksessa*.
- **Yleisen kielitieteen** opetusta on tarjolla joissakin yliopistoissa kaikilla tasoilla ja useissa muissa on sivuaineen tasoita opetusta. **Soveltavaa kielitiedettä** opetetaan Jyväskylän yliopistossa (JY) pääaineen tasolla, jonka lisäksi siellä on opetusta ja tutkimusta mm. teknologia-avusteisessa kielenopetuksessa. Kieliaineiden yhteydessä opetetaan myös useita kieliteknologian kannalta tärkeitä asioita, esim. keskustelututkimusta.



- **Käännöstiedettä** eri muodoissa opetetaan kääntäjänkoulutuksen yhteydessä monissa yliopistoissa.
- **Kieliaineita** kuten englantia, saksaa, ruotsia ja suomen kieltä opetetaan useimmissa yliopistoissa ainakin kielitaidon tarpeisiin, mutta useissa monialaisissa yliopistoissa myös pääaineena.
- **Fonetiikan** arvosanaopetusta on enää vain HY:ssä ja Oulun yliopistossa (OY) ja sen lisäksi yksittäisiä kursseja monissa muissa yliopistoissa. Fonetiikan opetus yliopistoissa lähtee useimmiten kielentutkimuksen ja -opetuksen tarpeista. Lisäksi TKK:ssa opetetaan **akustiikkaa**.
- **Digitaalista signaalinkäsittelyä** opetetaan teknillisissä korkeakouluissa syventävien opintojen tasoisena ja lisäksi joissain yliopistoissa on tarjolla yksittäisiä kursseja tietojenkäsittelyn yhteydessä.
- **Puheteknologiaa** on ollut tarjolla toistaiseksi yksittäisten kurssien tasoisena.
- **Kognitiotiedettä** on täysimittaisena vain HY:ssä ja sielläkin määräaikalaisilla resursseilla hoidettuna. *Käytettävyys* liittyy kognitiotieteen lisäksi Taideteollisen korkeakouluun (TAIK) ja mm. TKK:n *kognitiivisen teknologiaan* ja eri yliopistoissa opetettiin käyttöliittymiin tietotekniikan osana.
- **Viestintää**, *puheoppia*, *puhe-* ja *organisaatioviestintää* sekä *tiedotusoppia* opetetaan monissa yliopistoissa.
- **Informaatiotutkimusta** opetetaan pääaineena Tampereen yliopistossa (TaY), OY:ssä ja Åbo Akademiassa (ÅA).

## Luku 3

# Kieliteknologian tilanteen ja mahdollisuuksien arviointia

Suomalaisen kieliteknologian *vahvuuksina* voidaan pitää toisaalta laadukasta alan tutkimusta, jo käynnistynyttä kaupallisesti lupaavaa kieliteknologian yritystoimintaa, sekä maan hyvin vahvaa ja vakiintunutta monikielisyyttä ja vieraiden kielten opetusta sekä tietotekniikan korkeata tasoa. Tutkimuksen ja sen myötä tulleiden kansainvälisten yhteyksien varaan on hyvä rakentaa yliopisto-opetusta. Yritystoiminta motivoi opiskelijoita hakeutumaan alalle ja helpottaa esim. opetusta tukevien kieliteknologisten hankkeiden saamista yliopistomaailmaan.

Nykytilanteen *heikkoutena* voidaan nähdä kieliteknologian kouluttajien niukkuus, tutkimusyhteisön ja opetusta antavien ryhmien hajanaisuus sekä toisaalta se, että jo käsitteenä kieliteknologia on edelleen verraten vähän tunnettu. Kouluttajien määrän rajallisuus lähtötilanteessa rajoittaa alussa osaltaan opetuksen kasvamista ja samanaikaisesti käynnistettävien erilaisten toimintojen määrää. Alueellinen hajanaisuus asettaa omia vaatimuksiaan yhteistyölle. Alan tuntemattomuus vähentää opiskelijoiden hakeutumista kieliteknologian piiriin.

Kieliteknologian yliopisto-opetuksen kehittämisen ja voimavarojen yhdistämisen avulla on kuitenkin *mahdollista* kehittää kansallista kieliteknologiaan perustuvaa ja sitä hyödyntävää tietotekniikan teollisuutta kansainvälisesti kilpailukykyiseksi ja ennen kaikkea laajuudeltaan merkittäväksi. Esimerkiksi Tekes on juuri käynnistämällään USIX-nimisellä kansallisella teknologiaohjelmalla osoittanut tukevansa kieliteknologian tutkimus- ja kehitystyötä tällaisilla tavoitteilla. Riittäväillä kieliteknologian panostuksilla voidaan myös turvata kansallisten kulttuuriarvojen ja kielten säilymistä sekä osaltaan edistää kansainvälistymistä, jossa oma ja vieraat kulttuurit ovat tasapainoisesti tukemassa toisiaan.

Kieliteknologian yliopisto-opetuksen jääminen nykytasolle haittaisi ja hidastaisi kotimaisilla kielillä toimivien tietokoneiden käyttöliittymien kehittymistä ja saataville tuloa. Osaltaan tämä voisi johtaa myös ikääntyvien, vammaisten ja vähemmän koulutettujen kan-

salaisten elämänlaadun heikkenemiseen ja syrjäytymiseen tietoyhteiskunnasta, ellei tele-  
tekniikan ja tietoverkkojen palveluja saada toteutetuksi myös kotimaisille kielille. Jo nyt  
yleisesti tarvittavat palvelut edellyttävät yhä enemmän tietoteknistä koulutusta ja ilman  
asianmukaista kieliteknologiaa tilanne voi pahentua. Kieliteknologian ammattilaisten niuk-  
kuus voi myös haitata vähemmistökielten (esim. saamen, romanin ja viittomakielen) ase-  
maa, ellei näiden kaupallisesti kannattamattomien kielten kieliteknologialle riitä tekijöitä.

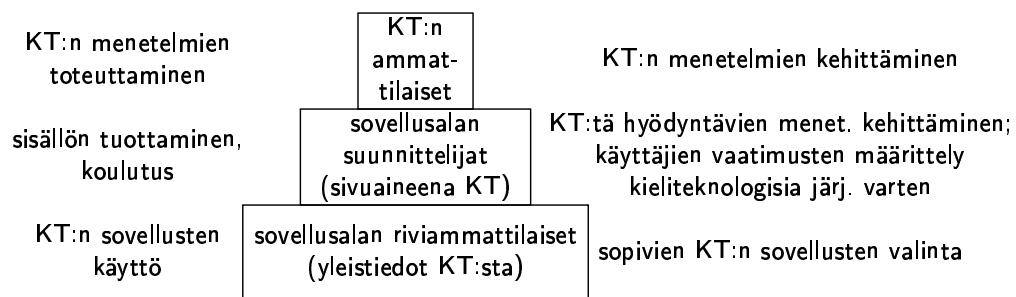
# Luku 4

## Kieliteknologian koulutuksen tarve

Kieliteknologian yleistietoja tarvitaan toisaalta laajalti, mutta toisaalta eriasteisesti. Yhä useammalla alalla valmistuvien ammattilaisten olisi hyvä olla selvillä kieleen ja sen käsittelemiseen liittyvistä ongelmista sekä siitä, mitä kieliteknologian menetelmillä pääpiirteittäin voidaan tehdä. Tällainen yleistuntemus on enenevässä määrin tarpeen oman alan oikeiden työtapojen ja välineiden valinnassa.

Monet eri alojen ammattilaiset osallistuvat kieliteknologiaa soveltavien järjestelmien tai tuotteiden rakentamiseen ja tarvitsevat olennaisesti laajempia tietoja kieliteknologiasta, esim. sivuaineena perus- ja ehkä aineopinnot. Heidän tulee pystyä toisaalta tuottamaan sisältöä järjestelmiin, osallistumaan järjestelmien suunnitteluun ja esimerkiksi täsmentämään järjestelmän tulevien käyttäjien vaatimuksia (user requirements) toteutettavalle järjestelmälle. Heidän tehtävänä on myös kehittää omalla sovellusalallaan olevia kieliteknologiaa hyödyntäviä menetelmiä.

Eniten kieliteknologian tietoja tarvitsevat toki ne kieliteknologian ammattilaiset, jotka toteuttavat uusia menetelmiä tai järjestelmiä—saati sitten ne, jotka tutkijoina pyrkivät kehittämään uusia ja entistä parempia kieliteknologian menetelmiä. Nämä ammattilaiset ottavat kieliteknologian pääaineekseen ja mielellään sopivan sovellusalan sivuaineekseen.



Kuva 4.1: Kieliteknologian (KT) koulutuksen pyramidi

Kieliteknologiaa voidaan soveltaa useisiin erityyppisiin tehtäviin. Seuraavassa hahmotellaan tärkeimpiä tällaisia ja kullakin alueella tarvittavaa korkeakouluopetusta. Useilla näistä aloista voidaan tunnistaa kuvan 4.1 pyramidimainen asetelma, jossa alhaalla laajemmille piireille tulisi tarjota yleistietoja kieliteknologiasta, jotta nämä voisivat hyödyntää omassa työssään erilaisia kieliteknologisia välineitä tai valmiita tuotteita. Ylhäällä pyramidissa on pienempi joukko kieliteknologiaan erikoistuneita ammattilaisia, jotka pystyvät tekemään näitä sovelluksia ja kehittämään menetelmiä. Keskivaiheilla voi olla mm. tulevia eriasteisia sisällön tuottajia tällaisiin sovelluksiin sekä alansa kehittäjiä menetelmille, jotka hyödyntävät kieliteknologisia välineitä.

Kieliteknologia oli aluksi voittopuolisesti pitkän tähtäyksen tutkimusta, jota yliopistot, Euroopan unioni ja eräät suuryritykset kuten Xerox, IBM ja ATT harjoittivat. Muutos käytännössä merkittäväksi ja monilla aloilla arkisissa asioissa tärkeäksi teknologiaksi on itse asiassa jo tapahtunut. Nyt on meneillään nopean kasvun ja haltuunoton vaihe. Eri-laisten verkkopalvelujen ja sisällöntuotannon alueella kyky omaksua kieliteknologian uudet menetelmät voi olla olennainen osa markkinoita vallattaessa tai niillä kilpailtaessa.

Kieliteknologian ensimmäiset sovellukset ovat jo lähes jokaisen tekstinkäsittelyä harrastavan kansalaisen työpöydällä oikeinkirjoituksen tarkistus- ja tavutusalgoritmeina. Menetelmät valtaavat nopeasti uusia alueita, jotka koskettavat pian väestön kaikkia kerroksia. Niitä palveluita, joissa kieliteknologian avulla voidaan päästä olennaisesti parempiin tuloksiin, on runsaasti: kielten opiskelu, digitaaliset kirjastot, sähköinen kaupankäynti, kansalaisten opastaminen erilaisissa yhteiskunnan palveluissa, avoin pääsy tietoyhteiskunnan tietoihin jne.

## 4.1 Puheteknologian opetus

Puheteknologian kehittyminen on ollut voimakasta viimeisten 15 vuoden aikana. Kehityksen on aiheuttanut digitaalisessa tietoliikennetekniikassa tapahtunut räjähdysmäinen kasvu, joka erityisesti matkapuhelinviestinnässä on tehnyt mahdolliseksi uusien puheteknologian menetelmien laajamuotoisen kaupallisen hyödyntämisen. Tämä kehitys tulee entisestäänkin voimistumaan lähivuosina, jota todistaa esimerkiksi se, että matkaviestinnässä tulee tapahtumaan merkittävä sukupolven vaihdos, kun nykyisestä GSM-järjestelmästä siirrytään seuraavaan vaiheeseen, UMTS:ään (Universal Mobile Telephone System). Kieliteknologian kannalta erittäin merkittävää on se, että tulevat tietoliikennejärjestelmät tulevat mahdollistamaan laajakaistaisen tiedonsiirron, joka sisältää sekä kuvaa että puhetta. Tällöin on mahdollista implementoida laajassa kaupallisessa käytössä oleviin tietoliikennepäätelaitteisiin nykyistä monipuolisempia puheteknologian palveluja.

Puheteknologian voi katsoa olevan tärkeä osa tietoyhteiskunnan yhteiskunnallisten ulottuvuuksien tasapuolisesta toteuttamisesta, koska tällä tekniikalla suurempi osa väestöstä, ja myös suuri osa muuten syrjäytyvistä voisi päästä palvelujen piiriin.

Puheteknologia on selvästi monitieteinen alue, johon erikoistuvalla opiskelijalla olisi hyö-

dyllistä saada opetusta usealta eri alueelta, esimerkiksi:

1. **Matematiikkaa** (Fourier-transformaatiot, todennäköisyyslaskenta, matriisilaskenta, diskreetti matematiikka ym.) signaalinkäsittelyn ja puheentunnistuksen menetelmien ymmärtämiseksi. Tätä on tarjolla teknillisissä korkeakouluissa laajahkoina kursseilla, jotka valmistavat opiskelijoita muunkinlaisiin sovelluksiin. Yliopistopohjalta tuleville tarvittaisiin erityisesti sovitettuja ja suppeampia kursseja.
2. **Digitaalista signaalinkäsittelyä**, jonka menetelmien avulla voidaan esim. eristää puhesignaalista piirteitä puheentunnistusta varten tai tuottaa synteettistä puhetta. TKK ja TTKK opettavat laajahkoja kursseja, jotka valmistavat opiskelijoita muunkinlaiseen signaalinkäsittelyyn, mutta esim. HY:ssa on myös joitakin kursseja. Yliopistopohjaisille opiskelijoille tarvittaisiin erityisesti sovitettuja kursseja.
3. **Fonetiikkaa tai akustiikkaa** puhesignaalin olemuksen ymmärtämiseksi ja menetelmien kehittämisen pohjaksi. Kursseja tarjolla ainakin HY:ssa, TKK:ssa ja TaY:ssa. Tarvetta saattaa olla puheteknologiaan erityisesti sovitetuille kursseille.
4. **Puheentunnistuksen ja puhesynteesin** menetelmiä, joilla käsitellystä signaalista voidaan saada esim. ehdotuksia mahdollisiksi äännejonoiksi. Kursseja tarjolla TKK:ssa ja TTKK:ssa.
5. **Kielen rakennetta** ja kielen rakenteen **tunnistusmenetelmiä**, joita tarvitaan puheentunnistuksessa ja puhesynteesissä. Taustatietoa insinööritieteilijöille saatavissa esim. yleisen kielitieteen kursseina yliopistoista ja kieliteknologian johdantokurssina HY:sta. Tarvittaisiin ehkä puheteknologiaan sovitettuja syvempiä kursseja.

Pelkkä tällaisten kurssien tarjonta ei riitä, vaan kursseja pitää muokata ja sovittaa kokonaisuuteen yhteensopiviksi osiksi. Äänen, kuvan ja tekstin käsittelyn integrointi on myös otettava huomioon koulutuksen suuntautumisen vaihtoehtoisissa.

Lista hyödyllisestä opetettavasta asiasta on pitkä, eikä ole kohtuullista odottaa, että jokainen puheteknologiaa opiskeleva hallitsisi yhtä hyvin kaikkia näitä alueita. Opetusta tulisi kuitenkin olla saatavilla kaikilta alueilta, vaikka esim. teknillisillä aloilla painotettaisiin fonetiikkaa ja kielen rakennetta ja analysointia vähemmän ja vastaavasti fonetiikan ja kielitieteen puolelta asiaa lähestyvät jäisivät pintapuolisemmalle ymmärrykselle matematiikan ja digitaalisen signaalinkäsittelyn suhteen. Joidenkin tulisi paneutua koko alueeseen menetelmien eteenpäin kehittämiseksi.

## 4.2 Käyttöliittymien kieliteknologian opetus

Tietokoneiden ja verkkojen käyttäjistä jo suurin osa on ei-ammattilaisia tavallisia ihmisiä, jolloin ihmisen ja koneen vuorovaikutuksesta muodostuu entistä tärkeämpi lenkki. Enää ei

voida olettaa että palvelujen käyttäjät koulutettaisiin erityisesti, vaan palvelujen tulee olla riittävän ilmeisiä ja helppokäyttöisiä, jotta kaikki pystyisivät niitä käyttämään.

Myös luonnollisen kielen käyttöliittymät muodostavat monitieteisen kokonaisuuden, jossa voidaan hyödyntää monenlaista oppia, esim.:

1. **Käytettävyyttä, käyttöliittymiä ja niiden ohjelmointitekniikkaa**, josta opetusta on saatavissa tietojenkäsittelyä opettavissa yliopistoissa. Erikoistumisohjelmia on myös käynnistymässä (TKK, TaY, HY:n kognitiotiede).
2. **Kognitiivista mallintamista**, eli ihmisen ja tietokoneen välisen vuorovaikutuksen hahmottamista. Tällaista opetusta lienee saatavilla ainakin HY:n kognitiotieteessä, sekä TKK:n kognitiivisen teknologian oppiaineessa.
3. **Puheteknologiaa** ääneen perustuvien käyttöliittymien kehittämiseksi (ks. yllä).
4. **Luonnollisen kielen jäsentämistä ja generointia**, jolla ihmisen viestit tulkitaan ja koneen tuottamat viestit muutetaan selväkielisiksi. Kurseja tarjolla vain HY:ssa. Tähän tarkoitukseen tulisi ehkä sovittaa erityisiä kurseja.
5. **Diskurssin** eli keskustelun mallintamista. (Diskurssia tutkitaan mm. HY:n suomen kielen laitoksella ja JY:ssa soveltavan kielentutkimuksen keskuksessa sekä TaY:n sosiologian ja sosiaalipsykologian laitoksella.) Sopivia kurseja olisi kehitettävä.

### 4.3 Dokumenttien hallinnan ja tiedon haun opetus

Tiedon tulva on jo tämän hetken todellisuutta verkossa, mutta tulossa on ilmeisesti paljon mittavampi epäsuhta tarjolla olevan ja sulatettavissa olevan tiedon suhteen. Suomen kieli (kuin myös jossain määrin monet muutkin eurooppalaiset kielet) ovat erilaisia kuin englannin kieli, jossa sanat eivät juurikaan taivu, eikä yhdyssanoja kirjoiteta yhteen ilman sananväliä. Jo tällaisten tekijöiden neutralointi edellyttää melko pitkälle meneviä kieliteknologian menetelmiä. Vaativampi tehtävä on pyrkiä eristämään oikeaa ja olennaista tietoa silloin, kun sanontatavat ja ilmaukset vaihtelevat. Voidaan ennakoida, että ilman parhaita välineitä tiedon tulvasta ei tulla selviämään kunnialla.

Digitaalisiin dokumentteihin sisältyvän tiedon tehokas ja monipuolinen käyttö edellyttää monitieteistä osaamista, missä luonnolliseen kieleen liittyvä teknologia muodostaa yhden tärkeän osa-alueen. Muita osa-alueita ovat formaalit kielet, verkko-, tietokanta-, multimedia- ja käyttöliittymäteknologiat sekä ryhmätyön ja viestinnän, tilastotieteen sekä tietämyksen eristämisen (data mining) menetelmät.

Dokumenttien hallinnassa ja tiedonhaussa tarvitaan mm. seuraavanlaista opetusta:

1. **Informaatiotutkimusta**, joka muodostaa erään kulmakiven dokumenttien relevanssin, etsimisen, luokittelun ja indeksoinnin ymmärtämiselle. Opetusta tarjolla TaY:ssa

ja OY:ssa. Tätä olisi pystyttävä hyödyntämään verkostossa ja kenties kehitettävä lisää kieliteknologisiin sovelluksiin orientoituja kursseja.

2. **Rakenteisten dokumenttien**, lähinnä dokumenttien rakenteen merkitsemiseen käytettyjen SGML- ja XML-kielten tuntemusta. Kursseja tarjolla ainakin JY:ssa ja HY:ssa. Tarvittaisiin myös laajoille opiskelijajoukoille sovitettuja kursseja ja sellaisia, joissa opetetaan suoraan hyödyntämään olemassa olevia työkaluja.
3. **Luonnollisen kielen jäsenysmenetelmien** tuntemusta tarvitaan kieliteknologian soveltamiseksi tiedonhallintaprosessiin. Kursseja tarjolla lähinnä HY:ssa.
4. **Tilastollisten ym. menetelmien** hallintaa relevanttien ja irrelevanttien dokumenttien erottamiseksi toisistaan. Tilastomenetelmien yleistä opetusta on tarjolla useissa yliopistoissa. Erityisiä tiedonhakuun sovitettuja tilastomenetelmien kursseja on TaY:ssa ja HY:ssa, mutta myös osa hahmontunnistuksen ja neuroverkkojen kursseista on tähän tarkoitukseen sovitettuja esim. TKK:ssa ja TAIK:ssa.

## 4.4 Kielenopettamisen ja -oppimisen teknologia

Kielenoppimisen ja -opettamisen teknologia (engl. technology enhanced language teaching and learning) tarkoittaa tässä tietokoneavusteiseen kielenoppimiseen (engl. Computer Aided Language Learning, CALL) pedagogisesti ja teknologisesti laajentunutta lähestymistapaa kielenopetukseen ja -oppimiseen. Tässä keskitytään käsittelemään erityisesti kieliteknologiaan luontevasti liittyvää osaa alasta.

Teknologiset edellytykset kieliteknologian laajamittaiselle hyödyntämiselle kielen oppimislussa ovat olemassa, mutta hyödyntäminen edellyttää varsin monitieteisiä taitoja esim. kasvatustieteiden, kielenopetuksen, kielitieteiden ja multimedian aloilta. Näissä tehtävissä tarvitaan toisaalta kieliteknologian ammattilaisia ja sisällön tuottajia, sekä toisaalta viestinnän ammattilaisten ja opettajien olisi hyödyllistä hankkia tarvittavat yleistiedot menetelmistä.

Kielenopiskelun ja -opettamisen menetelmillä on kieliteknologian kanssa yhteistä aluetta runsaasti. Sen vuoksi kieliteknologian ammattilaisten tulisi voida suuntautua ja erikoistua kielenopetuksen menetelmiin ja päinvastoin ja saada monipuolista opetusta mm.:

1. **Kielenopetuksen** eri osa-alueiden tuntemusta. Opetusta tarjolla JY:ssa soveltavassa kielitieteessä sekä kääntäjänkoulutuksen yhteydessä mm. HY:ssa ja TaY:ssa.
2. **Opetettavan kielen** rakenteen tuntemusta. Opetusta tarjolla useimmissa korkeakouluissa.
3. **Uusmedian** ja tietoverkkojen tuntemusta opetuksen näkökulmasta. Opetusta tarjolla ainakin Jyväskylässä, Tampereella ja Helsingin seudulla.



4. **Luonnollisen kielen jäsennessmenetelmien** tuntemusta joko opetusohjelmien sisäl-  
lön esikäsittelemiseksi tai jäsentimien liittämistä osaksi opetusohjelmaa. Opetusta  
tarjolla HY:ssa.
5. **Kasvatustiedettä ja kielipedagogiikkaa** eli mitä kieltä opetettaessa tulee huo-  
mioida. Tarvittavaa opetusta ja kompetenssia lienee JY:ssa ja HY:ssa ja opettajan-  
koulutuslaitoksilla (joihin pitäisi luoda tiiviimpiä yhteyksiä).
6. Erilaisten **tiedonhaun prosessien** tuntemusta. Opetusta tarjolla ainakin TaY:ssa,  
HY:ssa ja TKK:ssa.

## 4.5 Kääntämisen apuvälineiden opetus

Kielenkääntämisen tarve ja merkitys kasvaa Suomessa mm. Euroopan unionin ja sen tule-  
van laajenemisen vuoksi. Kääntämisen apuna käytetään nykyään runsaasti sekä tietotek-  
nisiä että kieliteknologisia apuvälineitä, kuten elektronisia sanakirjoja, terminologiatyöka-  
luja jne. sekä raakakäännösten tekemisessä myös automaattisen kielenkääntämisen mene-  
telmiä. Kääntäjänkoulutuksessa oleville olisi tarjottava mahdollisuuksia hankkia riittävät  
tiedot ja tarvittaessa erikoistua tällaisiin menetelmiin. Tarvittavia aineksia ovat:

1. **Käännöstieteen, yleisen kielitieteen, terminologian tutkimuksen ja leksi-  
kografian** opetusta, joita on saatavilla eri korkeakouluissa ja kääntäjänkoulutuslai-  
toksissa.
2. **Tietotekniikan tai tietojenkäsittelyn** opetusta, jota on myös saatavilla yleisesti,  
mutta erityisesti sovitettuja kursseja olisi myös tarpeen kehittää.
3. **Kieliteknologian** menetelmien, erityisesti sananmuotojen analysoinnin (eli morfolo-  
gisen analyysin) ja automaattisen lauseenjäsennyksen menetelmät sekä tietokonemu-  
toisten leksikoiden rakenne sekä tekstikorpusten käsittelymenetelmät yleensä ovat  
keskeisiä. Näitä on tarjolla lähinnä HY:ssa.

## 4.6 Kieliteknologian menetelmien ja teorian opetus

Yllä mainituille kieliteknologian sovellusaloille tarvitaan opettajia, sovellusalojen käyttöön  
tulisi kehittää yhä parempia menetelmiä ja kullekin alalle tarvitaan kieliteknologian eri-  
koistaitajia mukaan hankkeisiin ja tutkimusryhmiin. Tätä varten pitää lisätä pääaineenaan  
kieliteknologiaa (tietokone-lingvistiikkaa) opiskelevien valmistumistavoitteita olennaisesti,  
esim. 10–15 FM-tutkinnoksi vuodessa. Kieliteknologian pääaineopiskelijoiden rekrytointi  
tapahtuisi suurimmaksi osaksi muiden oppiaineiden opiskelijoista, jotka ovat jo suorittaneet  
esim. aineopinnot sopivassa aineessa.

Paitsi pääaineena kuvan 4.1 (ks. sivu 11) pyramidin ylintä osaa varten, tarvitaan kieliteknologian perus- ja aineopintojen tasoista arvosanaopetusta myös pyramidin keskikerrosta varten. Tällaiset arvosanat tulisivat sivuaineen opintoina osaksi eri yliopistoissa suoritettaviin FM- tai DI-tutkintoihin, joiden pääaineet vaihtelevat. Valtakunnallinen tarve tällaisille n. 20 ov laajuisille sivuaine-arvosanoille arvioidaan noin 50:ksi vuosittain.

## 4.7 Työmarkkinatilanne

Kieliteknologian asiantuntijoiden työllistäjinä ovat ja voivat olla:

- Kieliteknologia-yritykset (kuten Lingsoft Oy, Kielikone Oy jne.) joissa arvellaan tällä hetkellä olevan noin 40–60 henkeä pääosin kieliteknologiaan liittyvissä tehtävissä. Heidän koulutuksensa vaihtelee tietotekniikasta ja kieliaineista tietokone-lingvistiikan pääaineopintoihin. Ala on kasvanut ja kasvaa ripeästi ja tarve on ehkä noin 10–20 henkeä vuosittain, joista osalle kieliteknologia ensisijaisena ja osalla tietotekninen tai kielitieteellinen koulutus ensisijaisena.
- Telealan teollisuus (kuten Nokia) ja teleoperaattorit (kuten Sonera, HPY, Radiolinja, Telia) ovat kiinnostuneet puheteknologisista tuotteista ja sovelluksista. Niillä voidaan arvioida olevan satoja puheteknologian alalla toimivia työntekijöitä ja vuosittaisen kasvun voi arvioida olevan erittäin nopeata, 20–50 uutta työntekijää ja viiden vuoden päästä olennaisesti enemmän. Suuri osa tarpeesta on ensisijaisesti signaalinkäsittelyn taustan omaavia, ja osa yhä kattavamman kieliteknologisen koulutuksen saaneita. Erityisesti luonnollisten kielten puhekäyttöliittymien yleistyminen lisää monipuolisesti koulutettujen kieliteknologien tarvetta.
- Internetissä kauppaa käyvät yritykset tarvitsevat sekä luonnollisen kielen käyttöliittymien että monikielisyiden tukea verkkomyyntiinsä. Verkkokauppa on luontaisesti globaalia ja siten monikielistä. Nämä yritykset tai niille tietoteknisiä sovelluksia tekevät ATK-palvelufirmat tarvitsevat kieliteknologisesti koulutettuja työntekijöitä, osin kieliteknologian ammattilaisia, osin ATK-ammattilaisia, jotka ovat selvillä kieliteknologian perusteista ja mahdollisuuksista. Tällä hetkellä näissä tehtävissä on muutamia kieliteknologisia työntekijöitä.
- Kielenkääntämisen yritykset (belgialainen Lernout & Hauspie ym.) siirtyvät nopeasti kieliteknologisten menetelmien laajamittaiseen hyödyntämiseen. Työssä on tällä hetkellä useita kieliteknologisesti koulutettuja ja kasvun voi odottaa lähtevän nousuun. Ammattimaisten kieliteknologioiden lisäksi tarvitaan hyvin laajalti yleistietoja ja sivuaineen laajuisia tietoja kieliteknologiasta oikeiden välineiden valitsemiseksi ja käänösprosessin kehittämiseksi.
- Erilaiset kirjojen, sanakirjojen ja lehtien kustantajat sekä tarvitsevat kasvavassa määrin tekstitiedon järjestämistä, hakemistojen tekemistä sekä tiedonhakua varten kie-

liteknologisia menetelmiä. Kustannusalaan lähellä ovat myös opetussovellusten kehittäjät ja muu uusmediatoiminta. Tällä hetkellä muutamia henkilöitä on tällaisissa kieliteknologisissa tehtävissä.

- Suuret (enimmäkseen kansainväliset) tietokonealan yritykset (IBM, Xerox, Microsoft, Oracle, jne) rekrytoivat kieliteknologeja järjestelmiensä lokalisoiminnin ja monikielisten palvelujen kehittämiseksi. Muutamia suomalaisia on tällä hetkellä näissä tehtävissä.
- Teollisuusyritykset ja virastot tarvitsevat vastaisuudessa työntekijöitä huolehtimaan dokumentaatiostaan ja terminologiastaan, joka usein on monikielistä. Näissä tehtävissä on jo nyt useita kieliteknologeja ja alueelle on myös syntynyt uusia yrityksiä, jotka hakevat lisätyövoimaa. Erityisesti teknisen dokumentoinnin alueella on tarvetta henkilöistä, joilla on koulutusta sekä dokumenttien hallinnan että kieliteknologian alueilta. Esimerkiksi eri sovellusalueiden kontrolloitujen kielten kehittämiseen ja monikielisyys hallintaan tarvittaisiin lähitulevaisuudessa kymmeniä kieliteknologian asiantuntijoita.

# Luku 5

## Ehdotukset toimenpiteiksi

Ehdotetut toimenpiteet voidaan hahmottaa aiemmin sivulla 11 kuvassa 4.1 esitetyn pyramidimallin pohjalta:

1. kieliteknologian ammattilaisten koulutus,
2. kieliteknologiaa olennaisesti hyödyntävien alojen suunnittelijoiden ja menetelmien kehittäjien koulutus sekä
3. kieliteknologiaa hyödyntävien tai sitä lähellä olevien alojen tavanomaisten ammattilaisten koulutus.

Ensimmäisen ryhmän koulutus koostuu olennaisesti ylemmän akateemisen loppututkinnon suorittamisesta pääaineenaan kieliteknologia. Toisen ryhmän koulutus koostuu olennaiselta osaltaan oman pääaineen mukaisesta loppututkinnosta, johon sisällytetään kieliteknologian sivuainearvosana joko perus- tai aineopintojen laajuuisena. Kolmannen ryhmän koulutus on ensi sijassa oman pääaineen mukaista koulutusta, johon osana (tai vapaina opintoina) liittyy kieliteknologian yleistietoja esitteleviä kursseja sekä mahdollisia omaa erikoisalaa täydentäviä kieliteknologian erikoiskursseja opintojen eri vaiheissa.

**Työryhmä ehdottaa, että perustetaan yksi valtakunnallinen kieliteknologian opetuksen verkosto**, joka sisältää kieliteknologian maisteriohjelman. Verkosto toimii maisteriohjelman tapaan ryhmiä 1 ja 2 varten ja kurssien yhteiskäyttöä edistävän verkoston tapaan ryhmää 3 varten.

Erikoistuminen kieliteknologian ytimeen, sen eri tahoille suuntautuviin keskeisiin menetelmiin ja kieliteknologiaa hyödyntäviin eri aloihin muodostaa jatkumon. Opiskelijat voivat sovellusaloillaan hankkia eriasteisesti kieliteknologisia tietoja ja taitoja ja toisaalta kieliteknologiaa opiskelevat voivat eriasteisesti suuntautua jonkin sovellusalan menetelmiin.

## 5.1 Kieliteknologian ammattilaisten koulutuksen järjestäminen

Ryhmän 1 eli kieliteknologian ammattilaisten koulutus järjestetään siten, että opiskelijat suorittavat loppututkintonsa pääaineenaan kieliteknologia, mutta pääosa näistä opiskelijoista on aloittanut opintonsa kieliteknologiaa lähellä olevissa aineissa ja siirtyy verkostoon kuuluvaan maisterikouluun suoritettuaan noin 60 ov tällaisia muita opintoja. (Tällä hetkellä asetuksen mukaan HY:n humanistinen tiedekunta saa antaa kieliteknologian syventäviä opintoja.)

Maisterikoulu merkitsee ryhmän 1 opiskelijoille lähinnä kieliteknologian pääaineopiskelua tavoitteena FM-tutkinto. Kieliteknologian pääaineopiskelijoiden tulee voida ottaa verkoston piirissä olevista laitoksista yksittäisiä sovellusalojen kursseja pääaineensa erikoistumis-kursseiksi ja tarvittaessa sivuaineen laajuisia arvosanoja tällaisista aineista.

## 5.2 Kieliteknologiaa hyödyntävien ammattilaisten koulutus

Ryhmän 2 koulutus järjestetään hajautetusti eri yliopistoissa siten, että opiskelijat valmistuvat niillä laitoksilla, joilla pääaineopintojaan suorittavat, mutta suorittavat kieliteknologian kokonaisen arvosanan (tai kokonaisia arvosanoja) verkoston puitteissa. Kieliteknologian sivuaine-arvosanat suoritetaan kuitenkin Helsingin yliopiston yleisen kielitieteen laitoksella sen hyväksymien tutkintovaatimusten mukaisesti. Yksittäiset kurssit, joista arvosanat koostuvat, voivat olla myös verkoston piirissä olevilta muilta laitoksista.

Kokonaisten arvosanojen suorittaminen eri yliopistossa on nykyäänkin jossakin määrin mahdollista esim. JOO-sopimuksen puitteissa. Tällainen kokonaisten arvosanojen suorittaminen toisessa yliopistossa lienee tässä tapauksessa tarkoituksenmukaista, mutta sen toteuttaminen saattaa edellyttää lisäsopimuksia.

## 5.3 Kieliteknologiaa lähellä olevien alojen koulutus

Kaikki yhteistyössä olevat laitokset muodostavat yhdessä kieliteknologian valtakunnallisen verkoston, jonka puitteissa yhteistyössä mukana olevien eri korkeakouluissa olevien laitosten opiskelijat voivat osallistua yksittäisille toisten yhteistyölaitosten kieliteknologisille kursseille, joiden suoritukset hyväksytään osaksi omassa laitoksessa suoritettavia opintoja.

Johdantokursseja ja muuta laajoille kuulijakunnille tarkoitettuja kursseja pyritään järjestämään mahdollisuuksien mukaan erikseen eri yliopistokaupungeissa. Erikoiskurssit järjestetään yleensä hajautetusti, kukin kurssi yhdessä paikassa siten, että verkoston muiden

laitosten opiskelijoilla on käypä mahdollisuus osallistua niille.

Yliopistojen väliset yhteistyösopimukset antavat mahdollisuuden yksittäisten kurssien suorittamiseen toisessa yliopistossa. Sopimukset eivät tällä hetkellä kata kaikkia kyseeseen tulevia laitoksia. Verkoston laitosten kesken tulee tehdä mahdollisesti tarvittava erillinen sopimus, joka varmistaa opiskelijoiden mahdollisuuden hyödyntää verkoston toisten laitosten kieliteknologisia kursseja.

## 5.4 Kieliteknologian opetuksen ohjaus ja koordinointi

Kieliteknologian maisteriohjelman ja verkoston puitteissa nimetään kieliteknologian kunkin sovellus/osa-alueen koordinointiin jokin verkoston laitoksista vastuulliseksi kokoamaan tiedot saatavilla olevista kursseista, ohjaamaan neuvotteluja kurssien muotoilemiseksi eri tahojen koulutukseen yhteensopiviksi, suunnittelemaan osa-alueen opetusta sekä tiedottamaan tarjonnasta.

Koska suuri osa opetuksesta tapahtuu maantieteellisesti eri paikoissa, osapuolten tulee pyrkiä siihen, että kurssit järjestetään esim. tiiviinä periodina tai muuten sellaisina, että verkoston toisten laitosten opiskelijat voivat niihin kohtuullisesti osallistua. Tietoverkon pyritään hyödyntämään opetuksen tukena ristiinopiskelun helpottamiseksi.

Kieliteknologian valtakunnallisella maisterikoululla ja verkostolla tulee olla siihen osallistuvia laitoksia edustava johtoryhmä. Johtoryhmä ohjaa verkoston organisoitumista ja päättää esim. verkostolle yhteisesti myönnettyjen resurssien käytöstä.

Verkostolla tulee olla koordinaattori ja kullakin osallistuvalla laitoksella nimetty vastuuhenkilö.

## 5.5 Kurssien yhteisen hyödyntämisen periaatteista

Verkoston periaatteena (erityisesti 3. ryhmän opetuksen osalta) on se, että osallistuvat laitokset toimivat kahdessa roolissa: opiskelijoita lähettävinä ja vastaanottavina laitoksina. Vastaanottava laitos tarjoaa mahdollisuuksiensa mukaan saapuville opiskelijoille opetuksen ja tulevista kursseista hyvissä ajoin tiedon verkoston toisille osapuolille. Vastaanottava osapuoli ei (välttämättä) kirjaa opiskelijoita läsnäoleviksi opiskelijoiksi, mutta antaa opetuksen ja arvostelee ja hyväksyy suoritukset ja toimittaa niiden tulokset lähettävälle laitokselle.

Lähettävän laitoksen roolissa laitos puolestaan pitää huolen siitä, että tutkintovaatimuksissa on asianmukaiset sijat muualta saataville kursseille. Laitos tiedottaa muualla pidettävistä kursseista opiskelijoilleen ja täsmentää mihin ja minkä laajuisina kurssit kelpaavat tutkintovaatimukseen (ja mahdollisista päällekkäisyyttä koskevista rajoituksista).

Pyrkimyksenä on päästä verkostossa summittaiseen vastavuoroisuuteen. Vaihto on ilmeisen

edullista osapuolille, vaikka määrät eivät menisikään tasan.

Verkostolle perustetaan WWW-kotisivu ja sopiva elektroninen tiedotus (esim. uutisryhmä, sähköpostilista tai elektroninen lehti). Kukin osallistuva laitos huolehtii omalta osaltaan yhteisesti kiinnostavien tietojen tarjoamisesta omilla WWW-sivuillaan ja tarvittavien viitetietojen toimittamisesta yhteisen sivun ylläpitäjälle ja tiedotusta koordinoivalle taholle.

## 5.6 Tarvittavia panostuksia

Kieliteknologian yliopistotasoisien koulutuksen laajentaminen ja uudistaminen edellyttää paitsi tieteiden välistä yhteistyötä, myös konkreettisia panostuksia:

1. Opetuksen tarjonnan lisääminen kieliteknologian ja siihen liittyvien alojen ammattilaisten valmistamiseksi nykyistä ja tulevaa tarvetta vastaavasti. Lisäpanostuksen perusteena ovat paitsi koulutuksen määrän lisääminen, erityisesti myös kurssien onnistunut yhteiskäyttö.
2. Verkoston ja siihen sisältyvän maisteriohjelman koordinointi, suunnittelu ja tiedottaminen.
3. Osa-alueiden koordinointi, suunnittelu, yhteydenpito ja tiedotus.
4. Opiskelijoiden liikkuvuuden tukeminen, s.o. matka- ja majoituskuluja yhteisille kursseille osallistumisen mahdollistamiseksi. Erilaisen taustan ja koulutuksen omaavien opiskelijoiden yhteinen työskentely edistää myös eri tieteiden välistä yhteistyötä ja ymmärrystä.
5. Tietoverkkoa hyödyntävien etäopetuksen keinojen edistämistä kurssien yhteiskäytön helpottamiseksi. Paitsi joitakin laitteita tämä merkitsee opetuksen uusien muotojen kokeilua, ohjeistusta ja käyttöönottoa. Alkuvaiheessa ilmeisesti nopeiten voidaan hyödyntää sähköpostia ja kuvallisia yhteyksiä lähinnä harjoitustöiden yhteydessä sekä mahdollisesti erilaisia keskustelujärjestelmiä.
6. Eri tieteiden ja laitosten välisiä yhteisiä tutkimushankkeita, joiden avulla eri tieteiden välinen yhteistyö operationalisoidaan. Kieliteknologian, erityisesti suomen kielen käsittelyyn liittyvää tutkimusta tulee vahvistaa, jotta kieliteknologiaa voidaan täysimääräisesti soveltaa paikallisiin oloihin.
7. Kieliteknologian opetuksen ja tutkimuksen yhteisiä aineistoja ja ohjelmia on syytä kerätä ja dokumentoida esim. CSC Tieteellinen laskenta Oy:hyn perustetun kielipankin puitteisiin. Ohjelmien ja aineistojen käyttöönotto ja dokumentointi edellyttää työpanosta osin CSC:ltä ja osin verkoston piirissä olevilta laitoksilta.

## 5.7 Kustannusarvio

Kieliteknologian yliopistotasaisen opetuksen laajentaminen toteutetaan toisaalta vahvistamalla kieliteknologian nykyistä arvosanaopetusta ja pääaineopiskelua ja toisaalta tuomalla kieliteknologian hyödyntämiseen suuntautuneita opetusresursseja kieliteknologiaa lähellä oleville aloille.

Kieliteknologian arvosanaopetuksen resurssien vahvistamiseksi esitetään 4 opetushenkilöä tai opetusta tukevaa henkilöä, joista 1-2 professorintasoisia opettajia ynnä varoja tarvittavaa tuntiopetusta, sekä infrastruktuuria varten. Näillä resursseilla tavoitteena on nostaa kieliteknologia pääaineenaan vuosittain valmistuvien maisterien määrä 10–15:een. Samalla kieliteknologian sivuaine-arvosanan suorittavien opiskelijoiden määrä nostetaan noin 50:een, joista pääosa suorittaa tutkintonsa verkoston muissa yliopistoissa.

Kieliteknologiaa lähellä olevien aloille esitetään 8 opetushenkilöä, joista 3 professorintasoisia sijoitettavaksi eri yliopistoissa oleviin laitoksiin.

Koulutuksen laajennus merkitsisi noin 10 miljoonan markan lisäresurssitarvetta vuosittain koko verkostolle. Työryhmä esittää että kieliteknologian koulutuksen laajentaminen otetaan valtakunnalliseksi ohjelmaksi yliopistojen tulossopimuskaudella 2001–2003.

Ohjelman toteuttaminen pitäisi saada alkuun jo vuonna 2000, mitä varten Helsingin yliopistolle tulisi ohjata voimavaroja noin 500 000 markkaa toiminnan käynnistämiseksi.

Varat olisi pyrittävä ohjaamaan siten, että kieliteknologian koulutusta voitaisiin hyödyntää valtakunnallisesti ja monitieteisesti, jotta panostus edistäisi alan kehittymistä parhaalla tavalla.