

KYMENLAAKSON AMMATTIKORKEAKOULU

Liiketalouden koulutusohjelma / Kansainvälinen assistentti

Maija Kosonen

KIELITEKNOLOGIA, MITÄ SE ON? – OPETUSMATERIAALIA KYMENLAAK-
SON AMMATTIKORKEAKOULULLE

Opinnäytetyö 2009

TIIVISTELMÄ

KYMENLAAKSON AMMATTIKORKEAKOULU

Liiketalouden koulutusohjelma

KOSONEN, MAIJA	Kieliteknologia, mitä se on? Opetusmateriaalia Kymenlaakson ammattikorkeakoululle
Opinnäytetyö	45 sivua + 7 liitesivua
Työn ohjaaja	Lehtori Hannele Lappalainen
Toimeksiantaja	Kymenlaakson ammattikorkeakoulu
Lokakuu 2009	
Avainsanat	kieliteknologia, kieli, kääntäminen, puhe

Kymenlaakson ammattikorkeakoulussa käynnistyy toista kertaa johdon assistenttityön ja kielten koulutusohjelma. Tässä uudessa koulutusohjelmassa kieleen ja kulttuuriin suuntautuneet opiskelijat opiskelevat suuntautumisopinnoissaan kurssin kieliteknologiasta. Tämän opinnäytetyön tarkoituksena on tuottaa opetusmateriaalia Kymenlaakson ammattikorkeakoulun käyttöön. Tarkoitus on lisäksi selvittää, mitä hyötyä kieliteknologian opiskelusta on kielten ja kulttuurin opiskelijoille, sekä mikä merkitys kieliteknologialla on työelämässä.

Menetelmänä työn tarkoitusten saavuttamiseksi keskityttiin työssä asioihin, joiden katsottiin olevan merkityksellisiä työn tarkoitusten kannalta. Tämän johdosta työssä keskityttiin kieliteknologian perusteisiin, kieliteknologiaan Suomessa ja maailmalla sekä kieliteknologian sovelluksiin. Lisäksi tarkasteltiin kuinka kieliteknologia liittyy työelämään sekä kuinka Euroopan unioni käyttää kieliteknologiaa omassa toiminnassaan. Työn teoriassa käytettiin lähteenä lähinnä kieliteknologian tutkijoiden materiaalia, joka lisää tiedon luotettavuutta.

Opinnäytetyö toteutettiin toiminnallisena opinnäytetyönä, ja koko työ toimii tietopakettina kieliteknologian perusteista. Teoriaosuuden lisäksi työtä havainnollistaa PowerPoint-esitys. PowerPoint-esityksessä käydään läpi teoriaosan tärkeimpiä asioita. PowerPoint-esityksen tarkoitus on olla selkeä ja kiinnostava. Tarkoituksen saavuttamiseksi dioissa käydään läpi teoriaosan tärkeimpiä asioita, ja diojen ulkoasuun on kiinnitetty huomiota. Työ toteutettiin sopivaksi opetuskäyttöön, ja näkökulmassa otettiin huomioon se, että esitys on suunnattu kielten ja kulttuurin opiskelijoille.

Tuloksena saatiin työstä tiivis esitettävä tietopaketti kieliteknologian perusteista. Työtä voidaan käyttää välineenä tutustua kieliteknologiaan myös muille aiheeseen aiemmin tutustumattomille. Todettiin myös että kieliteknologialla on kasvava merkitys kansainvälistyvässä työelämässä. Kielten ja kulttuurin opiskelijoille on hyötyä kieliteknologian opiskelusta valmistuttuaan ja kansainvälisessä työympäristössä työskennellessään.

ABSTRACT

KYMENLAAKSON AMMATTIKORKEAKOULU

University of Applied Sciences

International Assistant

KOSONEN, MAIJA

Language technology, what is it? Educational material for
Kymenlaakso University of Applied Sciences

Bachelor's Thesis

45 pages + 7 pages of appendices

Supervisor

Hannele Lappalainen, Senior Lecturer

Commissioned by

Kymenlaakso University of Applied Sciences

October 2009

Keywords

language technology, language, translation, speech

This is the second time when a new programme, The Degree Programme of Multilingual Management Assistants, starts at Kymenlaakso University of Applied Sciences. In this programme the students who choose language and culture as their elective studies will study a course from language technology.

The purpose of this thesis was to produce educational material for the use of this course. The purpose was also to sort out the advantages of this course for the students and find out what role language technology plays in working life.

The method to carry out the research was to focus on the matters which have some meaning when considering the purposes of the thesis. Due to this, the main focus was on the basics of language technology, language technology in Finland and in other countries and the main applications in language technology. In addition the focus was also on how language technology links to working life and how the European Union uses language technology. The source of the material used in the thesis was mainly written by well-known researchers. This adds to the reliability of the thesis. The thesis was carried out as a functional thesis and the whole work can be used as an information package of the basics of language technology. The work was also illustrated with PowerPoint-presentation. The work was directed to the students of language and culture.

As a result the work was accomplished as a compact information package of the basics of language technology. The work can be also used as a tool to get to know language technology for the people who are not familiar with the issue. Language technology has a growing role in working life. There are advantages for the language and culture students for learning the basics of language technology when graduating and stepping into international working life.

SISÄLLYS

TIIVISTELMÄ	2
ABSTRACT	3
1 JOHDANTO	6
2 KIELITEKNOLOGIA	7
2.1 Mitä on kieliteknologia?	7
2.2 Kieliteknologian historia	9
2.3 Suomen kielen erityispiirteitä	10
2.4 Kieliteknologia muissa maissa	10
3 KIELITEKNOLOGIAN SOVELLUKSIA	12
3.1 Kirjoittajan työkalut	12
3.1.1 Oikeinkirjoituksen tarkistus	12
3.1.2 Kieliopin tarkistus	13
3.1.3 Tavutus	13
3.1.4 Synonyymisanasto	13
3.1.5 Elektroniset sanakirjat	14
3.1.6 Luettavuuden arviointi	15
3.2 Tiedonhaku ja siihen liittyvät sovellukset	15
3.2.1 Tiedonhaku	15
3.2.2 Dokumenttien automaattinen luokittelu	17
3.2.3 Tekstin tiivistäminen	18
3.2.4 Automaattinen hakemistojen muodostaminen	18
3.3 Tietokone apuna kääntämisessä	18
3.3.1 Käännösmuistiohjelmat	19
3.3.2 Termipankit	19
3.3.3 Konekääntäminen	20
3.4 Puheteknologia	21
3.4.1 Puheen olemuksesta	22

3.4.2 Puhesynteesi	23
3.4.3 Puheentunnistus	23
3.4.4 Puheteknologia Suomessa	25
4 KIELITEKNOLOGIA KÄYTÄNNÖSSÄ	25
4.1 Suomalaisia kieliteknologiayrityksiä	26
4.2 Kieliteknologian hyödyt työelämässä	27
4.3 Kieliteknologia EU:ssa	28
4.4 Euroopan komission käännöspalvelu	31
5 OPETUSMATERIAALIA KYMENLAAKSON AMMATTIKORKEAKOULULLE	34
5.1 Johdon assistenttityön ja kielten koulutusohjelma	34
5.2 Kieliteknologian kurssi	35
5.3 PowerPoint-esitys ja sen ulkoasu	35
5.4 Sisältö ja runko	37
6 JOHTOPÄÄTÖKSET	37
LÄHTEET	40
LIITTEET	
Liite 1. PowerPoint-esitys	

1 JOHDANTO

Tämän opinnäytetyön tarkoituksena on tuottaa tietomateriaalia kieliteknologiasta Kymenlaakson ammattikorkeakoulun käyttöön. Työn taustalla on uusi Kymenlaakson ammattikorkeakoulun johdon assistenttityön ja kielten koulutusohjelma. Tässä koulutusohjelmassa kielten ja kulttuurin opintoihin suuntautuneet opiskelijat opiskelevat suuntautumisopinnoissaan kurssin kieliteknologiasta. Tämä on ensimmäinen kerta kun kieliteknologiaa opiskellaan Kymenlaakson ammattikorkeakoulussa.

Oma kiinnostukseni kieliteknologiaa kohtaan heräsi ollessani vaihdossa Unkarissa. Tapasin ulkomaalaisia opiskelijoita, jotka käyttivät konekäännösohjelmia koulutehtäviä tehdessään. Heidän kielillään konekäännösohjelmilla sai parempaa jälkeä kuin suomen kiemuraisemmalla kielellä, mutta sain kosketuksen ohjelmiin silloin ensimmäisen kerran. Kun palasin takaisin Suomeen, menin töihin yritykseen, jossa yksi työtehtävistäni oli puhelinvaihteeseen vastaaminen. Yrityksessä mietittiin puhelinjärjestelmän uusimista, ja tutustuin silloin ensimmäisen kerran sovelluksiin joita puheteknologia tarjoaa. Vaikka yritys on pieni, katkaisee puhelimeen vastaaminen aina muiden työtehtävieni tekoa. Yritykseen ei otettu käyttöön puheteknologiaa käyttävää puhelinjärjestelmä, mutta ymmärsin silti, kuinka suuri vaikutus puhelinvaihteen siirtämisellä automaattiseksi voi olla.

Toteutan työni toiminnallisena opinnäytetyönä. Työ toimii kokonaisuudessaan tietopakettina kieliteknologian perusteista. Teoriaosuuden lisäksi työtä havainnollistaa PowerPoint-esitys Kieliteknologia, mitä se on? jossa esitellään kieliteknologiaa ja sen sovelluksia. Esitys suunniteltiin opetuskäyttöön sopivaksi. Luon myös katsauksen kieliteknologiaan työelämässä, johon uudesta johdon assistenttityön ja kielten koulutusohjelmasta valmistuvat mitä todennäköisimmin sijoittuvat. Esityksen tavoite on myös herättää mielenkiintoa kieliteknologiaa kohtaan.

Kieliteknologia käsitteenä on vielä kovin tuntematon. Opinnäytetyötä kirjoittaessani törmäsin usein aiheeni ihmettelyihin, eikä juuri kukaan ei tuntunut tietävän mitä kieliteknologia tarkoittaa. Nämä kokemukset ovat taustalla opinnäytetyöni kysymyksen muodossa olevassa otsikossa. Haluan työlläni lisätä kieliteknologian tunnettavuutta sekä ymmärrystä siitä, mitä hyötyä kieliteknologian sovelluksista voi olla ja mitä niillä voi tehdä. Työn tarkoitus on myös pohtia uuden kieliteknologia kurssin tarpeellisuutta

kielten ja kulttuurin opiskelijoille. Tarkoituksena on luoda tiivis helposti esitettävä tietopaketti kieliteknologiasta, jota voidaan käyttää välineenä kieliteknologiaan tutustumiseen myös aiheeseen aiemmin tutustumattomille.

2 KIELITEKNOLOGIA

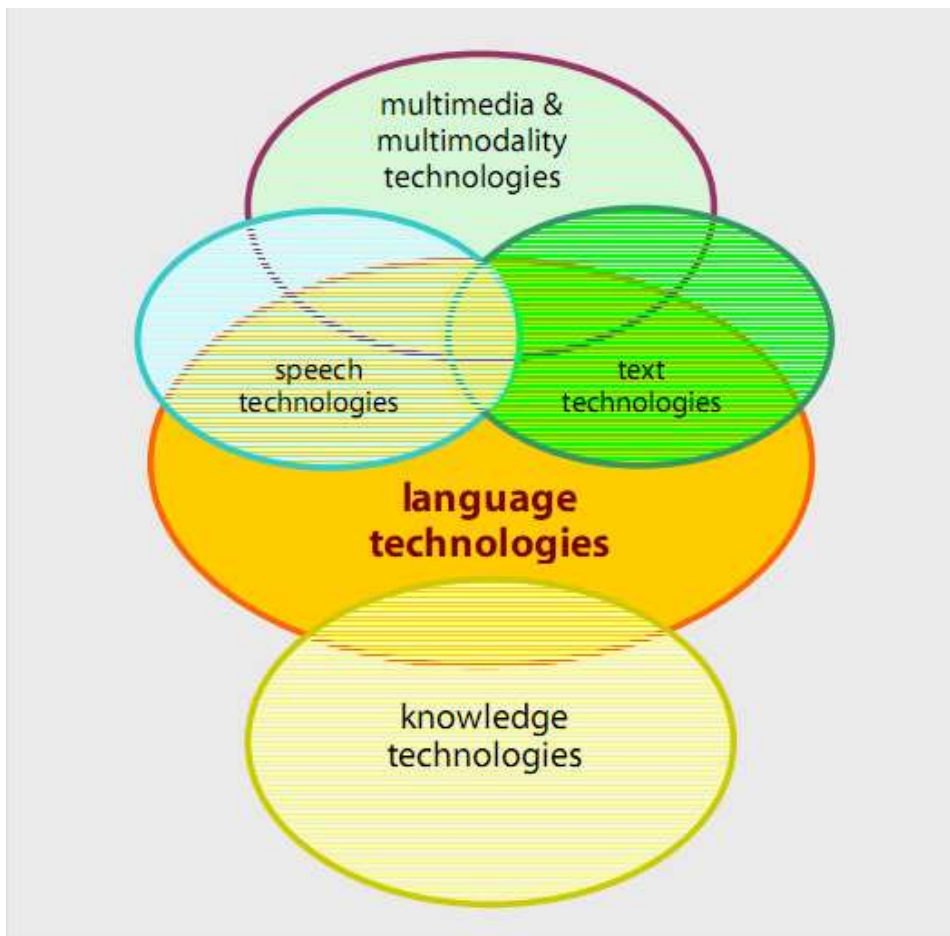
2.1 Mitä on kieliteknologia?

Kieliteknologia tutkii yhtä maailman monimutkaisinta ilmiötä, ihmisen käyttämää eli ihmisen luonnollista kieltä. Ihminen tuottaa kieltä sekä suullisesti puheena että kirjoitetussa muodossa. Puhe on ihmisten vanhin ja luonnollisin tapa kommunikoida, mutta paljon ihmisen tuottamaan tietoa säilytetään ja tuotetaan myös kirjoitetussa muodossa. Kieliteknologia käsittelee näitä molempia ihmisen kommunikointitapoja. Kieliteknologiaan kuuluu myös sovelluksia, joita käytetään sekä puheessa että kirjoituksessa. Näitä ovat esimerkiksi elektroniset sanakirjat sekä iso osa kielioppia. (Uszkoreit 2009, 1.)

Kieliteknologia sovelluksissa muokataan ihmisen luonnollista kieltä tietotekniikan avulla. Tietokone määrittää tuottamaan ja ymmärtämään ihmisen puhetta ja tekstiä. Tietokone voi käsitellä samassa sovelluksessa yhtä tai useampaa kieltä. Kielitaidon määrittäminen tietokoneelle ei ole aivan yksinkertaista. Ihmisen kielitaito ei ole vain kielen osaamista, vaan perustuu tilannetajuun, ihmisen tietoon maailmasta ja elämäkokemuksiin. Tietokone ei voi täysin ymmärtää kieltä ihmisen lailla. (Carlson 2005, 237–238.) Tutkijoilla ei vielä ole tietoa siitä, kuinka kieli, tieto ja ajattelu toimivat ihmisten aivoissa. Tämä tuottaa vaikeuksia yrittäessä opettaa ihmisen kieltä tietokoneelle. Kieliteknologia pyrkii kuitenkin tuottamaan ja jäljittelemään ihmisen kieltä niin, että se toimisi oikeassa maailmassa. (Uszkoreit 2009, 1.)

Kieliteknologia voi auttaa ihmisiä kommunikoimaan keskenään. Ihmisillä on aina ollut ongelmia ymmärtää eri äidinkieltä puhuvien kieltä. Yksi kieliteknologian alkupe-
räisistä tarkoituksista onkin aina ollut täysin automaattinen kielen kääntäminen toiselle kielelle. Tutkijat ovat kuitenkin joutuneet huomaamaan, että nykyiset sovellukset ovat vielä kaukana tästä tarkoituksesta. Niistä voi kuitenkin muuten olla suurta apua ihmisten välisessä kommunikaatiossa. (Uszkoreit 2009, 2.)

Kun kommunikoimme keskenämme, emme vain puhu tai kirjoita, vaan sekoitamme eri kielen muotoja ja muita viestimistapoja. Kuva 1 kuvastaa tätä kommunikaatiotapojen sekoittumista. Yhdistämme kommunikoinnissamme puhetta elekieleen ja kasvojen ilmaisiin. Myös sähköisiin teksteihin liitetään kuvia ja ääniä, ja esimerkiksi elokuvat voivat sisältää kieltä sekä puhutussa että kirjoitetussa muodossa. Näin ollen eri kieli-tekniikan sovellusalueet ovat vuorovaikutuksessa muiden teknologioiden ja kielen kanssa. (Uszkoreit 2009, 1.)



Kuva 1. Yhdistämme kommunikoinnissamme monia kielen eri esiintymismuotoja ja teknologioita (Uszkoreit 2009, 1)

Tietokoneen käyttö ihmisten työvälineenä on nykyisin erittäin yleistä alasta tai ammattista riippumatta. Suuri osa tietokonetta työvälineenä käyttävistä ei ole kiinnostunut tietokoneesta ja sen toiminnasta sen syvällisemmin, vaan tietokonetta pidetään välineenä joka auttaa päivittäisten työtehtävien suorittamisessa. Ihmisten ja tietokoneiden välinen työnjako tulisi hoitaa niin, että tietokoneet suorittavat rutiininomaiset, nopeut- ta ja laskutaitoa vaativat työt. Ihmiset voivat siten keskittyä enemmän ylläpitoa, val-

vontaa ja päätöksentekoa koskeviin työtehtäviin. Näin voidaan lisätä ihmisten tekemän työn mielekkyyttä. (Nurmesniemi 2001, 203–204.)

Iso osa yritysten ja organisaatioiden tiedoista sekä aineistoista sijaitsee tietokoneilla ja Internetissä. Kansainvälistyvä maailma asettaa ihmisille vaatimuksia oppia ja käyttää työssä muitakin kieliä kuin äidinkieltään. Kieliteknologia auttaa ymmärtämään vieraskielistä tekstiä ja helpottaa vieraan kielen oppimista. Tietokonetta voi myös käyttää ajasta ja paikasta riippumatta. (Koskenniemi 2002, 2.)

Globaalit työmarkkinat asettavat entistä suurempia haasteita myös yritysten työntekijöille. Vieraalla kielellä kommunikointi ja monikulttuurisuus ovat yhä tiiviimmin läsnä päivittäisessä työnteossa. Sosiaalinen media, sähköposti ja blogit ovat kiihdyttäneet tätä muutosta entisestään. Samaan aikaan monia aiemmin kielen ammattilaisten työtehtäviä on siirretty muille työntekijöille, kuten myyntineuvottelijoille, lakimiehille, kirjanpitäjille ja tuotepäälliköille. (Kites symposium 2009.) Nämä ovat myös mahdollisia valmistuneen tradenomin tehtäviä.

Kieliteknologian avulla voidaan lisätä myös kielellistä tasa-arvoa. Teknologian yhteydessä puhutaan yleensä aina asioiden kaupallistamisista, ja näin voidaan puhua myös kieliteknologian saralla. Kieltä ei silti usein haluta nähdä työkaluna ja teknologiana, vaan se halutaan usein kuvailla vain esteettiseksi tekijäksi ja identiteetin perustaksi. Kielen avulla voidaan kuitenkin saada voittoa tuottavaa toimintaa ja nostaa taloutta. Kieliteknologian kehittyessä tulevaisuudessa voidaan kohdata uusia merkittäviä kysymyksiä kieleen liittyen, kuten tuleeko ihmisen itse ylipäänsä osata kieliä, jos kieliteknologia korvaa henkilökohtaisen kielitaidon. (Reiman 2007.)

2.2 Kieliteknologian historia

Idea siitä, että tietokone tuottaisi ihmisen kieltä, on yhtä vanha kuin tietokoneet itse. Kieliteknologian ensimmäinen sovellus oli konekäännös, jonka kehittäminen alkoi toisen maailmansodan jälkeen. Kehitystä edisti kylmän sodan aikainen tarve ymmärtää venäjää ja englantia. Ensimmäiset käännösohjelmat korvasivat lähdekielen sanan kohdekielisellä vastikkeella. Alun menestyksen jälkeen kehitys kuitenkin hiipui. Yksi ensimmäisistä, edelleen toiminnassa olevista konekäännösohjelmista on Systran, jota EU käyttää omissa käännöstoiminnoissaan. (Carlson 2005, 247)

Tietokoneiden nopea yleistyminen 1980-luvulla mullisti kieliteknologian kehitystä. Tekstinkäsittelyohjelmista tuli jokapäiväinen työskentelyväline, ja informaatiota alettiin tuottaa ja tallentaa yhä enemmän elektronisessa muodossa. Ensimmäiset tavutusohjelmat sekä oikeinkirjoituksen- ja kieliopintarkistimet otettiin käyttöön 1980-luvun lopulla. Myös elektroniset sanakirjat kehitettiin samoihin aikoihin. (Arppe 2001.)

Lopullisen sysäyksen kieliteknologian tutkimukselle ja käytölle loi tietotekniikan kehityksen luoma informaatiomäärän räjähtänyt kasvu. Kieliteknologian sovelluksille luo kysyntää myös nykypäivän kansanvälinen työympäristö, jossa usein kommunikoidaan monilla kielillä. Toisaalta halutaan silti pitää kiinni omasta kielestä ja kulttuurista. Kauppaa käydään paljon ulkomaisten kanssa, ja myös ihmisten liikkuvuus on lisääntynyt. Yritykset tavoittelevat toiminnassaan tehokkuutta, laatua sekä automatisointia. Monia yritysten toimintoja, mitkä vain on mahdollista siirtää ihmiseltä pois, siirretään usein tehokkaille koneille. Nämä luovat kysyntää myös kieliteknologiasovelluksille. (Arppe 2001a.)

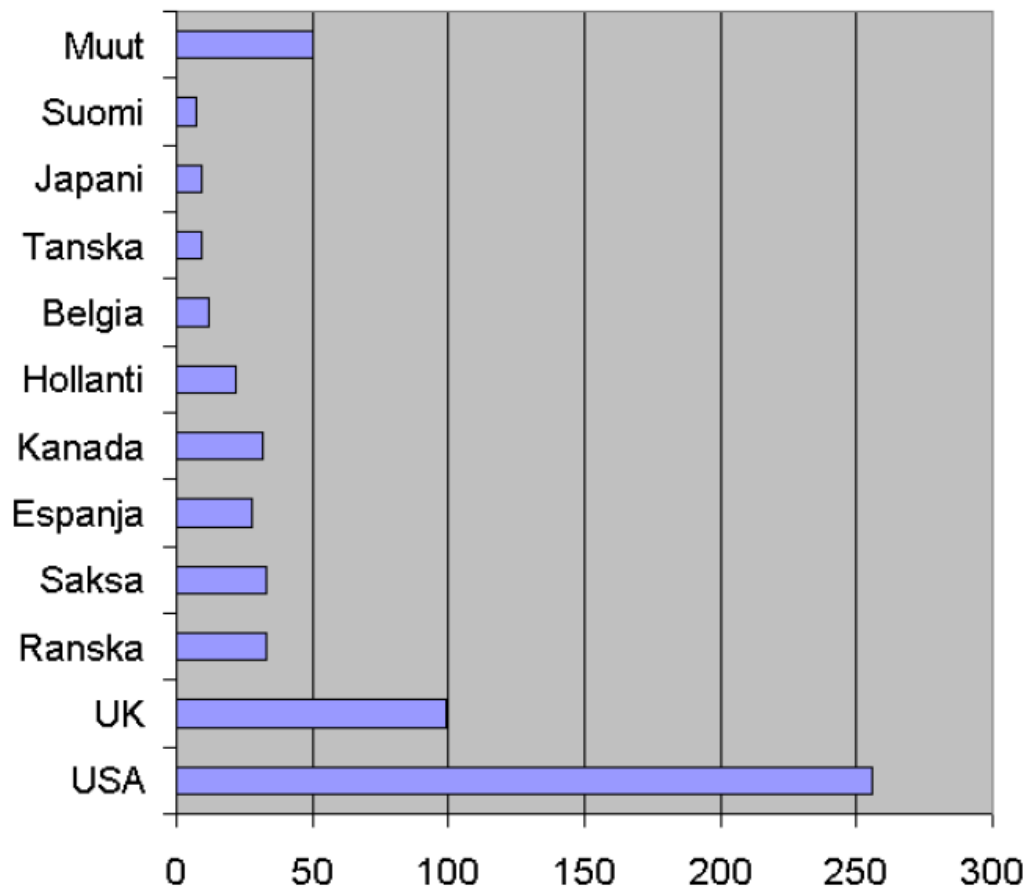
2.3 Suomen kielen erityispiirteitä

Suomalaiset kieliteknologit pitävät suomen kieltä tärkeänä synnyinlahjana. Suomen kielessä on muun muassa paljon eri taivutusmuotoja sekä yhdyssanojen käyttö on yleistä. Suomen kielen rakenteellinen rikkaus tuokin omia erityispiirteitä kieliteknologian suomenkielisiin sovelluksiin. Englanti on ollut hallitsevassa asemassa kieliteknologian alalla. Englannin kieli on rakenteeltaan yksinkertaisempaa kuin suomen kieli, ja kun englantia varten pystytään kehittämään kielenkäsittelyohjelmiston noin sadalla tuhannella sanalla, suomen kieleen tarvitaan samalla tekniikalla vähintään kymmeniä miljoonia sanamuotoja. Kieliteknologian sovellukset, jotka sopivat englannin kieleen, eivät välttämättä siis sovellu suomen kieleen. (Arppe 2008.)

2.4 Kieliteknologia muissa maissa

Varhaisimmat kaupalliset kieliteknologiayritykset syntyivät Yhdysvalloissa. Kuvasta 2 voidaan nähdä, että Yhdysvalloissa sijaitsee edelleen suuri osa kieliteknologiayrityksistä. Yhdysvalloissa ollaan kieliteknologian kehityksessä muita maita edellä, ja esimerkiksi puheteknologian sovellukset ovat arkipäivää. Osasyynä tähän on myös se, että englannin kieltä on tutkittu paljon. (Arppe 2001.)

Organisaatiot maittain Language Engineering Directoryssä 1996



Kuva 2. Kieliteknologia yritykset maittain vuonna 1996 (Arppe 2001a)

Myös pohjoismaissa kieliteknologiaa on tutkittu kauan. Korkeakouluopetusta järjestetään kaikissa pohjoismaissa, joista Ruotsissa yliopistotasoisessa kieliteknologian koulutuksessa on eniten opiskelupaikkoja verrattuna muihin pohjoismaihin. Saksassa panostetaan paljon erilaisiin kieliteknologian tutkimusohjelmiin. Belgiassa toimii suuri Lernout & Houspie -niminen kieliteknologia yritys, joka on koko maailman kieliteknologia yrityksiä verrattaessa huomattavan suuren kokoinen. Lernout & Hosupien palveluksessa on noin 1700 työntekijää ja yritys kehittää monipuolisesti kieli- ja puheteknologiaa. (Kieliteknologian korkeakouluopetus: Tarve, tarjonta ja ehdotukset toimenpiteiksi 1999, 5.)

3 KIELITEKNOLOGIAN SOVELLUKSIA

3.1 Kirjoittajan työkalut

Kirjoittaessamme tekstiä kirjoitustyötä ohjaa kyseiselle kielelle ominaiset kielioppisäännöt ja normit. Kirjoittajan työkalut auttavat kirjoittajaa kirjoittamaan näiden sääntöjen mukaista tekstiä. Työkalut huomaavat kielioppivirheiden lisäksi myös tahattomia sorminäppäilyvirheitä. Monesti sanotaan, että suomen kieltä kirjoitetaan niin kuin äännetään, ja sen vuoksi suomalaiset tietävät, kuinka kirjoittaa oikein. Suomen kielessä on kuitenkin paljon sopimuksenvaraisia ja opittavia asioita. Monissa muissa kielissä ääntäminen ja kirjoitus ovat erilaisia toisistaan kuin suomessa, ja myös kirjoitus- ja lukutaito heikompaa. Tällöin kirjoitetaan helposti tekstiä niin, kuin se kuulostaisi oikealta luettuna. Ajatus tekstissä voi olla oikein, mutta oikeinkirjoitus ja kielioppi eivät välttämättä ole normien mukaista. Kirjoittajan työkalut auttavat myös kirjoittamaan vieraskielistä tekstiä oikein. Kirjoittajan työkalut ovat yleisimpiä käytettäviä kieliteknologian sovelluksia. Ne helpottavat useiden ihmisten päivittäistä kirjoitustyötä. (Koskeniemi 2002b, 1-2.)

3.1.1 Oikeinkirjoituksen tarkistus

Kirjoittaessa tekstiä tietokoneella voi välillä tehdä kirjainten lyöntivirheitä tai kirjoittaa sanoja väärin. Oikeinkirjoituksen tarkistus on tietokoneohjelma tai sen sovellus joka tarkistaa, että sanat on kirjoitettu oikein. Tarkistin vertaa kirjoitettua sanaa sanastoihin. Jos sanaa ei löydy sanastosta, tulkitsee tarkistin sanan väärinkirjoitetuksi ja ilmoittaa siitä useissa ohjelmissa värillisellä alleviivauksella. Ohjelma voi myös antaa ehdotuksia oikeaksi sanaksi. Oikeinkirjoituksen tarkistusohjelmat ovat laajalti käytössä ja ne kuuluvat automaattisesti mukaan moniin tekstinkäsittelyohjelmiin. (Koskeniemi 2002b, 2-3.)

Kaikki tunnistamattomat sanat eivät ole virheitä, vaan ne voivat olla organisaatioiden tai ihmisten nimiä, lyhenteitä tai muuta erikoissanastoa, joita tarkistin ei tunnista. Käyttäjä voi tarvittaessa lisätä usein käyttämänsä erikoissanat sanastoihin. Oikeinkirjoituksen ongelmana voi olla, että sana on kirjoitettu muodollisesti oikein, mutta se ei kuulu asiayhteyteen. (Koskeniemi 2002b, 2.)

3.1.2 Kieliopin tarkistus

Kieliopin tarkistus auttaa etsimään kirjoitetusta tekstistä kielioppivirheitä ja ehdottaa korjauksia. Virheitä voi syntyä lyöntivirheistä tai tekstinosia siirreltäessä, jolloin tekstiin voi jäädä väärä sananmuoto. Kieliopin tarkistusta käytetään myös kirjoitettaessa vieraita kieliä. (Koskenniemi 2002b, 3.)

Kieliopin tarkistus on hyvä työkalu vajavaisesti vierasta kieltä osaavalle. Se auttaa saamaan tekstin ilmaisusta luontevampaa ja selkeämpää. Kehittyneimmät tarkistimet ymmärtävät yhä monimutkaisempia ja pidempiä lauserakenteita. Kieliopin tarkistin voi huomauttaa käyttäjälle sanajärjestyksestä, vääristä tai puuttuvista prepositioista ja artikkeleista sekä puuttuvista pilkuista. (Kielenhuolto 2009b, 3.)

3.1.3 Tavutus

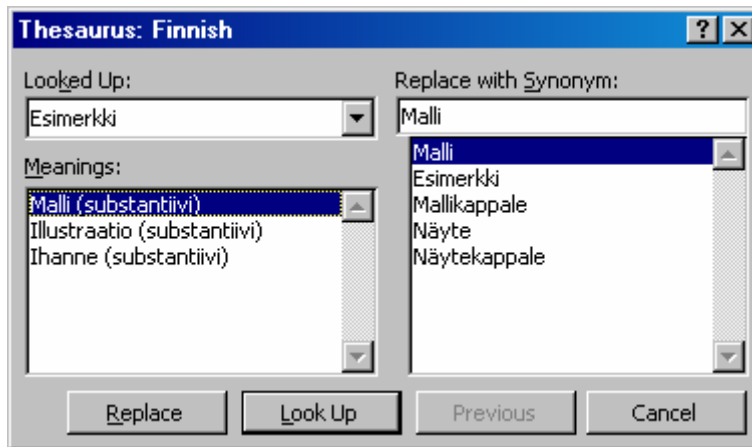
Tavutusta käytettäessä sana, jota ei mahdu riville, tavutetaan. Tällöin tekstiin ei jää välejä ja palstoista tulee tasaisia. Käyttäjä voi useissa tavutusohjelmissa itse määrittää tavutusasetuksia ja estää tarvittaessa tavutuksen käytön tietyissä sanoissa. Tavutusta käyttäessä tulee aina ottaa huomioon kieli jota käyttää. Suomalaisissa sanoissa ovat erilaiset tavutussäännöt kuin muunkielisissä teksteissä. (Tietoja tavutuksesta 2009.)

Useissa kielissä on sääntöjä siitä, kuinka tavutus tulee tehdä. Sääntöjen tarkoitus on saada sanojen jako sellaiseksi, että lukija pystyy mahdollisimman helposti hahmottamaan jaetun sanan oikein. Suomen kielessä tavutetaan ääntämyksen mukaan, kun taas esimerkiksi ruotsissa tavutetaan sekä ääntämyksen että sanojen muodon perusteella. Englannissa on käytössä tavutussanakirjoja, jotka määräävät sanojen sopivat tavutuskohdat. (Koskenniemi 200b, 3.)

3.1.4 Synonyymisanasto

Synonyymisanastolla voidaan korvata sana toisella, joka tarkoittaa samaa asiaa. Näin tekstissä voidaan ehkäistä häiritsevää saman sanan toistoa ja saada tekstistä elävämää. Synonyymisanastot osaavat usein myös löytää sanat taivutusmuotoisena, jolloin ne voidaan siirtää suoraan tekstiin. (Koskenniemi 2002b, 3.) Kuvasta 3 näkyy, kuinka

synonyymisanasto tuo sanasta listan synonyymeja. Synonyymisanaston avulla voidaan hakea myös toista asiaa paremmin kuvaavaa sanaa. (Tekstinkäsittelytyökaluja 2003.)



Kuva 3. Suomalainen synonyymisanasto (Tekstinkäsittelytyökaluja 2003)

3.1.5 Elektroniset sanakirjat

Elektroniset sanakirjat ovat sähköisessä muodossa olevia sanakirjoja. Ne ovat helposti sekä nopeasti käytettäviä, ja niihin voidaan yhdistää kieliteknologiasovelluksia, joka parantaa niiden käytettävyyttä. Elektroninen sanakirja mahdollistaa suuretkin aineistot, mikä on ongelmana kirjallisissa sanakirjoissa. Useisiin elektronisiin sanakirjoihin sanan voi syöttää sellaisenaan, eikä sitä tarvitse muuttaa perusmuotoon kuten painetussa sanakirjassa. Sanakirja voi myös kääntää sanan suoraan syötetyn sanan taivutusmuotoon. Elektroninen sanakirja tarjoaa usein käännöksiä hakusanan alamerkityksille. Kehittyneimmät sanakirjat osaavat arvioida alamerkityksen sopivuutta ympäröivän tekstin perusteella. (Koskenniemi 2002a, 3-4.)

Isoissa yrityksissä on tärkeää, että yrityksen viestintä sujuu mahdollisimman helposti ja vaivattomasti helposti eri kielillä. Yritykset voivat hankkia yleiskielen sanakirjoja, tekniikan ja kaupan sanakirjoja sekä yrityksen toimialalle tärkeiden erikoisalojen sanakirjoja helpottamaan viestintää. Kansainvälisen yrityksen työntekijät joutuvat usein matkustamaan työssään paljon. Sanakirjat voidaan myös asentaa kannettaviin tietokoneeseen, jolloin paljon matkustava pystyy työskentelemään tehokkaasti myös matkojen aikana. Sanakirjoja voi asentaa myös matkapuhelimiin, jolloin sanakirja on käytettävissä myös silloin, kun tietokone ei ole saavilla. Elektronisten sanakirjojen yhteyteen

saadaan usein liitettyä myös termisanasoja, jotka lisäävät entisestään sanakirjojen käytettävyyttä. (Suuryritys 2009.)

3.1.6 Luettavuuden arviointi

Luettavuuden arvioinnilla voidaan arvioida, onko kirjoitettu teksti helppolukuista. Ohjelma mittaa tekstin luettavuutta tekstin pituuden, rakenteiden ja sanojen perusteella. Jos esimerkiksi teksti sisältää tavanomaista enemmän harvinaisia sanoja, se voi olla vaikeasti luettavaa. Luettavuuden arviointi auttaa saamaan varsinkin vieraskielisestä tekstistä helppolukuista. (Koskenniemi 2002b, 4.)

3.2 Tiedonhaku ja siihen liittyvät sovellukset

Kirjallisen tiedon saatavissa oleva määrä on kasvanut valtavasti lyhyessä ajassa (Koskenniemi 2002d, 1). Etenkin internet on jatkuvasti kasvava ja muuttuva tiedonlähde. Internetsivujen tietosisältö lisääntyy joka päivä, uusia sivuja tehdään ja vanhoja häviää ja sivujen sisältö muuttuu nopeasti. Kun internetiä käytetään tiedonhakuun, voidaan hakupalveluiden avulla kohdistaa haut moniin internetin sisältöihin, kuten www-sivuihin, uutisiin, artikkeleihin, sähköpostiosoitteisiin sekä erilaisiin tiedon esittämismuotoihin kuten videotiedostot. Kun tietoa etsitään tavallisista kirjastoista, aineistot ovat tarkasti luokiteltuja ja helposti löydettävissä. Ne ovat myös ylittäneet tiedon julkaisukynnyksen. Internetiin voi taas kuka tahansa tuottaa tietoa, joten internet lähteissä tulee aina olla kriittinen tiedon suhteen. (Tiedonhaku internetistä 2009.)

Lisääntynyt tietomassa tuo mukanaan ongelman, kuinka löytää suuren tietomassan sisältä itseä kiinnostavaa tietoa. Tietojen etsimiseksi on kehitelty monia välineitä, jotta pärjättäisiin suurten tietomäärien kanssa. (Koskenniemi, 2002d, 1.)

3.2.1 Tiedonhaku

Pitkään aikaan ei ole ollut ongelmana, että tietoa ei löytyisi, vaan se, että tiedon haku tuottaa liikaa tuloksia. Tiedonhaulla tarkoitetaan yksinkertaisesti sen tiedon löytymistä, jota etsitään. Ohjelmien, jolla tietoa haetaan, olisi hyvä tarjota myös parhaiten aiheeseen sopivaa tietoa ensimmäiseksi. Tietoa haetaan useimmiten aiheeseen sopivien sanamuotojen perusteella, koska se on helppoa toteuttaa. Monissa kielissä hakusanat

esiintyvät useissa taivutusmuodoissa, ja voi olla vaikea laatia haku, jossa kaikki sanan taivutusmuodot löytyisivät. Hakuun voi myös liittyä sanoja, joilla on erilaisia alamerkityksiä, jotka eivät kaikki välttämättä liity haettavaan asiaan. Näitä ongelmia voidaan ratkaista kieliteknologian avulla. (Koskenniemi 2002d, 2.)

Tiedonhakuun liittyy kaksi tärkeää ominaisuutta, saanti ja tarkkuus. Saanti kuvaa sitä prosenttiosuutta olennaisesta tiedosta, joka tuli esille haun avulla kaikesta halutusta tiedosta. Tarkkuus kuvaa sitä, kuinka suuri prosenttiosuus haun tuomista tiedosta oli merkityksellistä. Tavoitteena on, että haku onnistuisi sataprosenttisesti molemmilla ominaisuuksilla, mutta käytännössä tällaiseen ei voida ikinä täysin päästä. Jos halutaan saada mahdollisimman korkea saantitulos hausta, saadaan paljon tietoa, mutta huono tiedon tarkkuus. Jos taas halutaan saada mahdollisimman tarkkaa tietoa, saadaan vain yksi tietolähde. Siksi kumpikin ääritapaus on useimmiten käyttäjälle hyödytön. (Koskenniemi 2002d, 2-3.)

Haku suoritetaan yleensä haettavassa tiedossa esiintyvien sanojen perusteella niin, että hakutulokseen hyväksytään vain sellaista tietoa, jossa vaaditut sanat esiintyvät. Arvioimiseen kuinka merkityksellistä tieto on, käytetään erilaisia tilastollisia menetelmiä sen mukaan, kuinka hyvin esimerkiksi haussa esiintyvät sanat yleensä erottelevat tietoa tai mitä lisätietoa hakuun saadaan parhaiten varsinaisten hakusanojen lisäksi. Haut voivat myös etsiä ja esittää vaihtoehtoja hakusanoille ja mahdollistaa kieltoja. Internetissä on vapaasti käytettävä erilaisia hakupalveluita. Osa näistä tarjoaa etukäteen laaditun luokituksen mukaan tietoa, osa etsii tietoa eri puolilta verkkoa sivujen sisältämien sanojen perusteella. Haun sisältämien sanojen perusteella hakeminen perustuu hakurobottiin eli ohjelmaan, joka kulkee ympäri internetiä ja kerää sivuja, joissa hakusanat esiintyvät. Tällä tavalla löydetystä verkkosivuista laitetaan haun pohjaksi haussa esiintyvät sanamuodot. Tavallinen haku koostuu siis joukosta sanamuotoja, joiden edellytetään esiintyvät haettavassa tiedossa. Vaativimmissa hauissa määritellään haettavassa tiedossa vaadittavat sanamuotojen yhdistelmät. Tavoitteena on, että hakukone löytää haun perusteella joukon tietolähteitä, joiden joukossa hakijalle oikeaa tietoa sisältäviä tietolähteitä. (Koskenniemi 2002d, 2-3.) Yksi tunnetuimmista hakukoneista on Google, joka on siis kieliteknologinen sovellus sekkin.

Internetistä löytyvät hakukoneet ovat yleensä ilmaisia ja helppokäyttöisiä. Ulkomaiset hakukoneet ovat Suomessakin suosituimpia, mutta haettaessa tietoa Suomesta, suoma-

laiset hakukoneet saattavat olla jopa parempia osumatarkkuudessa kuin ulkomaiset. Pelkkä osumien määrä ei myöskään aina kerro sitä, onko hakukone hyvä, tärkeämpää on se, että sen antamat linkit ovat ajantasaisia. Lisäksi on hyvä muistaa, että tehokas internetin hakupalvelu on hyödytön, ellei osaa valita oikeita hakusanoja. Perushaun sijasta kannatta myös käyttää laajennettua hakua ja niiden mahdollisia rajausmahdollisuuksia. Eri hakukoneiden hakutavat vaihtelevat, joten on tärkeää tutustua käytettävän hakupalvelun ohjeisiin varmistaakseen mahdollisimman onnistuneen haun. (Tiedonhaku internetistä 2009.)

3.2.2 Dokumenttien automaattinen luokittelu

Dokumenttien automaattinen luokittelu on samantapainen tehtävä kuin tiedonhaku. Dokumenttien automaattisessa luokittelussa dokumentti luokitellaan sisältönsä perusteella. Tästä on hyötyä esimerkiksi lehden toimituksessa, jonne saapuu valtavat määrät uutisviestejä. Viestit ovat tärkeitä vain tietyille toimittajille, ja viestit pitää osata ohjata oikealle henkilölle. Myös esimerkiksi tietoja arkistoidessa halutaan liittää dokumentteihin luokituksia, jotta aineisto löytyy helposti myöhemmin tarvittaessa. Dokumenttien luokitus tehdään yleensä dokumentissa esiintyvien sanamuotojen perusteella. Englanninkielisissä dokumenteissa tämä toimiikin hyvin, sillä englannin kielessä sanat taipuvat vähän. Muissa kielissä, joissa sanat taipuvat enemmän, hakusanat voivat olla sanamuodon sijasta parempi vaihtoehto, koska erilaisia hakusanoja saattaa olla vähemmän kuin sanamuotoja, ja samaa asiaa käsittelevällä lyhyellä dokumentilla on todennäköisemmin yhteinen hakusana kuin yhteinen sanamuoto. (Koskenniemi 2002d, 3.)

Yksi suosituista useita ihmisiä koskettavasta dokumenttien automaattisesta luokittelusta liittyy roskapostiin, jota tulvii sähköpostiin aina vain enemmän. Aiemmin viestijä on yritetty estää niiden alkuperän mukaan, mutta nykyään roskapostin lähettäjät osaavat väärentää viestien tietoja. Sisällön perusteella tapahtuva roskapostien torjunta voi toimia hyvin postin karsinnassa, varsinkin jos ohjelmalla on ajantasainen malli tyypillisimmistä roskaposteista. (Koskenniemi 2002d, 4.)

3.2.3 Tekstin tiivistäminen

Automaattinen tiivistelmän tekeminen tekstistä on tarpeellinen ja haluttu toiminto. Pitkästä tekstistä halutaan usein saada tietoa siitä, mitä dokumentti pääasiassa käsittelee. Automaattinen tekstin tiivistäminen tehdään usein niin, että tekstistä eristetään sellainen osa, joka kriteerien mukaan sisältää todennäköisesti tärkeää asiaa tekstistä. Kriteerinä voidaan pitää esimerkiksi sitä, että tekstin ensimmäinen ja viimeinen kappale ovat usein tärkeitä, tai valitsemalla kappale, jossa on runsaasti eri tekstejä toisistaan erottavia sanamuotoja tai hakusanoja. Kriteerin tulee olla tekstille tärkeä ja kuvaileva. Tekstin automaattisen tiivistämisen avulla voidaan käyttää hyväksi esimerkiksi tietoa haettaessa, jolloin voidaan tiivistelmän perusteella tehdä päätös, haluaako tekstiä lukea vai ohittaa sen. (Koskenniemi 2002d, 4.)

3.2.4 Automaattinen hakemistojen muodostaminen

Hakemistojen muodostaminen on tavanomaisista kirjoista tuttu asia. Tällöin tekstistä halutaan tunnistaa sisällön kannalta tärkeät termit, ja ottaa tekstistä huomioon ne kohdat, joista termiin liittyvässä yhteydessä puhutaan paljon. Automaattisella hakemiston muodostamisella voidaan päästä käsintehtyä hakemistoa parempaan laatuun. Hakemistoon pitäisi valita sellaisia termejä, jotka sisältävät paljon informaatiota. Tätä voidaan arvioida yksinkertaisesti vertaamalla, kuinka usein termi esiintyy kaikkien dokumenttien joukossa. Paljon informaatiota sisältävät termit keskittyvät osaan dokumentista. Lisäksi termien informaatiota voi mitata myös ilmaisujen esiintymisellä tietyissä lauseissa tai tietyissä ilmauksissa. (Koskenniemi 2002d, 4.)

3.3 Tietokone apuna kääntämisessä

Maailmassa käytetään tuhansia eri kieliä. Kieli ja kulttuuri ovat olennainen osa toisistaan. Kääntäjän tulee ottaa huomioon käännettävän kielen kulttuuri ja yhteiskuntaan liittyvät tyypilliset piirteet. Näitä voivat olla esimerkiksi koulujärjestelmä ja sosiaaliturva. Joissain kielissä ei välttämättä edes tunneta sanoja jonkun toisen kielen kulttuurin ilmaisulle. Eri kulttuureissa on myös erilaisia tapoja. Toisissa kulttuureissa mennään suoraan asiaa, kun toisissa taas noudatetaan tiettyjä muodollisuuksia. Kääntämisen vaikeus riippuu myös siitä, kuinka lähellä kohdekieli on lähtökieltä. Tekstejä toi-

selle käännettäessä tulee ottaa huomioon ihmisen kielen moniselitteisyys. (Koskeniemi 2002a, 1-3.)

3.3.1 Käännösmuistiohjelmat

Käännösmuistit ovat tietokantoja, joihin voi tallentaa tehdyn käännöksen myöhempää käyttöä varten. Käännösmuistista voidaan etsiä tallennetuista teksteistä uuteen käännettävään tekstiin vastaavuuksia. Osa ohjelmista tunnistaa lauseen vain, jos se on tallennetun käännöksen täydellinen vastine, osa tunnistaa lauseen, jos niissä on vain osa samaa. (Alho 2005, 283.)

Käännösmuistiohjelmat soveltuvat parhaiten laajojen ja usein toistuvien dokumenttien kääntämiseen. Ohjelman käyttö säästää aikaa ja kustannuksia. Käännösmuistia voidaan myös muokata ja tallentaa käyttäjän omien tarpeiden mukaan. Käännösmuistit tuovat tekstiin kielellistä yhdenmukaisuutta sekä yhtenäistä laatua varsinkin usean käyttäjän työympäristössä. (Nurmesniemi 2001, 207–212.)

Ohjelmaan tallennetut aineistot vaativat säännöllistä päivitystä ja tarkistusta. Päivitykseen voi isoissa organisaatioissa kulua paljon aikaa. Muistiin tallennettu virhe toistaa jokaisella käyttökerralla samanlaisena, jos sitä ei korjata. Käännösmuistiohjelmaa käytettäessä tulee aina tarkistaa lopullinen tulos ja tarvittaessa täydentää käännöstä. Käännösmuistiohjelmiin voidaan usein liittää myös muita kääntämistä helpottavia ohjelmia. (Nurmesniemi 2001, 208–211.)

3.3.2 Termipankit

Tekstissä käytettävät termit ovat olennainen osa käännoistyötä. Jotta käännoös olisi laadukasta, kääntäjä etsii lähdetekstille parhaiten sopivat termit. Yksi kieliteknologian sovelluksesta ovat termipankit, jotka auttavat kääntäjää oikeiden termien löytämisessä. (Koskeniemi 2002a, 3.)

Termipankkien tarkoitus on poimia käännettävästä tekstistä termejä, joiden vastineita sisältyy termipankkiin. Tällöin kääntäjän ei tarvitse etsiä termejä ja termit pysyvät yhdenmukaisena tekstissä. Tekstin yhdenmukaisuus on tärkeää myös, jos useat henkilöt

kääntävät samaa tekstiä. Termipankit ovat usein monikielisiä ja ne voivat olla omana ohjelmanaan tai osana käännösohjelmaa. (Alho 2005, 284–286.)

Termipankkien käyttö yhtenäistää yrityksen viestintää. Ne voidaan usein yhdistää myös sanakirjoihin, ja ne voidaan mukauttaa asiakkaan mukaan. Termipankit voivat sisältää termien lisäksi niiden synonyymejä, määritelmiä, kielioppitietoa ja käyttöesimerkkejä. Termipankkien tietoja voidaan usein myös muokata ja lisätä tarpeen mukaan. (Termien hallinta 2009.)

3.3.3 Konekääntäminen

Konekäännin on ohjelma, joka kääntää kokonaisia lauseita ja tekstejä kohdekielelle. Konekäännin eroaa käännösmuistiohjelmasta siinä, että se tuottaa käännöksiä myös entuudestaan tuntemattomille lauseille. (Carlson 2005, 240)

Konekääntämisessä tietokone toimii kääntäjänä. Ihminen toimii vain alkuperäisen tekstin kirjoittajana ja tietokoneen kääntämän tekstin tarkastajana. Konekääntämisen tarkoitus on, että ohjelma sisältää kaikki tiedot, joita tarvitaan tekstin kääntämiseen ilman ihmisen apua. Ohjelma käyttää hyväksi tietokoneen laskentakykyä, sekä analysoi lähtökielen lauseiden rakenteita ja ilmaisuja. Konekäännin pyrkii myös säilyttämään lähtötekstin tyylin sekä lauseiden merkitykset. (Alho 2005, 282–283.)

Ihmisen tuottamalle kielelle on tyypillistä, että ilmaisuilla on monia merkityksiä ja kulloinenkin merkitys riippuu tilanteesta. Konekäännin on vaikeuksissa tämän ihmiskielen monitulkintaisuuden ja tilannetaujan ymmärtämisen kanssa. Kone ei voi ymmärtää kieltä ihmisen tavoin. (Carlson 2005, 242.)

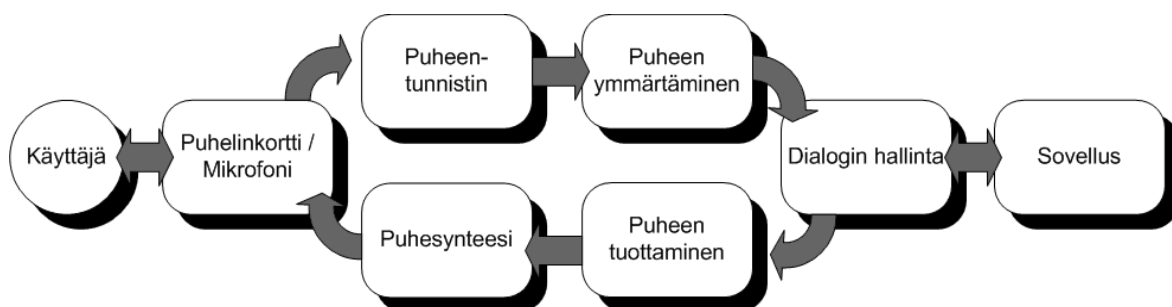
Nykyiset konekäännösohjelmat pystyvät parhaiten käsittelemään yleiskielisiä tekstejä, joissa lauserakenteet ovat yksinkertaisia ja joka on kirjoitettu kieliopillisesti hyvin. Parhaiten koneen käännettäväksi soveltuvat tekstit, joiden asiasisältö on tyyliä tärkeämpi, kuten käyttöohjeet ja tekniset tekstit. Konekäännin käsittelee usein virkkeen kerrallaan ja koneelle voivat olla hankalia virkkeiden väliset viittaukset. Sanonnat, runot ja sanaleikit ovat konekääntimelle mahdottomia kääntää. Sama koskee myös puhekieltä, murteita ja ammattitermejä. Käännösohjelmalle voidaan kuitenkin opettaa esimerkiksi tietyn alan erikoissanastoa. Konekäännösohjelman käyttö voi soveltua

myös suuryrityksen kuukausittain useilla kielillä julkaistavien raporttien kääntämiseen. Ihmisen tulee kuitenkin aina tarkistaa konekäännös. Konekäännin voi olla hyvä apu myös vieraskielisen tekstin sisällön ymmärtämisessä. Konekäännösohjelmiä käyttävät sekä virastot että yritykset. Yksi suurimmista käyttäjistä on EU, jossa dokumentit pitää kääntää jokaisen jäsenmaan omalle kielelle. (Honkonen 2007.)

Konekääntämisen etuja ovat nopeus ja sanastokuri. Tietokone kulkee helposti mukana joten suuristakin tekstimassoista saa käännösluonnoksen ajasta ja paikasta riippumatta. Konekäännösohjelmaa käytettäessä on tekstin sanastollinen yhtenäisyys helppo säilyttää, sillä ohjelma käyttää käännösvastaineita johdonmukaisesti eikä tee lyöntivirheitä. (Konekääntäminen 2009.)

3.4 Puheteknologia

Puheteknologia tarkoittaa puheen käsittelyä tietokoneen avulla. Puheteknologian sovelluksia ovat puheentunnistus, puhesynteesi ja puhujantunnistus. Keinotekoista, eli synteettistä puhetta, on käytetty jo kauan erityisesti näkövammaisille suunnatuissa sovelluksissa. Synteettinen puhe on tullut tunnetuksi myös televisio-ohjelmissa, joissa juontaja on korvattu puhetta tuottavalla automaatilla. (Turunen & Hakulinen 2005, 35.)



Kuva 4. Puheteknologian sovellusten osat ja kiertokulku (Turunen & Hakulinen 2005, 48)

Useimmat puheteknologian sovellukset toteutetaan kahden tai useamman ihmisen vuoropuheluna eli dialogina. Näissä dialogijärjestelmissä esiintyvät yleensä jossain muodossa kuvassa 4 esiintyvät osat liittyen puheentunnistukseen, luonnollisen kielen ymmärtämiseen, dialoginhallintaan ja puhesynteesiin. Kuvan esittämät osat jakaantuvat usein pienimpiin osiin, jotka hoitavat vielä enemmän erikoistuneita tehtäviä. (Tu-

runen & Hakulinen 2005, 47.) Esimerkiksi käyttäjä tuottaa puhetta mikrofonin kautta, jota tunnistetaan puheentunnistimen avulla, jonka jälkeen puheteknologiohjelma analysoi puhetta ja tuottaa sen jälkeen halutun toiminnon.

3.4.1 Puheen olemuksesta

Ihminen oppii yleensä aina puhumaan ennen kuin kirjoittamaan. Kaikissa kulttuureissa kirjoitustaito ei ole vieläkään edes tärkeää. Puhe on usein kertakäyttöisempää ja vähemmän suunniteltua kuin kirjoitettu kieli. Puheen ymmärtäminen on ihmisille luontaista, mutta on osoittautunut vaikeaksi analysoida tietokoneella. Nykyiset kielitekniologian menetelmät ovat kehittyneet valtavasti, vaikka ne eivät aina pystykään tuottamaan puhetta ihmisen tapaan. (Koskenniemi 2002c, 2.)

Puhuminen on osa ihmisten luontaista käyttäytymistä joka erottaa meidät muista eläimistä. Ihmiset ovat ajansaatossa sopeutuneet puhumiseen, ja se on luonut hermostoomme ainutlaatuisia kytköksiä. Siksi puhuminen voi myös ulkoisesti ilmetä monella tavalla. Kun puhe lähtee ihmisen sisältä, kieli on enemmänkin sosiaalinen ilmiö. Puhkekyvyn kehittyminen mahdollisti ihmisten kehityksessä myös kognitiivisen kehityksemme. (Aaltonen 2007.)

Puhe on tehokkain ja käytetyin menetelmä ihmisten välisessä vuorovaikutuksessa. Puhetta pidetään luonnollisena tapana kommunikoida. Puhetekniologian sanotaankin tekevän ihmisen ja tietokoneen vuorovaikutuksesta luonnollisempaa. Puheen luonnollisuus tuo kuitenkin monia vaikeuksia puhetekniologian sovelluksille. Asiat, joita ihmiset pitävät luonnollisena puheelle, kuten päällekkäin puhuminen, aiheuttaa ongelmia puheentunnistukselle. Ihminen ei myöskään kohtele tietokonetta kuten toista ihmistä, vaan esimerkiksi katkaisee helpommin puhelun automaattiselle äänelle. Puheen avulla on kuitenkin mahdollista ilmaista asioita, joita ei muuten voisi ilmaista. Puheella voi olla myös turvallisuutta lisäävä vaikutus. Esimerkiksi autolla ajaessa kädet ja silmät ovat käytettävissä, mutta samalla voidaan kuunnella puhetta. Puhe ei välttämättä aina ole tehokkain vuorovaikutusmenetelmä, mutta käyttäjien kannalta se voi olla suosituin. Ihmiset ratkaisevat asioita puheen avulla, ja myös esimerkiksi vanhemmille ihmisille puheen käyttö voi olla luontevampaa kuin tietokoneen muu käyttö. (Turunen & Hakulinen 2005, 36–37.)

Puhe välittää myös merkityksiä, jotka eivät liity vain sanoihin. Merkityksellisempää voi olla se, miten sanotaan, kuin mitä sanotaan. Puhujan äänenvoimakkuus, paino, puhenopeus ja sävelkorkeuden vaihtelu paljastaa, kuinka puhuja suhtautuu asiaan, josta puhutaan. Näin voidaan päätellä, onko puhuja esimerkiksi onnellinen, vihainen tai jännittynyt. Tutkijat ovatkin alkaneet kehitellä puheteknologian menetelmiä, jolla ihmisten puheen merkityksiä voitaisiin tunnistaa puheesta tietokoneiden avulla. Näin voitaisiin esimerkiksi yrityksissä erottaa uhkaussoitot nykyistä helpommin. (Karppinen & Toivanen 2005.)

3.4.2 Puhesynteesi

Puhesynteesillä tarkoitetaan tietokoneella tuotettavaa ihmispuhetta jäljittelevää ääntä. Puhesynteesiä käytetään, kun halutaan automatisoida laitteita tai palveluita. Synteesiä käytetään esimerkiksi junakuulutuksissa, tiedotuksissa, sähköpostin ääneen lukemisessa ja näkövammaisille suunnatuissa tekstin ääntenlukuohjelmissa. (Koskenniemi 2002c, 5-7.)

Puhesynteesiä on tutkittu ja sovellettu jo satoja vuosia. Ensimmäiset ihmisääntä muistuttavat laitteet kehitettiin jo 1700-luvulla. Nykymenetelmillä tuotettu puhe on ymmärrettävää ja lähes luonnollisen kuuloista. Puhesynteesin ongelmana on saada tekstiin mukaan asiayhteyteen sopivia tunneilmaisuja, ihmisen puheen tauot, lyhenteet, numeroilmaisut ja muu ihmisen normaaliin puheeseen liittyvä sanaton informaatio. (Honkonen 2005.)

Puhesynteesin sovellukset kehittyvät huimaa vauhtia. Jo nyt puhesynteesiä soveltaviin laitteisiin on lisätty suomen lisäksi muita kieliä. Tulevaisuudessa puhesynteesillä toivotaan saavan esimerkiksi matkittua puhujan ääntä, ja kääntämään omalla äänellä puhuttu puhe toiselle kielelle. Yrityksiä puhesynteesin sovelluksissa kiinnostaa halu säästää rahaa sekä automatisoida toimintoja. (Nenonen 2007, 15)

3.4.3 Puheentunnistus

Puheentunnistus tarkoittaa ihmisen puheen tunnistamista tietokoneen avulla. Puheteknologiaa käyttävät automaattiset puheentunnistussovellukset ovat kaikki verrattain uusia. (Koskenniemi 2002c, 9.)

Yksi suosittu puheentunnistuksen sovellusalue on asiakaspalvelun automatisointi. Tällöin puhelimen asiakaspalveluna toimii puheentunnistusta hyödyntävä automaatti. Automaatti jaksaa käsitellä väsymättä asiakkaiden puheluita. (Asiakaspalvelun automatisointi 2009.) Näitä sovelluksia ovat esimerkiksi puhelinvaihteen automatisointi, jossa automaatti yhdistää puhelun nimen, tittelin tai osaston perusteella, tai ajanvarauspalvelu, jossa automaatti voi ohjata soittajan henkilöasiakaspalveluun tai hoitaa itse aikojen siirrot, peruutukset ja tiedustelut. (Esimerkkejä perussovelluksista 2009.)

Automaattisen asiakaspalvelun etuja on se että soittaja on vapaa soittamaan, milloin hänelle parhaiten käy, ja jonotusajat lyhenevät. Soittajan ei myöskään tarvitse jonottaa asiakaspalveluun. Puhelinsovellus kysyy soittajalta perustietoja, kuten henkilötunnuksen, jonka avulla soittajan asia etenee. Jos soittajan asia on monimutkainen, puhelinsovellus selvittää, mitä asia koskee, ja välittää puhelun eteenpäin asiakaspalvelijalle. Puheentunnistus helpottaa sekä soittajan että asiakaspalvelijan tilannetta. Soittaja saa parempaa palvelua ja asiakaspalvelijan työ helpottuu, kun automaatti on kerännyt soittajan perustiedot valmiiksi. Tämä tekee myös asiakaspalvelijan työstä mielekkäämpää. (Sovelluskohteita 2009.)

Puheentunnistuksen avulla voidaan tunnistaa ihmisen puheen perusteella ääniaalloista kuka puhuja on. Tämä teknologia mahdollistaa ihmisen tunnistamisen esimerkiksi hoitaessa pankkiasioita puhelimella, ostosten tekemisen puhelimella, puhelinvastaajan ja tietokoneen käytön etäyhteyden avulla. Puheentunnistuksen sovelluksista on myös tärkeää apua rikosten selvittämisissä. (Furui 2009, 34.) Muita puheentunnistuksen sovelluksia ovat puheella komennettavat laitteet, kuten matkapuhelin, jota voi ohjata puheella, numerotiedustelut sekä sanelusovellukset esimerkiksi lääkäreiden käyttöön (Koskenniemi 2002c, 9-10).

Vaikka puheentunnistuksen avulla on saavutettu merkittäviä sovelluksia, ja kehitys on edistynyt nopeasti, puheentunnistukseen liittyy vielä monia ongelmia. Useimmat ongelmista liittyvät vaihteluihin puheessa sekä lähetyksessä ja äänitteessä. Vaihtelut johtuvat puhujasta itsestään sekä myös taustäänistä. Puhe ei ole koskaan täysin samantyyppistä, vaan muuttuu jatkuvasti. Olisi tärkeää, että tulevaisuudessa puheentunnistinkin ymmärtäisi puhua ja muuttaa puhettaan enemmän tilanteeseen sopivaksi. (Furui 2009, 36.)

3.4.4 Puheteknologia Suomessa

Puheteknologiaa on tutkittu Suomessa jo 1960-luvulla. Kaupallisilla markkinoilla on ollut saatavilla puheteknologiatuotteita 1990-luvulta lähtien, mutta nämä tuotteet ovat suunnattu erityisryhmille, kuten näkövammaisille. Vuosituhannen vaihteessa käynnistyi merkittäviä puheteknologian tutkimuksia, jotka ovat mahdollistaneet puheteknologiatuotteiden leviämisen myös suuren yleisön ja yritysten käyttöön. (Arppe 2008a.)

Suomen kieli tuo omia haasteitaan myös puheteknologian tutkimukseen. Suomen kielen ääntäminen eroaa huomattavasta muista indoeurooppalaisista kielistä kuten englannista ja saksasta. Suomen kieltä on myös yleisesti tutkittu vähemmän kuin esimerkiksi englantia. Haasteista huolimatta suomalaiset puheteknologian tutkimukset ja sovellukset ovat saaneet myös kansainvälistä tunnettavuutta. (Arppe 2008a.)

Suomen kieli on rikas myös murteiltaan. Osaan puheteknologian sovelluksista on kerätty akustista mallia tuhansilta suomalaisilta ääninäytteillä ympäri maan. Näin sovellukset tunnistavat ja ymmärtävät myös murteita sekä myös esimerkiksi ulkosuomalaisten puhumaa heikkoa suomen kieltä. (Tikkanen 2009.) Toisaalta puheteknologiaa ei saa huijata liikaa puhumatta ylikorostetusti murretta, tai muuten yrittämällä huijata puheentunnistinta, jolloin ohjelma ei välttämättä ymmärrä, mitä puhuja tarkoittaa (Reiman 2009).

4 KIELITEKNOLOGIA KÄYTÄNNÖSSÄ

Kieliteknologialla on Suomessa vahvat akateemiset juuret. Suurimmalla osalla kieliteknologiayrityksistä on takanaan tutkijoita suomalaisista yliopistoista tai korkeakouluista. Muuten kieliteknologiayritykset eroavat toistaan melko paljon. Osa yrityksistä on löytänyt itselleen selkeän ja kannattavan liiketoiminta-alueen, kun toiset vielä haavevat paikkaansa. Historiaa tarkasteltaessa näyttää siltä, että ne yritykset, jotka ovat nähneet itsensä ensisijaisesti ohjelmistoyrityksinä ja vasta toissijaisesti kieliteknologiayrityksenä, ovat parhaiten menestyneitä. Suomalaiset kieliteknologiayritykset työllistivät vuonna 2001 vajaat sata ihmistä, ja yritysten liikevaihdot liikkuvat suurimmillaan muutamassa miljoonassa eurossa. Suomalaiset kieliteknologiayritykset ovat kansallisesti menestyneet hyvin, mutta mitään suurta kansainvälistä läpimurtoa emme ole ainakaan vielä päässeet näkemään. (Arppe 2008.)

4.1 Suomalaisia kieliteknologiayrityksiä

Suomessa on monia kymmeniä kieliteknologiayrityksiä, jotka erikoistuvat hieman erilaisiin kieliteknologian sovelluksiin. Tarkastelen seuraavassa tarkemmin kolmea hieman eri kieliteknologian sovelluksia käyttäviä yrityksiä, joiden kanssa olen ollut yhteydessä joko sähköpostitse tai puhelimen välityksellä.

Kaksi Suomen suurimpia kieliteknologiayrityksiä Lingsoft ja Kielikone ovat akateemisen taustan omaavia (Arppe 2008). Lingsoft Oy on vuonna 1986 perustettu monia kieliteknologisia sovelluksia tuottava kieliteknologiayritys. Lingsoftin tuotteet ja palvelut liittyvät puhutun ja kirjoitetun kielen analysoimiseen, käsittelemiseen ja hyödyntämiseen. Sen pääalueita ovat käännökset, puhe-sovellukset, kielentarkistimet, sähköiset sanakirjat, kielityökalut ja sanastot. Lingsoftin asiakaskunta vaihtelee suurista pieniin, mutta sillä on esimerkiksi asiakkaina suuria yrityksiä kuten Microsoft, EU ja Suomen valtio. (Lingsoft on täyden palvelun kielitalo 2009.)

Kielikone Oy on vuonna 1987 perustettu digitaalisten sanakirjojen markkinajohtaja Suomessa, sekä yksi maailman johtavista asiakkaiden tarpeisiin pohjautuvista elektronisten sanakirjojen toimittajista. Kielikone on tunnettu MOT-kielipalveluistaan. MOT on laajassa käytössä yrityksissä, julkishallinnossa, oppilaitoksissa ja yksityisillä ihmisillä. Näiden lisäksi Kielikoneen palveluvalikoimaan kuuluu konekäännös- ja kielentarkistuspalvelut. Kielikoneella on asiakkaita Suomen lisäksi yli 100 maassa. (Kielikone Oy 2009.)

Lisäksi Suomessa on vain esimerkiksi puheteknologiaan erikoistuneita yrityksiä, kuten Suomen Puheentunnistus Oy. Suomen Puheentunnistus Oy on verrattain nuori yritys, se on perustettu vuonna 2004. Suomen Puheentunnistus Oy tarjoamat kieliteknologian sovellukset liittyvät puhesynteesiin sekä puheentunnistukseen. Muita yrityksen palveluita ovat puhelinasiakaspalvelun konsultointi ja kehittäminen, vaihde- ja puhelinjärjestelmäpalvelut sekä puhelinasiakaspalvelun äänityöt. Suomen Puheentunnistus Oy:lla on oma studio, tuottajat ja äänitalli, joka tekee ääniä sekä omiin että muiden sovelluksiin. Suomen Puheentunnistus Oy:n tekemissä asiakasratkaisuissa käytetään kielenä pääasiassa suomen, mutta heiltä löytyy myös kieliä Keski-Euroopan kielille, sekä myös venäjän ja kiinan kielille. Yrityksen suosituimpia kieliteknologian sovelluksia ovat terveydenhuollon ja työttömyyskassa-alan sovellukset. Suomen Puheen-

tunnistus Oy:n suurimpia asiakkaita ovat yritykset ja julkinen sektori ja asiakkaiden koko vaihtelee laidasta laitaan. (Tikkanen 2009.)

4.2 Kieliteknologian hyödyt työelämässä

Kieliteknologian avulla voidaan saavuttaa monia hyötyjä työelämässä. Kieliteknologian sovellukset auttavat yritystä parantamaan kilpailukykyään globaalistuneilla markkinoilla. Automatisoidut sovellukset parantavat ja nopeuttavat julkisen hallinnon ja palveluiden palvelua. Kieliteknologia auttaa levittämään informaation saatavuutta tietojärjestelmien käytön helpottuessa. Yrityksissä voidaan myös säästää työntekijöiden aikaa, kun älykkäät tietokonejärjestelmät toimivat agentteina. Kieliteknologisten sovellusten avulla voidaan myös kohentaa yritysten tallennetun informaation laatua. Informaatiota voidaan myös suodattaa tarvittaessa paremmin. Kun yritykset pystyvät helpommin kieliteknologian sovellusten avulla kommunikoimaan myös muilla kielillä, edistää se tehokkaampaa kansainvälistä yhteystyötä. Myös yhteydenpitomahdollisuudet ulkomaisten liikekumppaneiden kanssa helpottuvat. Yritykset saavat myös helpommin tietoa toisten maiden tiedoista, palveluista, työmahdollisuuksista, säästä ja tieolosuhteista. Kun yhteydenpito ulkomaille paranee, tulee myös keskinäisestä riippuvuudesta helpompi ja palkitsevampi sekä toimiva vuorovaikutussuhde. Kieliteknologian avulla voidaan myös merkittävästi parantaa vammautuneiden mahdollisuutta osallistua jokapäiväiseen työhön ja sosiaaliseen toimintaan. (Arppe 2001a.)

Lingsoftin toimitusjohtajan Juhani Reimanin mukaan kieliteknologian avulla voidaan saada yrityksen viestinnästä johdonmukaista ja saavuttaa kustannussäästöjä. Tiedonhaun saralla voidaan lisätä yritysten löydettävyyttä netissä, ja markkinoida yritystä ja lisätä yrityksen näkyvyyttä. Reimanin mukaan kieliteknologian sovelluksilla on suuri vaikutus yrityksen tehokkuuden ja kilpailuarvon kannalta. Kieliteknologian avulla voidaan myös lisätä automaatiota. Elektroniset sanakirjat auttavat työntekijöitä työssä, ja automaattisesti tekstiä puheesta muodostavat sanelukoneet helpottavat paljon esimerkiksi lääkäreiden työtä. (Reiman 2009.)

Suomen Puheentunnistus Oy:n avainasiakaspäällikön Antti Tikkasen mukaan puhe-
teknologian sovelluksilla voidaan mahdollistaa ympärivuorokautinen puhelinasia-
palvelu. Ne myös parantavat työssä viihtymistä, kun rutiiniasiat hoidetaan puhelin it-

sepalvelussa ja vasta monimutkaisemmat asiat kuormittavat asiakaspalvelua. (Tikkanen 2009.)

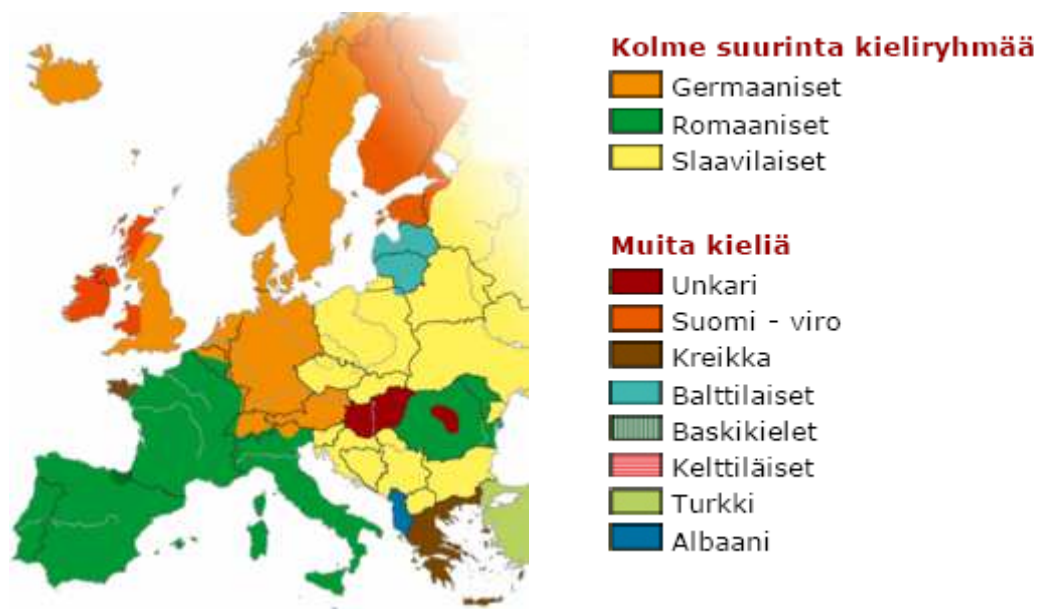
Yksi esimerkki yrityksestä, jossa kieliteknologian sovelluksia on käytetty monipuolisesti, on Rautaruukki Oyj. Rautaruukissa on panostettu kieliteknologiaan jo 1980-luvulta lähtien. Nykyään Rautaruukissa on käytössä Kielikoneen kehittämät TranSmart-konekäännösohjelma, MOT-sanakirjasto sekä WinVirku-kielenhuolto-ohjelma. TranSmart-konekäännösohjelma on Kielikoneen kehittämä suomalainen huipputeknologiatuote, joka on herättänyt myös kansainvälistä huomiota. TranSmartin etu on nopeus ja suuristakin tekstimassoista saadaan raakakäännös vaivattomasti. WinVirku on suomen kielenhuolto-ohjelma, jolla voidaan oikolukea tekstiä, tarkistaa tavutus sekä muita kieliopillisia ja tyylillisiä seikkoja. Lisäksi se sisältää suomen kielen synonyymisanakirjan, ja sille voi opettaa uusia sanoja. (Laitinen 2008.)

TranSmart-käännösohjelmaa hyödyntää Rautaruukissa noin 500 käyttäjää, jotka ovat pääasiassa teknisiä ja kaupallisia toimihenkilöitä. He käyttävät ohjelmaa apuna esimerkiksi kirjeenvaihdossa ulkomaisten toimittajien kanssa. Käännösohjelman avulla jokainen pystyy tuottamaan suomenkielisestä tekstistä englanninkielisen raakakäännöksen, ja vastaamaan itse siitä, mitä kirjoittaa. WinVirku-kielenhuolto-ohjelman avulla tekstistä saadaan selkeää ja ymmärrettävää suomea. Konsernin verkossa toimiva MOT-sanakirjaohjelma on tärkeä esimerkiksi markkinointihenkilöstölle. Sanasto sisältää normaalin sanaston lisäksi noin 14 000 Rautaruukin omaa sanaa. Järjestelmän ansiosta Rautaruukin sanat ja termit ovat yhdenmukaiset koko konsernissa sekä suomeksi että englanniksi. TranSmart-käännösohjelma helpottaa työntekijöiden yhteydenpitoa, ja työaika säästyy, kun käännökset saadaan nopeasti. Tämän lisäksi työntekijöiden ei tarvitse käyttää enää suomalaista agenttia, kun he voivat itse kommunikoida suoraan päämiehien kanssa. Käännösjärjestelmän avulla voidaan välttää myös väärinkäsityksiä, koska kaikki puhuvat asioista samoilla termeillä. (Laitinen 2008.)

4.3 Kieliteknologia EU:ssa

Euroopan unionin valtioissa puhutaan monia kieliä. Euroopan unioniin kuuluu nykyään jo 27 jäsenvaltiota ja 23 virallista kieltä. Yhteensä EU:n alueella puhutaan noin 60 kieltä. Kuvassa 5 näkyy Euroopan maita kieliryhmittäin. Kuvasta voidaan havaita, että EU:n alueella on useita eri kieliryhmiä. EU-maiden kuuluminen eri kieliryhmiin tuo

mukanaan haasteita kielten erilaisuuksien takia. Osaa EU:n kielistä puhutaan erityisalueilla tai erityisryhmissä. Myös maahanmuuttajat tuovat mukanaan lisää uusia kieliä. Euroopassa elävien maailmasta on tullut kansainvälisempää ja monikielistä. (Monikielisyys: Euroopan voimavara ja samalla myös yhteinen sitoumus 2008, 3–5.)



Kuva 5. Euroopan maat ja kieliryhmät (EU:n kieliä – Kielikartta 2009)

Kielet määrittävät ihmisten identiteettiä, mutta voivat myös avata yhteyksiä muihin maihin ja kulttuureihin. EU:n tavoitteena on onnistunut monikielisyyspolitiikka, joka takaa kaikkien kielten rauhallisen rinnakkaiselon. Onnistunut monikielisyyspolitiikka voi parantaa EU:n kansalaisten työllistymistä, helpottaa palveluiden saamista, edistää kulttuurienvälistä toimintaa sekä lisätä yhteenkuuluvuuden tunnetta. (Monikielisyys: Euroopan voimavara ja samalla myös yhteinen sitoumus 2008, 3–5.)

Kielten osaaminen voi olla kilpailuetu EU:ssa oleville yrityksille. Monikieliset yritykset ovat näyttäneet, että panostamalla kielelliseen monimuotoisuuteen, kielitaitoon sekä kulttuurien tuntemukseen voidaan saada monia etuja ja saavuttaa menestystä. Muutamia eurooppalaisia kieliä puhutaan myös muualla maailmassa, ja tämä voi olla yrityksille arvokas vientiväline. Kehittyvien talousmaiden kuten Brasilian, Venäjän, Intian ja Kiinan kielen ymmärtäminen on tullut yhä tärkeämmäksi EU:ssa oleville yrityksillä, ja näiden riittävä kielitaito on tärkeää markkinoilla kilpailtaessa. Kieliteknologi-

an sovellukset voivat olla apuna vieraan kielen ymmärtämisessä. (Monikielisyys: Euroopan voimavara ja samalla myös yhteinen sitoumus 2008, 8.)

Euroopan unionin komission tekemässä tutkimuksessa kävi ilmi, että 11 % vientiä harjoittavista EU:n pk-yrityksistä voi menettää liiketoimintamahdollisuuksia kielellisten esteiden vuoksi. Vaikka englanti on tärkeä ja johtava kieli liike-elämän kannalta, tarjoavat myös muut kielet EU:n yrityksille kilpailuetua ja mahdollisuuksia valloittaa uusia markkinoita. Hyvällä kielitaidolla on myynnin ja markkinoinnin lisäksi vaikutusta yrityksen kaikkeen toimintaan. Monilta, varsinkin pieniltä yrityksiltä, puuttuu tieto ja voimavarat, jolla kielet voidaan sisällyttää yritysten liiketoimintasuunnitelmiin. Näissä astuvat esiin kieliteknologian sovellukset, jotka tarjoavat yrityksille menetelmiä kielistrategioiden parantamiseksi. (Monikielisyys: Euroopan voimavara ja samalla myös yhteinen sitoumus 2008, 8.)

Kielen ja muiden viestintävälineiden käyttäminen on tärkeää monikielisessä Euroopassa. Jokaisella kansalaisella on oltava mahdollisuus myös itse viestiä laajentuneessa EU:ssa. Tämä ei koske vain monia kieliä osaavia, vaan myös yksikielisiä tai muutenheitä, joiden kielitaito on vähäinen. Tiedotusvälineet, uudet kieliteknologian sovellukset ja automaattiset käännöspalvelut voivat tuoda EU:n laajan kieli- ja kulttuurivalikoiman lähemmäs tavallisia kansalaisia ja auttaa murtamaan kielimuureja. Myös internetissä on yhä enemmän tietoa monilla kielillä, ja on tärkeää, että ihmiset voivat käyttää näitä tietoja ja palveluita kielellisistä esteistä riippumatta. EU onkin kehittänyt kieliteknologisia tekniikoita, joilla voidaan helpottaa näitä kielellisiä esteitä. (Monikielisyys: Euroopan voimavara ja samalla myös yhteinen sitoumus 2008, 12.)

Kieliteknologian avulla kääntäminen on myös tärkeä osa EU:n monikielisyyspolitiikkaa. Automaattinen kääntäminen voi helpottaa tiedon kulkua kansallisten viranomaisien välillä, ja parantaa myös hallinnon yhteistyötä eri maiden välillä. Parhailaan EU kehittää esimerkiksi sisämarkkinoiden tietojenvaihtojärjestelmää, joka antaa jäsenvaltioille mahdollisuuden vaihtaa tietoja kaikilla EU:n virallisilla kielillä. (Monikielisyys: Euroopan voimavara ja samalla myös yhteinen sitoumus 2008, 12–13.)

EU:ssa on käännetty suomeksi jo satoja tuhansia EU:n toimintaan liittyviä tekstejä. Suomessa tätä kutsutaan yleisesti EU-kääntämiseksi. EU eroaa muista kansainvälisistä järjestöistä siinä, että EU:ssa kaikkien jäsenvaltioiden viralliset kielet ovat myös unio-

nin virallisia kieliä ja tasavertaisia keskenään. EU:n tasavertaisuuden kannalta on tärkeää, että sen kansalaiset voivat lukea tekstejä ja lähestyä toimieliimiä omalla kielellään. EU:n monikielisyys on merkittävä poliittinen kysymys. (Vuorinen 2001, 109–110.)

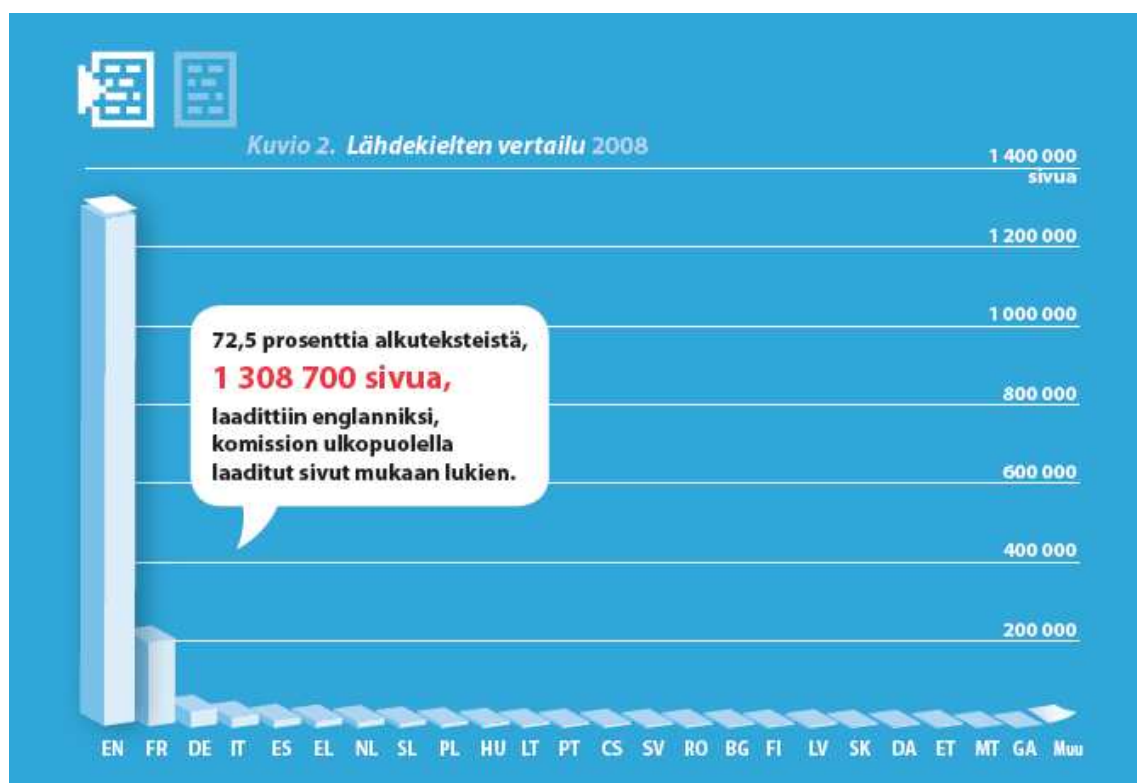
EU kääntämisessä erikoista tekee sen, että tekstien on puhuteltava kaikkien EU-maiden hallintoelinten kansalaisia samaan aikaan. Tämä on olennaista lainsäädännön kannalta, sillä lainsäädännöltä putoaisi pohja, jos lakia sovellettaisiin erilailla eri maissa. Käännöksiä ei siis voi muokata maan kulttuurin tai tyylin mukaan. EU onkin joutunut kehittämään omia käsitteitä ja puhetapoja. Tämän takia kansalaisten on ymmärrettävä EU:n periaatteita ja sanastoa, jotta ymmärtää mikä teksteissä puhutaan. (Vuorinen 2001, 113–114.)

Euroopan unioni on yksi maailman johtavista kieliteknologian käyttäjistä ja kehittäjistä. EU:ssa työskentelee kielen ja tietotekniikan asiantuntijoita kaikkialta Euroopasta. Kieliteknologian kehittäminen on ollut jo jonkin aikaa EU:n toimintapoliittinen painopiste. Kieliteknologia on keino myös säilyttää Euroopan rikasta kulttuuriperintöä ja yksi tulevaisuuden taloudellisen kasvun lähde. (Kieliteknologia 2006.)

Lingsoftin toimitusjohtajan Juhani Reimanin mukaan Suomen EU-jäsenyydellä on ollut paljon vaikutusta myös Suomen kieliteknologian kehitykseen. EU:ssa käännetään valtavat massat aineistoja, ja myös kieliteknologiafirmat kilpailevat keskenään, kuka myy EU:lle omia palveluitaan. (Reiman 2009.)

4.4 Euroopan komission käännöspalvelu

Euroopan komission käännöspalvelu on maailman suurin käännösorganisaatio. Suuri organisaatio mahdollistaa ja edellyttää panostuksen myös kääntämisen apuvälineisiin. Kieliteknologian sovellusten kehitys onkin ollut nopeaa EU:ssa, ja kääntäjien käytössä on monia elektronisia termi- ja tekstitietokantoja. Suurissa, samoja toistuvia osia sisältävissä teksteissä voidaan käyttää apuna esimerkiksi käännösmuistiohjelmia. EU:lla on käytössä myös oma Systran käännösjärjestelmä. (Vuorinen 2001, 115.)



Kuva 6. Käännöstoimen alkuteksteistä suurin osa on laadittu englanniksi (Kääntäminen monikielisessä yhteisössä 2009, 7)

EU:n käännöstoimen pääosasto käänsi vuonna 2008 yhteensä 1 805689 sivua. Kuvasta 6 voi nähdä, että suurin osa alkuteksteistä oli laadittu englanniksi. Yleisin tekstien alkuperäkieli on nykyisin ranskan sijasta englanti. Asiakirjojen käännöstyössä sekä muissa kieleen liittyvissä tehtävissä työskentelee EU:ssa 1750 kääntäjää. Näiden lisäksi noin 600 työntekijää toimii erilaisissa hallintoon, sihteeritoimeen, viestintään, tietotekniikkaan ja koulutukseen liittyvissä tehtävissä. (Kääntäminen monikielisessä yhteisössä 2009, 5–6.)

Komission kääntäjillä on käytössä suuri valikoima sähköisiä käännösohjelmia. Joitakin näistä voivat käyttää myös muut komission virkamiehet ja muiden toimielinten kääntäjät. Lisäksi EU:ssa on kehitetty joukko muita välineitä työn organisointiin, jotta pystytään hallitsemaan yli miljoonan sivun vuosituotannon säilyttämistä ja seurantaan. Yksi komission käyttämistä kieliteknologian sovelluksista on Translator's Workbench eli TWB, joka on käännösmuistiohjelma, johon voidaan tallentaa ja josta voidaan hakea aineistoa kaikilla virallisilla kielillä. Komission käyttämään versioon on sisällytetty komission tarvitsemia erikoisvaatimuksia. Ohjelmaa on käytetty vuodesta 1997 lähtien. Käännösmuistin avulla etsitään aiemmin käännettyissä teksteissä esiintyviä sa-

manlaisia tekstejä, ja löydetyt käännökset sisällytetään uuteen käännökseen. TWB-ohjelmaa pidetään erittäin hyödyllisenä työkaluna, sillä suuri osa komissiossa käännettyistä teksteistä perustuu aiempiin teksteihin tai lakiin. Aikaisemmin käännettyjen tekstien käyttäminen säästää kääntäjien aikaa ja pitää termistön yhdenmukaisena. (Kääntäminen monikielisessä yhteisössä 2009, 13.)

Toinen komission käyttämistä kieliteknologian sovelluksista on Euramis, joka on komissiossa kehitetty järjestelmä. Ohjelma koostuu joukosta verkkosovelluksia jotka ovat yhdistetty sähköpostiin, ja ohjelman avulla on mahdollista käyttää useita kielen käsittelyyn käytettäviä palveluja. Euramis liittyy myös yhteen kaikki käännöstoimen pääosaston kieliteknologiset käännösjärjestelmät. Yksi sovelluksen tärkeimmistä palveluista on keskuskäännösmuisti. Palvelu toimii niin, että kun toimeksiantona saatu käännös hyväksytään, lähetetään alkuperäinen asiakirja automaattisesti Euramikseen, ja jos teksti on jo aiemmin käännetty, poimii ohjelma käännöksen keskusmuistista. Jos teksti tai osa siitä ei löydy keskusmuistista, palvelu voidaan yhdistää muihin sovelluksiin, kuten konekääntämiseen. Kun käännös on valmis, viedään se Euramiksen keskuskäännösmuistiin, jolloin se on taas käytettävissä mahdollista seuraavaa tekstiä käännettäessä. (Kääntäminen monikielisessä yhteisössä 2009, 13.)

Komissiossa on käytetty konekääntämistä jo vuodesta 1976 lähtien. Komission konekäännösohjelma toimii 18 kieliparilla, ja se pystyy tuottamaan 2000 sivua raakakäännöstä tunnissa. Konekääntäminen on kaikkien komission virkamiesten käytettävissä sähköpostin välityksellä. Myös muut EU:n toimielimet käyttävät komission konekäännöspalvelua. Konekääntämistä on viime vuosina käytetty paljon, esimerkiksi vuonna 2008 käännettäväksi toimitettiin 1 963991 sivua. Konekäännösohjelmaa räätälöidään käännöstoimen tarpeisiin lisäämällä siihen sopivaa erikoistermistöä. Erityisen pitkälle räätälöinti on viety kielipareissa englanti-espanja ja ranska-espanja, jotka ovat kääntäjien keskuudessa eniten käytetyt kieliparit. Vasta kolmanneksi suosituin kielipari on englanti-ranska. Konekäännöksen avulla kääntäjät saavat tekstistä raakaversioon, jonka he voivat muokata valmiiksi. Komission konekäännöksen käyttäjistä suurimman ryhmän muodostavat muiden osastojen virkamiehet, jotka käyttävät eniten englannin, ranskan ja saksan kielen välisiä käännöksiä. He hyödyntävät konekääntämistä tekstin selailun apuvälineenä, mutta myös silloin, jos käännöksiä tarvitaan nopeasti käyttöön. Virkamiehille on tarjolla myös editointipalveluita, joissa konekäännöksen avulla tehdystä raakakäännöksistä tehdään lopullinen muokaus. Jos tekstiä tehdään sisäiseen

käyttöön, jolloin tärkeintä on nopeus ja täsmällisyys, on konekäännös oiva väline. Jos tekstit on tarkoitettu ulkoiseen jakeluun, tulee myös tyyliin ja termistöön kiinnittää huomiota ja raakakäännös aina tarkistaa ja muokata. (Kääntäminen monikielisessä yhteisössä 2009, 14.)

Komission käännöstoimella on oma kirjasto, joka tarjoaa laajan valikoiman sähköisiä sanakirjoja ja muita lähteitä, jotka ovat käytettävissä käännöstoimen intranetin kautta. Lisäksi käännöstoimen kieliosasto vastaa termityöstä. Tähän kuuluu esimerkiksi vastaaminen kääntäjien ja muiden virkamiesten esittämiin termikysymyksiin ja termien syöttäminen käännöstoimen tietokantaan, IATE:hen. IATE on maailman suurin monikielinen termitietokanta, ja siihen kootaan termejä kaikista EU:n toimielimistä. IATE on ollut käytössä Euroopan komissiossa vuoden 2005 alusta lähtien. Se sisältää yli kahdeksan miljoonaa termiä ja 560 000 lyhennettä. IATE käsittää kaikki EU:n viralliset kielet sekä lisäksi latinan. IATE:n termistö kattaa kaikki EU:n toimielinten toiminta-alat. (Kääntäminen monikielisessä yhteisössä 2009, 14–15.)

5 OPETUSMATERIAALIA KYMENLAAKSON AMMATTIKORKEAKOULULLE

5.1 Johdon assistenttityön ja kielten koulutusohjelma

Johdon assistenttityön ja kielten koulutusohjelma järjestetään Kymenlaakson ammattikorkeakoulussa toista kertaa. Tutkinnon laajuus on 210 opintopistettä. Koulutusohjelmasta opiskelijat valmistuvat kielten ja kulttuurien sekä yritys- ja yhteisöviestinnän asiantuntijoiksi tutkintonimikkeellä tradenomi. Koulutusohjelma valmentaa opiskelijoita kansainvälisiin työtehtäviin yritys-elämän, kaupan, teollisuuden ja julkishallinnon aloilla. Koulutusohjelma yhdistää eri alojen, liiketoiminnan, kielten ja kulttuurin sekä viestinnän osaamista. Koulutusohjelman perustana on edistää vahvaa kielitaitoa, eri kulttuurien ja kulttuurienvälistä viestintää, talouselämän ja yritysviestinnän tuntemus sekä ammattimainen viestintä- ja vuorovaikutusosaaminen. Opiskelijat saavat valmiudet toimia valmistuttuaan esimerkiksi johdon assistentteina, tiedottajina, vientiassistenttina, yrittäjänä tai projektiassistenttina. Koulutusohjelma tarjoaa hyvät mahdollisuudet myös työharjoitteluun tai opiskelijavaihtoon ulkomailla. (Johdon assistenttityön ja kielten koulutusohjelma 2009.)

Koulutusohjelmassa opiskelijat voivat suuntautua joko kielten ja kulttuurin opintoihin tai yritys- ja yhteisöviestinnän opintoihin. Yritys- ja yhteisöviestinnän suuntautumisvaihtoehto antaa valmiudet toimia viestintä- ja vuorovaikutustaitoja vaativissa työtehtävissä yrityksissä ja yhteisöissä. Kielten ja kulttuurin suuntautumisvaihtoehto antaa valmiudet toimia kansainvälisessä yrityselämässä. (Johdon assistenttityön ja kielten koulutusohjelma 2009.)

5.2 Kieliteknologian kurssi

Kielten ja kulttuurien suuntautumisvaihtoehdossa yhtenä pakollisista kursseista on Kieliteknologia. Kurssi kuuluu suuntautumisvaihtoehdon suuntaaviin ammattiopintoihin ja suoritetaan opintojen neljäntenä vuotena. Kurssi on laajuudeltaan 3 opintopistettä. Kurssin sisältöön kuuluu tietotekniset työkalut kääntäminen, kirjoittaminen ja puhe. Kurssin tavoitteena on, että opiskelija osaa hyödyntää tietotekniikkaa kirjoittamisen ja puheen tuottamisessa. (Johdon assistenttityön ja kielten koulutusohjelma 2009.)

Kieliteknologia on ennen ollut voimakkaasti akateemiseen koulutukseen suuntautunut aihealue. Siksi onkin mielenkiintoista, kun nyt myös ammattikorkeakoulun opiskelijoilla on mahdollisuus opiskella ennen yliopisto-opintoihin suuntautunutta ainetta. Muutos kertoo myös siitä, että kansainvälistyvässä maailmassa myös kaupallisen alan työntekijöiden on osattava kommunikoida tehokkaasti monilla kielillä ja toki myös tekniikan kehittymisestä.

5.3 PowerPoint-esitys ja sen ulkoasu

PowerPoint-ohjelmalla voidaan laatia esityskalvoja sekä muita multimediaesityksiä käytettäväksi erilaisissa tilaisuuksissa. PowerPoint-esityksen tulee olla selkeä, jotta se välittää kuulijalle halutun sanoman. Esityskalvoissa kannattaa välttää suurta määrää yksityiskohtia eikä kalvojakaan kannata olla liikaa. Paras esitys syntyy, kun käytetään riittävän suurta kirjasintyyliä, erikoistehosteita käytetään harkiten ja muistetaan säilyttää selkeä tyyli työssä alusta loppuun asti. (Ensio, Kentala, Laiho & Väisänen 2000, 65.)

Valitsin opetusmateriaalin esitystavaksi PowerPoint-esityksen. PowerPoint on nykyaikainen ja helppokäyttöinen tapa esittää asioita havainnollisesti. PowerPoint-tiedosto

on sähköisessä muodossa oleva materiaali. Sähköisessä muodossa oleva materiaalia on helppo käyttää paikasta riippumatta, ja esitys on tarvittaessa helppo lähettää esimerkiksi opiskelijoille. Sähköisen materiaalin voi myös aina tulostaa paperille, kun paperitulostetta on vaikeampi saada sähköiseen muotoon.

Halusin saada omasta PowerPoint-esityksestäni persoonallisen ja kiinnostavan. Esityksen tavoitteena on herättää mielenkiintoa kieliteknologiaa kohtaan. Siksi en käyttänyt ohjelman omia suunnittelumalleja, vaan suunnittelin diojen pohjan itse. Toteutin pohjan taustan Adobe Photoshop-kuvankäsittelyohjelmalla, jonka käyttöä olen opiskellut koulussa muutaman kurssin verran. Yhtenäisen ja selkeän yleisilmeen varmistamiseksi käytin samaa sivurakennetta kaikilla sivuilla.

Lisäsin dioihin kuvia lisäämään mielenkiintoa, vaihtelevuutta ja konkretisoimaan tekstejä. Käytin kuvia dioissa harkiten, ja halusin, että jokaisella kuvalla on jokin ajatus tai tarkoitus. Pidin kuvien koot pieninä, jotta esityksen käytettävyys pysyisi hyvänä. Kuvat dioihin hankin pääasiassa internetin ilmaisista kuvapalveluista. Varmistin aina, että kuvia voi käyttää omiin tarkoituksiin. Käytin myös muutamia taulukoita ja ruutukopioita havainnollistamaan teoriaa. Rajasin muutamia kuvia, joiden tausta ei ollut valkoinen, tummanvihreällä värillä, joka toi selkeyttä ja kontrastia dioihin.

Valitsin diojen teksteissä käytettävän vain yhtä fonttia, Tekton Pro Condia. Fontti on mielenkiintoinen ja sopii taustan rakenteeseen. Käyttämällä samaa fonttia kaikissa teksteissä halusin pitää yllä selkeyttä. Vaihtelevuutta toin otsikoiden ja kappaleiden fonttikoon muuttamisella. Fontin väriksi valitsin normaalin mustan, joka erottuu hyvin valkoiselta taustalta ja tuo asiallisuutta dioihin.

Useissa kulttuureissa vihreä kuvaa elämää ja kasvua, ja se liitetään vehreään luontoon. Vihreää väriä pidetään usein neutraalina ja passiivisena, eli taustavärinä. Vihreä on rauhoittava väri, joka saa ihmiset tuntemaan olonsa turvalliseksi. Vihreä väri koetaan usein myös positiivisena asiana, toisin kuin punainen. Tämän asetelman huomaa esimerkiksi liikennevaloissa. Vihreä on myös symboli uudelle. (Hintsanen 2009.)

Valitsin PowerPoint-työni taustaväriksi vihreän sen positiivisen ja uutta tarkoittavan vaikutuksen takia. Osasyynä värivalintaani oli myös Kymenlaakson ammattikorkeakoulun kotisivuilla ja logossa käyttämä vihreä värimaailma, jonka kanssa halusin

työni olevan samassa linjassa. Halusin, ettei vihreä väri olisi kuitenkaan liian huomiota herättävässä tai keskeisessä asemassa. Siksi sijoitin vihreän värin reunuksiin, ja varsinaisen tekstipohjan halusin pitää valkoisena selkeyden takia. Halusin esitykseen persoonallisen ilmeen, joten tein vihreän ja valkoisen rajasta rosoisen, mutta silti selkeän.

5.4 Sisältö ja runko

PowerPoint-esityksen sisältö seuraa samaa runkoa kuin opinnäytetyöni. Tarkoituksena on, että asiat käsitellään loogisessa järjestyksessä. Ensin esityksessä kerrotaan kieliteknologiasta yleisesti, sitten historiasta ja tämän jälkeen kieliteknologiasta Suomessa ja maailmalla. Tämän jälkeen käyn läpi kieliteknologian sovelluksia. Sovellusten esittelyjen yhteydessä esitän asioita itse ottamien ruutukopiokuvien avulla selvittääkseni diojen sanomaa. Sovellusten esittelyjen jälkeen kerron kieliteknologiasta käytännössä, ja selvitän kieliteknologiaa ja sen hyötyjä työelämän kannalta, sekä EU:n kieliteknologiaa. Loppuun laitoin muutaman mielenkiintoisen linkin, joihin asiasta kiinnostuneet voivat tutustua ja saada lisää tietoa kieliteknologiasta. Esityksen selkeyden ja mielenkiinnon ylläpitämiseksi pidin tekstin määrän dioissa kohtuullisena.

6 JOHTOPÄÄTÖKSET

Tämän opinnäytetyön tekeminen on ollut minulle suuri matka kieliteknologian maailmaan. Itse en ennen työn aloittamista tuntenut kieliteknologia käsitettä juuri ollenkaan, vaikka olinkin tietämättäni törmännyt kieliteknologian sovelluksiin. Tämän työn tekeminen on ollut silmiä avaava ja opettava kokemus.

Työn aloittamisessa ongelmia tuotti se, että kieliteknologia on alana suhteellisen uusi ja tietoa sekä kirjallisuutta ei ole saatavilla valtavia määriä. Koska työn tarkoituksena oli tuottaa tietoa johdon assistenttityön ja kielten koulutusohjelman opiskelijoille, jotka eivät ole aikaisemmin tutustuneet aiheeseen, en halunnut tiedosta liian teoreettista, vaan perustuvan enemmän ymmärtämiseen, käytäntöön ja sovelluksiin. Kieliteknologiaa opetetaan Suomessa yliopistoissa, ja osa heidän teorianaan käyttämistä teoksista oli mielestäni liian kielenopiskelijoille ja kääntäjille suunnattua materiaalia, eikä soveltunut käytettäväksi liiketalouden opiskelijoille suunnatussa opinnäytetyössäni. Poinnin näistä lähteistä tietoa aina ajatellen lopullista kohderyhmää. Käytin työni teoriaosuudessa pääasiassa kieliteknologian tutkijoiden kirjoittamia aineistoja, ja pidin

tärkeänä, että tieto on ajantasaista sekä virheetöntä. Tämä lisää työni tiedon luotettavuutta. Tietoa etsiessäni kiinnitin huomiota myös siihen, että uusia puheteknologiayrityksiä on ilmestynyt paljon markkinoille viimeaikoina. Puheteknologia tuntuu olevan kasvava ala, kun taas esimerkiksi kirjoittajan apuvälineet ovat arkipäivää ja välttämättömiä välineitä päivittäisessä työskentelyssä.

Mielestäni onnistuin tavoitteessani tehdä tiivis tietopaketti kieliteknologiasta, jonka kohderyhmänä ovat kielen ja kulttuurin suuntautumisvaihtoehdon valitsevat opiskelijat. Työni esittää kieliteknologian perusteita ja PowerPoint-esitys selittää asioita selkeästi ja käytännöllisesti kieliteknologiaa asiaan aiemmin perehtymättömälle. Työni teoriaosaa voidaan pitää PowerPoint-esitystä syventävänä tietona. Työtä voidaan käyttää myös muille kieliteknologiasta kiinnostuneille, sekä esimerkiksi kielten ja kulttuurin suuntautumisvaihtoehdon markkinointivälineenä.

Suomen kuuluminen EU:hun luo vaatimuksia työelämässä. EU:n normeja ja lakeja tulee soveltaa kaikissa EU-maissa. Olen myös itse törmännyt tähän työelämässä, kun lähetämme työpaikassamme EU:lle esimerkiksi kuukausittain tietoa tuontituotteistamme. Halusin työssäni käsitellä EU:ta myös tämän takia. Suurin syy EU:n käsittelemiseen työssäni oli silti, että EU on suuri kieliteknologian käyttäjä maailmassa. EU on myös valmistuneen tradenomin mahdollinen työpaikka.

Valitsin teoriaan myös esimerkin kuinka Rautaruukki Oyj käyttää kieliteknologiaa toiminnassaan. Halusin tämän avulla auttaa opiskelijoita ymmärtämään käytännössä kuinka tradenomi voi törmätä kieliteknologiaan työelämässä. Halusin myös selittää kuinka paljon ihmiset työskentelet kieliteknologian kanssa, vaikka eivät välttämättä sitä aina itse tiedostaisikaan. Kymenlaakson ammattikorkeakoulun opiskelijatkin käyttävät kieliteknologian sovelluksia, kuten koulun tarjoamaa MOT-sanakirjaa, eivätkä välttämättä heti ajattele, että kyseessä on kieliteknologian sovellus.

Työssäni tuli aiemmin selkeästi ilmi kieliteknologian hyödyt työelämässä. Kieliteknologialla on kasvava merkitys kansainvälistyvässä työelämässä. Kieliteknologiaan työn tekemisen myötä paremmin tutustuneena voin kuvitella hyötyjä, joita se kielten ja kulttuurin opiskelijoille voi antaa. Enää vain harvoissa työpaikoissa ei törmää vieraisiin kieliin ja kulttuureihin. Suomeen tuodaan paljon tavaraa ulkomailta, ja ulkomaisien toimittajien kanssa kommunikointi on elintärkeää. Valmistuneet kielten ja kulttuu-

rin opiskelijat voivat kieliteknologian avulla helpottaa sekä suomella että vieraalla kielellä kommunikoimista, ja viimeistellyllä kielellä antaa hyvää kuvaa itsestään ja yrityksestään.

Kieliteknologian kurssi soveltuu mielestäni erinomaisesti juuri kielten ja kulttuurin suuntautumisvaihtoehtoon. Kieliteknologia kurssin laajuus on 3 opintopistettä, ja mielestäni tämä on riittävä laajuus kieliteknologian opiskelemiselle liiketalouden opinnoissa. Kielten ja kulttuurin opiskelijoiden ei tarvitse mielestäni perehtyä niinkään kieliteknologian teoriaan, vaan keskittyä enemmän kieliteknologian soveltamiseen. Mielestäni onkin hyvä, että myös kurssilla keskitytään kieliteknologian työkaluihin. Ehdottaisin kurssilla opiskeltavan kieliteknologiaa työelämälähtöisesti. Näin uskon opiskelijoiden saavan parhaan hyödyn kurssista. Opiskelijat voivat mielestäni parantaa kieliteknologian tietämyksellään omia mahdollisuuksiaan saada valmistuttuaan työpaikka.

LÄHTEET

- Aaltonen, O. 2007. Miten yhdistää puheen tuottaminen ja puheen havaitseminen. Suomen kieliteknologian dokumentaatiokeskus. Helsingin yliopiston yleisen kielitieteen laitos. 2.2.2007. Saatavissa: <https://kitwiki.csc.fi/twiki/pub/KitWiki/FoneV102/FONEV102-2007k-08.pdf> [viitattu 18.10.2009].
- Alho, M. 2005. Tietokoneavusteinen kääntäminen – teoriaa, käytäntöä ja opetusta. Teoksessa: Yli-Jokipii, H. (toim.). Kielen matkassa multimediaan. Näkökulmia kääntämisen tutkimiseen ja opiskelemiseen. Helsinki: Käännöstieteen laitos.
- Arppe, A. 2001. Kaupallinen kieliteknologia Suomessa, Pohjoismaissa ja maailmalla. Helsingin yliopisto. 1.10.2001. Saatavissa: <http://www.ling.helsinki.fi/kit/2001s/ct1310com/011001/luento-011001-2.htm> [viitattu 13.9.2009].
- Arppe, A. 2001a. Mitä on kaupallinen kieliteknologia?. Helsingin yliopisto. 1.10.2001. Saatavissa: <http://www.ling.helsinki.fi/kit/2001s/ct1310com/011001/luento-011001-1.htm> [viitattu 29.9.2009].
- Arppe, A. 2008. Ei yhtä ainoa polkua - Suomalaisia kokemuksia matkalla kieliteknologisesta tutkimuksesta liiketoimintaan. Suomen kieliteknologian dokumentaatiokeskus. Helsingin yliopiston yleisen kielitieteen laitos. Saatavissa: <https://kitwiki.csc.fi/twiki/bin/view/FiLT/ArppeFi> [viitattu 28.9.2009].
- Arppe, A. 2008a. Kärki on kapea mutta kärjen tuntumassa ollaan – puheteknologiaa suomalaisittain. Suomen kieliteknologian dokumentaatiokeskus. Helsingin yleisen kielitieteen laitos. 14.4.2008. Saatavissa: <https://kitwiki.csc.fi/twiki/bin/view/FiLT/PuheteknologiaFi> [viitattu 1.10.2009].
- Asiakaspalvelun automatisointi. Lingsoft Oy. Saatavissa: http://www.lingsoft.fi/?doc_id=427 [viitattu 4.10.2009].

Carlson, L. 2005. Käännösteknologia. Teoksessa: Yli-Jokipii, H. (toim.). Kielen matkassa multimediaan. Näkökulmia kääntämisen tutkimiseen ja opiskelemiseen. Helsinki: Käännöstieteen laitos.

Ensio, S., Kentala, J-P., Laiho, R., & Väisänen, K. 2000. Microsoft Office 2000-opas. Jyväskylä: Gummerus Kirjapaino Oy.
Esimerkkejä perussovelluksista. Suomen puheentunnistus Oy. Saatavissa:
<http://www.puheentunnistus.fi/fi/tuotteet-ja-palvelut/esimerkkeja-perussovelluksista>
[viitattu 4.10.2009].

EU:n kieliä – Kielikartta. Nitobete-keskus. Saatavissa:
<http://fi.nitobe.info/ld/lingvoj/lingvomapo.php> [viitattu 18.10.2009].

Furui, S. Speaker Recognition. Teoksessa: Cole, R., Mariani, J., Uszkoreit, H., Zerenen, A. & Zue, V. (ed.). Survey of the State of the Art in Human Language Technology. Language Technology World. Saatavissa: http://www.lt-world.org/HLT_Survey/master.pdf [viitattu 1.10.2009].

Hintsanen, P. Vihreä. Cloria. Saatavissa: <http://www.coloria.net/varit/vihrea.htm> [viitattu 21.10.2009].

Honkonen, J. 2005. Missä viipyvät puhuvat tietokoneet? Tiede-lehti. 2.11.2005. Saatavissa: <http://www.tiede.fi/arkisto/artikkeli.php?id=586&vl=2005> [viitattu 4.10.2009].

Honkonen, J. 2007. Tietokone kääntäjänä: Kieli sujuu, vaikka ymmärrys pätkii. Tiede-lehti. 7.11.2007. Saatavissa:
<http://www.tiede.fi/arkisto/artikkeli.php?id=862&vl=2007> [viitattu 27.9.2009].

Johdon assistenttityön ja kielten koulutusohjelma. Opetussuunnitelma. Kymenlaakson ammattikorkeakoulu. Saatavissa:
http://soleops.kyamk.fi/opsnet/dis/fi/ops_KoulOhjSel/tab/tab/sea?koulohj_id=2569306&ryhmytyyp=1&stack=push [viitattu 22.10.2009].

Karppinen, K. & Toivanen, J. 2005. Tietokone kuulostelee tunteita. Tiede-lehti. 5.7.2005. Saatavissa: <http://www.tiede.fi/arkisto/artikkeli.php?id=481&vl=2005> [viitattu 4.10.2009].

Kielenhuolto. Kielikone Oy. Saatavissa: <http://www.kielikone.fi/fi/Tuotteet/Kielenhuolto/> [viitattu 22.2.2009].

Kielikone Oy. Saatavissa: <http://www.kielikone.fi/fi/Yritys/> [viitattu 18.10.2009].

Kieliteknologia. Euroopan unioni. 21.12.2006. Saatavissa: <http://europa.eu/languages/fi/chapter/17> [viitattu 7.1.0.2009].

Kieliteknologian korkeakouluopetus: Tarve, tarjonta ja ehdotukset toimenpiteiksi. Helsingin yliopiston yleisen kielitieteen laitos. Opetusministeriön työryhmän muistio. 1.9.1999. Saatavissa: <http://www.ling.helsinki.fi/~koskenni/kieliteknologia/opm-raportti.pdf> [viitattu 18.10.2009].

Kites symposium 2009. Monikielisen ja – kulttuurisen vuorovaikutuksen foorumi. Kites ry. Esite. 8.10.2009.

Konekääntäminen. Kielikone Oy. Saatavissa: <http://www.kielikone.fi/fi/Tuotteet/Konek%C3%A4%C3%A4nt%C3%A4minen/> [viitattu 29.9.2009].

Koskenniemi, K. 2002. Johdatus kieliteknologiaan. Luku 1: Johdanto. Helsingin yliopisto. 28.10.2002. Saatavissa: <http://www.ling.helsinki.fi/kit/2002s/ctl190net/kt-johd/kt-johd-yl-art.pdf> [viitattu 13.9.2009].

Koskenniemi, K. 2002a. Johdatus kieliteknologiaan. Luku 5: Kielen kääntämisen apuvälineet ja automaattinen kielen kääntäminen. Helsingin yliopisto. 18.11.2002. Saatavissa: <http://www.ling.helsinki.fi/kit/2002s/ctl190net/kt-johd/kt-johd-mt-art.pdf> [viitattu 12.7.2009].

Koskenniemi, K. 2002b. Johdatus kieliteknologiaan. Luku 5: Kirjoittajan apuvälineet. Helsingin yliopisto. 26.11.2002. Saatavissa:

<http://www.ling.helsinki.fi/kit/2002s/ctl190net/kt-johd/kt-johd-at-art.pdf> [viitattu 22.2.2009].

Koskenniemi, K. 2002c. Johdatus kieliteknologiaan. Luku 2: Puheteknologia. Helsingin yliopisto. 28.10.2002. Saatavissa:

<http://www.ling.helsinki.fi/kit/2002s/ctl190net/kt-johd/kt-johd-pt-art.pdf> [viitattu 4.10.2009].

Koskenniemi, K. 2002d. Johdatus kieliteknologiaan. Luku 4: Tiedon haku ja siihen liittyvät sovellukset. 12.11.2002. Saatavissa:

<http://www.ling.helsinki.fi/kit/2002s/ctl190net/kt-johd/kt-johd-ir-art.pdf> [viitattu 17.10.2009].

Kääntäminen monikielisessä yhteisössä. Euroopan komissio. Luxemburg: Euroopan yhteisöjen virallisten julkaisujen toimisto. 1/2009. Saatavissa:

http://ec.europa.eu/dgs/translation/publications/brochures/translating_eu_brochure_fi.pdf [viitattu 18.9.2009].

Laitinen, H-R. 2008. Euromap HTL Case Study: TransSmart-konekäännösjärjestelmä Rautaruukissa. Suomen kieliteknologian dokumentaatiokeskus. Helsingin yliopiston yleisen kielitieteen laitos. 22.4.2008. Saatavissa:

<https://kitwiki.csc.fi/twiki/bin/view/FiLT/RautaruukkiFi> [viitattu 18.10.2009].

Lingsoft on täyden palvelun kielitalo. Lingsoft Oy. Saatavissa:

http://www.lingsoft.fi/?doc_id=35&lang=fi [viitattu 18.10.2009].

Monikielisyys: Euroopan voimavara ja samalla myös yhteinen sitoumus. Komission tiedonanto Euroopan parlamentille, neuvostolle, Euroopan talous- ja sosiaalikomitealle ja alueiden komitealle. Bryssel: Euroopan yhteisöjen komissio. 18.9.2008.

Nenonen, H. 2007. Kännykästä puhekaveri. Kauppalehti Vip, nro 12, 3.9.2007.

Nurmesniemi, H. 2001. Käännösohjelmat – uusi ulottuvuus. Teoksessa: Oittinen, R. & Mäkinen, P. (toim.). Alussa oli käännös. Tampere: Tampere University Press.

Reiman, J. 2007. Teknologia monikielisuuden apuvälineenä. Suomi ja Viro monikielissä Euroopassa: haasteita ja mahdollisuuksia. Seminaari. Turun yliopisto. 20.11.2007. Saatavissa: http://www.tkk.utu.fi/alueosasto/kv/sunrise/Sunriseseminaari_Turussa_20112007.pdf [viitattu 16.10.2009].

Reiman, J. Lingsoft Oy. Puhelinhaastattelu 13.10.2009.

Sovelluskohteita. Suomen puheentunnistus Oy. Saatavissa: <http://www.puheentunnistus.fi/fi/tuotteet-ja-palvelut/sovelluskohteita> [viitattu 4.10.2009].

Suuryritys. Kielikone Oy. Saatavissa: <http://www.kielikone.fi/fi/Tuotteet/Yritysratkaisut/Suuryritys/> [viitattu 18.10.2009].

Tekstinkäsittelytyökaluja. Informaatioteknologia. Jyväskylän yliopiston IT-tiedekunta ja avoin yliopisto. 23.9.2003. Saatavissa: <http://appro.mit.jyu.fi/doc/tyovaline/tekstinkasittely/index6.html#TOC12> [viitattu 18.10.2009].

Termien hallinta. Kielikone Oy. Saatavissa: <http://www.kielikone.fi/fi/Tuotteet/Termien+hallinta/> [viitattu 29.9.2009].

Tiedonhaku internetistä. Oulun seudun ammattikorkeakoulu. Raahen tekniikan ja talouden yksikkö. Saatavissa: <http://www.ratol.fi/opensource/tiedonhaku/> [viitattu 20.10.2009].

Tietoja tavutuksesta. 2009. Microsoft Corporation. Saatavissa: <http://office.microsoft.com/fi-fi/word/HP052290961035.aspx> [viitattu 5.7.2009].

Tikkanen, A. Suomen Puheentunnistus Oy. Sähköpostikeskustelu 15.10.2009.

Turunen, M. & Hakulinen, J. 2005. Vuorovaikutteiset puhe-sovellukset. Tampereen yliopisto. Tietojenkäsittelytieteiden laitos. Tietojenkäsittelytiede. 24.12.2005. Saatavissa: <http://www.tkts.fi/lehti/a24/35-51.pdf> [viitattu 3.10.2009].

Uszkoreit, H. Language technology. A first overview. The German Research Center for Artificial Intelligence. Saatavissa: <http://www.dfki.de/~hansu/LT.pdf> [viitattu 1.10.2009].

Vuorinen, E. 2001. Kielitoisintoja Euroopasta - EU-kääntämisen erityispiirteitä. Teoksessa: Oittinen, Riitta & Mäkinen, Pirjo (toim.). Alussa oli käänös. Tampere: Tampere University Press.

Kieliteknologia, mitä se on?




Maija Kosonen | 25.10.2009

Mitä on kieliteknologia?

- tutkii ihmisen luonnollista kieltä
- ihminen tuottaa kieltä sekä puheena että kirjoituksessa
- tietokone määrittää tuottamaan ja ymmärtämään ihmisen kieltä
- auttaa ihmisiä kommunikoimaan keskenään



- tietokoneen käyttö työssä yleistä
- tietokoneen käyttäminen ajasta ja paikasta riippumaton
- kieliteknologia auttaa kansainvälisessä kommunikoimisessa
- kielten ammattilaisten työtehtävät siirtyvät yhä enemmän muille työntekijöille




Lyhyt katsaus historiaan

- ensimmäinen sovellus konekäännös
- kylmä sota edisti kehitystä
- tietokoneiden yleistyminen 1980-luvulla mullisti kehitystä
- 1980-luvulla ensimmäiset oikeinkirjoituksen- ja kieliopintarkistimet
- tietotekniikan räjähdysmäinen kasvu lopullinen sysäys



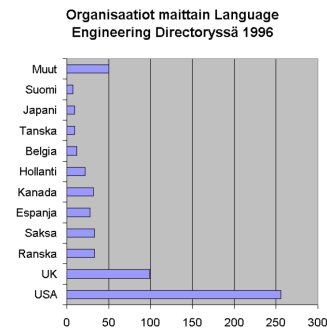
Suomen kieli ja kieliteknologia

- suomen kielen rakenteellinen rikkaus tuo erityispiirteitä kieliteknologia suomenkielisiin sovelluksiin
- paljon eri taivutusmuotoja
- yhdyssanojen käyttö yleistä



Kieliteknologia maailmalla

- varhaisimmat yritykset Yhdysvalloissa
- Yhdysvalloissa edelleen eniten kieliteknologiayrityksiä
- englannin kieltä tutkittu kielistä eniten



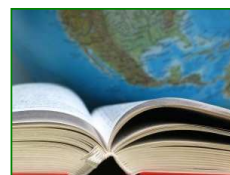
Kieliteknologian sovelluksia

- kirjoittajan työkalut
- tiedonhaku
- tietokone apuna kääntämisessä
- puheteknologia




Kirjoittajan työkalut

- oikeinkirjoituksen tarkistus
- kieliopin tarkistus
- tavutus
- synonyymisanastot
- elektroniset sanakirjat
- luettavuuden arviointi




Tiedonhaku

- kirjallisen tiedon määrä kasvanut valtavasti
- internet jatkuvasti kasvava ja muuttuva tiedon lähde
- hakuohjelmat käyttävät kieliteknologiaa
- dokumenttien automaattinen luokittelu
- tekstin tiivistäminen
- automaattinen hakemistojen muodostaminen



Tietokone apuna kääntämisessä

- käännösmuistiohjelmat
 - tietokantoja, joihin voi tallentaa tehdyn käännöksen
- termipankit
- konekääntäminen
 - tietokone kääntää kokonaisia lauseita ja tekstejä kohdekielelle
 - Ihminen tarkistaa konekäännöksen



Puheteknologia

- puheen käsittelyä tietokoneen avulla
- ihminen yleensä oppii puhumaan ennekuin kirjoittamaan
- puhe tehokkain ja käytetyin kommunikaatiokeino
- puhe välittää merkityksiä jotka eivät liity vain sanoihin



Puhesynteesi

- tietokoneella tuotettava ihmispuhetta jäljittelevä ääni
- esimerkkejä:
 - junakuulutukset
 - tiedotukset
 - sähköpostin ääneen lukeminen
 - “puhuva pää”
 - näkövammaisten äänenlukuohjelmat



Puheentunnistus

- ihmisen puheen tunnistamista tietokoneen avulla
- Esimerkkejä:
 - puheella komennettavat laitteet (esim. matkapuhelin)
 - puhelinvaihteen automatisointi
 - automaattinen ajanvarauspalvelu
 - pankkipalvelut (puhujan tunnistaminen äänestä)
 - sanelusovellukset



Kieliteknologia käytännössä

- suomessa kieliteknologialla vahvat akateemiset juuret
- hyötyjä työelämässä:
 - parempi kilpailukyky globaaleilla markkinoilla
 - automatisointi parantaa nopeutta ja palvelua
 - säästää työntekijöiden aikaa
 - edistää tehokasta kansainvälistä yhteydenpitoa ja yhteistyötä
 - tiedon saatavuus paranee

- hyötyjä työelämässä:
 - viestinnän johdonmukaisuus
 - kustannussäästöt
 - parantaa työntekijöiden työssä viihtymistä



Kieliteknologia ja EU

- 27 jäsenvaltiota ja 23 virallista kieltä
- EU:n alueella puhutaan noin 60 kieltä
- jäsenvaltioiden viralliset kielet ovat myös unionin virallisia kieliä ja tasavertaisia keskenään
- EU johtava kieliteknologia käyttäjä ja kehittäjä maailmassa
- kielen osaamisella etuja EU:ssa oleville yrityksille



Euroopan komission käännöspalvelu

- maailman suurin käännösorganisaatio
- kääntäjillä ja virkamiehillä käytössä suuri valikoima kieliteknologian sovelluksia
 - käännösjärjestelmä- ja muistiohjelmiä
 - konekäännösohjelmiä
 - elektroniset sanakirjat ja termipankit



Linkkejä

- Puhuva panda:
<http://www.bitlips.fi/pixnet/>
- Kieliteknologian dokumentaatiokeskus FiLT:
<https://kitwiki.csc.fi/twiki/bin/view/FiLT/FiLTWikiEn>