

”SOKEA IDIOOTTI”  
KONEKÄÄNTÄMISEN ONGELMIA JA MAHDOLLISUUKSIA

Anni Malmivaara  
Tampereen yliopisto  
Kieli ja käännöstieteiden laitos  
Käännöstiede (saksa)  
Pro gardu -tutkielma  
Toukokuu 2007

Tampereen yliopisto  
Käännöstiede (saksa)  
Kieli- ja käännöstieteiden laitos

MALMIVAARA, Anni: ”Sokea idiootti” – Konekääntämisen ongelmia ja mahdollisuuksia

Pro gradu -tutkielma, 72 sivua + liitteet (3 kpl) + saksankielinen lyhennelmä, 13 sivua.  
Toukokuu 2007

---

Tässä tutkielmassa tarkastellaan konekääntämisen mahdollisuuksia ja ongelmia. Konekääntämisestä on sen alkuajoista, siis noin 1950-luvulta, lähtien povattu tulevaisuuden suurta menestystä, joka korvaa ihmiskääntäjät täysin. Kuitenkin on osoittautunut, että luonnolliselta kielelle toiselle kääntävää ohjelmaa on käytännössä hyvin vaikea toteuttaa.

Kääntäminen on ihmisellekin hankalaa; se vaatii kahden tai useamman kielen erinomaista osaamista sekä laajaa kulttuurien tuntemusta. Ihmisen käyttämä kieli on monimutkaista ja välillä monitulkintaista. Kone ei ymmärrä kieltä, saatikka että se osaisi kääntää toiselle kielelle.

Konekäännökset sisältävät usein vääriä käännöksiä, sanajärjestysvirheitä sekä kääntämättä jääneitä sanoja. Koneelle on lähes mahdotonta syöttää monitulkintaisia sanoja, kaikkien sanojen eri taivutusmuotoja ja synonyymejä. Elektroniset sanakirjat paisuisivat järjettömän suuriksi, mikä johtaisi hitaampiin käännösohjelmistoihin.

Tutkin konekääntämisen erilaisia menetelmiä ja taustoja. Pohdin myös, mihin konekäännöksiä voidaan käyttää laatuongelmista huolimatta. Konekääntämisen tutkijat ovat perustelleet, että konekääntäminen on erinomainen tapa selvittää vieraalla kielellä kirjoitetun tekstin sisältö. Asiantuntijalle usein riittää, että tekstin ydinasiat tulevat selviksi, vaikka sanajärjestys- tai kielioppivirheet estäisivätkin tekstin julkaisemisen sellaisenaan.

Konekäännöksiä käytetään yleisesti esimerkiksi Euroopan komission käännöspalvelussa nopeuttamaan ihmiskääntäjien työtä. Vieraskielisestä tekstistä tehdään ensin konekäännös, jonka pohjalta ihminen muokkaa julkaisukelpoisen tekstin. Joskus virkamiehet teettävät konekäännöksiä itselleen, saadakseen selville mitä teksti käsittelee.

Internetissä on tarjolla kaupallisia konekäännösohjelmia, joista osa tarjoaa käännöksiä ilmaiseksi. Tutkimuksessani käännsin kolme erilaista teknisen alan tekstiä Babel Fish -nimisellä ohjelmalla tarkoitukseni selvittää konekäännösten laatua. Tulokseni ei ollut yllättävä, vaan päinvastoin tuki ennakkokäsitystä siitä, että nykyiset konekäännösohjelmat eivät sovellu kääntämiseen sellaisenaan. Koneen tekemät käännökset täytyy ihmisen toimesta muokata jälkikäteen. Konekäännökset sisälsivät paljon vääriä sanajärjestyksiä, kääntämättä jääneitä sanoja sekä kielioppivirheitä.

Avainsanat: konekääntäminen, konekäännös, ihmiskääntäjä, luonnollinen kieli

1.	JOHDANTO .....	5
2.	KÄÄNTÄJÄN TOIMENKUVA .....	8
3.	KONEKÄÄNTÄMINEN .....	11
3.1	HISTORIAA .....	11
3.2	KONEKÄÄNTÄMISEN PERUSMENETELMÄT .....	12
3.3	AUSTERMÜHLIN JAOTTELU .....	14
3.4	BERGENIN (BOWKERIN) JAOTTELU .....	15
3.5	ELEKTRONISET SANAKIRJAT OSANA KÄÄNNÖSOHJELMAA .....	16
4.	MIKSI KONEKÄÄNTÄMISTÄ? .....	19
4.1	LIIKAA KÄÄNNETTÄVIÄ TEKSTEJÄ .....	20
4.2	TEKNISET TEKSTIT OVAT TYLSIÄ .....	21
4.3	TERMINOLOGIAN YHDENMUKAISUUS .....	22
4.4	PIDEMPIÄ TEKSTEJÄ NOPEAMMIN .....	22
4.5	HUIPPULAATUA EI AINA TARVITA .....	23
4.6	YRITYKSET HALUAVAT LASKEA KÄÄNNÖSTEN KUSTANNUKSIA .....	24
5.	KONEKÄÄNTÄMISEN MENETELMIÄ .....	25
5.1	SUORA KÄÄNTÄMINEN .....	25
5.2	SIIRTOMENETELMÄ .....	25
5.3	INTERLINGVAALINEN MENETELMÄ .....	26
5.4	KÄÄNNÖSÄLY .....	26
5.5	ONGELMIA .....	27
6.	KONEKÄÄNNÖKSEN KÄÄNNÖSPROSESSI .....	29
6.1	LÄHTÖTEKSTIN EDITOINTI .....	29
6.2	LÄHTÖTEKSTIN ANALYYSI .....	30
6.2.1	<i>Morfologinen analyysi</i> .....	31
6.2.2	<i>Syntaktinen analyysi</i> .....	32
6.2.3	<i>Semanttinen analyysi</i> .....	32
6.3	KÄÄNNÖSTRATEGIAN VALINTA .....	33
6.4	INTERAKTIIVISET VAIHEET .....	34
6.5	KÄÄNNÖKSEN JÄLKIEDITOINTI .....	36
7.	LUONNOLLINEN KIELI JA SEN YMMÄRTÄMISEN ONGELMA .....	38
7.1	KIELEN SUHDE AJATTELUUN JA TODELLISUUDEN HAHMOTTAMISEEN .....	39
7.2	MONISELITTEISET ILMAUKSET .....	40
7.2.1	<i>Homonymia</i> .....	40
7.2.2	<i>Polysemia</i> .....	41
7.3	KIELEN RAKENTEELLISET ONGELMAT .....	42
7.4	VERTAUSKUVALLINEN KIELENKÄYTTÖ .....	43
7.5	KIELEN VIITTAUSSUHTEET .....	44
8.	KONEKÄÄNNÖSOHJELMIA .....	46
8.1	EUOTRA .....	47
8.2	SYSTRAN .....	47
8.3	KAUPALLISET KÄÄNNÖSOHJELMAT .....	48

9. BABEL FISH.....	49
9.1 AINEISTON JA TUTKIMUSMENETELMÄN ESITTELY.....	50
9.2 AINEISTO 1.....	51
9.3 AINEISTO 2.....	56
9.4 AINEISTO 3.....	61
10. BABEL FISH -KÄÄNNÖSTEN ANALYYSIN YHTEENVETO .....	68
11. KONEKÄÄNTÄMISEN TULEVAISUUS .....	70
LÄHDELUETTELO .....	71
LIITTEET.....	75
DEUTSCHE KURZFASSUNG	

# 1. Johdanto

Maallikoilla tuntuu olevan yleisesti käsitys, että käännösala on täysin automatisoitu. Kun käännöstoimistolta pyydetään 24 000 sanan käännös huomiseksi, on kysyjä yllätynyt, kun vastaus on ”Ei onnistu”. Tällöin maallikko kysyy huolestuneena: ”Kai teillä siellä jotkut käännöskoneet on?” Todellisuus on kuitenkin vielä kaukana tästä. Tämänhetkiset käännöskoneet eivät hallitse kieltä eivätkä kääntämistä niin sujuvasti, että ne voisivat täysin korvata ihmisen. Toki raakakäännöksiä voidaan tehdä koneilla, ja toki kääntäjät ja muutkin tekstinkäsittelijät saattavat käyttää konekäännöksiä apunaan luodessaan uusia tekstejä. Mutta siihen se useimmiten jää. Tällä hetkellä ei ole olemassa niin kehittyntä konekäännösohjelmaa, että esimerkiksi 24 000 sanan käännöksen tekeminen onnistuisi ilman ihmisen puuttumista käännösprosessiin tai ihmisen tekemää käännöksen viimeistelyä. Kääntäminen on monimutkainen prosessi, joka vaatii inhimillistä älykkyyttä, kykyä meta-ajatteluun sekä luovuutta.

Tässä tutkielmassa tarkastelen konekääntämistä, sen menetelmiä, historiaa, ongelmia ja mahdollisuuksia. Muun muassa Jorma Tommola (2004: 9) on kirjoittanut, että edes kahden kielen sujuva osaaminen ei välttämättä tarkoita sujuvaa kääntämistä. Kääntäminen ei ole pelkästään sanojen kääntämistä toiselle kielelle, vaan on otettava huomioon monia seikkoja, jotta käännös olisi samalla tavalla ymmärrettävä kuin lähtöteksti.

Konekääntäminen on periaatteeltaan yksinkertaista: Lähtötekstin sanat muutetaan kohdekielisiksi, ja koneelle syötetyn kieliopin avulla käännöksestä tulee ymmärrettävää kieltä. Jokaisella asiaan perehtyneellä tuntuu olevan mielipide joko puolesta tai vastaan. John Hutchins (2005: 5–6) perustelee kantaansa konekäännösten puolesta kuudella argumentilla:

1. Käännettäviä tekstejä on liikaa
2. Ne ovat liian tylsiä ihmiskääntäjille
3. Koneen on helpompi pysytellä orjallisesti yhdenmukaisessa sanastossa
4. Kone on nopeampi kuin ihminen
5. Huippulaatu ei ole aina tarpeen

## 6. Koneen edulliset kustannukset verrattuna ihmiskääntäjään.

Konekäännösohjelman tärkein osa on sen sanakirja. Sanakirjat ovat usein puutteellisia, koska kaikkia kielen sanojen kaikkia muotoja on mahdoton listata. Nämä puutteet näkyvät konekäännöksissä kääntämättä jääneinä sanoina ja väärinä käännösvastineina. Ihminen ymmärtää, että *menen* on taivutusmuoto verbistä *mennä*, mutta kone ei tällaista päätelmää osaa automaattisesti tehdä.

Ihminen osaa myös tehdä lähtötekstianalyysia ja harkita erilaisia käännösmenetelmiä, mutta koneella ei ole muuta mahdollisuutta kuin valita ne menetelmät, jotka sille on syötetty. Luonnollinen kieli on osa ihmistä. Sen opettaminen koneelle on mahdotonta, sillä kielen ymmärtäminen ei rakennu pelkästään yksittäisten sanojen ymmärtämiseen tai kieliopin analysoinnin varaan, vaan sanojen ja ilmausten tulkinta edellyttää monipuolista sanojen välisten merkityssuhteiden ja myös epäsuorien merkitysten ymmärtämistä. Kielen ymmärtäminen edellyttää myös paljon kielenulkoista käytännön tietoa ympäröivästä yhteiskunnasta. Koneelle tuottaa suuria ongelmia käsitellä esimerkiksi moniselitteisiä ilmaisuja, kielen viittaussuhteita sekä eri kielten kieliopillisia rakenteita.

Konekääntäminen on myös teknisesti vielä haasteellista. Insinöörit eivät ole onnistuneet kehittämään ohjelmaa, joka osaisi tuottaa muuta kuin kömpelöä ja epäluontevaa kieltä. Erikoisalojen kääntäminen saattaa onnistua paremmin, mutta mitä yleisemmällä tasolla kieli on, sitä vaikeampaa koneen on kääntää sitä toiselle kielelle.

Markkinoilla on kuitenkin saatavilla kohtuullisen hyviä konekäännösohjelmia, joita esimerkiksi eri viranomaiset käyttävät raakakäännösten tekemiseen. Kehitteillä on myös niin kutsuttu käännösäly. Käännösäly perustuu tekoälytutkimukseen, ja sitä pidetään käännösmuistiohjelmien seuraavan sukupolven edustajana.

Tutkielman lopussa esittelen oman tutkimusaineistoni ja -tulokset. Käänsin kolme erilaista teknistä tekstiä Internetissä ilmaiseksi saatavilla olevalla Babel Fish -käännösohjelmalla. Tulokset eivät yllättäneet, vaan pikemminkin tukivat ajatusta siitä, että kaupalliset

konekäännösohjelmat tuottavat ainoastaan kelvotonta materiaalia. Konekäännökset vilisivät vääriä sanajärjestyksiä, kieliopillisesti virheellisiä lauseita sekä kääntämättä jääneitä sanoja. Parempi tietysti olisi ollut, jos olisin voinut tutkia aineiston jollain vakavasti otettavalla ohjelmalla, mutta sellaista mahdollisuutta ei käytännössä valitettavasti ollut. Uskon, että laadukas käännösohjelma olisi osannut kääntää tekstit paremmin, jopa ymmärrettävään, joskaan ei välttämättä julkaisukelpoiseen muotoon. Tämä on kuitenkin vain tuntumaa, eikä perustu tutkimustuloksiin.

## 2. Kääntäjän toimenkuva

Kääntäminen ei ole yksinkertainen asia, eikä sitä voida kuvailla yksinkertaisesti. Vaikka osaisi kieliä sujuvasti, ei ole itsestään selvää, että osaa kääntää sujuvasti. Kääntäminen on vaikeaa – ihmisellekin: pitää osata kahta tai useampaa kieltä todella hyvin ja osata tuottaa hyvää tekstiä. Sellainenkaan kaksikielisyys, joka riittää päivittäisten asioiden hoitamiseen välttävästi tai jopa erinomaisesti, ei automaattisesti tarkoita sitä, että osaa kääntää sujuvasti. Kääntäminen ei ole tavanomaista kielenkäyttöä. Jorma Tommolain (2004: 9) mukaan se vaatii erityisvalmiuksia ja tekniikoita, jotka eivät luonnostaan seuraa kielitaidosta. Ne vaativat kulttuurien, erilaisten elämäntapojen ja aihepiirien tavanomaista syvempää tuntemusta. Kääntäjältä vaaditaan kykyä analysoida, arvioida ja kehittää omaa suoritustaan. Tämän lisäksi yhteiskunnan globalisaatiosta johtuen monissa prosesseissa jo suorastaan vaaditaan monikielisyyttä. Kääntämisestä tulee entistä vaativampi asiantuntijatehtävä. (Mts.)

Kääntäjän tärkein työkalu on oma äidinkieli. Käännöstieteen piirissä ajatellaan yleisesti, että kääntäjän pitäisi kääntää vain omalle äidinkielelleen, koska siitä ja sen kulttuurista hänellä on kattavimmat tiedot. Kuten Ulla Palomäki (2004: 21) kirjoittaa: ”Kääntäjän ammattitaidon tärkeä mittari on juuri äidinkielen kohdeteksti ja äidinkielen taito.” Kääntäjän on hallittava äidinkieltä kattavasti: normitettu yleiskieli ja sen rakenne oikeakielisyys- ja oikeinkirjoitusvaatimuksineen sekä välimerkkien käyttö, eri tyylilajit, sanontatavat, idiomit, kielen vaihtelut tilanteesta toiseen eli rekisterit, murteet ja niin edelleen. (Mts.) Myös lähtökielestä tulisi olla erittäin hyvät tiedot, jotta merkitysten siirtäminen toiselle kielelle onnistuu. Useimmiten kyseessä ei ole vain kieleltä toiselle kääntäminen, vaan kielen vaihtuessa vaihtuu myös kulttuuri. Tarkan käännöksen tekemiseksi pitää tietää konteksti ja tuntee sekä lähtö- että kohdekielen rakenteet ja säännöt. On osattava verrata lähtökulttuuria kohdekulttuuriin.

Kääntäminen ei voi tapahtua muuttamalla konemaisesti lähtökielen sanoja suoraan kohdekielisiksi, vaan käännettäessä on otettava huomioon monia seikkoja. Kääntämisessä korostetaan merkityksen siirtämisen ensisijaisuutta. Minna Häyrysen (1992: 4–5) mukaan vaikeimpana ongelmana voidaan pitää lähtökielen ilmauksen merkityksen siirtämistä



mahdollisimman samanlaisena toiseen kielijärjestelmään. Keskeisenä ongelmana voidaan myös pitää ekvivalenssin eli vastaavuuden sekä vertailukriteerien ongelmaa. Sågvall Hein (2004: 7) sanoo laatukriteereiden olevan korkealla, kun on kyse ihmisen tekemästä käännöksestä. Käännökseltä odotetaan samaa kielellistä tasoa kuin lähtötekstiltä sekä samaa informaation sisältöä. Toisaalta muun muassa Pertti Nuutila (2005: 7) kirjoittaa, että konekäännöksiä tehdään paljon myös sellaisista teksteistä, joita ei ilman mahdollisuutta saada ilmaista ja nopeaa raakakäännöstä tehtäisi ollenkaan.

Jukka Pohjolan (2004) mukaan ”tietokone on ottanut miltei kaikessa kääntämisessä pääroolin ja jättänyt ihmiskääntäjän lähes statistin asemaan.” Tietoyhteiskunta on muuttanut kääntäjän työtä enemmän kuin monien muiden alojen ammattilaisten työtä. Tietoa on saatavilla enemmän kuin ennen ja elektroniset työkalut helpottavat työn tekemistä. Kääntäjän tietotekniikan hallintaa pidetään itsestäänselvytenä.

Kääntäjän elektroniset työkalut eivät tarkoita yksinomaan konekäännösohjelmia, vaan on olemassa paljon muitakin teknisiä apuvälineitä, esimerkiksi termipankkeja tai käännösmuisteja. Koska konekääntämisessä ei oikeastaan ole kyse kommunikatiivisten, kulttuuristen tai sanakirjamaisen suorien käännösten tuottamisesta, Frank Austermühl (2001: 1) perustellusti kysyy, voidaanko tässä yhteydessä edes puhua varsinaisesta kääntämisestä.

Elektroniset työkalut eivät ole ihmiskääntäjän korvike, vaan ne pikemminkin toimivat sen tukena. Mikään ohjelma ei pysty syrjäyttämään ihmistä, mutta laadukkaat käännösohjelmat tarjoavat huomattavia etuja kääntämiseen. Oikeat työkalut auttavat kääntäjää tekemään työtään ja parantamaan työn laatua. Tänä päivänä tietotekniset taidot ovat edellytyksenä lähes jokaisessa, varsinkin kielten parissa tapahtuvassa työssä. David Bergen (2004: 141) huomauttaa, että käännösteknologiaksi voidaan luokitella lähes mikä tahansa teknologia, joka on kehittyneempi kuin kynä ja paperi. Ammattimaisella kääntäjällä ei ole enää vaihtoehtoa olla käyttämättä Internetiä ja verkostoja työnsä apuna. Kääntäjä, joka työskentelee pelkästään kynien ja kirjojen parissa, ei ole enää realiteetti, vaan nykyajan kääntäjä käyttää sähköpostia toimeksiantojen vastaanottamiseen ja valmiiden käännösten lähettämiseen, elektronisia sanakirjoja termien hakuun, käännösmuistiohjelmia, elektronista arkistointia ja niin edelleen. Perinteinen posti tai faksi ei vastaa näihin vaatimuksiin. (Austermühl 2001: 19.)

Nykyajan kääntäjän tärkein elektroninen työkalu lienee käännösmuisti. Se ei ole automaattinen käännösohjelma, sillä se ei anna mitään käännöksiä valmiina. Käännösmuistiohjelma tallentaa kääntäjän tekemät käännökset lause tai niin sanottu *segmentti* kerrallaan. Ohjelma etsii muistista vastaavan tai lähes vastaavan käännösvastineen ja ehdottaa sitä kääntäjälle, jolloin kääntäjä päättää, käyttääkö hän vanhaa käännöstä sellaisenaan vai muuttaako hän sitä toisenlaiseksi. Jos vastaavaa käännöstä ei löydy, täytyy segmentti kääntää manuaalisesti. Uusi käännös tallentuu muistiin. Kääntäessä muisti siis päivittyy jatkuvasti.

## 3. Konekääntäminen

### 3.1 *Historiaa*

Hutchinsin (2005b) mukaan jo ennen kuin tietokoneet oli keksitty, oli olemassa ajatus luonnollista kieltä kääntävästä automatiikasta. Kuitenkin vasta 1900-luvulla ajatusta voitiin käytännössä pitää realistisena, ja kun 1940-luvulla matemaattisia laskelmia tekevä tietokone oli onnistuttu rakentamaan, ryhdyttiin miettimään myös luonnollisen kielen kääntämistä koneella. Ensimmäiset konekäännösohjelmat sisälsivät kaksikielisen sanakirjan, joista ne tarjosivat lähtökielen sanalle yhden tai useamman kohdekielisen vastineen, ja mahdollisesti myös vaihtoehdon sanajärjestyksen muuttamiseksi kohdekielen konventioiden mukaiseksi. (Hutchins 2005b.) Vuonna 1954 Yhdysvalloissa esiteltiin ensimmäinen suppea konekäännösjärjestelmän prototyyppi, joka käänsi englannista venäjään. Sen seurauksena Yhdysvalloissa saatiin konekäännöksen tutkimukseen mittava rahoitus ja se synnytti lukuisia tutkimusryhmiä eri puolilla maailmaa. (Arppe 1994.)

Seuraavat 10 vuotta oltiin hyvin optimistisia ja konekääntämisen ennustettiin tekevän läpimurtoja pianikin. Optimismi karisi kuitenkin nopeasti, kun tutkijat törmäsivät kielen ”semanttisiin rajoituksiin” eikä konekäännösten laatu parantunut yhtä nopeasti kuin oli odotettu (Hutchins 2005b).

Konekääntämisen alkuaikoina tietokoneet olivat hitaita ja niiden muisti oli hyvin rajallinen verrattuna nykytietokoneisiin. Tuolloin puhuttiin paljon sanakirjan laajuudesta ja hakusanojen lataamisen nopeudesta. (Hutchins 1986: 40–41.) Vuonna 1966 ilmestyneessä kuuluisassa ALPAC:in (Automatic Language Processing Advisory Committee) raportissa (National Academy of Sciences 1966) todettiin konekääntämisen olevan hitaampaa, laadultaan huonompaa ja tuplasti niin kallista kuin ihmiskääntäminen. Raportissa konekääntämistä pidettiin saavuttamattomana tavoitteena. ALPAC:in raportin antama tuomio lopetti käytännössä konekääntämisen rahoituksen Yhdysvalloissa 1960- ja 1970-luvuilla. Raportin keskeinen sisältö oli, ettei konekäännösjärjestelmiltä olisi koskaan odotettavissa riittävää

parannusta käännoistyössä, eikä niitä välttämättä edes tarvittaisi, sillä ihmiskääntäjiä uskottiin riittävän tyydyttämään kaiken käännöstarpeen. (Arppe 1994.) Konekääntämisen tutkimisen sijaan raportissa suositeltiin tietokoneavusteisen kääntämisen, kuten automaattisten sanakirjojen, sekä kieliteknologian kehittämistä (Hutchins 2005b). Raportin julkaisemisen jälkeen konekääntämisen tutkimuksen rahoitus ehtyi etenkin Yhdysvalloissa ja sen painopiste siirtyi Eurooppaan. 1970-luvun loppupuolella ensimmäiset vieläkin toimivat konekäännösjärjestelmät otettiin käyttöön: *Systran* Euroopan Yhteisön komissiossa ja *Météo* Kanadan sääpalvelussa. (Arppe 1994; Hutchins 2005b.) *Météo* on konekäännösohjelma, joka kääntää sää tiedotuksia ja -ennustuksia englannin ja ranskan välillä. Sen sanasto käsittelee niin suppeaa aihealuetta, että käännökset ovat suhteellisen hyvälaatuisia (Kay 2001).

1980-luvulla tutkittiin kehittyneempiä menetelmiä ja tekniikoita: tilastoihin ja esimerkkeihin pohjautuvat menetelmät, niin sanotut korpuspohjaiset tai tilastolliset ratkaisut. Ne perustuvat suurten elektronisten tekstimäärien ja niiden käännösten olemassaoloon, analysointiin ja systemaattiseen kasvattamiseen (Arppe 1994). Vielä 1980-luvulla tutkijoiden tavoitteena oli kehittää ohjelma, joka kääntäisi tekstejä sujuvasti luonnolliselta kieleltä toiselle (Hutchins 1986: 15).

### **3.2 Konekääntämisen perusmenetelmät**

Tietokoneisiin liittyy lukemattomia ennusteita, joista suurin osa ei ole koskaan toteutunut. Vielä 1960-luvulla kuviteltiin, että kun tietokoneelle syötetään kielioppi ja sanakirja, ne syrjäyttävät ihmiskääntäjät nopeasti. Mutta kun käännös ohjelma käänsi ”*out of sight - out of mind*” muotoon ”*sokea idiootti*”, huomattiin, ettei konekääntäminen ollutkaan niin yksinkertaista kuin oli luultu. Jo vuonna 1996 Petteri Järvinen kirjoitti: ”Tämän päivän koneet pystyvät kääntämään kohtuullisella tarkkuudella yksinkertaisia tekstejä, mutta lopullisen hiomisen tekee aina ihminen.” (Järvinen 1996.)

Konekääntämisen perusidea on Star Trekin universaali kielenkääntäjä tai mekaaninen versio Douglas Adamsin Babel Fishistä: kone, joka muuttaa lähtötekstin täydelliseen kohdekieliseen tekstiin ilman minkäänlaista ihmisen vuorovaikutusta. Alun perin käännösten toimeksiantajien

unelma (tai kääntäjien painajainen) oli täysin automaattinen konekääntäminen, jossa ihmisen työpanosta ei tarvittaisi lainkaan (Bergen 2004: 143). Konekääntäminen on aihe, josta puhutaan yhä enemmän. Elektronisista käännöstyökaluista se herättää kenties eniten huomiota julkisuudessa, varsinkin maallikkojen keskuudessa (Austermühl 2001: 153).

Konekääntämiseksi kutsutaan sitä, kun käännösohjelma kääntää kokonaisen tekstin automaattisesti ja esittää valmiin lopputuloksen. Kun ihminen kääntää tekstin, hän ehkä käyttää apunaan käännösmuistia tai elektronista termipankkia, mutta se ei ole konekääntämistä. Ihmisen tekemän käännöstyön ja konekääntämisen välissä on harmaata aluetta: kone saattaa ehdottaa kokonaisia aikaisemmin käännettyjä lauseita tai tarvittaessa tehdä pienehköjä muutoksia. Kuitenkin tällä harmaallakin alueella jokainen lause on alun perin joko ihmisen tai käännösohjelman kääntämä. Konekäännöksiksi voidaan kutsua vain sellaisia käännöksiä, jotka ovat kokonaan käännösohjelman tuottamia. (Melby 2001.)

Konekääntäminen terminä on vakiintunut tarkoittamaan yleistä nimitystä järjestelmälle, joka käyttää tietokonetta muuttamaan lähtökielisen tekstin kohdekieliseksi (Hutchins 1986: 1). Se siis pyrkii jäljittelemään ihmistä kääntäjänä (Häyrynen 1992: 9). Nykyiset kehittyneimmätkään ohjelmat eivät käännä yhtä sujuvasti kuin äidinkielen puhuja tai ammattimainen kääntäjä. Häyrysen (mts.) mukaan ihmiskääntäjän ja konekäännösohjelman lähtökohdat ja päämäärät ovat pitkälti samanlaiset. Molemmat pyrkivät tuottamaan lähtökielen sanomalle kohdekielille siten, että sanoman merkitys pysyy niin lähellä alkuperäistä kuin mahdollista. Automaattinen kielen kääntäminen on hyvin vaikea toteuttaa, ja suurin ongelma on sanojen merkityksen eli semantiikan kuvaaminen koneelle. Merkitys riippuu käyttöyhteydestä eli kontekstista. Vaikka insinöörit ja kielitieteilijät ratkovat näitä ongelmia, lienee tällä hetkellä mahdotonta, että kukaan pystyisi tarjoamaan nopeaa ja täysin saumatonta konekäännöspalvelua.

Konekäännösosalalla kehitys ei ole ollut niin huikeaa kuin aluksi luultiin. Toisaalta paljon asioita, joita pidettiin tietokoneille mahdottomina, saatiin toteutettua 1950–1970-luvuilla. Teknologian kehittyminen on muuttanut maailmaa ja ihmisten elämää nopeammin kuin olisi voitu ennustaa. On vaikea kuvitella, mitä tapahtuu seuraavina vuosikymmeninä, kun elektroniikka on monta kertaa nykyistä nopeampaa ja halvempaa. Journalisti Matti Lintulahti

(2003) kirjoittaa, että jos tietokoneiden tietojenkäsittelykyky kasvaa nykyvauhtia, vuoden 2030 tienoilla voi olla mahdollista rakentaa oikeasti älykäs ja itsestään tietoinen robotti.

Historia on myös osoittanut, että se, mitä on pidetty mahdottomana, voi muuttua yllättävänkin nopeasti mahdolliseksi. Viime vuosina mielenkiinto tietokoneen avulla tapahtuvaan kääntämiseen on kasvanut. Konekääntäminen tunnetaan kuitenkin valitettavasti useimmiten sen tekemien virheiden, eikä niinkään sen onnistumisien vuoksi.

### **3.3 Austermühlin jaottelu**

Koneen ja ihmisen osallistumisen astetta kääntämiseen Frank Austermühl (2001: 157–158) kuvaa seuraavasti:

FAHQT	Fully Automatic High Quality Translation (automaattinen hyvälaatuinen kääntäminen)
FAMT	Fully Automatic Machine Translation (automaattinen konekääntäminen)
HAMT	Human-Aided Machine Translation (ihmisavusteinen konekääntäminen)
MAHT	Machine-Aided Human Translation (tietokoneavusteinen ihmiskääntäminen)
HT	Human Translation (ihmiskääntäminen)
MT	Machine Translation (konekääntäminen)

Näistä voidaan heti jättää pois HT sekä FAHQT, koska ne eivät ole nykyaikaa (Austermühl 2001: 11). Jäljelle jäävät MAHT (elektronisten sanakirjojen ja terminologiatietokantojen käyttö sekä tekstinkäsittelyohjelmat), HAMT (käännösmuistit, jotka ovat aluksi tyhjiä ja jotka täyttyvät käännöksiä tehdessä) ja MT eli konekääntäminen, joka vaatii tekstin esi- ja jälkieditointia.

Käytettyään sievoisen summan rahaa konekäännösohjelmien kehittelyyn tutkijat ovat luopuneet täysin automaattisen korkealaatuisen käännösohjelman (FAHQT) toivosta. Sen sijaan on keskitytty käyttämään konekäännösohjelmia tietyn rajoitetun alan tekstien

kääntämiseen. Täysin automaattista kääntämistä on toki jo tarjolla Internetissä. Näiden palveluiden heikkous on David Bergenin (2004: 143) mukaan se, että laatu ei yleensä ole ”*high quality*”.

### **3.4 Bergenin (Bowkerin) jaottelu**

David Bergenin (2004: 144) mukaan kaikkea kääntämistä voitaisiin kutsua ihmiskääntämiseksi (HT), koska mikään tietokone ei osaa luoda omaa käännosmuistiaan tai tietokantaansa ilman ihmisen väliintuloa. Sinänsä tietokone ei suomenkielisestä nimestään huolimatta tiedä mitään eikä se myöskään käsittele tietoa samassa mielessä kuin ihminen. Tietokoneelle voidaan syöttää merkkijonoja, informaatiota, jota se mekaanisesti käsittelee siihen ohjelmoitujen käskyjen mukaisesti.

Yhtä lailla voidaan sanoa, että nykyään kaikki kääntäminen on jossain määrin tietokoneavusteista. Sen takia Bergen (mts.) ehdottaakin, että ihmiskääntämistä (HT) ei pitäisi enää luokitella erikseen. Myös Bergenin (2004: 147) lainaama Bowker (2002) käyttää samankaltaista jakoa ja jättää perinteisen ihmiskääntämisen (HT) pois laskuista kokonaan. Hän jakaa käännosmenetelmät kolmeen osa-alueeseen seuraavasti:

1. HT (Human translation): tekstinkäsittelyohjelmat, kieliäsun tarkistus, Internet ja muut elektroniset lähteet
2. CAT (Computer-Aided Translation): skannerit, termipankit, käännosmuistit
3. MT (Machine Translation): konekäännohjelmat

Vaikka Austermühl (2001: 157–158) sanookin, että perinteinen ihmiskääntäminen (HT) ei ole nykyaikaa, hän haluaa silti edelleen erottaa sen tietokoneavusteisesta ihmiskääntämisestä (MAHT). Bowker luokittelee yhteen ihmisavusteisen konekääntämisen ja tietokoneavusteisen ihmiskääntämisen ja kutsuu sitä kokonaisuudessaan ihmiskääntämiseksi (HT).

### 3.5 **Elektroniset sanakirjat osana käännösohjelmia**

Elektronisen sanakirjan luominen ja täydentäminen on yksi konekääntämisen tärkeimmistä tehtävistä. Tällaiset sanakirjat olivat alkuaikoina konekääntämisen kehittäjien keskeisin huolenaihe, ja ne ovat edelleen ratkaisevan tärkeässä asemassa. Kuten muutkaan sanakirjat, myöskään elektroniset sanakirjat eivät ole täydellisiä: niistä puuttuu sanoja ja merkityksiä. Joskus joukkoon voi myös livahtaa suorastaan vääriä käännösvastineita. Tällaiset seikat johtavat virheellisiin käännöksiin.

Hakumuotona käytetään sanan perusmuotoa, koska monissa kielissä sanat esiintyvät useissa eri muodoissa. Ohjelman sanakirjoihin voitaisiin tietysti lisätä sanojen kaikki taivutusmuodot, mutta Hutchinsin (1986: 44) mukaan on täysin tarpeetonta sisällyttää kaikkien sanojen kaikkia muotoja, koska sanakirjat paisuisivat turhan suuriksi. Esimerkiksi nykysuomen sanakirjassa on tällä hetkellä noin 210 000 perusmuotoista hakusanaa. Yksikielisistä sanakirjoista jätetään joskus pois itsestäänselvyksiä, jotta mukaan mahtuisi enemmän sanoja (Westman 1998: 85). Kone käyttää kaksikielisiä sanakirjoja, ja niistä ei tietenkään voida jättää pois mitään, sillä sille ei mikään ole itsestään selvää. Kone ei ymmärrä sanojen taivutusmuotoja, vaan käsittelee samankin sanavartalon eri muodot erillisinä yksikköinä, mikä hankaloittaa sanakirjan käyttöä ja päivittämistä (Nuutila 2005: 21).

Sanakirjojen päivitys ei onnistu samassa tahdissa kuin yleiskielen sanaston uusiutuminen. Kieli muuttuu koko ajan. Sanat ja niiden merkitykset muuttuvat tilanteesta toiseen. Uusia sanoja voi muodostaa vanhan, vakiintuneen mallin mukaan, ja samalla elvyttää epäproduktiivisiksi muuttuneita elementtejä. (Häkkinen 1998: 140; Nuutila 2005: 19.) Ihmiset keksivät uusia sanoja jatkuvasti, joista osa jää kieleen pysyvästi ja toiset elävät vain hetken; entisiä sanoja jää pois käytöstä ja uusia voidaan vaikkapa lainata muista kielistä. Johtamalla ja yhdistämällä saadaan helposti myös tilapäisiä sanoja, jotka eivät tule yleisesti tunnetuiksi eivätkä vakiinnu kieleen (Häkkinen 1998: 143). Sana voi myös vähitellen siirtyä tyylilajista toiseen. Esimerkiksi vuosia sitten sanaa *kännykkä* käytettiin ainoastaan leikillisissä yhteyksissä, mutta nykyään talousuutisissakin puhutaan jo *kännykkämarkkinoista*. (Palomäki 2004: 23.)



Elektroniset sanakirjat eivät osaa kovin hyvin luoda uusia yhdyssanoja tai sanaliittoja. Tavallisiin yksikielisiin sanakirjoihin ei ole tapana kirjata sellaisia yhdyssanoja, jotka ihminen pystyy päättelemään, esimerkiksi *keltaoranssi* tai *sohvaostos*. Konekäännösohjelmalle tarkoitettussa sanakirjassa kaikki vaihtoehdot tulisi luetella, sillä kuten seuraava esimerkki osoittaa, konekäännösohjelma epäonnistuu sanaliittojen tunnistamisessa ja kääntämisessä. Kyseessä on TransAction-käännösohjelmalla tehty käännös englannista saksaan (Schmid 1991: 134):

Healthy living by the numbers: The vaunted benefits of the **eat-right, stay-fit movement** have always been slippery when applied to individuals.

Die gepriesenen Vorteile **des essen-richtig, sind-bleiben-fite Bewegung** immer heikel wenn gewesen auf Individuen angewandt.

Esimerkin käännöksessä on muitakin virheitä, mutta uusien sanaliittojen kääntäminen toiselle kielelle on epäonnistunut.

Elektronisten sanakirjojen avulla käännösohjelma tekee sanasta-sanaan-käännöksen, jossa sanajärjestys on useammin väärä kuin oikea (Hutchins 1986: 41). Tällaisia alkeellisia ohjelmia ovat tosin enää kaupalliset ilmaisohjelmat, jotka käyttävät suoran kääntämisen menetelmää. Tästä kerron tarkemmin kappaleessa 5.1.

Joskus teksteissä on käytetty sekaisin kahden tai useamman kielen sanoja. Jos vaikkapa ruotsinkielisen tekstin kirjoittaja ei muista tai keksi sopivaa ruotsinkielistä ilmausta, hän saattaa käyttää esimerkiksi englannin-, saksan- tai ranskankielistä vastinetta. Saattaa olla myös niin, että kyseistä termiä ei ruotsin kielessä ole olemassa. Tällöin konekäännösohjelmalle pitäisi olla annettu useamman kielen sanakirjat käyttöön, mikä taas kuormittaa muistikapasiteettia ja usein hidastaa ohjelman käyttämistä.

Sanakirjojen yleinen ongelma on se, että luonnollisissa kielissä ei aina ole täydellisiä vastineita sanoille. Lähtökielen sanalla saattaa olla kohdekielessä monta käännösvaihtoehtoa. Tällöin käännösohjelma joko antaa kaikki vaihtoehdot tai koettaa valikoida niistä sopivimman

kyseiseen kohtaan. (Hutchins 1986: 41–42.) Tarkastelen moniselitteisiä ilmauksia tarkemmin kappaleessa 7.2.

## 4. Miksi konekääntämistä?

Konekääntäminen herättää paljon mielipiteitä puolesta ja vastaan. Teksasin yliopiston lingvistiikan lehtori Jonathan Slocum kirjoitti jo vuonna 1984, että tuskin kukaan voi pysyä mielipiteissään täysin neutraalina kun kyse on konekääntämisestä. Useimmilla asiaan perehtyneillä ihmisillä – varsinkin kääntäjillä – on mielipide, jota he puolustavat. Melbyn (1995) mukaan toiset uskovat, että ihmisen ja tietokoneen välillä ei ole mitään perustavaa eroa: He ovat sitä mieltä, että tulevaisuudessa konekäännösten laatu tulee saavuttamaan ihmisen tekemien käännösten laadun. Tietokone osaa tehdä laskutoimituksia nopeammin ja tarkemmin kuin ihminen, ja matematiikka on monille ihmisille vaikeampaa kuin lingvistiikka. Lisäksi heidän argumenttinaan on, että ihmisaivot ovat tietynlainen tietokone. Näiden ihmisten mukaan on vain ajan kysymys, milloin tullaan kehittämään uudenlainen tietokone, joka toimii kuin ihmisaivot, mutta nopeammin ja paremmin, ja joka ylittää ihmisen suorituskyvyt kielentuottamisen alueella. Tekniikan lisensiaatti Risto Kaivola (2006) kirjoittaa Tiede-lehdessä, että tietokoneet voivat hyvinkin alkaa pärjätä aivoille, jos numerolaskentaan ja valmiiseen ohjelmakoodiin perustuvan tekoälyn sijaan saadaan tietokone itse oppimaan ja ymmärtämään.

Toiset ovat vastakkaisella kannalla: Heidän mukaansa ihmisaivot ja tietokone toimivat niin eri tavoin, ettei tietokone milloinkaan tule saavuttamaan ihmisen käännöstaitoja (Melby 1995). Ihmisen aivot ovat paljon monimutkaisempi ja hienompi kuin parhainkaan tietokone, ja ajatus siitä, että tietokone voisi oppia ajattelemaan samalla tavalla kuin ihminen, on pelottava ja kauhistuttava. Perusteena voidaan myös pitää sitä tosiasiaa, että tietokoneen piirikytkentä ei ole aivojen kaltainen systeemi (Kaivola 2006). Toki tietokoneet ovat tänä päivänä hyvinkin älykkäitä, mutta niiden älyä ei voi verrata ihmisen älyyn. Kääntäminen vaatii korkeatasoista kokonaisuuden hallintaa, jota kone ei voi saavuttaa lyhyessä ajassa.

Konekäännösohjelmien interaktiiviset vaiheet yrittävät kiertää ongelmat käyttämällä hyväksi ihmisen tietoa ja ymmärrystä. Tällaiset vaiheet kuitenkin hidastavat kääntämistä. Voidaan

myös pohtia, onko kyseessä enää automaattinen konekääntäminen, jos ihminen osallistuu käännösprosessiin ja valitsee käännösvaihtoehdot.

John Hutchinsin (2005a: 5–6) mukaan on olemassa ainakin kuusi syytä, joiden takia konekäännöksiä tulisi käyttää:

1. Käännettäviä tekstejä on aivan liikaa.
2. Tekniset tekstit ovat ihmiskääntäjille liian tylsiä.
3. Terminologian yhdenmukaisuus on taattava.
4. Konekääntäminen mahdollistaa paljon pidempien tekstien kääntämisen paljon nopeammin kuin mihin ihmiskääntäjät pystyvät.
5. Ihmiskääntäjien tuottamaan huippulaatua ei aina tarvita.
6. Yritykset haluavat laskea käännösten kustannuksia.

Käsittelen seuraavaksi näitä kohtia yksitellen hieman tarkemmin.

#### **4.1 Liikaa käännettäviä tekstejä**

Ihmiskääntäjiä ei ole tarpeeksi. Internetin kautta ihmiset törmäävät vieraskielisiin teksteihin enemmän kuin koskaan ennen. He seuraavat linkkejä kiinnostaville sivuille ja löytävät tekstiä, joiden kieltä he eivät ymmärrä. Sivujen sisällön ymmärtämiseksi on kehitetty online-käännösohjelmia, kuten Babel Fish. Tällaiset ilmaisohjelmat eivät kuitenkaan ole ammattimaiselle kääntäjälle tarpeeksi laadukkaita, vaikka ne maallikolle kelpaavatkin. Myös teollisuudessa ja yritysmaailmassa käännettäviä tekstejä on aiempaa enemmän. Koneet kääntävät tällä hetkellä jo määrällisesti enemmän tekstejä kuin ihminen (Lisa 2004). Globalisoitumisen myötä maailma ”kutistuu” kovaa vauhtia: englannista tulee yhä vallitsevampi kieli ja yhä useammat kansainväliset yritykset käyttävät ainoastaan englantia myös sisäisessä dokumentoinnissa ja viestinnässä. Englannista tulee universaali käyttökieli, uusi *lingua franca*. Englannin yleistymisen voi toisaalta osaltaan vähentää käännettävien tekstien määrää, kun kaikkien oletetaan ymmärtävän englanninkieliset lähtötekstit.

## **4.2      *Tekniset tekstit ovat tylsiä***

Kääntäjät eivät yleensä pidä teknisten tekstien kääntämisestä, ja siksi avuksi tarvitaan konekäännösohjelmiä. Realistisimmillaan konekääntäminen tehostaa kääntämisen toistuvia rutiineja. Melbyn (1995) mukaan etenkin teknisen tekstin kääntäminen voi onnistua hyvin käännösohjelmalla, jos se käsittelee vain suppeaa aihealuetta ja on vivahteeton. Yleiskielinen teksti voi olla kiinnostavampi ihmiskääntäjälle ja usein yleiskielisistä konekäännöksistä tulee verrattain huonoja. (Emt.) Sen sijaan ammattitaitoinen ihmiskääntäjä voi käyttää mielikuvitustaan ja luovuuttaan ja saada aikaan todella hyviä käännöksiä.

Tekniseksi tekstiksi voidaan määritellä sellainen teksti, joka käsittelee teknistä sisältöä, eli puhuttaessa teknisestä laitteesta tai teknisestä puheenaiheesta yleensä (Anderson et al. 1983; Andrews 1996). Toiset taas katsovat, että kaikki teknisessä tai teollisessa ympäristössä esiintyvä diskurssi on teknistä kieltä (Ulijn 1996). Kolmas määrittelyyn vaikuttava seikka on rajaus: tekniikan diskurssi menee osittain päällekkäin varsinkin kaupan kielen kanssa (Yli-Jokipii 2004: 82). Vaikka tekniikan kielen määrittely onkin vaikeaa, ollaan Yli-Jokipiin (mts.) mukaan kohtalaisen yksimielisiä siitä, että tekniikan kieli on ammattikieltä ja erikoiskieltä. Ammattikieltä käytettäessä vain toinen saman ammatin edustaja saattaa täysin ymmärtää, mistä toinen puhuu tai kirjoittaa. Erikoiskieli on erotettu yleiskielestä. (mts.)

Konekäännösohjelmien tekemien käännösten yksitoikkoisen loogisuuden vuoksi konekääntämistä ei tule käyttää kaunokirjallisuuden, lyriikan, näytelmäkäsikirjoitusten, puheiden tai muiden yleiskielisten tekstien kääntämiseen. Aina tulee siis olemaan tekstejä, joiden kääntämiseen kykenee vain ihminen. On kuitenkin mahdollista, että konekäännösohjelman käyttäjä ei tunne käännettävän tekstin tekstilajia, ja käännättää vaikkapa kaunokirjallisen tekstin koneella halutessaan tietää sen asiasisällön (Nuutila 2005: 19).

### **4.3 Terminologian yhdenmukaisuus**

Ihmiset käyttävät vaihtelun vuoksi tai joskus huomaamattaankin synonyymeja. Varsinkin jos pitkä teksti on jaettu osiin, joita kääntävät useat eri kääntäjät, on varmaa, etteivät kaikki käytä samoja käännösvastineita. Ihmiskääntäjällä on usein tarve luoda vaihtelua yksitoikkisiin teksteihin. Teknisissä teksteissä terminologian yhdenmukaisuus on kuitenkin erittäin tärkeää. Koneet ovat luotettavampia kuin ihmiset, koska ne eivät käytä samasta termistä erilaisia käännöksiä. Kone osaa säilyttää terminologian yhdenmukaisena ja se osaa kääntää tekstejä, joissa toistuvat samat, ennalta tutut asiat. Käännösohjelma myös rakentaa lähtökielen rakenteiden mukaisen käännöksen, jota ihminen korjaa enemmän kohdekielen mukaiseksi. (Melby 1995.)

### **4.4 Pidempiä tekstejä nopeammin**

Yritykset haluavat saada käännökset heti, mieluiten samana tai viimeistään seuraavana päivänä ja tällöin ne ovat valmiita myös tinkimään käännöksen laadusta. Tutkimustiedot osoittavat, että konekääntäminen säästää aikaa noin 50–70 % verrattuna ihmisen tarvitsemaan aikaan (Sågvall Hein 2004: 8). Varsinkin laajojen tekstikokonaisuuksien kääntämisessä konekäännösohjelmasta on tuntuva hyötyä. Ihmiskääntäjä käyttää käännösprosessin aikana yleensä noin kolmanneksen kaikesta tarvitsemastaan ajasta sopivien käännösvastineiden etsimiseen (Melby 1995). Haettua sanaa ei aina löydy sanakirjasta tai tietokannasta. Tällaisessa tapauksessa tietokone loistaa nopeudellaan. Toisaalta koneen nopeudella ei ole niin suurta merkitystä, kun se voi tehdä työtä vaikka kellon ympäri eikä se tarvitse lepoa. Jos käännöksellä ei ole kiire, koneen voi jättää kääntämään vaikka viikonlopuksi, mihin ihminen tuskin suostuisi. Tällöin ei ole väliä kestääkö kääntäminen tunnin vai 48 tuntia.

Tietokoneilla on paljon parempi muisti kuin ihmisellä, mutta ne ovat huonoja tekemään päätöksiä kahden tai useamman vaihtoehdon välillä (mts.).

#### **4.5        *Huippulaatua ei aina tarvita***

Usein käännöksen ei tarvitse olla täydellinen täyttääkseen tehtävänsä. Raakakäännöstä ei ole tarkoitukseen julkaista. Monet ihmiset ovat sitä mieltä, että koska koneen tekemät käännökset eivät ole laadukkaita, ne ovat hyödyttömiä, mutta näin ei aina ole. Kun on kyse ihmisen tekemästä käännöksestä, ovat laatuvaatimukset korkealla. Yleensä odotetaan, että käännöksen laatu vastaa lähtötekstin laatua, ja lisäksi käännöksen ja lähtötekstin informaatioisisällön on vastattava toisiaan. Sen sijaan konekäännöksistä puhuttaessa ovat laatukriteerit hajanaisempia, koska koneelta ei odoteta vastaavaa älykkyyttä eikä luovuutta.

Anna Sågvall Hein (2004: 8) jakaa käännösten laadun kolmeen kategoriaan: julkaisulaatuun (*publiceringskvalitet*), ymmärrettävyyksilaatuun (*begriplighetskvalitet*) ja muokkauslaatuun (*redigeringskvalitet*). Julkaisulaatu tarkoittaa valmista, loppuun asti hiottua tekstiä, joka voidaan julkaista sellaisenaan. Konekäännösohjelmat eivät tällä hetkellä kuitenkaan kykene tuottamaan tämän tason tekstiä. Yksi konekäännösten tavoitteista on saada raakakäännös, joka antaa pääpiirteittäisen käsityksen lähtötekstin sisällöstä ja informaatiosta. Tällaisessa tapauksessa voidaan jättää huomiotta kielelliset heikkoudet, kuten kongruenssivirheet, heikko artikkelien käyttö, tai virheellinen sanajärjestys. Jos koneen tekemässä käännöksessä ei ole muita virheitä, kannattaa muokata käännös manuaalisesti. Tällöin Sågvall Hein (mts.) puhuu muokkauslaadusta.

Kulo (1991: 75) huomauttaakin, että konekäännöksiä arvioitaessa pitää ottaa huomioon ennen kaikkea niiden kohderyhmä ja käyttötarkoitus. Koneen tekemä käännös saattaa olla riittävän hyvä ja ymmärrettävä tietyn alan asiantuntijalle, vaikka käännös ei maallikon mielestä kovin tasokkaalta näyttäisikään. Mikäli käännös on riittävän hyvä toimimaan esimerkiksi tiedon lähteenä, voidaan konekääntämistä pitää tarkoituksenmukaisena. (Mts.) Voidaan myös

ajatella, että koneen tekemän raakakäännöksen perusteella käyttäjä päättää, haluaako hän käännättää tekstin ihmiskääntäjällä.

Konekäännösten laatua arvioidaan usein kielioppi- tai sanajärjestysvirheiden määrän perustella, mutta huomattavasti vaikeampaa on punnita sen luettavuutta tai uskollisuutta lähtötekstille (Kulo 1991: 76). Konekäännökset ovat selkeitä käänösvirheitä lukuun ottamatta ainakin uskollisia lähtötekstille. Täytyy myös ottaa huomioon se tosiseikka, että kukaan ei pysty sanomaan, mikä käänös on ”oikein” ja mikä ”väärin” – laadun kriteerit ovat aina riippuvaisia käänöksen käyttökontekstista ja jokainen käänös on jonkinlainen lähtötekstin tulkinta.

#### **4.6 Yritykset haluavat laskea käänösten kustannuksia**

Konekäännökset tulevat kyseeseen, koska ne ovat halvempia kuin ihmisten tekemät käänökset. Konekääntämisen kustannukset koostuvat alkuinvestoinneista, järjestelmän käyttö- ja ylläpitokustannuksista sekä editointikustannuksista (Kulo 1991: 80). Joskus alhaiset kustannukset ovat tärkeämpiä kuin käänöksen huippulaatu.



## **5. Konekääntämisen menetelmiä**

Konekääntämisessä on käytössä erilaisia menetelmiä, jotka voidaan jakaa karkeasti suoraan kääntämiseen (Direct Translation System), siirtomenetelmään (Transfer Machine Translation) ja interlingvaaliseen rakenteeseen (Interlingual Machine Translation).

Käsittelen kutakin menetelmää seuraavaksi tarkemmin.

### **5.1 Suora kääntäminen**

Käännöskoneiden ensimmäisen sukupolven ohjelmat toimivat suoran kääntämisen menetelmän periaatteella: käännösprosessissa ei ole välivaiheita, vaan lähtökielinen sana korvataan kohdekielen vastineella. Suoran kääntämisen menetelmä perustuu kaksikieliseen sanakirjaan ja kääntäminen tapahtuu mekaanisesti sana sanalta tai fraasi fraasilta. (Arppe 1994). Tällaisen systeemin rajoituksena on yksisuuntaisuus; se ei voi sisältää useita kieliä, sillä sanakirjat ja kielioppisäännöt perustuvat vain ennalta ohjelmoituihin kieliin (Austermühl 2001: 159). Suora kääntäminen on konekääntämisen yksinkertaisin ja vanhin menetelmä, mutta tästä menetelmästä on kuitenkin käytännössä jo luovuttu (Arppe 1994). Suora käännösstrategia osoittautui käytännössä riittämättömäksi, koska se ei käytä lingvistisiä teorioita lähtökielen analysoimiseen (Häyrynen 1992: 34).

### **5.2 Siirtomenetelmä**

Siirto- eli transformaatiopohjainen menetelmä käyttää hyväkseen tietoja kielten välisistä eroista. Siirtomallissa käännös tapahtuu kolmessa vaiheessa: ensin ohjelma analysoi lähtötekstin lauseenjäsennyksen rakenteen, seuraavaksi se siirtää lähtö- ja kohdekielten lauseenjäsennyksen rakenteen, ja viimeisenä vaiheena tuottaa kohdekielen vastaavan rakenteen. Varsinainen kääntäminen tapahtuu keskimmaisessä vaiheessa. (Sågvall Hein

2004: 9.) Siirtomenetelmä on kieliparikohtainen, sillä siirrot on suunniteltava jokaisen kieliparin kohdalla erikseen (Arppe 1994).

### **5.3 Interlingvaalinen menetelmä**

Siirtomenetelmän vaihtoehtona voidaan pitää interlingvaalista menetelmää. Kieli analysoidaan täydellisesti yleiselle, kielestä riippumattomalle universaalille välikielelle, eli niin sanotulle interlingualle. Välikieli ilmaisee lähtötekstin merkityksen, ja siitä voidaan tuottaa mikä tahansa kohdekieli (Arppe 1994). Interlinguan eduksi voitaneen laskea sen käytettävyys periaatteessa minkä tahansa kieliparin kanssa. Austermühlin (2001: 162) mukaan on kuitenkin käytännössä osoittautunut vaikeaksi luoda kaikille kielille sopiva malli.

### **5.4 Käännösäly**

Tietokonelingvistiikka tutkii luonnollisen kielen rakennetta ja toimintaa tietojenkäsittelyn menetelmien ja teorioiden avulla. Tekoälyä on tutkittu noin 55 vuotta, ja tavoitteena on ollut saada kone toimimaan ihmisen tavoin, siis älykkäästi, sekä pohtia ihmisen ja tietokoneen eroja ja yhtäläisyyksiä. Tietokone luonnollisen kielen käyttäjänä on ollut tekoälytutkimuksen keskeisiä kohteita alusta lähtien. (Seppänen 2001: 3.) On tutkittu, miten kone käsittelee kieltä ja tulkitsee kirjoitettua tekstiä.

Fred Karlsson (1993) jakaa tekoälyn käyttämisen luonnollisen kielen käsittelyssä viiteen eri vaiheeseen: morfologiseen analyysiin, morfologisten moniselitteisyyksien ratkomiseen, lauseiden syntaktiseen jäsennykseen, kirjaimellisen merkityksen tulkintaan sekä kontekstimerkitysten tulkintaan. Morfologisesta analyysistä kirjoitan tarkemmin luvussa 6.2.1 ja kontekstimerkitysten tulkinnasta eli semanttisesta analyysistä luvussa 6.2.2. Jos käännösälyä onnistuttaisiin kehittämään ja hyödyntämään konekääntämisessä tarpeeksi hyvin, voitaisiin olla todella lähellä laadukasta automaattista kääntämistä.

Kieliteknologiayritys Master's Innovations Ltd. Oy kehittää ohjelmatyökaluja, joilla voidaan hyödyntää käännösälyä erilaisten rutiininomaisten käännöstöiden tehostamiseksi. Käännösälyä pidetään käännösmuistiohjelmien seuraavan sukupolven edustajana. Toisin kuin konekäännös tai käännösmuisti, käännösälyn tietämuskantaa voi laajentaa ja se on itseoppiva. Tämän ansiosta järjestelmä tehostuu jatkuvasti ja sopeutuu käyttäjän kieleen. Perinteiset konekäännösjärjestelmät, joissa vain sanastoa voi kasvattaa, ovat paljon kankeampia. Masterin ohjelmistolla voidaan kääntää esimerkiksi vuosikertomuksia, lehdistötiedotteita, esitemateriaalia, teknisiä kuvauksia, käyttöohjeita ja muuta materiaalia suomen ja englannin kielten välillä. Ohjelma ei toimi automaattisesti, vaan se ehdottaa kääntäjälle vaihtoehtoja ja tarkkailee valintoja. (Grönroos 2006.)

Master's Innovations Ltd:n omassa julkaisussa annetaan käännösälystä seuraava esimerkki:

”Jos kääntäjä haluaa suomentaa 'the man is singing in the rain' ja antaa vastineeksi 'mies laulaa sateessa', järjestelmä oppii tämän, ja lisäksi tietämystään älykkäästi soveltaen se osaa sitten kääntää vaikkapa sellaisia lauseita kuin 'nainen kävelee puistossa' tai 'the bird is singing in the forest'. Vaikka järjestelmä tallensi vain yhden esimerkin, se itse asiassa oppi kääntämään yli 25 000 000 000 rakennetta molempiin suuntiin. Yltääkseen samaan perinteisen käännösmuistin pitäisi tallentaa ne kaikki.” (Master's Innovation 2006.)

## **5.5 Ongelmia**

Erään varhaisen, englannin ja venäjän välillä toimineen konekäännösohjelman väitetään muuntaneen lauseen ”*The spirit is willing but the flesh is weak*” muotoon ”*The vodka is good but the meat is lousy*”, kun se käännettiin ensin englannista venäjäksi ja sitten takaisin englanniksi. (Austermühl 2001: 153.) John Hutchins (1995) tosin epäilee tämän olevan paikkaansa pitämätön legenda, koska se on peräisin julkaisusta ajalta, jolloin ei vielä ollut olemassa englannista venäjään kääntävää konekäännösohjelmaa.

Anna Sångvall Heinin (2001) mukaan konekääntämiseen kohdennetaan usein epärealistisia odotuksia. Lisäksi koneet tekevät toistaiseksi huonoja käännöksiä ja muun työn oheen on vaikea integroida konekääntämistä. Käännösohjelmat tuottavat kömpelöä ja epäluontevaa

kieltä, joiden käyttö hidastaa ymmärtämistä. Konekääntäminen kuitenkin tuottaa käännöksiä nopeammin ja halvemmin kustannuksin kuin ihmisen tekemänä. Koneet tekevät myös teknisesti johdonmukaisempia ja täsmällisempiä käännöksiä. Saadaksemme parempia konekäännöksiä tarvitaan ihmisen ja koneen välistä vuorovaikutusta, parempia ohjelmia tutkimuksen ja kehityksen tuloksena, lähtötekstin tarkastelua, kohdetekstin laadun varmistamista ja realistisia odotuksia. (emt.)

Käännöstä ei voida tehdä ilman lähtötekstiä, eli käännettävää dokumenttia. Jotta käännösohjelma pystyisi kääntämään tekstin, sen on oltava koneelle ymmärrettävässä, siis sähköisessä muodossa. (Melby 2001.) Tämä ei nykyään liene suurikaan ongelma, sillä useimmiten tekstit ovat saatavilla sähköisessä muodossa.

Joissakin tapauksissa virheiden korjaaminen voi tuntua vähäpätöiseltä, mutta jälkieditointi on silti tehtävä. Jos esimerkiksi lähtötekstissä on kirjoitusvirhe, mikä voi tulla kyseeseen, vaikka kieliasun tarkistus olisikin tehty sekä manuaalisesti että automaattisesti, käännösohjelma ei tunnista sanaa, mikä saattaa vaikuttaa loppuvirkkeen kääntämiseen. Pienikin kirjoitusvirhe — esimerkiksi sanat *from* ja *form*, jotka molemmat ovat ihan tavanomaisia englanninkielisiä sanoja — aiheuttaa ongelmia käännöksen tekemiseen. Myös puutteellinen välimerkkien käyttö vaikeuttaa kääntämistä. (Hutchins 2005a: 12–13.)

## 6. Konekäännöksen käänösprosessi

Kielen monimutkaisuuden takia ja tämänhetkisen teknisen kehityksen tuloksena ei ole vielä onnistuttu kehittämään täysin automaattista konekäännösohjelmaa, joka tuottaisi julkaisuvalmista tekstiä ilman ihmisen puuttumista käänösprosessiin.

Ennen käännettävän tekstin syöttämistä koneelle se on syytä käydä läpi huolellisesti ja muokata se käänösohjelmalle ymmärrettävään muotoon. Myös kääntämisen aikana saattaa olla interaktiivisia vaiheita, jos ohjelma ei osaa kääntää jotain kohtaa ja pysähtyy odottamaan ihmisen ratkaisua. Tällöin käyttäjän eli ihmisen täytyy ratkaista käänösongelma. Kääntämisen jälkeen teksti on muokattava julkaisuvalmiiseen kuntoon, sillä kone tekee vain raakakäännöksen.

Seuraavaksi tarkastelen yksittäisiä vaiheita tarkemmin.

### 6.1 *Lähtötekstin editointi*

Elektroninen sanakirja on päivitettävä ennen kuin lähtöteksti syötetään koneelle. Se saattaa sisältää virheitä ja poikkeavuuksia, jotka on korjattava, jotteivät samat virheet toistuisi jatkossa. (Austermühl 2001: 163–165.)

Lähtöteksti täytyy käydä läpi huolellisesti ja se on muokattava käänösohjelmalle ymmärrettävään muotoon muun muassa poistamalla idiomeja ja ellipsejä ja yksinkertaistamalla lauserakenteita. Tekstin yksinkertaistamisella tarkoitetaan luettavaa, mutta yksityiskohtaista tekstiä. Se on tarkastettava vähintäänkin kielioppivirheiden poistamiseksi, sillä järjestelmät eivät pysty käsittelemään huonoa kieltä (Arppe 1994). Käänösohjelmalle aiheutuu huomattavasti ongelmia, jos tekstissä on esimerkiksi väärin kirjoitettu tai tuntematon lyhenne, kuten *te. x.*, kun pitäisi olla *t. ex.* (Knutsson 2002: 29–30). Ohjelmalle voidaan mahdollisesti myös kertoa, mihin sanaluokkaan tietyt sanat kuuluvat, esimerkiksi onko *convict* tässä

kohdassa verbi vai substantiivi. Käännettävä teksti on rajattava, sillä usein kaikkea ei ole tarpeen kääntää, esimerkiksi osoitteita ei käännetä. Voidaan osoittaa tarvitseeko nimiä kääntää, jottei *John White* ole käänöksessä *Johann Weiss* tai ettei *Dear Bill* muutu *Liebe Rechnung*iksi. Myös yhdyssanoja voidaan merkitä ohjelmalle valmiiksi. Välimerkkejä voidaan lisätä, jotta ohjelma ymmärtää lauseiden viittaussuhteita, esimerkiksi lauseeseen *There are he says two options* voidaan lisätä pilkut seuraavasti: *There are, he says, two options*. (Hutchins 2005a: 15–16.) Pitkät virkkeet voidaan myös pilkkoa lyhyemmiksi. Lopuksi teksti on vielä luettava läpi ja tarkastettava, että kaikki lauseet ovat täydellisiä eikä sanoja puutu. (Austermühl 2001: 163–165.) Teoriassa ei ole tarpeellista tuntea kohdekieltä, mutta käytännössä on hyödyllistä tietää, että tietyissä kielissä tietyt sanonnat aiheuttavat ongelmia. (Hutchins 2005a: 15–16.) Parhaiten käännösohjelmalle soveltuu yksiselitteinen ja huolellisesti kirjoitettu teksti, jonka kielioppi- ja kirjoitusvirheet on tarkistettu, ja jonka sisältö on selkeästi esitetty. Nuutila (2005: 70) muistuttaa, että lähtötekstin editointia voidaan vähentää ja helpottaa, jos jo tekstin kirjoitusvaiheessa tiedetään, että se tullaan kääntämään joko automaattisesti tai ihmisvoimin.

## **6.2 Lähtötekstin analyysi**

Kun ihminen saa käännettäväkseen tekstin, hän ensin lukee sen läpi ja yrittää ymmärtää mitä siinä sanotaan, mikä on tekstin viesti. Tietokone ei osaa ajatella mitä se on tekemässä. Se ei ymmärrä tekstiä eikä kykene ottamaan huomioon kohdeyleisöä, eikä se tiedä mitään käännöksen funktiosta tai viestintätilanteesta. Lähtötekstin tyyli ja tekstilaji jäävät konekäännösohjelmalta havaitsematta. Se ei ymmärrä edes sanojen merkitystä. Kehittynyt konekäännösohjelma voi kuitenkin osata analysoida kielen rakennetta, sanaluokkia ja taivutusmuotoja.

Häkkisen (1998: 16) mukaan kielen kuvaamiseen tarvitaan sekä teoriaa että malleja. Teorian perusteella voidaan muodostaa malli, johon luonnollisesta kielestä saatu materiaali lokeroidaan. Konkreettisimmillaan sanat järjestellään mekaanisesti tiettyjen tuntomerkkien perusteella, mutta malli voi myös olla monitasoinen systeemi, jonka sisäisiä suhteita on kuvattava hyvinkin abstraktien sääntöjen avulla. Lauseita ja virkkeitä on monenlaisia, eivätkä niiden keskinäisiä suhteita kuvaavat mallitkaan voi olla kovin konkreettisia. (Mts.)

Konekäännösohjelma tarvitsee malleja analysoimaan lähtötekstiä. Sen pitäisi osata lokeroida sanat tiettyihin kategorioihin, esimerkiksi sanaluokkiin, osatakseen kääntää ne oikein.

### 6.2.1 Morfologinen analyysi

Voidakseen määritellä käännettävän tekstin sanojen sanaluokat, on käännösohjelman ensin tehtävä morfologinen analyysi. Morfologinen analyysi on välttämätön osa käännösprosessia (Häyrynen 1992: 36). Ensimmäinen askel on lähtökielisen tekstin sanojen sanaluokkien määrittäminen. Tämä on suhteellisen helppoa englannin kielessä ja useimmissa eurooppalaisissa kielissä, joissa sanat erotetaan toisistaan välilyönneillä, toisin kuin esimerkiksi kiinan tai thain kielissä.

Morfologinen analyysi tarkoittaa sanojen palauttamista perusmuotoon ja taivutusmuotojen tietojen ilmoittaminen. Aluksi erotetaan sanojen vartalot ja suffiksit, minkä jälkeen kone tekee kieliopillisen luokittelun. Suomen kielessä yhdellä sanalla voi olla jopa satoja muotoja, joten perusmuotoon palauttaminen on tarpeen piirteiden vähentämiseksi. Esimerkiksi sananmuoto ”tulisi” on ”tulla”-verbin konditionaalin preesensin 3. persoonan muoto (”hän tul+isi”) tai sen kieltomuoto (”hän ei tul+isi”).

Häyrynen (1992: 37) mainitsee ongelmaksi sen, että useimmat sanojen päätteet ovat monimerkityksisiä, esimerkiksi englannin kielessä -s voi tarkoittaa yksikön kolmannen persoonan verbin preesensmuotoa tai substantiivin monikkomuotoa (esimerkiksi *analyses*). Koneen pitäisi myös osata ottaa huomioon epäsäännölliset taivutusmuodot sekä erilaiset johdannaismuodot prefiksien ja suffiksien avulla (mts).

Kaisa Häkkisen (1998: 28) mukaan Kimmo Koskenniemi (1983) esitti väitöskirjassaan jo vuonna 1983 tietokonejärjestelmän kielen morfologian analysoimiseen. Koskenniemen kaksitasomalli (*two-level model*, TWOL) kykenee antamaan kieliopillisen kuvauksen suomen – tai periaatteessa minkä tahansa kielen – sanojen kaikista taivutusmuodoista pelkästään muotoon itseensä sisältyvien rakenteellisten vihjeiden perusteella. Järjestelmä ehdottaa

monitulkintaisille muodoille vaihtoehtoista tulkintaa, joista lauseyhteyden antamien vihjeiden perusteella voidaan valita todennäköisimmin oikea. Mallia on sovellettu jo noin 30 kielen kuvaukseen. (Häkkinen 1998: 28.)

### **6.2.2 Syntaktinen analyysi**

Syntaktinen analyysi tarkoittaa lauseopillista analyysia. Morfologisesti analysoiduista lauseista tehdään syntaktinen analyysi eli jäsenitys, jossa määritellään kunkin sananmuodon syntaktinen tehtävä lauseessa. Syntaksi kuvaa sitä, miten kielen rakenteen mukaisia lauseita muodostetaan. Konekääntämisen analyysiprosessissa lauseen syntaktinen analyysi on vaativampaa kuin morfologinen analyysi.

Hutchins (1986: 45) jakaa syntaktisen analyysin tavoitteet kolmeen:

1. Sanaluokkien järjestys
2. Sanaluokkien ryhmittely
3. Riippuvuussuhteiden tunnistaminen

Syntaktisessa analyysissa joukko sanoja analysoidaan käyttäen hyväksi käsiteltävän kielen kielioppia. Analysoinnin tuloksena saadaan lauseen rakenne. Termi ”semanttinen analyysi” ei varsinaisesti tähän yhteyteen sovi, sillä analyysia ei voi tehdä ilman kontekstia, ja kääntäminen tapahtuu aina kontekstissa.

### **6.2.3 Semanttinen analyysi**

Semantiikka eli merkitysoppi kuvaa lauseiden merkitystä, toisin sanoen niihin liittyvää abstraktia merkitystä ilman kontekstia (Hiirikoski 2004: 32). Kun pelkistä sanoista ja lauseista edetään merkitykseen, kielen automaattinen käsittely vaikeutuu huomattavasti. Ratkaisemalla



lauseen rakenne ja siihen sisältyvien yksittäisten sanojen merkitys saadaan selville osittainen kuvaus koko lauseen merkityksestä.

Semanttinen analyysi jaetaan yksittäisten sanojen ja kokonaisten lauseiden merkityksen analyysiin. Sanasemantiikka analysoi kielellisen käsitejärjestelmän rakennetta, kun taas lausesemantiikka tarkastelee sanojen yhdistelemisessä tapahtuvaa merkitysten muotoutumista. (Häkkinen 1998: 167.)

Semanttisen analyysin tekeminen automaattisesti on hyvin vaikeaa, ellei mahdotonta; merkitystä ei aina voi ymmärtää pelkkien sanojen perusteella, vaan tarvitaan yleismaailmallista tietämystä asioista. Koneella ei ole muuta tietoa kuin mitä sille on syötetty, joten sen on hyvin vaikea tehdä semanttista analyysia yksittäisistä käännettävistä sanoista tai kokonaisista lauseista.

### **6.3 Käännösstrategian valinta**

Ihmiskääntäjä valitsee käännösstrategian lähtö- ja kohdekielisten tekstilajien ja niiden erojen tuntemuksen perusteella (Vehmas-Lehto 2002: 41). Käännösstrategian valitseminen tarkoittaa koko käännettävään tekstiin suhtautumista ja on eri asia kuin yksittäiset käännösratkaisut. Käännösstrategia voi olla esimerkiksi kotouttava tai vieraannuttava. Vaikkapa kaunokirjallisia tekstejä käännettäessä tavoitteena voi olla alkuperäisen tekstin kieliasun ja tyylin jäljitteleminen, jolloin käänнос on vieraannuttava, kun taas kotouttava käänносstrategia palauttaisi lukija mieleen omasta kielestä ja ympäristöstä tuttuja asioita.

Konekäännösohjelma ei voi valita erilaisten käänносstrategioiden välillä. Se kääntää sana tai lause kerrallaan, eikä se osaa muokata lausejärjestyistä tai lisätä käännökseen sellaisia elementtejä, joita ei lähtötekstissä ole. Kone ei myöskään ymmärrä tekstiä kokonaisuutena. Se kääntää orjallisesti annetun sanakirjan ja kieliopin mukaan. Käännösohjelma ei ota huomioon kohdeyleisöä, se vaan noudattaa sääntöjä sokeasti (Melby 1995).

Käännösohjelmalla ei ole mahdollisuutta antaa kulttuurierojen vaikuttaa käännösratkaisuihin. Koneella ei ole kulttuuritaustaa, johon se voisi tukeutua (Hutchins 2005a: 34). Käännösohjelma ei voi myöskään käyttää muita apuvälineitä kuin sille syötettyä dataa. Ihmiskääntäjät usein turvautuvat asiantuntijoihin tai muihin ulkopuolisiin lähteisiin, kun he tarvitsevat lisätietoa. Koneella tätä mahdollisuutta ei ole. Koska kone ei voi valita erilaisten käännösmenetelmien välillä, sen pitäisi tuottaa kaikenlaisiin tarkoituksiin soveltuvaa tekstiä. Käytännössä kuitenkin käynee usein niin, että sellainen teksti ei sovellu mihinkään tarkoitukseen. Tämä on ongelma konekääntämisessä.

## **6.4        *Interaktiiviset vaiheet***

Annemarie Schmid (1991) on tutkimusraportissaan koonnut tietoja erään konekäännösohjelman interaktiivisista vaiheista. Kyseessä on TransActive-niminen käännösohjelma, joka vaatii käyttäjältä toimenpiteitä käännöksen aikana, eli se ei siis ole täysin automaattinen ohjelma. TransActive saattaa kysyä kääntämisen aikana käyttäjältä kysymyksiä, jotka voivat liittyä käännettävän tekstin semantiikkaan, syntaksiin tai morfologiaan. Käyttäjä, yleensä kääntäjä, voi joutua valitsemaan muun muassa seuraavanlaisien vaihtoehtojen välillä: käännösvaihtoehdot, elektronisen sanakirjan päivittäminen tai vaikkapa tietyn tekstipätkän kääntäminen manuaalisesti. (Schmid 1991: 118.)

Semanttinen interaktiivisuus tarkoittaa Schmidin (emt.) mukaan sitä, että elektroniseen sanakirjaan tallennetut käännösvaihtoehdot näkyvät tietokoneen näytöllä, ja kääntäjä valitsee niistä sopivimman. Näin konekäännösohjelma ei joudu itse tekemään päätöstä, joka voisi olla täysin väärä. Schmid (1991: 119) antaa kaksi esimerkkiä TransAction-ohjelman semanttisista interaktiivisista vaiheista. Kyseessä on englanninkielisen tekstin kääntäminen saksan kielelle.

Translate: to  
**a** zu  
**b** &&dative (dative case)  
**c** in  
**d** in (accusative case)  
**e** an  
**f** an (accusative case)  
**g** auf  
**h** auf (accusative case)

**i** nach  
**j** bis  
**k** in Richtung  
**l** für  
**m** mit

Translate: launch  
**a** schleudern  
**b** vom Stapel lassen (Naut)  
**c** ab.schießen  
**d** in Gang setzen

Näissä kahdessa esimerkissä ohjelma antaa käyttäjälle mahdollisuuden valita useammasta esimerkistä sen, joka vastaa parhaiten lähtötekstiä.

Käyttäjä voi joutua tekemään myös syntaktisia tai morfologisia päätöksiä, mutta Schmidin (emt.) mukaan semanttisia interaktiivisia vaiheita tulee vastaan eniten. Hän antaa kuitenkin esimerkkejä myös TransActiven syntaktisista ja morfologisista interaktiivisista vaiheista:

Is the text speaking of:  
**a** comprising the multiple telescopes  
**b** multiple telescopes which are comprising  
**c** multiple telescopes for comprising

Select a preferred form of translation: durch.führen  
**a** infinitive (...zu...)  
**b** infinitive (um...zu...)  
**c** phrase (zum...)  
**d** phrase (zur...)  
**e** clause (der/die/das ... können)  
**f** clause (der/die/das ... sollen)  
**g** clause (der/die/das ... dürfen)  
**h** clause (der/die/das ... müssen)  
**i** clause (der/die/das ... wollen)  
**j** clause (der/die/das ...)

Translate : 's  
**a** \*\* deferred section\*\*  
or translate : be  
**b** \*\* deferred section\*\*  
or translate: have  
**c** \*\* deferred section\*\*  
or translate: have.at  
**d** los.gehen auf (acc)  
**e** auf den Leib rücken  
**f** an.greifen

Ohjelma työstää käyttäjän valitseman vaihtoehdon kohdekieliseen käännökseen. Schmid (1991: 121) toteaa, että interaktiiviset vaiheet tekevät konekääntämisestä hitaampaa kuin pelkkä ihmiskääntäminen, koska jälkieditointi on interaktiivisista vaiheista huolimatta välttämätöntä. Hitaus riippuu tietysti tekstin pituudesta ja interaktiivisten vaiheiden määrästä. Näinkin monen vaihtoehdon läpikäyminen vaatii sekä lähtötekstiin että kohdetekstiin tutustumista, ja todennäköisesti näinkin irrallisten sanojen, kuten prepositioiden, tutkiminen vaatii paljon aikaa ja vaivaa.

## **6.5 Käännöksen jälkieditointi**

Yksikään tähän mennessä kehitelty konekäännös menetelmä ei vielä selviä julkaisuvalmiin tekstin tuottamisesta yksinään, vaan käännös vaatii ihmisen tekemää muokkausta. Ihminen joutuu korjaamaan koneen tekemän käännöksen virheet – jälkieditointi on Hutchinsin (2005a: 12) mukaan aina suositeltavaa. Koneen tekemät virheet ovat yleisimmin luonnollisen kielen ymmärtämisestä johtuvia ongelmia, mikä johtaa editoinnin ja tarkistuksen tarpeeseen. Hutchins (2005a: 15) mainitsee, että käännös ohjelmien haittapuolena on se, että kone toistaa johdonmukaisesti samoja, joskus vähäpätöisiäkin, virheitä.

Jälkieditoinnin määrä voi kuitenkin vaihdella huomattavasti. Mitä ”standardisoidumpia” tekstit ovat, toisin sanoen mitä enemmän ne sisältävät erikoisanalan sanastoa ja vakiintuneita sanontoja, mitä tavanomaisempia ja vähemmän omaperäisiä ne ovat, sitä tarkempi on koneen tekemän käännöksen tulos ja sitä vähemmän ihmisen täytyy tehdä korjauksia tekstiin jälkeenpäin. (Hutchins 2005a: 14.) Koneen kääntämissä teksteissä esiintyy vääriä taivutusmuotoja ja sanavalintoja, ja lisäksi käännöksessä saattaa olla aukkoja niiden sanojen kohdalla, joita ohjelma ei ole löytänyt sanakirjastaan. Jälkieditoinnin tekee kääntäjä. Tämä ei ollut konekäännösten pioneerien tarkoitus: heidän tavoitteenaan oli tehdä ohjelma, jonka lopputuloksen voi kuka tahansa kohdekieltä osaava henkilö tarkistaa. (Hutchins 2005a: 15.) Antti Arppe kirjoitti vuonna 1994: ”Korkealuokkaista, täysin automatisoitua konekäännöstä pidetään toistaiseksi mahdottomana saavuttaa”. Tulevaisuus näyttää edelleen, 13 vuotta myöhemmin, samalta.

Kaikkien näiden vaiheiden – lähtötekstin editoinnin, sen analysoinnin, käännösstrategian valinnan, interaktiivisten vaiheiden ja käännöksen jälkieditoinnin – jälkeen konekääntäminen kuulostaa perin työläältä. Ainakin lyhyemmän tekstin kohdalla ja ainakin käännösohjelman kehittelyvaiheessa ihmiskääntäjä luultavasti kääntäisi tekstin sujuvammin ja nopeammin ilman konekäännösohjelman apua. Kun tähän vielä lisätään luonnollisen kielen ymmärtämisen ongelmat, on kone auttamattomasti ihmisaivoja jäljessä. Tarkastelen näitä ongelmia tarkemmin seuraavassa kappaleessa.

## 7. Luonnollinen kieli ja sen ymmärtämisen ongelma

Konekäännösohjelmien käännösongelmat eivät aina johdu pelkästään teknisistä seikoista tai kehityksen asettamista rajoista, vaan usein niiden takana ovat luonnollisen kielen monimutkaisuudet. Luonnollinen kieli tarkoittaa ihmisten kieltä, eli mitä tahansa ihmisten käyttämää ilmaisuvoimaista kieltä, jota ei ole muodostettu keinotekoisesti (Lammi 2004). Kielen ymmärtäminen ei rakennu pelkästään yksittäisten sanojen ymmärtämisen tai kieliopin analysoinnin varaan, vaan sanojen ja ilmausten tulkinta edellyttää monipuolista sanojen välisten merkityssuhteiden ja myös epäsuorien merkitysten ymmärtämistä. Kielen käyttöön liittyvät käytännöt edellyttävät lisäksi paljon kielenulkoista käytännön tietoa ympäröivästä yhteiskunnasta. Kaisa Häkkinen (1998: 178) kirjoittaa lauseen muodollisen rakenteen ja merkityssisällöstä. Lause voi olla kieliopillisesti täysin korrekti, mutta semanttisesti vailla järkeä ja näin käyttökelvoton. Esimerkiksi hän antaa lauseen *Karvaiset vihamielisyydet rasvasivat hampaitaan*. Kieliopilliset suhteet voidaan määritellä, vaikka jokainen sana olisi vieras, kuten Häkkisen (mts.) seuraavassa esimerkissä: *Kuukerot latustavat resukkeita*.

Luonnollisen kielen kieliopin ja käännösohjelmalle syötetyn tiedon välillä on armottoman suuri kuilu. Ilman mitään kielioppisääntöjä kaikenlaiset sanojen yhdistelmät olisivat koneelle yhtä todennäköisiä. Tuottaakseen sujuvaa tekstiä koneen pitää tunnistaa sanaluokat ja lauseenjäsenet, osata taivuttaa ja johtaa sanoja, tunnistaa lauseiden perusrakenteet ja lauseiden väliset suhteet sekä hallita oikeinkirjoitussäännöt. Tällaisten tietojen syöttäminen koneen muistiin olisi kuitenkin melkoinen urakka.

Kielenkäytön sääntöjen lisäksi ohjelmalle on syötettävä suuria tekstimääriä lähtömateriaaliksi (Knutsson 2002: 29). Ihmiskääntäjä käyttää tietolähteinä yksi- tai monikielisiä sanakirjoja, jotka sisältävät monenlaista tietoa sanojen merkityksistä ja erilaisista käyttötilanteista. Sanakirjat eivät kuitenkaan sellaisinaan usein riitä, vaan ihmisellä on kyky löytää kontekstiin sopiva merkitys ja ymmärtää jopa huonoja selityksiä. Sanojen merkityssisällön hahmottaminen voi olla monivivahteinen ja vaikeasti määriteltävissä. Siihen vaikuttavat esimerkiksi niin sanotusti rivien välissä olevat tai taka-ajatuksina esitetyt vihjaavat merkitykset, muut

hienovaraiset sävyt kuten ironia ja sarkasmi, metaforat tai sanavalintojen lisämerkitykset (Tommola 2004: 10–11). Toisinaan ihminen vain tietää, miten joku sana pitää kääntää tietyssä yhteydessä, vaikka sille ei olisi mitään valmista sanakirjakäännöstä olemassa. Kääntäjä siis luottaa intuitioonsa. Joskus taas viestin kommunikatiivinen tarkoitus välittyy, vaikka lauseen rakenne olisi kieliopillisesti puutteellinen. Häkkinen (1998: 178) antaa esimerkiksi sanakirjan avulla puhuvan ulkomaalaisen.

Tietokone on täysin riippuvainen sille annetusta tiedosta, eikä osaa muokata sitä tilanteeseen sopivaksi. Vaikeutena on sanojen semantiikan kuvaaminen käännohjelmalle (Häyrynen 1999: 10). Tietokone ei voi ymmärtää sanojen merkitystä samalla tavalla kuin ihminen. Ongelma on siinä, että konekäännösohjelmalla ei ole äidinkieltä; sille syötetään luonnollisen kielen kielioppi ja sanastot, mutta se ei kykene oppimaan kielen vivahteita ja yksityiskohtia kuten ihminen (Melby 1995). Luonnollinen kieli on ääretön, ja sen opettaminen koneelle on mahdotonta. Jos ihmisaivojen toiminta olisi helppo kuvata, olisi automaattinen konekäännösohjelma varmaankin jo osattu rakentaa (Nuutila 2005: 51).

## **7.1 Kielen suhde ajatteluun ja todellisuuden hahmottamiseen**

Kielen käyttö on sosiaalinen kommunikaatiokeino: ihmisten välinen viestintä tapahtuu pitkälti kielen välityksellä, sillä se on ihmisen tärkein kommunikointiväline. Kielen ja ajattelun suhde on hyvin monimutkainen. Pelkkien sanojen ymmärtäminen ei riitä, vaan tarvitaan myös sosiaalinen ympäristö sekä kyky hahmottaa oma itsensä ja ympäröivä maailma. Vaikka tietokone pystyykin tunnistamaan eri sanaluokkien sanoja ja taivutusmuotoja, tällainen mekaaninen analyysi ei riitä lauseiden kokonaisuuden hahmottamiseen. (Seppänen 2001: 2.)

Jotta kone voisi osata ajatella tai käyttää luonnollista kieltä, sillä täytyisi olla muistoja sekä käsitys nykyhetkestä ja tulevaisuudesta. Sillä pitäisi olla tavoitteita ja toimintavaihtoehtoja niihin pääsemiseksi. Sen täytyisi osata etukäteen suunnitella, kuvitella ja ennakoida toimintansa mahdollisia lopputuloksia. (Kaivola 2006.) Koska kehityksen tässä vaiheessa kone ei voi ymmärtää ympäröivää maailmaa, se ei myöskään kykene rationaaliseen ajatteluun tai

kielen tuottamiseen. Tietokone ei osaa ajatella itsenäisesti vaan se tietää vain sen, mitä sille ohjelmoitu. Ihmisen kokemusmaailma mahdollistaa uusien asioiden oppimisen nopeasti. Kun eteen tulee tilanne, jollaista henkilö ei ole aikaisemmin kokenut, hän pyrkii toimimaan samalla tavalla kuin aikaisemmin kenties samankaltaisessa tilanteessa välttääkseen ongelman syntymisen tai ratkaistakseen sen. (Nuutila 2005: 18.) Koneella ei ole ihmisen kykyä käyttää semanttisten merkityksien ratkaisuun omaa elämäkokemustaan (Häyrynen 1999: 9).

Tietokoneen kannalta kielen moninaisuus ja tulkinnanvaraisuus ovat ongelmallisia. Se saattaa helposti vaikuttaa hyvin älykkäältä, kun se nykyään osaa tuottaa ymmärrettävää tekstiä. Ei kuitenkaan lähellekään niin monipuolista kuin ihminen. Strukturalismin perustaja Ferdinand de Saussure vertasi kielen merkitystä ja muotoa saman paperin kahteen eri puoleen: kumpaakin on mahdotonta tarkastella erikseen, mutta niitä on mahdotonta irrottaa toisistaan. Kieleen kuuluu sekä merkitys että muoto. (Häkkinen 1998: 186.)

## **7.2 Moniselitteiset ilmaukset**

Luonnollinen kieli sisältää paljon sanoja, joilla on useampi kuin yksi merkitys. Tällöin puhutaan sanojen merkityksen monitulkintaisuudesta. Monitulkintaisuudessa voi olla kyse joko homonymiasta tai polysemiasta. Homonymiassa kahdella (tai useammalla) sanalla on sama muoto, mutta eri merkitys. Polysemiassa yhdellä sanalla on useita merkityksiä.

### **7.2.1 Homonymia**

Homonyymeillä tarkoitetaan sanoja, joilla on sama asu mutta eri merkitys, esimerkiksi suomen kielen sana *kuusi*, joka tarkoittaa sekä puuta että lukusanaa. Jukka Korpela (2003) kirjoittaa, että homonymia voidaan jakaa tarkemmin täydelliseen homonymiaan, kirjoituksen homonymiaan ja puheen homonymiaan. Täydellisessä homonymiassa sanoilla on sama kirjoitus- ja äänneasu, mutta ne kuuluvat eri sanaluokkiin ja niillä on eri merkitys. Esimerkkinä täydellisestä homonymiasta voi olla vaikka edellä mainittu sana *kuusi*. Kirjoituksen



homonymiassa (homografiassa) sanoilla on sama kirjoitusasu mutta eri äänneasu ja merkitys, esimerkiksi englannin *lead* [liid] ”johtaa” ja *lead* [led] ”lyijy”. Puheen homonymiassa taas sanoilla on sama äänneasu mutta eri kirjoitusasu ja eri merkitys, esimerkiksi englannin *key* [kii] ”avain” ja *quay* [kii] ”laituri”. (Korpela 2003.) Puheen homonymia (homofonia) on konekääntämisen kannalta epäolennainen ongelma, koska konekääntäminen ei toistaiseksi koske puhuttua kieltä. Toki ihmiskääntäjäkään ei aina tiedä, milloin sanalla on kaksi tai useampi merkitys, ja joskus hänkin valitsee väärän käännösvaihtoehdon. Homonymiat ovat siksi vaikeita, että niitä kaikkia ei voida luetteloida ja syöttää käännösohjelman sanakirjaan. Homonymia pitäisi ensin tunnistaa ja sitten osata valita oikea käännös. (Melby 1995.) Monimerkityksiset sanat ovat aina riippuvaisia kontekstista.

## 7.2.2 Polysemia

Kahden eri kielen samoiksi miellettyjen sanojen välillä saattaa olla suuriakin merkityseroja. Polysemiassa sanojen tarkoitteet ovat usein jollain lailla samankaltaisia, ilmiö on syntynyt usein metaforan kautta. Polyseemit kuuluvat samaan sanaluokkaan, mutta niillä on useampia merkityksiä (Austermühl 2001: 171). Suomen kielessä monimerkityksisiä sanoja ovat esimerkiksi *kieli*, joka voi tarkoittaa ruumiinosaa, puhuttua kieltä, soittimen osaa tai kielekettä, ja *selkä*, jota käytetään tarkoittamaan ruumiinosaa, kirjan selkää, yön selkää tai järven selkää.

Austermühlin (emt.) mukaan konekääntämisessä polyseemit ovat relevantteja silloin, kun sanan merkitys vaihtuu käännöksessä. Käännösohjelmalle voidaan tietysti opettaa, että esimerkiksi englanninkieliseen verbiin *drink* täytyy liittyä elollinen olento, mutta on melko vaikeaa suunnitella kaikkiin tilanteisiin soveltuvia osia. Lingvistinen tieto ei aina riitä. Usein tarvitaan myös yleismaailmallista tietoa, jotta lause voidaan ymmärtää oikein. Esimerkiksi seuraava lause:

*”The man saw the horse with the telescope.”*

Tässä tapauksessa täytyy tietää, että hevosella ei voi olla kaukoputkea, vaan sen on oltava miehellä. (Austermühl 2001: 171–172.)

Homonymian ja polysemian erottaminen toisistaan ei ole aina mahdollista. Käytännössä on käännohjelman kannalta kuitenkin samantekevää, onko kyseessä homonyymi vai polyseemi, sillä molemmat aiheuttavat järjestelmälle saman ongelman. Monimerkityksisten sanojen kääntämisessä ratkaisevaa on luonnollisesti konteksti, jossa sana esiintyy.

### **7.3 Kielen rakenteelliset ongelmat**

Jotkut lauserakenteet ovat vaikeatajuisia äidinkieliselle lukijallekin, puhumattakaan jos teksti pitäisi kääntää vierasalle kielelle. Jos käännöksen tekee ihminen, hän varmaankin muokkaa lauseet kohdekieleen sopiviksi, mutta kone ei osaa muokata lähtötekstiä, ei edes omiin tarpeisiinsa.

Kielen rakenteen analysoiminen on koneelle vaikeaa. Rakenteen analysoimisessa sattuneet virheet ovatkin tyypillisimpiä, joita käännohjelmille sattuu. Jos lauseen analysointi epäonnistuu, ohjelma joutuu kääntämään lauseen sana sanalta ja kopioimaan lähtökielen rakenteen kohdekieleen. Tällöin syntyy kieliopillisesti virheellisiä käännöksiä. Austermühl (2001: 172) antaa esimerkiksi seuraavan lauseen:

- A) „Schalten Sie einen ProPrint 100 am Schalter, der sich am PC befindet, aus.“
- B) „Switch off a ProPrint 100 at the switch, which is situated at the PC, from.“

Tässä lauseessa upotettu relatiivilause häiritsee analyysiä merkittävästi, koska kone ei ymmärrä eriävän yhdysverbin prefiksin (aus) kuuluvan verbiin. Analyysin epäonnistuttua kone kääntää lauseen sanasta sanaan ja tulos on kieliopillisesti virheellinen. (emt.)

Ruotsinkielisen lauseen *Bilen tvättas också av mig* sanasanainen käännös olisi suurin piirtein *Auto pestään myös minulta*. Tätä ei kuitenkaan voida pitää suomenkielisenä lauseena. Lause on käännetty sanasta sanaan ja sillä on ruotsin kielen rakenne. (Palomäki 2004: 27.)

Muita lukemista hankaloittavia lauserakenteita ovat Terho Itkosen Uuden Kielioppaan (2000) mukaan muun muassa

- Partisiippiattribuutit, esimerkiksi ”Tästä velasta pankin tai muun rahalaitoksen kautta vallitsevan korkokannan mukaisesti vuoden loppuun mennessä maksetut korot”.
- Samasijaisuus, esimerkiksi ”Tällöin lisätään palkkaan sopimukseen perustuvat korot”.
- Kiilalauseet, esimerkiksi ”Summan, joka ylittää 10 000 mk, viemiseen maasta tarvitaan Suomen Pankin lupa”.
- Niin kutsuttu substantiivitauti, esimerkiksi ”Velkaantumisen lisäys erityisesti tänä vuonna on suuri menneeseen kehitykseen verrattuna ja ennusteen toteutuminen edellyttää, että ulkomaalaisten lainojen saatavuus ei aseta esteitä tällaiselle kehitykselle”. (Itkonen 2000: 87–93.)

Tämänkaltaiset tekstin ymmärtämistä vaikeuttavat lauserakenteet ovat iso ongelma kääntäjälle, varsinkin jos kääntäjä on kone. Ammattitaitoisen ihmiskääntäjän käsittelyssä teksti usein paranee, kun hän muokkaa käännöksen luettavampaan muotoon, mutta kone ei tällaista parannusta osaa tehdä.

## **7.4 Vertauskuvallinen kielenkäyttö**

Aina asioita ei sanota suoraan. Kiertoilmaisuja käytetään esimerkiksi tabujen ja muiden herkkien asioiden kaunistelemiseen. Rakenne ja merkitys ovat etäimmällä toisistaan silloin, kun kieltä käytetään metaforisesti. Kielen metaforisella käytöllä on ikivanhat juuret, ja tämä heijastuu monin tavoin kielellisessä kansanperinteessä. Metaforien käyttö johtaa usein pysyviin muutoksiin kielessä. (Häkkinen 1998: 181.)

Idiomi on vertauskuvallinen ilmaus, joka on vakiintunut metaforasta sanonnaksi. Idiomiä ei voi täysin ymmärtää pelkästään sanojen perusteella; esimerkiksi *saada rukkaset tai jäädä lehdellä soittelemaan*. Käännösohjelmien on erityisen vaikeaa käsitellä idiomeja ja siksi idiomeille tulisi olla sanakirjassa erillinen osasto ja niitä pitäisi käsitellä erillisinä yksikköinä. Tämä

toteutetaan usein lisäämällä niitä käännohjelman sanakirjaan sellaisinaan. Varmempaa ja nopeampaa olisi välttää idiomien käyttöä kokonaan. (Austermühl 2001: 172.)

Metonymia on puhekuvio, jossa varsinaisen käsitteen tilalle asetetaan toinen sen kanssa paikallisessa, ajallisessa, loogisessa tai muussa yhteydessä oleva (Korpela 2003).

Metonymioita ovat esimerkiksi

*Bussit* ovat lakossa. (bussien kuljettajat)

Hän lukee *Shakespearea*. (Shakespearen teoksia).

Metaforia ei tulkita kirjaimellisesti, vaan kielikuvina. Suomen kielessä voidaan esimerkiksi isokokoista, rotevaa miestä kutsua karhuksi tai kaapiksi.

Monet sanonnat on annettava käännohjelmalle manuaalisesti, mutta tällaisia sanontoja on niin paljon, että on mahdotonta syöttää kaikkia. Vakiintuneita sanontoja voivat olla esimerkiksi ihan arkipäiväiset sanonnat, kuten ”Here you are”, jota ei yleensä käännetä suomeksi *Tässä sinä olet*. Kielissä voi myös olla vakiintuneita, kielikohtaisia tapoja sanojen yhdistämiseksi. Nämä ovat niin sanottuja konventionaalisia syntagmoja, missä konventionaalisuus merkitsee sitä, että monien periaatteessa mahdollisten vaihtoehtojen joukosta valitaan jokin tietty. Esimerkiksi suomen kielessä sanotaan *parvi kaloja*, *lauma lampaita* ja *tokka poroja*, mutta ei sanota *parvi poroja*, *lauma kaloja* tai *tokka lampaita*. Suomalainen voi myös pestä hampaat ja ajaa pyörällä, mutta englanniksi hampaat harjataan (*brush your teeth*) ja pyörällä ratsastetaan (*ride a bike*). (Häkkinen 1998: 180.)

## **7.5 Kielen viittaussuhteet**

Verbistä riippuu koko lauseen rakenne. Verbi ilmaisee prosessia ja tilaa, sen avulla voidaan luontevasti ilmoittaa, mitä nimetty olio tekee tai missä tilassa hän/se on (Häkkinen 1998: 179).

Verbien ohella erityisesti pronomien viittaussuhteet aiheuttavat käännskoneille ongelmia.

Tämä on vaikea opettaa koneelle varsinkin sellaisissa kielissä, joissa pronomini ilmentää myös sanan sukua, kuten saksan kielessä. Esimerkiksi seuraavissa lauseissa pronomini ”it “ viittaa eri sanoihin:

”The monkey ate the banana because it was hungry.”  
“The monkey ate the banana because it was ripe.”

Ensimmäisessä lauseessa “it” viittaa sanaan *monkey*, ja toisessa lauseessa sama ”it” viittaa sanaan *banana*. Saksan kieleen käännettäessä näiden lauseiden kääntämisessä ensimmäisessä kohdassa ”it” pitäisi kääntää *er*, koska se viittaa sanaan *der Affe*, ja toisessa kohdassa *sie*, koska se viittaa sanaan *die Banane*. Tehdäkseen oikean ratkaisun käännskoneen pitäisi käyttää lingvististä tietoa, jota sillä ei kuitenkaan ole. Esimerkin ensimmäisessä kohdassa ohjelman pitäisi tietää, että sanan *hungry* täytyy viitata elävään olentoon. (Austermühl 2001: 173.) Häkkinen (1998: 179) kirjoittaa, että toisiin verbeihin voidaan luontevasti liittää objekti, toisiin ei.

Myös seuraavassa lauseessa pronominin ”it” viittaussuhde on ongelmallinen:

”Put the paper in the printer. Then switch it on.”

Ymmärtääkseen viittaussuhteet täytyy tietää, että tulostimen voi kytkeä päälle, mutta paperia ei. Tämän kaltaisen tiedon syöttäminen koneelle on erityisen vaikeaa. Nykytekniikan avulla ei voida kattavasti opettaa koneelle lauseiden sisäisiä viittaussuhteita, puhumattakaan lauseiden välisistä viittauksista, joiden opettaminen lienee tällä hetkellä mahdotonta. (Austermühl 2001: 173.) Ihmiskääntäjä tekee harvoin, jos koskaan, esitetynkaltaisia virheitä.

## 8. Konekäännösohjelmia

Euroopan Unioni (entinen Euroopan Yhteisö) on käyttänyt konekäännöksiä pisimpään, lukuun ottamatta USA:n ilmavoimia, ja käyttää niitä luultavasti eniten maailmassa. EU:ssa konekäännöksiä käyttävät kääntäjät, jotka tekevät koneen tekemästä raakakäännöksestä julkaisuvalmiin version. Useimmiten konekäännöksiä käytetään sisäiseen dokumentointiin. Kääntäjät eivät käytä konekäännöksiä silloin, kun käännös on tarkoitus julkaista ulkopuolisille. (Hutchins 2005a: 11.)

Kääntäjät eivät kuitenkaan ole EU:n ahkerimpia konekäännösohjelma Systranin käyttäjiä: useimmin konekäännöksiä tarvitsevat EU:n komission virkamiehet. He saavat satoja dokumentteja, joiden sisällöstä heidän täytyy olla tietoisia. He käyttävät konekäännöksiä saadakseen karkean yleiskuvan tekstistä ja kyetäkseen päättämään, tarvitsevatko he ihmisen tekemää käännöstä. Virkamiehet myös pyytävät kääntäjiä tekemään konekäännöksistä pikaisia jälkieditointeja, joita voidaan levittää komission sisällä. Kolmas käyttötarkoitus konekäännöksille EU:n komission virkamiehille on dokumentin luonnoksen kirjoittaminen kielellä, jota osaavat vähän, mutta eivät sujuvasti. Tuoreen kehityksen mukaan myös EU:n komission tulkit käyttävät konekäännöksiä taustatietona EU:n parlamentin tai muissa kokouksissa. (Hutchins 2005a: 11.) EU:n komission tarve konekäännöksille on kasvanut alle kymmenessä vuodessa viisinkertaiseksi. Se on lisääntynyt yli 20 % vuodessa, ja määrä on kasvussa muutamien seuraavien vuosien aikana. Laatu ei ole parantunut yhtä nopeasti, mutta komissiossa ollaan sitä mieltä, että niin kauan kuin käännös voidaan luokitella aiheen tai dokumentin tyyppin mukaan, voidaan myös saavuttaa laadun parannuksia. (Hutchins 2005a: 11.)

Esittelen seuraavaksi kaksi niin sanottua vakavasti otettavaa konekäännösohjelmaa. Lisäksi on olemassa kaupallisia ohjelmia, jotka eivät tee yhtä laadukasta jälkeä. Eurotra ja Systran ovat markkinoiden tunnetuimpia käännösohjelmia ja niillä on laaja käyttäjäkunta.

## **8.1 EUROTRA**

EUROTRA kehitettiin vuosina 1985–1993 yhdeksän EU-kielen konekäännösohjelmaksi. Järjestelmä auttaa ammattikäntäjiä siinä mielessä, että turhaa käännöstä syntyy vähemmän. Kielten tasavertaisuus EU:ssa ei tarkoita sitä, että kaikkia kieliä käännettäisiin kaikkiin. Komission käntäjien käsittelemistä teksteistä enemmistössä on käytetty englantia tai ranskaa joko lähtö- tai kohdekielenä. EUROTRA:n tavoitteena 1980-luvulla oli saavuttaa suurempaa tasapuolisuutta, mutta jo silloisetkin 72 kieliparia ja 144 eri käännessuuntaa osoittautuivat liian vaikeiksi. (Carlson 1999.)

## **8.2 SYSTRAN**

Systrania alettiin kehittää 1950-luvulla, ja se on sekä käntäjien että hallintovirkamiesten käytössä. Sitä käytetään apuvälineenä kolmessa tarkoituksessa:

- Silmäily: SYSTRAN pystyy käntämään jopa 2000 sivua tunnissa, joten se voi antaa nopeasti kuvan teksteistä, joiden lähtökieltä tilaaja ei tunne.
- Asiakirjojen laatiminen muulla kuin tilaajan äidinkielellä: Jotkut virkamiehet kirjoittavat tekstin mieluiten ensin äidinkielellään, tilaavat sitten konekäännöksen ja viimeistelevät tuloksen.
- Käntäminen: Konekääntämisen ensisijainen käyttötarkoitus. Koneen tekemää raakakäännöstä voidaan pitää pohjana, jota muokkaamalla saadaan lopullinen käänös. Korjausten määrä riippuu tekstilajista.

*SYSTRAN*in kokonaistuotanto vuonna 2001 oli lähes 800 000 sivua, joista käänöspalvelun osuus oli noin 40 % ja loput koostui hallintovirkamiesten tilaamista konekäännöksistä. Konekääntämistä käytetään komissiossa siis sekä varsinaisena käntämisen apuvälineenä että hallinnollisena apuvälineenä. (Euroopan komissio 2002.)

Eurotran jälkeen komissio on keskittynyt käyttämään jo 50-luvulla perustettua Systran-konekään-nösjärjestelmää ja tukee sen kehitystä. Systran tulee sanoista SYStem TRANslation. Sen ensimmäinen versio oli saatavilla jo vuonna 1969 venäjän ja englannin kielten välille. Nykyään se kattaa yli 40 kieliparia. Ohjelmasta on olemassa myös kaupallinen versio, joka on tarkoitettu pääasiassa yleiskielen kääntämiseen, mutta siihen voidaan lisätä erikoisalojen sanakirjoja. (Sågvall Hein 2004: 9.)

EU:n komissiolla on tällä hetkellä käytössä Systran seuraaville kielille: englantia, ranska, italia, saksa, hollanti, espanja, kreikka ja portugali. Yleisimmät kielet saatiin ensin. Seuraavina vuosina Itä-Euroopan maiden kielet liitetään ohjelmaan. Tšekki ja puolan kielet ovat jo työn alla. (Hutchins 2005a: 11.)

### **8.3 Kaupalliset käännösohjelmat**

Markkinoilla on tarjolla runsaasti halpoja kaupallisia käännösohjelmia. Vaikka tällaiset ohjelmat voivat olla hyödyksi silloin, kun halutaan saada tekstistä irti sen keskeinen sisältö, niitä ei voida kuitenkaan pitää ihmiskääntäjän vakavana kilpailijana. Nämä ohjelmat eivät pysty jäsentämään kieltä luotettavasti. Osa ohjelmista tuottaa kuitenkin jokseenkin ymmärrettäviä käännöksiä ja antaa vihiä tekstin sisällöstä sekä siitä, kannattaako käännös viimeistellä. Esimerkkejä tällaisista kaupallisista käännösohjelmista ovat muun muassa WebTranSmart, Sunda konekäännin sekä Babel Fish, jota käsittelen seuraavassa luvussa tarkemmin.



## 9. Babel Fish

Tässä luvussa tarkastelen kolmea konekäännöstä ja vertaan niitä ihmisen tekemiin käännöksiin. Laadukkaan konekäännösohjelman testaaminen ei valitettavasti ollut käytännössä mahdollista. Sen sijaan käytin Internetissä ilmaiseksi käännöksiä tarjoavaa Babel Fish -ohjelmaa. Seuraavassa kerron Babel Fishistä tarkemmin.

Babel Fish kuuluu SYSTRANin omistamaan konekäännösohjelmaan. Vuonna 1997 Alta Vista laittoi ohjelman sivuilleen ja tarjosi ensimmäisenä ilmaisia Internet-käännöksiä. Aluksi käännöksiä oli saatavilla vain kuudella kielellä: englanniksi, saksaksi, ranskaksi, espanjaksi, portugaliksi ja italiaksi. (Wikipedia.)

Puutteistaan huolimatta ohjelma saavutti suuren suosion ja vuonna 2001 sillä tehtiin jopa miljoona käännöstä joka päivä. Nykyään Babel Fish kääntää jo 19 kieliparia, muun muassa kiinaa (sekä perinteistä että yksinkertaistettua), venäjää ja japania. Käyttöliittymiä on useammallekin kielelle, muun muassa suomeksi. Babel Fishillä voidaan kääntää pieniä, noin 40 lauseen pituisia, tekstipätkiä tai kokonaisia, korkeintaan 5 kB:n kokoisia, verkkosivuja kieleltä toiselle. (Emt.)

Babel Fish on helppokäyttöinen, nopea ja tehokas. Raakakäännös saadaan aikaan nopeasti. Yksinkertaiset lauseet kääntyvät ohjelmalla yleensä oikein, mutta mitä monimutkaisempia ovat lähtötekstin lauseet, sitä todennäköisempää on, että käännöksessä on kielioppivirheitä, sanajärjestysvirheitä ja jopa täysin vääriä käännösvastineita.

Kritiikistä huolimatta täytyy muistaa, että Babel Fishin palvelut ovat käyttäjille ilmaisia. Käännökset eivät luonnollisesti ole yhtä hyviä tai laadukkaita kuin ihmisen tai kehittyneemmän konekäännösohjelman tekemät tekstit, mutta jos tarvitaan nopeasti tietoa vieraskielisen tekstin sisällöstä, Babel Fishistä on hyötyä.

Babel Fish on Internetissä osoitteessa <http://babelfish.altavista.com/>.

## **9.1 Aineiston ja tutkimusmenetelmän esittely**

Aineistona oli ote Nokian kännykän käyttöoppaasta, ote PR Newswire Association LLC:n julkaisemasta artikkelista sekä kappale konsulttiyritys Axonin yritysesityksestä. Tekstiotteet sopivat tutkimuksen aineistoksi, koska niissä on pääosin yleiskielistä tekstiä, mutta ne sisältävät kuitenkin myös tekniikan ja kaupan alan erikoistermejä. Tekstit sisältävät sekä pitkiä ja monimutkaisia että lyhyitä ja yksinkertaisia lauseita. Kielellisesti otteet muodostavat vaihtelevan aineiston, joka asettaa lukuisia haasteita konekäännösohjelmalle. Lähtötekstit on kirjoitettu englanniksi, ja verkossa on myös valmis ihmiskäännös kustakin otteesta. Lähtötekstit, valmiit ihmiskäännökset sekä Babel Fish -käännökset ovat luettavissa myös liitteessä.

Babel Fish ei ole kovinkaan kehittynyt käännösohjelma, eikä se osaa tehdä lähtötekstin analyysiä, kysyä käyttäjältä interaktiivisesti eri käännösvaihtoehtoista eikä valita erilaisia erikoissanakirjoja tai käännösstrategioita. Tästä johtuen syötin tässä tutkimuksessa tekstiotteet Babel Fish -konekäännösohjelmaan sellaisenaan, ilman muokkausta. Numeroin virkkeet jälkepäin, jotta käännösten analyysi olisi selkeämpää. En halunnut käsitellä tekstiä, koska toivoin saavani konekäännösohjelman mahdollisuuksista ja ongelmista mahdollisimman luonnollisen ja aidon kuvan. Kolmannen aineiston lähtötekstissä olleen kirjoitusvirheen korjasin ja syötin koneelle uudestaan, mutta käännätin ensin myös virheellisen lähtötekstin saadakseni vertailukelpoista aineistoa.

Babel Fish ei tarjoa käyttäjälle mahdollisuutta valita erilaisten sanakirjojen tai erikoissanastojen välillä. Ohjelmalle yksinkertaisesti syötetään lähtöteksti ja painetaan Käännä-painiketta. Sen jälkeen ohjelma antaa kääntämänsä tekstin ruudulle, josta käyttäjä voi manuaalisesti kopioida sen Word- tai muuhun tekstinkäsittelyohjelmaan. Tekstin kääntämistä olisi varmasti helpottanut, jos ohjelmalle olisi voinut kertoa ennalta, mihin sanaluokkiin sanat kuuluvat ja mitkä ovat sellaisia sanoja (esimerkiksi nimet), joita ei tule kääntää. Vielä parempi olisi tietysti, jos kone olisi itse voinut tehdä lähtötekstianalyysin ja luokitella sanat oikeisiin kategorioihin, mutta niinkin vaativaa ominaisuutta lienee ilmaisohjelmalta turha edellyttää.

Mikäli Babel Fish ei tunnista lähtö- tai kohdekieltä, se yksinkertaisesti jättää tekstin kääntämättä, ilmoittamatta siitä käyttäjälle. Tämä on iso puute, sillä se voi jossain tilanteessa jäädä käyttäjältä huomaamatta, varsinkin jos kohdekieli ei ole käyttäjälle tuttu.

Konekäännösohjelma Babel Fishin tuottamaa käännöstä ei luonnollisesti voida analysoida yhtä kriittisesti kuin ihmisen tekemää tekstiä. Koneelta ei odotetakaan julkaisuvalmista tekstiä, ainoastaan raakakäännöstä. Babel Fishin tekemät käännökset palvelevat tähän tarkoitukseen osittain hyvinkin, mutta joidenkin aineistossani olevien käännösten kohdalla epäilin, ymmärtäisikö kohdekielen lukija käännöstä, jollei lähtötekstiä olisi saatavilla tai jos hän ei ymmärtäisi siitä mitään. Kone ei osaa lisätä käännökseen mitään sellaisia selventäviä tai kuvaavia elementtejä, mitä lähtötekstissä ei ole. Se ei myöskään voi ymmärtää tekstiä tai edes kappaletta kokonaisuutena, vaan kääntää annetun sanakirjan ja kieliopin mukaan sana tai lause kerrallaan.

## 9.2 *Aineisto 1*

Englanninkielinen lähtöteksti

### **1. Charging and discharging**

2. Your phone is powered by a rechargeable battery.
3. Note that a new battery's full performance is achieved only after two or three complete charge and discharge cycles!
4. The battery can be charged and discharged hundreds of times but it will eventually wear out.
5. When the operating time (talk-time and standby time) is noticeably shorter than normal, it is time to buy a new battery.
6. Use only batteries approved by the phone manufacturer and recharge your battery only with the chargers approved by the manufacturer.
7. Unplug the charger when not in use.
8. Do not leave the battery connected to a charger for longer than a week, since overcharging may shorten its lifetime.
9. If left unused a fully charged battery will discharge itself over time.
10. Temperature extremes can affect the ability of your battery to charge.
11. Use the battery only for its intended purpose.
12. Never use any charger or battery which is damaged.

## Saksankielinen käännös

### **1. Laden und Entladen von Akkus**

2. Ihr Telefon wird von einem aufladbaren Akku gespeist.
3. Beachten Sie, dass ein neuer Akku erst seine volle Leistung erreicht, nachdem er zwei- oder dreimal vollständig ge- und entladen wurde!
4. Der Akku kann zwar mehrere hundert mal ge- und entladen werden, nutzt sich aber schließlich ab.
5. Wenn die Betriebszeit (Sprechzeit und Stand-by-Zeit) im Vergleich zur normalen Betriebszeit deutlich kürzer ist, sollten Sie einen neuen Akku kaufen.
6. Verwenden Sie nur Akkus und Ladegeräte, die vom Telefonhersteller zugelassen sind.
7. Trennen Sie das Ladegerät von der Steckdose, wenn Sie es nicht verwenden.
8. Lassen Sie den Akku nicht länger als eine Woche am Ladegerät angeschlossen, da eine zu starke Aufladung seine Lebensdauer verkürzen kann.
9. Bei Nichtgebrauch entlädt sich ein voll aufgeladener Akku mit der Zeit.
10. Extreme Temperaturen können einen nachteiligen Einfluss auf das Laden Ihres Akkus haben.
11. Benutzen Sie den Akku nur für den vorgesehenen Zweck.
12. Benutzen Sie keine schadhafte Ladegeräte oder Akkus.

## Babel Fish -käännös

### **1. Aufladung und Entladung**

2. Ihr Telefon wird durch eine nachladbare Batterie angetrieben.
3. Merken Sie, daß die volle Leistung einer neuen Batterie nur nach kompletter Aufladung zwei oder drei erzielt wird und entladen Sie Zyklen!
4. Die Batterie kann aufgeladen werden und entladene Hunderte Zeiten, aber sie trägt schließlich heraus.
5. Wenn die Anzugszeit (Sprechzeit und Bereitschaftszeit) merklich kürzer als Normal ist, ist es Zeit, eine neue Batterie zu kaufen.
6. Benutzen Sie die nur Batterien, die vom Telefonhersteller genehmigt werden und laden Sie Ihre Batterie nur mit den Aufladeeinheiten neu, die vom Hersteller genehmigt werden.
7. Trennen Sie die Aufladeeinheit wenn nicht im Gebrauch.
8. Lassen Sie nicht die Batterie angeschlossen an eine Aufladeeinheit für länger, als eine Woche, seit dem Überladen, seine Lebenszeit verkürzen kann.
9. Wenn sie unbenutzt verlassen wird, entlädt sich eine völlig belastete Batterie Überschusszeit.
10. Temperaturübermaße können die Fähigkeit Ihrer Batterie beeinflussen aufzuladen.
11. Benutzen Sie die Batterie nur für seinen beabsichtigten Zweck.
12. Benutzen Sie nie jede mögliche Aufladeeinheit oder Batterie, die beschädigt wird.

Ensilukemalta konekäännös vaikuttaa yllättävänkin ymmärrettävältä. Sisältö on sama kuin ihmiskääntäjän tekemässä käännöksessä, joskin suurin osa teknisistä ydintermeistä, kuten *charging, rechargeable, operating time, standby time* tai *battery*, on eri tavoin käännetty. Konekäännöksessä on myös selkeitä kielioppivirheitä, jollaisia ihmiskääntäjä ei tekisi. Ne eivät kuitenkaan ratkaisevasti haittaa ymmärtämistä.

Jos uuden, kalliin puhelimen mukana tulisi tällainen käyttöopas, se antaisi puhelimen valmistajasta melko epäluotettavan ja epäammattimaisen kuvan. Konekäännös täytyisi vähintäänkin tarkistaa ja kirjoittaa uudelleen. Samalla vaivalla sen luultavasti kääntäisi alusta loppuun itse, varsinkin jos apuna olisi käännösmuisti.

Koneella ei ole käsitystä missä kontekstissa tekstiä käytetään tai minkälaista kieltä tällaisissa yhteyksissä käytetään, ja siksi sen ehdottamat käännökset voivat olla hassun kuuloisia eivätkä sovellu puhelimen käyttöoppaaseen. Ihmiskääntäjä taas tietää kontekstin ja hänellä on mitä todennäköisimmin käytössään käännösmuisti, josta ydintermit ja muu sanasto tulevat automaattisesti. Näin käännös on yhdenmukainen aikaisemmin käännettyjen oppaiden kanssa. Toki konekäännös on yhdenmukaisempi, koska se kääntää samat sanat aina samalla tavalla, mutta tämänkaltainen ehdoton yhdenmukaisuus ei aina ole tärkeintä tällaisessa tekstissä. Tietysti ydintermien on oltava aina oikein eikä niissä saa soveltaa luovia ratkaisuja, mutta muun sanaston ei tarvitse olla konemaisen analoginen.

Käännösohjelma on välillä käyttänyt erikoisia ja monimutkaisia käännöksiä, kun ihminen on osannut kääntää saman asian yksinkertaisesti ja nokkelasti. Teknisessä tekstissä ei voi käyttää luovuutta samalla tavalla kuin esimerkiksi kaunokirjallisessa tekstissä, mutta mielestäni se ei tarkoita sitä, että tekstin täytyisi olla kuivaa ja sisältää pelkästään asiasanastoa. Asiat voidaan ilmaista yksinkertaisesti ja ymmärrettävästi olematta silti tylsä.

Konekäännöksessä puhutaan *Batteri*stä, vaikka yleisesti kännykän akuista käytetään saksaksi sanaa *Akku*. Konekäännöksessä lukee *nachladbar* ja ihmiskäännöksessä *aufkladbar*. Nämä ovat vakiintuneita kännykkäalan termejä saksan kielessä, joita olisi syytä käyttää oikein käyttöoppaassa.

Babel Fish ei näköjään käytä uusia oikeinkirjoitussääntöjä, vaikka ne ovat olleet käytössä jo lähes kymmenen vuotta. Konekäännöksen lauseessa 3 lukee edelleen *daß* kun pitäisi olla *dass*. Tällaisten pienten virheiden korjaaminen sanastoon ei luulisi olevan iso vaiva.

Neljäs virke on epäonnistunut täysin, koska ohjelma ei ole tunnistanut lauseenrakennetta. Englanninkielissä lähtötekstissä virke on seuraava: *The battery can be charged and discharged hundreds of times* ja saksantajan käännös menee näin: *Der Akku kann zwar mehrere hundert mal ge- und entladen werden*. Babel Fishin konekäännös on sekava eikä täysin ymmärrettävä: *Die Batterie kann aufgeladen werden und entladene Hunderte Zeiten*. Käännös on mielestäni täysin epäonnistunut eikä sitä voitaisi tällaisenaan käyttää. Rinnastus *charged and discharged* ei ole toteutunut eikä ohjelma ole tunnistanut sanontaa *hundreds of times*. Sanonnat tulisi syöttää koneelle erikseen, jottei se käännä niitä sanasta sanaan. Tällaisen tekstianalyysin tekeminen ei ole kuitenkaan Babel Fishissä mahdollista edes manuaalisesti.

Lauseessa 5 *Operating time* on epäonnistuneesti kääntynyt muotoon *Anzugszeit*, joka on huomattavasti harvinaisempi ja epäsovinmaisempi kuin ihmiskääntäjän käyttämä *Betriebszeit*. Myös *Talk-time* ja *Standby time* ovat kääntyneet vaikeasti ymmärrettäviin *Sprechenzeitiin* (ihmiskäännöksessä *Sprechzeit*) ja *Bereitschaftszeit* (pitäisi olla yleisesti käytössä oleva *Standby-Zeit*).

Konekäännöksen seuraavan kappaleen virke 6 sisältää omituisen ja täysin ylimääräisen artikkelin *die*, jonka lähde en pystynyt jäljittämään. Englanninkielinen lähtöteksti kuuluu: *Use only batteries approved by the phone manufacturer*. ja ihmiskääntäjän saksannos: *Verwenden Sie nur Akkus und Ladgeräte, die vom Telefonhersteller zugelassen sind*. Babel Fishin käännös on kieliopillisesti virheellinen ja sisältää siis ylimääräisen sanan: *Benutzen Sie die nur Batterien, die vom Telefonhersteller genehmigt werden*.

Konekäännös on sanasta sanaan uskollinen lähtötekstille, eikä siinä ole yhdistelty virkkeitä kuten ihmiskäännöksessä on tehty. Ihmiskääntäjän tekemässä saksannoksessa virkkeessä 6 *Use only batteries approved by the phone manufacturer and recharge your battery only with the chargers approved by the manufacturer* on yhdistetty lyhyesti ja ytimekkäästi muotoon:

*Verwenden Sie nur Akkus und Ladegeräte, die vom Telefonhersteller zugelassen sind.* Babel Fish ei arvatenkaan pysty muokkaamaan tekstiä samalla tavalla, ja sen tekemä käännös onkin kömpelö ja monimutkainen: *Benutzen Sie die nur Batterien, die vom Telefonhersteller genehmigt werden und laden Sie Ihre Batterie nur mit den Aufladeeinheiten neu, die vom Hersteller genehmigt werden.* Virke sisältää verbin *genehmigen* kaksi kertaa, tosin myös lähtötekstissä on toistettu *approved by the (phone) manufacturer* kahdesti. *Charger* eli laturi on käännetty ihmiskäännöksessä *Ladegerät*, mutta konekäännös on muuttanut sen muotoon *Aufladeeinheit*.

*When not in use* virkkeessä 7 on ihmiskäännöksessä mielestäni monimutkaisempi kuin konekäännöksessä. Ihminen on käyttänyt muotoa *wenn Sie es nicht verwenden* ja käännösohjelmassa on *wenn nicht im Gebrauch*. Molemmat ovat kieliopillisesti oikein, mutta mielestäni koneen tarjoama käännös on yleisempi ja yksinkertaisempi ja se sopii käyttöoppaaseen paremmin.

Konekäännöksen virkkeessä 8 on monia virheitä. *Lassen Sie nicht die Batterie angeschlossen an eine Aufladeeinheit für länger, als eine Woche, seit dem Überladen, seine Lebenszeit verkürzen kann* on kieliopillisesti virheellinen. Sanajärjestys on väärä (*nicht* on väärässä kohdassa) ja myös pilkutus on mennyt väärin. Lisäksi on käytetty outoja termejä, kuten *Aufladeeinheit*, *Überladen* ja *Lebenszeit*. Ihmiskääntäjä käänsi nämä sanat konventionaalisemmin *Ladegerät*, *Aufladung* ja *Lebensdauer*.

Virkkeessä 9 ihmiskääntäjä on keksinyt hieman neuvokkaamman tavan ilmaista englanninkielisen lähtöilmaisun *unused* kuin koneen ehdottama monimutkainen *Wenn sie unbenutzt verlassen wird*. Ihminen on käyttänyt yksinkertaisempaa muotoa *Bei Nichtgebrauch*. Saman virkkeen loppuosa on konekäännöksessä epäonnistunut niin pahasti, että epäilen sen ymmärrettävyyttä: *entlädt sich eine völlig belastete Batterie Überschlußzeit*. Virke on järjetön ja vailla sidonnaisuutta. Kone ei ole onnistunut välittämään englanninkielistä sanaliittoa *over time*, vaan se on kääntynyt muotoon *Überschlußzeit*. Myös *fully charged* on konekäännöksessä saanut negatiivisemmän sävyn *völlig belastet*. Ihmiskääntäjä oli kääntänyt sen *voll aufgeladen*. Puhelimen valmistaja tuskin haluaa käyttöoppaaseensa negatiivissävyisiä rakenteita, joten tämäntyyppisten konekäännöksiä käyttäminen ei olisi tässä yhteydessä mahdollista.

Virkkeessä 10 *Temperature extremes* on konekäännöksessä *Temperaturübermaße*, mikä on harvinainen ja erikoinen käännös, kun sama on ihmiskäännöksessä yksinkertaisesti *Extreme Temperatures*. Konekäännökseen on myös sattunut kielioppivirhe. Ohjelma ei ole tunnistanut englanninkielisen lähtötekstin rakennetta: *Temperature extremes can affect the ability of your battery to charge*. Konekäännöksessä on kaksi verbiä virkkeen lopussa. *Temperaturübermaße können die Fähigkeit Ihrer Batterie beeinflussen aufzuladen*. Käännös ei ole oikeaa saksaa eikä sitä voitaisi sellaisenaan käyttää oikeassa käyttöoppaassa, vaan ihmisen pitäisi korjata virheet ensin.

Viimeisen virkkeen (12) *Never use any charger or battery which is damaged* on myös jäänyt käännösohjelmalta tunnistamatta. Ihminen on kääntänyt sen seuraavasti: *Benutzen Sie keine schadhaften Ladegeräte oder Akkus*. Babel Fish sen sijaan on kääntänyt *any*-sanan näin: *jede mögliche*, mikä ei ole oikein. Konekäännös kuuluu virheellisesti: *Benutzen Sie nie jede mögliche Aufladeeinheit oder Batterie, die beschädigt wird*.

Kokonaisuutena mielestäni konekäännös on yllättävän hyvä, vaikkakin väärine sanajärjestyksineen ja virheellisine käännöksineen kömpelö ja paikoitellen vaikeasti ymmärrettävä. Sellaisenaan sitä ei voitaisi käyttää, mutta osittain se tarjosi mielestäni hyviä ratkaisuja, joita ihmiskääntäjäkin voisi ehdottaa.

### **9.3 Aineisto 2**

Englanninkielinen lähtöteksti

1. The research project that was presented at the RSNA Meeting in Chicago (U.S.) last week by Dr. W. A. Bashir et al. got enormous press coverage and spread like wildfire over the worldwide web.
2. The study confirmed with the help of a MRI technique that sitting in a slouched forward position or just upright in a 90degrees position was very unfavourable for the position of the nucleus and the intervertebral disk height.



3. By comparing a supine position with the position of a Formula 1 race car driver in his seat and then show how favourable this position is for the lumbar spine you can really make your point.
4. Of course, leaning back in your chair with full support of backbones and shoulders is more relaxing than any more erect sitting posture.

### Saksankielinen käännös

1. Das Forschungsprojekt, das letzte Woche auf dem RSNA-Kongress in Chicago (USA) von Dr. W. A. Bashir et al. präsentiert wurde, erhielt enormes Medienecho und verbreitete sich wie ein Lauffeuer im Internet.
2. Die Studie bestätigte mithilfe eines MRI-Verfahrens, dass Sitzen in gekrümmter Vorwärtsposition oder aufrecht in einer 90-Grad-Position äusserst ungünstig für die Kernposition und Bandscheibenhöhe ist.
3. Der Vergleich einer Rückenlage mit der Position eines Formel-1-Piloten in seinem Sitz und die anschliessende Demonstration, wie günstig sich diese Position auf die Lendenwirbelsäule auswirkt, vermitteln den Standpunkt wirklich gut.
4. Natürlich ist ein Zurücklehnen mit voller Unterstützung des Rückgrats und der Schultern entspannender als jede aufrechte Sitzposition.

### Babel Fish -käännös

1. Das Forschung Projekt, das wurde dargestellt bei der RSNA Sitzung in Chicago (US.) letzte Woche durch Dr. W. A. erhielten Bashir et al. enorme Pressedeckung und -verbreitung wie wildfire über dem weltweiten Netz.
2. Die Studie, die mit Hilfe einer MRI Technik bestätigt wurde, daß das Sitzen in a Vorwärtsposition oder gerade aufrecht in einer Position 90degrees slouched, war für die Position des Kernes und der intervertebralen Scheibe Höhe sehr ungünstig.
3. Indem Sie eine nachlässige Position mit der Position eines Rennen-Autotreibers der Formel 1 in seinem Sitz vergleichen und zeigen Sie dann, wie vorteilhaft diese Position für den lumbalen Dorn ist, den Sie Ihren Punkt wirklich bilden können.
4. Selbstverständlich zurück ist sich lehnen in Ihrem Stuhl mit voller Unterstützung des Rückgrats und der Schultern mehr aufgerichtete Sittinglage dann sich entspannen.

Ensimmäinen virke on ehkä kokonaisuudessaan ymmärrettävä, mutta ei hyvää saksaa. Jo ensimmäinen sana, *Forschung Projekt*, on kirjoitettu väärin. Sen pitäisi olla yhdyssana ja siitä puuttuu yhdistävä s-kirjain, mutta tässä Babel Fish on noudattanut englannin oikeinkirjoitussääntöjä ja kirjoittanut koko termin erikseen.

Virkkeessä sanajärjestys on väärä; verbit ovat väärissä kohdissa. *RSNA Meeting* on myös väärin kirjoitettu ilman yhdysviivaa, ja lisäksi on käytetty sanaa *Sitzung*, vaikka ihmiskäännöksessä käytetty *Kongress* kuulostaa paljon ammattimaisemmalta ja isommalta. Suluisissa oleva Yhdysvaltojen lyhenne on *US*, kun taas ihmiskääntäjä on käyttänyt yleisempää lyhennettä *USA*.

Tohtori W.A. Bashirin nimi on käännöksessä täysin epäonnistunut. Nimen keskelle on jostain syystä livahtanut verbi! Tällaista virhettä ihmiskääntäjä ei tekisi milloinkaan, eikä näin huonoa käännöstä voitaisi missään tapauksessa julkaista.

*Press coverage* on myös väärin tulkittu konekäännöksessä. Se on käännetty *Pressedeckungiksi*, mikä on täysin väärä käännös tähän asiayhteyteen. *Wildfire* ilmeisesti puuttuu Babel Fishin sanastosta, sillä se on jätetty kokonaan kääntämättä. Tällainen ratkaisu ei toimi tässä kohtaa, ja lukija ei välttämättä ymmärrä mistä on kyse.

Konekäännökseen on joutunut tavuviiva pilkun paikalle. Ilmeisesti on tarkoitettu ajatusviivaa, mikä sekin tuntuisi oudolle tässä kohtaa.

Seuraava sana *verbreitung* pitäisi olla verbi, mutta tässä se on väärin pienellä alkukirjaimella kirjoitettu substantiivi. Lähtötekstin sana *spread* on imperfektissä ja näyttää samalta kuin saman sanavartalon substantiivi. Homonyymien käännösongelmista olen kirjoittanut kappaleessa 7.2.1.

Vakiintunut termi *Internet* on konekäännöksessä *Weltweites Netz*, mikä kuulostaa todella oudolta ja saattaa herättää täysin väärä mielikuvia konkreettisesta verkosta. Uskon, että Babel Fish tuntee sanan Internet, mutta englannin kielessä käytetään yleisesti Internetin synonyymina *world wide webiä*, mikä taas on jäänyt ohjelman sanakirjaan päivittämättä.

Tohtorin nimen keskelle eksynyt verbi puuttuu loppulauseesta, ja näin ollen koko virke on kieliopillisesti virheellinen ja vain vaivoin ymmärrettävissä.

Toinen virke alkaa järkevästi ja kuulostaa ymmärrettävälle pienistä kirjoitus- ja käänösvirheistä huolimatta (*MRI Technik*, kun pitäisi olla *MRI-Verfahren*), kunnes yhtäkkiä sivulauseessa on englanninkielinen artikkelin *a* ja samalla rivillä on kääntämättä jäänyt sana *slouched*. *90degrees position* on myös jäänyt kääntämättä ja kirjoitustapakin olisi saksalaisittain epätyypillinen ellei peräti väärä.

Virkkeen 3 *the position of the nucleus and the intervertebral disk height* on konekäännöksessä epäonnistunut. Tekstipätkä sisältää erikoissanastoa kuten *nucleous* ja *intervertebral*, jotka ohjelma on kääntänyt eri tavalla kuin ihmiskääntäjä. Lauserakenne on tässä epäymmärrettävä: *die Position des Kernes und der intervertebralen Scheibe Höhe. Die Position des Kernes* ei vastaa lähtötekstin merkitystä.

Lähtöteksti on jaettu kahteen kappaleeseen ja niiden välissä on tyhjä rivi (ks. liite). Babel Fish ei osaa lyödä rivivälejä, ja kappalejakokin puuttuu. Tämä tietysti hankaloittaa lukemista, ja oikolukijan täytyisikin korjata tällaiset virheet.

Kolmannessa ja neljännessä virkkeessä verbit ovat osuneet oikeampiin kohtiin kuin alkuosan virkkeissä. *Formula 1 race car driver* on hämmästyttävän monimutkaisesti ilmaistu saksankielisessä konekäännöksessä: *Rennen-Autotreiber der Formel 1*. Ihmiskääntäjä on käyttänyt yleistä ja yksinkertaista käännöstä *Formel-1 Pilot*.

Seuraava kompastuskivi on *lumbar spine*, joka erikoisterminä pitäisi tietysti kääntää erikoissanaston avulla. Babel Fishissä ei ole saatavilla erilaisia sanastoja, vaan kaikki käännökset tehdään samaa yleissanakirjaa käyttäen. Näin ollen myös näinkin erikoiset lääketieteelliset termit haetaan yleissanastosta. Ihmiskääntäjä on käyttänyt tässä sanaa *Lendenwirbelsäule*, mikä tarkoittaa suomennettuna lannerankaa. Kone on kääntänyt sen näin: *lumbaler Dorn*, mikä suomennettaisiin suunnilleen lannepiikiksi. Nämä kaksi käännöstä ovat melko erilaisia ja tarkoittavat eri asioita. Tässä kohtaa ihmisen olisi tarkistettava lopullinen konekäännös ja korjattava se.

Virkkeen 3 *You can make your point* on sanonta, jonka kääntäminen on koneelle haasteellista. Ihminen on ilmaissut asian saksaksi *vermitteln den Standpunkt wirklich gut*, kun taas kone on

epäonnistunut kääntäessään kielikuvan sanatarkasti *den Sie Ihren Punkt wirklich bilden können*. Sanontojen kääntämisen ongelmista kirjoitin kappaleessa 7.4.

Virke 4 sisältää suunnilleen oikeat sanat, mutta väärässä järjestyksessä. Englanninkielisessä lähtötekstissä on harmillinen kirjoitusvirhe: *then* kun pitäisi olla *than*. Koska väärinkirjoitettu sana on myös oikea olemassa oleva englanninkielinen sana, ohjelma kääntää sen, mutta tietysti asiayhteyteen väärin. Lähtötekstin lukija ymmärtää lauseen kirjoitusvirheestä huolimatta ja ihmiskääntäjä on osannut välttää virheen siirtymisen käännökseensä, mutta konekäännöksessä virhe vaikuttaa. Kun korjasin lähtötekstin virheen ja käänsin tekstin uudelleen, oli lopputulos seuraava:

*Selbstverständlich zurück ist sich lehnen in Ihrem Stuhl mit voller Unterstützung des Rückgrats und der Schultern, entspannend als aufgerichtete Sittinglage.*

Babel Fish lisäsi pilkun, ja sen jälkeen tuleva käänös on hieman erilainen kuin aikaisempi. Käänös on edelleen huono eikä kovin ymmärrettävä.

Näinkin isot virheet käänöksissä vievät pohjan koko konekääntämiseltä ja antavat huonon kuvan konekääntämisen mahdollisuuksista. Näin monta kääntämätöntä termiä näin lyhyessä tekstikappaleessa vaikeuttavat ymmärtämistä. Tietenkään konekäännökseen ei kohdisteta samankaltaisia odotuksia kuin ihmiskääntäjän tekemään tekstiin, kuten kirjoitin kappaleessa 4.5, mutta näin huono lopputulos on varmasti yllätys.

Babel Fishiä on kuitenkin kehuttava siitä, että se tunnistaa saksan kielen taivutusmuodot, se osaa kirjoittaa esimerkiksi *in seinem Sitz*, kun lähtötekstissä on *his seat*. Ohjelma osaa myös teititellä ja kirjoittaa suoran puhuttelun isolla alkukirjaimella aivan kuten pitääkin: *bilden Sie Ihren Punkt*.

## 9.4 *Aineisto 3*

### Englanninkielinen lähtöteksti

#### 1. **Aerospace and defence**

2. There is no more demanding an industry than Aerospace and Defence, where failure can be catastrophic and the pressures of flight safety, government regulation, globalisation and product/program complexity are unrelenting.
3. Axon's experience, with consultants making a significant contribution to around 25% of SAP's go-lives in this sector world-wide, brings a unique combination of deep industry experience and SAP technology expertise to use SAP to transform Aerospace and Defence businesses such as:

- Cessna Aircraft
- Goodrich Corporation
- Lockheed Martin
- Northrop Grumman
- Pratt & Whitney
- Raytheon Corporation
- Rolls-Royce Aero Engines
- Sikorsky Aircraft
- US Army Armaments Research and Engineering

### Saksankielinen käännös

#### 1. **Luft-, Raumfahrt und Verteidigung**

2. Es gibt wohl keine anspruchsvollere Industrie als die der Luft- und Raumfahrt und Verteidigung – jeder Fehler kann katastrophale Auswirkungen haben und der Druck durch Faktoren wie Flugsicherheit, Regierungsaufgaben, Globalisierung und Produkt-/Programmkomplexität ist enorm.
3. Die Erfahrung von Axon und seinen Consultants, die bei rund 25 % aller SAP Systemfreigaben in diesem Sektor weltweit einen großen Beitrag leisten, setzt sich aus umfassender Branchenkenntnis und technologischer SAP Expertise zusammen. Mithilfe von SAP transformieren wir Unternehmen aus der Luft- und Raumfahrtbranche und der Verteidigungsindustrie wie:

- Cessna Aircraft
- Goodrich Corporation
- Lockheed Martin

- Pratt & Whitney
- Raytheon Corporation
- Rolls-Royce Aero Engines
- Sikorsky Aircraft
- US Army Armaments Research and Engineering

## Babel Fish -konekäännös

### 1. Aerospace und Verteidigung

2. Es gibt nicht mehr Verlangen einer Industrie, als Luftfahrt- und Verteidigung, in der Ausfall verhängnisvoll und der Druck der Flugsicherheit sein können, Regierung Regelung, globalisation und product/program Kompliziertheit unnachgiebig sind.
3. Erfahrung des Axons, wenn die Berater einen bedeutenden Beitrag bis herum 25% bilden, von Saft gehen-lebt in diesem weltweiten Sektor, holt eine einzigartige Kombination der tiefen Industrieerfahrung und SAFT-Technologiesachkenntnis zum Gebrauch SAFT, um Aerospace- und Verteidigungsgeschäfte wie umzuwandeln:

- ? Cessna Flugzeug
- ? Goodrich Corporation
- ? Lockheed Martin
- ? Northrop Grumman
- ? Pratt U. Whitney
- ? Raytheon Corporation
- ? Rolls-Royce-Royce Aero Maschinen
- ? Sikorsky Flugzeug
- ? US Armee-Bewaffnungen Forschung und Technik

Lähtötekstin tekstilaji on enemmän markkinointihenkinen kuin tekninen, mutta koska se sisältää myös teknisiä termejä, päätin ottaa sen esimerkiksi. Yli-Jokipiin (2004: 82) mukaan tekniikan diskurssi meneekin osittain päällekkäin kaupallisen kielen kanssa.

Ensilukemalta konekäännös vaikuttaa sekavalta: lauseiden sanajärjestys on omituinen, verbit ovat väärissä kohdissa ja muutamia sanoja on jäänyt kääntämättä. Toisella lukemalla tekstin sanoma alkaa etäisesti hahmottua. Konekäännös on silti niin huono, että ilman lähtötekstin tietoja sen sanomaa ei varmaankaan ymmärtäisi kokonaan.

Ammattisanasto on pääosin käännetty väärin. Vain suhteellisen yleiskielinen *flight safety* on oikein *Flugsicherheit*. Muut termit, kuten *government regulation*, *globalisation* tai lyhenne

SAP sekä tietysti otsikko *Aerospace and defence*, ovat epäonnistuneita käännöksiä. Käsittelen näitä termejä myöhemmin tarkemmin.

Jo otsikko (1) on hieman väärin käännetty. *Aerospace*-sanaa ei ole löytynyt ohjelman sanakirjasta ja Babel Fish on jättänyt sen kääntämättä. Puolittain käännetty, epäjohdonmukainen otsikko ei soveltuisi käyttöön sellaisenaan. Ihmiskääntäjä on käyttänyt otsikosta vakiintunutta sanamuotoa ja siksi sitä tulisi suosia.

Virke 2 on kokonaisuudessaan epäonnistunut. Lähtötekstin virke on pitkä, ja se sisältää monta sivulausetta, mikä on varmasti osaltaan vaikeuttanut kääntämistä. Konekäännösohjelma ei osaa pilkkoa virkkeitä tai yhdistää lauseita, vaan se seuraa uskollisesti lähtötekstin rakennetta. Upotettu sivulause tekee lähtötekstin ensimmäisestä virkkeestä raskaan lukea ja ymmärtää. Käännöksessä verbit ovat osuneet vääriin kohtiin, ja koko virkkeessä ei ole kieliopillisesti järkeä. Lähtötekstin *There is no more demanding an industry than Aerospace and Defence* ei ole tavanomaisen yksinkertainen lause, vaan sen rakenne on monimutkainen, ja asia on ilmaistu kieltomuodon avulla. Ehkä juuri siksi käännösohjelma ei ole onnistunut sen kääntämisessä. Ihmiskääntäjä on ratkaissut asian näin: *Es gibt wohl keine anspruchsvollere Industrie als die der Luft- und Raumfahrt und Verteidigung*, ja käännös on toimiva. Hän on lisännyt sanan *wohl* keventääkseen lausetta, jottei se kuulostaisi niin ehdottomalta.

Seuraavan lauseen adverbi *where* on konekäännöksessä väärin tulkittu pronominiksi. Homonymioiden kääntäminen on koneelle ymmärrettävästi vaikeaa, kuten kirjoitin kappaleessa 7.2.1. Jos kyseessä olisi laadukas konekäännösohjelma, olisi etukäteen voitu syöttää ohjelmaan tietoja tietyistä sanoista, jolloin virheiden määrä pienenesi. Babel Fish on kuitenkin niin alkeellinen automaattinen käännösohjelma, että siinä ei ole tällaista mahdollisuutta.

Saman virkkeen (2) *failure* on väärin ymmärretty ja käännetty *Ausfall*iksi. Ihminen on käyttänyt sanaa *Fehler*, mikä sopii tähän yhteyteen paljon paremmin. Siinä missä ihmiskääntäjä on kirjoittanut *jeder Fehler kann katastrophale Auswirkungen haben*, on konekäännös täysin epäymmärrettävä: *in der Ausfall verhängnisvoll und der Druck der Flugsicherheit sein können*. *Verhängnisvoll* on väärä sana tähän yhteyteen, ja myös verbit ovat

väärissä kohdissa. Lähtötekstissä *pressures* viittaa koko seuraavaan luetteloon, mutta konekäännöksessä *der Druck* viittaa vain sanaan *Flugsicherheit*, jolloin lauseesta jää merkitys pois.

Ammattitermi *government regulation* on konekäännöksessä täysin outo *Regierung Regelung*, jonka ymmärtäminen lienee mahdotonta. Ihmiskäännöksessä on käytetty muotoa *Regierungsauflagen*, joka sekin on harvinainen ilmaus.

*Globalisation* on jäänyt kääntämättä kokonaan. Saksalainen lukija kuitenkin varmasti ymmärtää sanan, vaikka se oikeammin onkin saksaksi *Globalisierung* (tosin nykyään saksassa on käytössä myös suora käännöslaina *Globalisation*).

Kone ei ilmeisesti ole tunnistanut lähtötekstissä virkkeessä 2 olevia tavallisia sanoja *product/program* niiden välissä olevan kauttaviivan takia. Tästä johtuen se ei ole osannut kääntää koko termiä, vaikka molemmat sanat lieneekin syötetty sanakirjaan. Lähtötekstin seuraava sana *complexity*, johon *product/program* viittaa, on ihmiskäännöksessä yhdistetty suoraan edelliseen sanaliittoon: *Produkt/Programmkomplexität*, kun kone on käyttänyt monimutkaisempaa ja hieman harvinaisempaa ilmaisua *Kompliziertheit*.

Koko virke 2 on konekäännöksessä epäjohdonmukainen. Siinä on seurattu tarkasti lähtötekstin rakennetta. Mihinpä muuhunkaan kone pystyisi? Ihminen on kuitenkin osannut lokalisoida käännöksen kohdekieleen sopivaksi muuttamalla lauserakennetta, vaihtamalla sanajärjestystä, käyttämällä esimerkiksi ajatusviivaa ja saksan kieleen sopivampia sanoja. Koneella ei ole kielitajua eikä se ole minkään kielen syntyperäinen puhuja. Siksi konekäännös on töksähtelevä ja paikoin jopa täysin epäymmärrettävä.

Virkkeiden 2 ja 3 välissä on lähtötekstissä kaksi välilyöntiä, mutta Babel Fish on osannut jättää virheen toistamatta. Tällaisten asioiden korjaaminen on koneelle helppo "opettaa", toisaalta ihminen kirjoittaa koko tekstin uudelleen eikä vain vaihda sanoja toisenkieliseksi, ja näin ollen hänkään ei siirtäisi tämänkaltaista virhettä edelleen käännökseen.



Myös virke 3 on kieliopillisesti virheellinen ja sisältää sanajärjestysvirheitä, käännösvirheitä ja typografisia virheitä. Muun muassa *Consultant* on ihmiskäännöksessä myös saksaksi *Consultant*, mutta konekäännöksessä mielestäni hieman arkisempi ja tyylittömämpi *Berater*. Luvun ja prosenttimerkin välissä pitäisi olla välilyönti, kuten ihmisen tekemässä käännöksessä on, mutta Babel Fish on kopioinut lähtötekstin tavan merkitä asia ja jättänyt välilyönnin pois. Ihmiskääntäjä on jakanut lähtötekstin virkkeen 3 käännöksessä kahdeksi virkkeeksi helpottaakseen ymmärtämistä. Ratkaisu toimii hyvin ja ihmiskäännös onkin kevyempi lukea ja ymmärtää kuin lähtöteksti tai konekäännös.

*SAP*-lyhenne on muuttunut *saffiksi*, mikä on tietysti erittäin harmillinen virhe eikä kyseinen käännös sovi tähän kontekstiin laisinkaan. Kaikenlisäksi lyhenne on kahdesti *SAFT* isoilla kirjaimilla ja kerran *saft* pienillä. Englanninkielisessä lähtötekstissä on yhdenmukaisesti käytetty muotoa *SAP*, joten käännöksessäkin konsistenssia olisi syytä vaatia. On myönnettävä, että *SAP*-lyhenteen kääntäminen on haasteellinen tehtävä. Lyhenteellä on monta mahdollista merkitystä; löysin ainakin 28 erilaista merkitystä sille, muun muassa seuraavat: Second Audio Program, Standard Accounting Program, Special Access Program, Santa Paula California, Statistical Analysis Plan, Secure Audio Plan, SIM Access Program, Social Alternative Politics, South African Party. Näin monesta vaihtoehdosta ihmisenkin olisi vaikea valita ja päättää, mikä on oikea lyhenne tai sen käännös. Ja koska kone ei tiedä tekstin kontekstista mitään, sen on vaikea ratkaista mikä näistä tai jostain muista merkityksistä sopisi parhaiten tähän yhteyteen. Koneen tarjoama käännösvastine *SAFT* tarkoittaa ainakin kahta asiaa: Swiss Army Fitness Test tai Simple Asynchronous File Transfer. Tässä tekstissä on tuskin kyse kummastakaan, joten *SAFT* on väärä lyhenne tähän kohtaan. Englanninkielinen sana *sap* tarkoittaa suomeksi mehua tai mahlaa, ja siksi kone lienee valinneen käännökseksi *Saft*, vaikka se ei selvästikään tähän kontekstiin sovi.

Tässä voi tietysti olla kyse myös yksinkertaisesti *SAP*-nimisen yrityksen sovelluksista.

Konekäännösohjelma on laittanut firman nimen eteen artikkelin (*Erfahrungen des Axons*), mikä ei ole oikea käytäntö. Ihmettelen, mistä artikkeli on tullut, sillä englanninkielisessä tekstissä artikkelia ei ole käytetty.

Virkkeen 3 *Go-live*-verbiä Babel Fish ei ole tunnistanut, vaan se on epäonnistuneesti käännetty suoraan *gehen-lebt*. Sanakirjoihin ei ole tapana kirjata sellaisia yhdyssanoja, jotka ihminen pystyy päättämään, mutta käännösohjelmien sanakirjoihin sellaisetkin olisi tarpeen syöttää. Se on kuitenkin mahdoton tehtävä, sillä kaikkien mahdollisten yhdyssanojen ja sanaliittojen syöttäminen koneelle olisi ensinnäkin loputon tehtävä ja toiseksi muistin täyttäminen äärimmilleen hidastaa koneen suoritusnopeutta. Tästä kirjoitin enemmän luvussa 3.5.

*Industry experience* on ihmiskäännöksessä elegantisti *Branchenkenntnis*, mutta konekäännöksessä hieman erikoinen *Industrieerfahrung*. Myös *SAP technology expertise to use SAP to transform Aerospace and Defence businesses* on konekäännöksessä väärin pilkottu osiin ja koko käännös on vailla järkeä: *SAFT-Technologiesachkenntnis zum Gebrauch SAFT, um Aerospace- und Verteidigungsgeschäfte*. Niinkin yleinen ja vakiintunut sanaliitto kuin *such as* on konekäännöksessä onnistunut menemään väärin: *wie umzuwandeln*.

Tekstinäytteen lopussa luettelomerkit ovat muuttuneet kysymysmerkeiksi, ja ne näyttävät hivenen hullunkurisilta. Ne pitäisi ilman muuta korjata ennen tekstin julkaisemista. Myös luettelossa olevat firman nimet pitäisi muuttaa takaisin alkuperäiseen muotoon, sillä niitä ei ole tarkoitus kääntää. Nämä olisi ollut hyvä merkitä ennen tekstin syöttämistä koneelle, mutta Babel Fishissä sellaista vaihtoehtoa ei ole tarjolla. Käännökset ovat kaikenlaisiksi virhellisiä. Esimerkiksi *Rolls-Royce Aero Engines* on konekäännöksessä *Rolls-Royce-Royce Aero Maschinen*. Tällaiset virheet pistävät silmään heti ja antavat todella huonon kuvan käännöksen laadusta.

Babel Fishin toinen suuri ongelma ainakin tämän tekstipätkän kohdalla on sille tuntemattomat termit, jotka se yksinkertaisesti jättää kääntämättä. Näin ei julkaistavaksi tarkoitettussa tekstissä tietenkään voisi tehdä. Käännösohjelman olisi vähintäänkin kysyttävä käyttäjältä, eli ihmiseltä, mitä näille termeille tehdään, sekä mahdollisesti ehdottaa muutamaa vaihtoehtoa. Jos ohjelman muistissa ei ole valmiita vaihtoehtoja, ihmisellä pitäisi olla mahdollisuus syöttää koneelle oikea käännösvastine. Nyt Babel Fishin käännöslaatu kärsii huomattavan paljon, kun saksankieliseksi tarkoitettu teksti vilisee englanninkielisiä sanoja.

Jos käännöksen haluaisi julkaista, se olisi pakko kirjoittaa lähes kokonaan uudelleen. Konekäännöstä ei voitaisi käyttää ilman huomattavaa jälkieditointia, ja voidaan mielestäni perustellusti kysyä, eikö näinkin mittavan korjausoperaation vaatimassa ajassa olisi parempi kirjoittaa koko käännös alun pitäen suoraan ihmisvoimin. Konekäännöksestä ei tämän tekstin kohdalla ole suurta ajallista, taloudellista tai muutakaan hyötyä. Näinkin markkinointisävytteinen teksti vaatii ihmisen luovuutta ja kykyä käyttää erilaisia käännösratkaisuja ja vaihdella sanastoa. Tässä ei ole tarkoitus pitäytyä orjallisesti yhdenmukaisessa sanastossa, ja jos yhdenmukaisuutta tarvitaankin, se saadaan käyttöön käännösmuistista.

## 10. Babel Fish -käännösten analyysin yhteenveto

Suurimmat ongelmat Babel Fishin tuottamissa käännöksissä olivat sanajärjestysvirheet ja kääntämättä jääneet termit. Tämä oli osittain odotettavissa, sillä kriteerit eivät ole korkealla kun on kyse ilmaiseksi käännöksiä tarjoavasta ohjelmasta. Osa ongelmista olisi varmasti voitu välttää editoimalla lähtötekstiä, muun muassa pilkkomalla pitkiä lauseita, korjaamalla kirjoitusvirhe, lisäämällä pilkkuja tai yksinkertaistamalla tekstin rakennetta.

Tekstin muokkaaminen ymmärrettäväksi jälkikäteen on varmasti mahdollista, mutta ainakin osassa tapauksista koen, että kaikki vaiva ei ole sen arvoista, vaan helpommalla olisi päästy jos ihminen kääntäisi tekstin alusta pitäen itse. Varsinkin kokenut teknisten tekstien kääntäjä, joka tunnistaa terminologian ja tekstilajin, tekisi varmasti parempaa jälkeä ilman koneen tekemää raakakäännöstä. Joissain tapauksissa koneen tekemä raakakäännös voi jopa hämätä tekstin jälkieditoijaa, niin että paremmat ja sopivimmat ratkaisut jäävät huomaamatta, kun kone ehdottaa valmiiksi huonompaa vaihtoehtoa. Epäilen, että käännösmuistia käyttämällä ihminen olisi saanut tekstin valmiiksi nopeammin ja laadukkaammin kuin konekäännösohjelman raakakäännöstä käyttäen.

Positiivistakin palautetta voin Babel Fishille antaa: Ensimmäisen aineiston virke 7 oli hyvin käännetty, mielestäni jopa paremmin kuin ihmiskääntäjän tekemä käännös. Joskus suora käännös voi toimia paremmin kuin luova. Toki kyseinen virke oli lyhyt ja näin ollen koneelle helppo käsitellä. Babel Fish osaa myös tuottaa saksankielen sanojen taivutusmuotoja sekä teititellä saksalaisten konventioiden mukaisesti isolla alkukirjaimella.

Kääntämisen nopeudessa ja edullisuudessa kone voittaa ihmisen. Ottaen huomioon kuinka nopeasti Babel Fish pystyy kääntämään suuria määriä tekstejä ja kuinka edulliset sen kustannukset ovat käyttäjälle, voisi vetää sen johtopäätöksen, että esi- ja jälkieditoinnin kanssakin käännöksen kustannukset jäisivät alemmas kuin käyttämällä pelkästään ihmiskääntämistä.

Tällainen huonolaatuinen raakakäännös palvelee apuvälineeksi tekstin silmäilyyn, sillä se antaa kuvan tekstistä, jonka lähtökieltä käyttäjä ei tunne, tai mallin asiakirjojen laatimiseen muulla kuin omalla äidinkielellä. Varsinaiseen tarkoitukseensa, eli käännöksen pohjaksi, se ei mielestäni nykymuodossaan sovellu. Laadukkaammat, pidemmälle kehitellyt käännösohjelmat ovat luku sinänsä. Ne varmasti tarjoavat laadukasta, ymmärrettävää tekstiä. Tämän tutkimuksen pohjalta en voi yleistää konekäännösten laatua suuntaan enkä toiseen, sillä tiedostan käyttäneeni tarjolla olevista ohjelmista kaikkein hankalinta: kaupallista ilmaisohjelmaa. Täten en katso voivani tämän tutkimuksen perusteella ottaa kantaa yleisesti käännösohjelmien menetelmiin, tuloksiin tai kannattavuuteen.

## 11. Konekääntämisen tulevaisuus

Käännösten teettäjät toivovat, että pian tulisi markkinoille tehokas käännösohjelma, joka kääntäisi minkä tahansa tekstin yhtä hyvin kuin ihmiskääntäjä, mutta nopeammin ja halvemmalla. Alan Melbyn (2001) mukaan tällaista ohjelmaa saadaan vielä odottaa. Konekääntäminen tarvitsee edelleen runsaasti tekoälytutkimusta (Lammi 2004). Konekääntämisessä on potentiaalia tulevalle tutkimukselle, kehitykselle ja paremmalle laadulle. Erilaisissa muodoissa se on tullut jäädäkseen. (Sågvall Hein 2004: 11.)

Olen tässä tutkimuksessa tullut siihen tulokseen, että konekääntämistä kannattaa ehdottomasti kehittää edelleen. Nykyisellään laadukkaat käännösohjelmat pystyvät tarjoamaan raakakäännöksen, eli pohjan ihmiselle julkaisukelpoisen käännöksen tekemiselle ja ne nopeuttavat ihmisen käännöstyötä, mutta aivan vielä ne eivät kykene tuottamaan automaattisesti hyvää luonnollista kieltä. Ammattitaitoinen kääntäjä – ihminen – hallitsee kääntämisen teorian, mallien, erilaisten käännösstrategioiden lisäksi elektronisten apuvälineiden käytön. Vaikka koneelta näistä ominaisuuksista on hallussa vain elektronisen datan omaksuminen, kannattaa mielestäni konekäännöksiä jatkossakin käyttää hyväksi. Konekääntäminen ei ole kääntäjien pelosta huolimatta näillä näkymin uhka kääntäjän toimenkuvalle, vaan se pitäisi pikemminkin nähdä hyödyllisenä apuvälineenä, aivan kuten elektroniset sanakirjat tai käännösmuisti. Mutta kuten Nuutila (2005: 102) huomauttaa, se että konekääntäminen on *mahdollista*, ei vielä välttämättä tarkoita, että se olisi sujuvaa tai tarkkaa. Ja vaikka konekääntäminen olisikin niin sanotusti automaattista, taustalla on aina ihminen. Ihminen on luonut konekäännösohjelmat ja syöttänyt sille kieliopit ja sanakirjat. Ihminen lopulta tarkistaa käännöksen ja mahdollisesti muokkaa sitä. Ainakin niin kauan kun kone ei osaa ajatella kuin ihminen.

## Lähdeluettelo

Anderson P.V., Brockmann R.J. & Miller, C.R. (toim.), 1983. *New Essays in Technical and Scientific Communication: Research, Theory, Practice*. Farmingdale, NY: Baywood.

Andrews D. (toim.), 1996. *International Dimensions of Technical Communication*. Arlington, VA: Society for Technical Communication.

Austermühl Frank, 2001. *Electronic Tools For Translators*. Translation Practices Explained. St. Jerome Publishing. Sivut 153–176.

Bergen David, 2004. Translation Technology and Translator Training. Teoksessa Tommola, Jorma (toim.) 2004. *Kieli, teksti ja kääntäminen. Language, Text and Translation*. Turun yliopisto. Painosalama Oy. Sivut 141–156.

Bowker L, 2002. *Computer-Aided Translation Technology: A Practical Introduction*. Ottawa. University of Ottawa Press.

Grönroos Mickel, 2005. Från översättningsminne till översättningsintelligens. *Språkbruk* 4/2005.

Hiirikoski Juhani, 2004. Lauserakenteet ja kääntäminen. Teoksessa Tommola, Jorma (toim.) 2004. *Kieli, teksti ja kääntäminen. Language, Text and Translation*. Turun yliopisto. Painosalama Oy. Sivut 31–46.

Hutchins John, 1986. *Machine Translation: Past, Present, Future*. Ellis Horwood Limited.

Hutchins John, 2005a. Current Commercial Machine Translation Systems and Computer-based Translation Tools: System Types and their Uses. *International Journal of Translation* Vol. 17, No. 1-2, Jan-Dec 2005. Sivut 5–38.

Häkkinen Kaisa, 1998. *Kielitieteen perusteet*. Suomalaisen kirjallisuuden seura. Tammer-Paino Oy, Tampere.

Häyrynen Minna, 1992. *Konekääntämisen keskeiset ongelmat*. Pro gradu –tutkielma. Tampereen yliopisto. Venäjän kielen kääntäminen ja tulkinta.

Itkonen Terho, 2000. *Uusi Kieliopas*. Kustannusosakeyhtiö Tammi, Helsinki. Sivut 87–93.

Kaivola Risto, 2006. Tietoinen kone tarvitsee lapsuuden. *Tiede* 6/2006.

Karlberg Peter, 2005. Handdator bryter språkbarriärer. *Teknik & Ekonomi* 24.11.2005.

- Karlsson Fred, 1993. Teoksessa *Tekoälyn ensyklopedia*. Gaudeamus, Hämeenlinna. Sivut 47–52.
- Knutsson Ola, 2002. Datorn som språkgranskare. *Språkvård* 1/2002. Sivut 26–33.
- Kulo Susanne, 1991. *Tietokone—tulevaisuuden kääntäjä?* Pro gradu -tutkielma. Tampereen yliopisto. Venäjän kielen kääntäminen ja tulkinta.
- Lammi Jari, 2004. *Luonnollisen kielen koneellinen ymmärtäminen*. Kandidaatintutkielma. Helsingin yliopisto. Tietojenkäsittelytieteiden laitos.
- Lintulahti Matti, 2003. Teknologian avulla voimme elää terveemmin, pitempään ja helpommin. *Apu-lehti* 45/03.
- National Academy of Sciences, National Research Council, 1966. *LANGUAGE AND MACHINES –COMPUTERS IN TRANSLATION AND LINGUISTICS*. A Report by the Automatic Language Processing Advisory Committee, Division of Behavioral Sciences, National Academy of Sciences, National Research Council. Washington D.C. Publication 1416.
- Nuutila Pertti, 2005. *Rough Machine Translation in the Communication Process*. Licentiate Thesis. University of Tampere. School of Modern Languages and Translation Studies.
- Pohjola Jukka, 2004. Jäikö kieli läppäriin väliin? *Tietokone-lehti* toukokuu 2004.
- Schank Roger C., 1996. “I’m sorry Dave, I’m afraid I can’t do that”: How could HAL Use Language. Stork, David G. (toim.) *HAL’s legacy: 2001’s computer as dream and reality*. MIT press. Sivut 171–190.
- Schmid Annemarie, 1991. Maschinенübersetzung: Möglichkeiten und Grenzen (Forschungsbericht). *Text-context* 6/1991. Sivut 115–156.
- Slocum Jonathan, 1984. A survey of machine translation: its history, current status, and future prospects; *Working Paper* LRC-84-3, Linguistic Research Center, University of Texas, Austin. Sivut 1–28.
- Sågvall Hein Anna, 2004. Maskinöversättningens mål och möjligheter. *Språkbruk* 4/2004. Sivut 7–11.
- Uljin J.M., 1996. *Translating the Culture of Technical Documents: Some Experimental Evidence*. Teoksessa Andrews D. (toim.) Sivut 69–86.
- Vehmas-Lehto Inkeri, 2002. Kääntäjän työ. Teoksessa Oittinen Riitta & Mäkinen Pirjo (toim.) 2002. *Alussa oli käänös*. Tampereen yliopistopaino Oy. Sivut 35–49.
- Westman Margareta, 1998. *Språkets lustgård och djungel*. Aktietryckeriet, Trondheim.



Yli-Jokipii Hilikka, 2004. Tekniikan kieli tutkimuksen ja kääntämisen näkökulmasta. Teoksessa Tommola, Jorma (toim.) 2004. *Kieli, teksti ja kääntäminen. Language, Text and Translation*. Turun yliopisto. Painosalama Oy. Sivut 81–94.

## Elektroniset lähteet

Arppe Antti, 1994. *Yritysstrategian erikoistyö Kielikone Oy:n kääntäjän työaseman markkinoinnista*. Saatavilla www-muodossa osoitteessa

<<http://www.ling.helsinki.fi/kit/2002s/ctl310com/991101/tausta-991101.html>> (Luettu 12.11.2005).

Carlson Lauri, 1999. *Konekin kääntää suomea*. Media-arkisto. Saatavilla www-muodossa osoitteessa <<http://mongbat.helsinki.fi/Media-arkisto/Studia/281099b.htm>> (Luettu 7.3.2006).

Euroopan komissio – Käännöspalvelu (SdT), 2002. Euroopan komission käännöspalvelun työvälineet ja työnkulku. Saatavilla pdf-muodossa osoitteessa

<[http://europa.eu.int/comm/dgs/translation/bookshelf/2002\\_tools\\_and\\_workflow\\_fi.pdf](http://europa.eu.int/comm/dgs/translation/bookshelf/2002_tools_and_workflow_fi.pdf)> (Luettu 28.3.2006).

Grönroos Mickel, 2006. *Käännösäly oppii käyttäjän kielestä*. Saatavilla www-muodossa osoitteessa

<<http://websrv2.tekes.fi/opencms/opencms/OhjelmaPortaali/Kaynnissa/FENIX/fi/system/utinen.html?id=2245&nav=Uutisia>> (luettu 11.4.2006).

Gupta Piklu & Schultze Mathias, 2000. ICT4LT Moduuli 3.5. *Kieliteknologia*. Saatavilla www-muodossa osoitteessa <[http://www.ict4lt.org/fi/fi\\_mod3-5.htm](http://www.ict4lt.org/fi/fi_mod3-5.htm)> (Luettu 12.11.05).

Hutchins John, 2005b. *The History of Machine Translation in a Nutshell*. Saatavilla www-muodossa osoitteessa <<http://ourworld.compuserve.com/homepages/WJHutchins/>> (Luettu 24.3.2007).

Järvinen Petteri, 1996. *Internet-muutostekijä*. WSOY. Saatavilla myös www-versiona osoitteessa <<http://www.pjoy.fi/kirjat/imuutos/luku02b.htm>> (Luettu 12.11.05).

Kay Martin, 2001. *Survey of the State of the Art in Human Language Technology*. Saatavilla www-muodossa osoitteessa <<http://cslu.cse.ogi.edu/HLTsurvey/indextop.html>> (Luettu 10.12.2006).

Korpela Jukka, 2003. Datatekniikka ja viestintä. *Jukan avoin tietosivusto*. Saatavilla www-muodossa osoitteessa <<http://www.cs.tut.fi/~jkorpela/orto.html>> (Luettu 2.2.2006).

Lisa Members in the News, 2005. *More Words Are Translated by Machine Translation Than by Humans*. Saatavilla www-muodossa osoitteessa

<<http://www.lisa.org/press/membernewsmore.html/383>> (Luettu 26.9.2005).

Master's Innovation Ltd, 2006. *Älykäs käännteknologia eli käänntsäly*. Saatavilla www-muodossa osoitteessa <<http://www.masterin.com/fi/itt.php>> (Luettu 1.2.2006).

Melby Alan, 1995. Why Can't a Computer Translate More Like a Person? Saatavilla www-muodossa osoitteessa <<http://www.fortunecity.com/business/reception/19/akmelby.htm>> (Luettu 28.3.2006).

Melby Alan, 2001. *Should I use machine translation?* TTT.org. Translation, Theory & Technology. Saatavilla www-muodossa osoitteessa <<http://www.ttt.org/theory/mt4me/index.html>> (Luettu 8.3.2006).

Saarikoski Harri M.T., 2004. *Semantic Webin ontologiat ja sanojen disambiguaatio*. Helsingin yliopisto / AAC Global. Älyä verkossa -symposio. Saatavilla pdf-muodossa osoitteessa <[http://www.seco.tkk.fi/events/2004/2004-09-02-web-intelligence/esitykset/harri\\_saarikoski\\_sanojen\\_merkitykset.pdf](http://www.seco.tkk.fi/events/2004/2004-09-02-web-intelligence/esitykset/harri_saarikoski_sanojen_merkitykset.pdf)> (Luettu 2.2.2006).

Seppänen Liisa, 2001. *Tietokone – luonnollisen kielen taitaja?* TKK, tietoliikenneohjelmistojen ja multimedian laboratorio. Saatavilla pdf-muodossa osoitteessa <<http://www.tml.tkk.fi/Opinnot/Tik-111.590/2001/paperit/seppanen.pdf>> (Luettu 12.11.2005).

Sågvall Hein Anna, 2001. *Automatisk översättning och översättningshjälpmedel*. Institutionen för lingvistik, Uppsala universitet. Saatavilla pdf-muodossa osoitteessa <[http://stp.ling.uu.se/educa/kurssidor/mosg/F1\\_Rosenbad.pdf](http://stp.ling.uu.se/educa/kurssidor/mosg/F1_Rosenbad.pdf)> (Luettu 12.11.2005).

Wikipedia. *Alta Vista*. Saatavilla www-muodossa osoitteessa <<http://fi.wikipedia.org/wiki/AltaVista>> (Luettu 1.1.2007).

# Liitteet

Aineisto 1

## **Nokia 7600 –puhelimien käyttöohjeesta kappaleesta 7. *Battery information: Charging and discharging***

Englanninkielinen lähtöteksti

### **Charging and discharging**

Your phone is powered by a rechargeable battery.

Note that a new battery's full performance is achieved only after two or three complete charge and discharge cycles!

The battery can be charged and discharged hundreds of times but it will eventually wear out.

When the operating time (talk-time and standby time) is noticeably shorter than normal, it is time to buy a new battery.

Use only batteries approved by the phone manufacturer and recharge your battery only with

the chargers approved by the manufacturer. Unplug the charger when not in use. Do not leave the battery connected to a charger for longer than a week, since overcharging may shorten its lifetime. If left unused a fully charged battery will discharge itself over time. Temperature extremes can affect the ability of your battery to charge.

Use the battery only for its intended purpose.

Never use any charger or battery which is damaged.

---

Saksankielinen käännös

### **Laden und Entladen von Akkus**

Ihr Telefon wird von einem aufladbaren Akku gespeist.

Beachten Sie, dass ein neuer Akku erst seine volle Leistung erreicht, nachdem er zwei- oder

dreimal vollständig ge- und entladen wurde!

Der Akku kann zwar mehrere hundert mal ge- und entladen werden, nutzt sich aber schließlich ab. Wenn die Betriebszeit (Sprechzeit und Stand-by-Zeit) im Vergleich zur normalen Betriebszeit deutlich kürzer ist, sollten Sie einen neuen Akku kaufen.

Verwenden Sie nur Akkus und Ladegeräte, die vom Telefonhersteller zugelassen sind.

Trennen Sie das Ladegerät von der Steckdose, wenn Sie es nicht verwenden. Lassen Sie den

Akku nicht länger als eine Woche am Ladegerät angeschlossen, da eine zu starke Aufladung

seine Lebensdauer verkürzen kann. Bei Nichtgebrauch entlädt sich ein voll aufgeladener Akku mit der Zeit.

Extreme Temperaturen können einen nachteiligen Einfluss auf das Laden Ihres Akkus haben.

Benutzen Sie den Akku nur für den vorgesehenen Zweck.

Benutzen Sie keine schadhafte Ladegeräte oder Akkus.

Babel Fish -käännös

### **Aufladung und Entladung**

Ihr Telefon wird durch eine nachladbare Batterie angetrieben. Merken Sie, daß die volle Leistung einer neuen Batterie nur nach kompletter Aufladung zwei oder drei erzielt wird und entladen Sie Zyklen! Die Batterie kann aufgeladen werden und entladene Hunderte Zeiten, aber sie trägt schließlich heraus. Wenn die Anzugszeit (Sprechzeit und Bereitschaftszeit) merklich kürzer als Normal ist, ist es Zeit, eine neue Batterie zu kaufen. Benutzen Sie die nur Batterien, die vom Telefonhersteller genehmigt werden und laden Sie Ihre Batterie nur mit den Aufladeeinheiten neu, die vom Hersteller genehmigt werden. Trennen Sie die Aufladeeinheit wenn nicht im Gebrauch. Lassen Sie nicht die Batterie angeschlossen an eine Aufladeeinheit für länger, als eine Woche, seit dem Überladen, seine Lebenszeit verkürzen kann. Wenn sie unbenutzt verlassen wird, entlädt sich eine völlig belastete Batterie Überschubzeit. Temperaturübermaße können die Fähigkeit Ihrer Batterie beeinflussen aufzuladen. Benutzen Sie die Batterie nur für seinen beabsichtigten Zweck. Benutzen Sie nie jede mögliche Aufladeeinheit oder Batterie, die beschädigt wird.

## Englanninkielinen lähtöteksti

The research project that was presented at the RSNA Meeting in Chicago (U.S.) last week by Dr. W. A. Bashir et al. got enormous press coverage and spread like wildfire over the worldwide web. The study confirmed with the help of a MRI technique that sitting in a slouched forward position or just upright in a 90degrees position was very unfavourable for the position of the nucleus and the intervertebral disk height.

By comparing a supine position with the position of a Formula 1 race car driver in his seat and then show how favourable this position is for the lumbar spine you can really make your point. Of course, leaning back in your chair with full support of backbones and shoulders is more relaxing then any more erect sitting posture.

<<http://www.prnewswire.co.uk/cgi/news/release?id=185951>>

---

## Saksankielinen käännös

Das Forschungsprojekt, das letzte Woche auf dem RSNA-Kongress in Chicago (USA) von Dr. W. A. Bashir et al. präsentiert wurde, erhielt enormes Medienecho und verbreitete sich wie ein Lauffeuer im Internet. Die Studie bestätigte mithilfe eines MRI-Verfahrens, dass Sitzen in gekrümmter Vorwärtsposition oder aufrecht in einer 90-Grad-Position äusserst ungünstig für die Kernposition und Bandscheibenhöhe ist.

Der Vergleich einer Rückenlage mit der Position eines Formel-1-Piloten in seinem Sitz und die anschliessende Demonstration, wie günstig sich diese Position auf die Lendenwirbelsäule auswirkt, vermitteln den Standpunkt wirklich gut. Natürlich ist ein Zurücklehnen mit voller Unterstützung des Rückgrats und der Schultern entspannender als jede aufrechte Sitzposition.

<<http://www.prnewswire.co.uk/cgi/news/release?id=185955>>

---

## Babel Fish -käännös

Das Forschung Projekt, das wurde dargestellt bei der RSNA Sitzung in Chicago (US.) letzte Woche durch Dr. W. A. erhielten Bashir et al. enorme Pressedeckung und -verbreitung wie wildfire über dem weltweiten Netz. Die Studie, die mit Hilfe einer MRI Technik bestätigt wurde, daß das Sitzen in a Vorwärtsposition oder gerade aufrecht in einer Position 90degrees slouched, war für die Position des Kernes und der intervertebralen Scheibe Höhe sehr ungünstig. Indem Sie eine nachlässige Position mit der Position eines Rennen-Autotreibers der Formel 1 in seinem Sitz vergleichen und zeigen Sie dann, wie vorteilhaft diese Position für den lumbalen Dorn ist, den Sie Ihren Punkt wirklich bilden können. Selbstverständlich zurück ist sich lehnen in Ihrem Stuhl mit voller

Unterstützung des Rückgrats und der Schultern mehr aufgerichtete Sittinglage dann sich entspannen.

## Englanninkielinen lähtöteksti

### **Aerospace and defence**

There is no more demanding an industry than Aerospace and Defence, where failure can be catastrophic and the pressures of flight safety, government regulation, globalisation and product/program complexity are unrelenting. Axon's experience, with consultants making a significant contribution to around 25% of SAP's go-lives in this sector world-wide, brings a unique combination of deep industry experience and SAP technology expertise to use SAP to transform Aerospace and Defence businesses such as:

- Cessna Aircraft
- Goodrich Corporation
- Lockheed Martin
- Northrop Grumman
- Pratt & Whitney
- Raytheon Corporation
- Rolls-Royce Aero Engines
- Sikorsky Aircraft
- US Army Armaments Research and Engineering

<[http://www.axon.co.uk/pages/sectors/aerospace\\_and\\_defence.asp](http://www.axon.co.uk/pages/sectors/aerospace_and_defence.asp)>

---

## Saksankielinen käännös

### **Luft-, Raumfahrt und Verteidigung**

Es gibt wohl keine anspruchsvollere Industrie als die der Luft- und Raumfahrt und Verteidigung – jeder Fehler kann katastrophale Auswirkungen haben und der Druck durch Faktoren wie Flugsicherheit, Regierungsaufgaben, Globalisierung und Produkt-/Programmkomplexität ist enorm. Die Erfahrung von Axon und seinen Consultants, die bei rund 25 % aller SAP Systemfreigaben in diesem Sektor weltweit einen großen Beitrag leisten, setzt sich aus umfassender Branchenkenntnis und technologischer SAP Expertise zusammen. Mithilfe von SAP transformieren wir Unternehmen aus der Luft- und Raumfahrtbranche und der Verteidigungsindustrie wie:

- Cessna Aircraft
- Goodrich Corporation
- Lockheed Martin
- Pratt & Whitney
- Raytheon Corporation
- Rolls-Royce Aero Engines

- Sikorsky Aircraft
- US Army Armaments Research and Engineering

<[http://www.axon.co.uk/deu/pages/sectors/aerospace\\_and\\_defense.asp](http://www.axon.co.uk/deu/pages/sectors/aerospace_and_defense.asp)>

---

## Babel Fish -konekäännös

### **Aerospace und Verteidigung**

Es gibt nicht mehr Verlangen einer Industrie, als Luftfahrt- und Verteidigung, in der Ausfall verhängnisvoll und der Druck der Flugsicherheit sein können, Regierung Regelung, globalisation und product/program Kompliziertheit unnachgiebig sind. Erfahrung des Axons, wenn die Berater einen bedeutenden Beitrag bis herum 25% bilden, von Saft gehen-lebt in diesem weltweiten Sektor, holt eine einzigartige Kombination der tiefen Industrieerfahrung und SAFT-Technologiesachkenntnis zum Gebrauch SAFT, um Aerospace- und Verteidigungsgeschäfte wie umzuwandeln:

- ? Cessna Flugzeug
- ? Goodrich Corporation
- ? Lockheed Martin
- ? Northrop Grumman
- ? Pratt U. Whitney
- ? Raytheon Corporation
- ? Rolls-Royce-Royce Aero Maschinen
- ? Sikorsky Flugzeug
- ? US Armee-Bewaffnungen Forschung und Technik



# Deutsche Kurzfassung

Universität Tampere  
Institut für Sprach- und Translationswissenschaften  
Translationswissenschaft (Finnisch-Deutsch)

MALMIVAARA, Anni: „Sokea idiootti“ – Möglichkeiten und Probleme des  
Maschinellen Übersetzens

MAGISTERARBEIT: 74 Seiten  
Anhang 6 Seiten, Deutsche Kurzfassung: 13 Seiten  
Mai 2007

---

## 1. Einleitung

MÜ bezeichnet die automatische Übersetzung von Texten aus einer Ausgangssprache in eine Zielsprache mit Hilfe eines Computerprogrammes. Dem Programm wird ein Text eingegeben und heraus kommt ohne menschliche Mithilfe eine Übersetzung in einer anderen natürlichen Sprache.

Maschinelle Übersetzung (MÜ) ist ein Thema, das immer mehr diskutiert wird. Es erregt viele Meinungen für und wider. Die meisten, vor allem Fachleute, haben eine Ansicht, die sie verteidigen. Laut Alan Melby (1995) glauben die einen, dass es keine entscheidenden Unterschiede zwischen Mensch und Maschine gibt. Sie meinen, Maschinen könnten in absehbarer Zeit die Qualität ‚menschlicher Übersetzungen‘ erreichen. Sie argumentieren, dass es nur eine Frage der Zeit sei, bis eine Maschine, die wie ein Mensch denken kann, entwickelt wird. Der Wissenschaft wird zugetraut, den Traum in absehbarer Zeit vollständig zu verwirklichen.

Die anderen dagegen sind der Meinung, dass durch eine Maschine die menschliche Intelligenz niemals erreicht werden kann. Sie führen an, dass die Leitungsvermittlung des Computers nicht so entwickelt ist wie das menschliche Gehirn (Kaivola 2006). Die immer wieder angekündigten großen technologischen Durchbrüche sind bisher nicht eingetreten.

Man hat zwar maschinelle Übersetzungen schon seit den 1970er Jahren machen können, aber die Qualität entspricht nicht den hohen Erwartungen. Der Mensch muss das Ergebnis der Maschine korrigieren, bevor man es publizieren kann.

## **2. Übersetzen und Maschinelles Übersetzen**

Das Übersetzen von einer natürlichen Sprache in eine andere ist ein komplizierter Prozess, der menschliche Intelligenz und Kreativität voraussetzt. Man kann den Prozess nicht einfach beschreiben. Auch wenn man zwei oder mehrere Sprachen ausgezeichnet beherrscht, heißt es nicht automatisch, dass man ein guter Übersetzer ist. Jorma Tammola (2004: 9) schreibt, dass das Übersetzen Sonderansprüche und Techniken verlangt, die nicht selbstverständlich ein gutes Sprachkenntnis folgen. Beim Übersetzen kann man nicht nur Worte aus einer Sprache in die andere übersetzen, sondern man muss auch die Kultur berücksichtigen. Am wichtigsten ist es, die Bedeutung der Ausgangssprache in die Zielsprache zu übertragen (Häyrynen 1992: 4–5). Eine zentrale Frage ist auch das Problem der Äquivalenz und die Kriterien der Vergleichung. Anna Säägvall Hein (2004: 7) sagt, dass die Qualitätskriterien sehr hoch sind, wenn man über Humanübersetzungen spricht. Es wird erwartet, dass die Übersetzung der gleichen Information und der gleichen sprachlichen Qualität wie der Ausgangstext entspricht.

John Hutchins (2005: 5–6) gibt sechs Gründe dafür, MÜ zu verwenden:

1. Es gibt einen großen Bedarf an Übersetzungsdiensten und nicht genug Humanübersetzer. Über das Internet kommen Leser ständig mit Texten in fremden Sprachen in Kontakt. Auch in der Industriebranche gibt es immer mehr Bedarf an Übersetzungen. Für Laien gibt es Online-Übersetzungsprogramme, zum Beispiel Babel Fish, die aber für den professionellen Übersetzer nicht qualitativ gut genug sind. Heute übersetzen die Übersetzungsmaschinen quantitativ schon mehr Texte als die Humanübersetzer (Lisa 2004).
2. Technische Texte sind langweilig. Humanübersetzer wollen diese langweiligen Texte nicht mehr übersetzen, sondern sie wollen ihre Kreativität erschließen. Die

- MÜ-Systeme produzieren eintönig logische Übersetzungen, im Gegensatz zum Menschen. Deswegen könnte man laut Hutchins (2005: 5–6) die MÜ-Systeme ohne Schwierigkeiten für technische Texte verwenden.
3. Die Terminologie muss konsistent sein. Die Menschen haben die Gewohnheit, zur Abwechslung oder manchmal auch aus Versehen, Synonyme oder anders klingende Formulierungen zu verwenden. Das ist aber bei technischen Texten nicht erwünscht. Das MÜ-System ist in dieser Hinsicht zuverlässiger als der Mensch, weil es die Terminologie konsistent hält. Laut Melby (1995) kann das MÜ-System einen technischen Text sehr wohl übersetzen, wenn die Terminologie nur ein beschränktes Fachgebiet behandelt und nuancenlos ist. Für den Humanübersetzer ist ein standardsprachlicher Text interessanter.
  4. Das MÜ-System kann längere Texte schneller übersetzen als der Mensch. Menschliche Übersetzung ist langsam. Viele Unternehmen brauchen Übersetzungen am liebsten heute und spätestens morgen. Wenn die Zeit knapp ist, sind sie auch bereit, auf Qualität zu verzichten. Untersuchungen zeigen, dass ein MÜ-System 50–70 % schneller übersetzt als der Mensch (Sågvall Hein 2004: 8). Besonders wenn größere Textmassen übersetzt werden, profitiert man merkbar von einem MÜ-System verglichen mit Humanübersetzen.
  5. Hohe Qualität ist nicht immer notwendig. Viele sind der Meinung, dass maschinelle Übersetzungen völlig untauglich wären, weil sie von einer niedrigeren Qualität sind. Oft kann man aber Maschinenübersetzungen sehr wohl benutzen, wenn sie nicht publiziert werden. Die Erwartungen sind oft auch nicht so hoch, wenn man von Maschinenübersetzungen spricht, denn die Maschine ist nicht intelligent oder kreativ. Es ist andererseits nicht eindeutig, was hohe Qualität bedeutet. Eine Maschinenübersetzung kann für den Fachmann ausreichend und plausibel sein, auch wenn der Laie nur die Grammatik- und Wortstellungsfehler sieht. (Kulo 1991: 75.) Die Qualität einer Übersetzung ist schwer objektiv messbar. Jede Übersetzung ist eine Interpretation, wer sagt was „richtig“ ist und was „falsch“?
  6. Unternehmen wollen die Kosten für Übersetzungen senken. Die Maschinenübersetzungen kommen in Frage, weil sie günstig sind. Die Kosten

eines MÜ-Systems bestehen aus Startinvestitionen, Betriebs- und Unterhaltskosten sowie Editions-kosten (Kulo 1991: 80). Manchmal sind niedrige Kosten wichtiger als Topqualität.

### 3. Geschichte des MÜ

Laut Hutchins (2005b) gab es schon – noch bevor es überhaupt Computer gegeben hat – die ersten Vorschlägen, natürliche Sprachen automatisch übersetzen zu lassen. In den 1940er Jahren wurde eine Maschine konstruiert, die mathematische Berechnungen automatisierte. Dadurch erschien es realistischer, dass auch eine natürliche Sprache maschinell übersetzt werden könnte. Es gab die Erwartung, dass MÜ in wenigen Jahren vollautomatisiert sein würde, die sprachlichen Probleme wurden stark unterschätzt. Eines der frühesten Projekte war ein Prototyp des russisch-englischen Übersetzungsprogramms, das im Jahr 1954 vorgestellt wurde. Die MÜ erlangte große Aufmerksamkeit. Auf maschinelle Übersetzung wurden hohe Hoffnungen gesetzt, und man nahm an, die Probleme würden bald gelöst sein. (Arppe 1994.)

Die ersten Übersetzungsprogramme basierten auf direkter MÜ, das heißt die Texte wurden Wort-für-Wort übersetzt mit minimaler Kontextanalyse. Das führte unvermeidbar zu schlechter Qualität. Der 1966 erstellte "ALPAC"-Bericht (National Academy of Sciences 1966) bescheinigte der MÜ grundsätzliche Unrealisierbarkeit und brachte mit einem Schlag die Forschung für fast 20 Jahre praktisch ganz zum Erliegen. Die wesentlichen Aussagen des Berichts waren, dass MÜ niemals die Qualität des Humanübersetzens erreichen wird, und das wäre auch nicht notwendig, denn es würde genug Humanübersetzer geben, um den Bedarf an Übersetzungen zu decken. (Arppe 1994.)

In der Mitte der 1970er Jahre wurden jedoch die ersten, immer noch im Gebrauch stehenden MÜ-Systeme in Betrieb genommen: *Systran* in der Europäischen Kommission und *Météo* in kanadischen Wetterstationen. (Arppe 1994 und Hutchins 2005b.)

## 4. Methoden

Die einzelnen Methoden unterscheiden sich erheblich voneinander. Ein völlig automatisiertes Programm macht alles selbstständig, ohne jegliche menschliche Interaktion. Dies ist aber heute noch ein Traum, denn völlig automatisierte Übersetzungsprogramme gibt es noch nicht – jedenfalls keine, die eine publikationsreife Qualität ohne Vor- und Nachbearbeitung schaffen könnten. (Austermühl 2001: 11.) Die drei Hauptmethoden der MÜ sind direkte MÜ (Direct Translation System), Transfer (Transfer Machine Translation) und Interlingua (Interlingual Machine Translation).

- **Direkte MÜ.** Die Wörter des Ausgangstextes werden mit Hilfe eines eingegebenen elektronischen Wörterbuchs Wort für Wort oder satzweise übertragen. Dies ist die älteste und einfachste MÜ-Methode. Man hat aber auf diese Methode in der Praxis schon frühzeitig verzichtet (Arppe 1994), weil die Ergebnisse von schlechter Qualität waren.
- **Transfer.** Die Transfer-Methode hat drei Schritte: Analyse der Ausgangssprache, Transfer der Satzstruktur und schließlich Generierung in die Zielsprache. Die eigentliche Übersetzung findet im zweiten Schritt statt. (Sågvall Hein 2004: 9.)
- **Interlingua.** Als Alternative zur Transfer-Methode kann man das Interlingua-Modell bezeichnen. Der Ausgangstext wird auf eine universale Repräsentationssprache, die sogenannte Interlingua, analysiert, aus der der Zieltext generiert wird. Die Interlingua enthält die Information des Ausgangstexts und im Prinzip ließe sich jede natürliche Sprache daraus reproduzieren (Arppe 1994.)

## 5. Sprache und MÜ

### *Elektronische Wörterbücher*

Ein wichtiger Teil der MÜ-Systeme sind die zweisprachigen Wörterbücher. Wie andere Wörterbücher sind auch elektronische Wörterbücher nicht vollständig; es fehlen Wörter oder die zielsprachliche Übersetzung ist nicht korrekt. Das führt zu falschen Übersetzungen. Als Suchwort wird normalerweise die Grundform verwendet; wenn in allen Wörterbüchern alle Formen abgespeichert wären, würden sie viel zu umfangreich werden. In den elektronischen Wörterbüchern, die das MÜ verwendet, müssen jedoch alle Formen enthalten sein. Man kann keine Formen weglassen, weil für die Maschine nichts selbstverständlich ist. Nuutila (2005: 21) schreibt, dass die Maschine die verschiedenen Flexionsformen nicht verstehen kann, sondern behandelt jede Form als eine eigene Einheit, was sehr kompliziert wird bei der Verwendung und Aktualisieren des Wörterbuches.

Die Sprache verändert sich schnell; neue Wörter werden erfunden und die Bedeutung alter Wörtern kann sich ändern. (Häkkinen 1998: 140 und Nuutila 2005: 19.) Es ist schwer, die Wörterbücher im gleichen Takt zu aktualisieren. Ein weiteres allgemeines Problem mit Wörterbüchern ist, dass es oft in Sprachen keine 100%igen Äquivalente gibt. In solchen Fällen kann das MÜ-System entweder alle möglichen Varianten in der Zielsprache angeben oder versuchen, die beste Alternative auszuwählen. (Hutchins 1986: 41–42).

### **5.1 Überarbeitung der Texte**

Es gibt bislang kein vollautomatisches MÜ-System, stattdessen müssen die maschinell übersetzten Texte durch einen menschlichen Übersetzer überarbeitet werden. Der Humanübersetzer muss als erstes die Basis für die Übersetzung, das Wörterbuch,

aktualisieren. Vor dem Übersetzen sollte man die Sprache des Ausgangstextes genau durchgehen, Idiome und Ellipsen entfernen und den Satzbau vereinfachen. Eventuell kann man auch eintragen, zu welcher Wortklasse bestimmte Wörter gehören, zum Beispiel ob das englische Wort *convict* hier ein Substantiv oder ein Verb ist. (Hutchins 2005: 15–16).

## **5.2 Schwierigkeit, die natürliche Sprache zu schildern**

Übersetzungsprobleme mit MÜ-Systemen resultieren nicht nur aus technischen Schwierigkeiten oder den Grenzen der Entwicklung. Als man mit einem frühen MÜs die Redewendung „*Out of sight - out of mind*“ ins Russische und dann zurück ins Englische übersetzte, war die Resultat „*Blind idiot*“ (Järvinen 1996).

Die Hauptschwierigkeit, mit der Übersetzungsprogramme zu kämpfen haben, ist die ungeheure Vielfalt der natürlichen Sprache. Eine natürliche Sprache kann niemals mit den Mitteln der Mathematik und der Logik vollständig berechnet oder geschildert werden. Eine natürliche Sprache ist unmessbar. Und eine Maschine kann nicht selbst denken, sondern sie weiß nur das, was ihr eingegeben wird. Eine große Schwäche ist, dass die Maschine keine Muttersprache hat. (Melby 1995.) Wenn man die menschliche Hirnfunktion einfach beschreiben könnte, hätte man wahrscheinlich schon ein vollautomatisches Übersetzungsprogramm konstruieren können (Nuutila 2005: 51).

Es ließ sich feststellen, dass das Übersetzen aus einer natürlichen Sprache in eine andere dem Computer erheblich schwerer fällt als die Berechnung komplizierter mathematischer Formeln. Auch die am meisten fortgeschrittenen Computerprogramme können heute nicht so gut wie ein Muttersprachler bzw. ein professioneller Übersetzer übersetzen. Automatisches Übersetzen ist schwer zu verwirklichen, und die Bedeutung von Wörtern und Phrasen sowie ihrer Verwendungsmöglichkeiten zu schildern, verursacht wahrscheinlich am meisten Probleme. (Häyrynen 1992: 9.) Auch wenn Ingenieure und Linguisten mit diesen Problemen kämpfen, scheint es heute unmöglich, dass es je ein vollständiges Übersetzungsprogramm geben wird (Järvinen 1996). Andererseits ist es jedoch unmöglich zu wissen, was die Zukunft bringt; heute ist schon vieles automatisiert,

was man vor einigen Jahrzehnten nicht hätte glauben können. Realistischer ist maschinelles Übersetzen in den Fachsprachen, wo das Vokabular eingeschränkt ist (Melby 1995).

### 5.3 Probleme der Sprache

Die natürliche Sprache enthält viele Worte, die mehr als eine Bedeutung haben (Homonymie, Polysemie etc.). Ein Homonym bezeichnet ein Wort, das für verschiedene Begriffe steht. Ein Beispiel ist das finnische Wort *kuusi*, das entweder „sechs“ oder „Fichte“ bedeuten kann. (Korpela 2003.) Homonyme sind für MÜ ein Problem, weil man nicht die Unterschiede der Bedeutungen im Wörterbuch einfach eintragen kann (Melby 1995).

Ein Polysem hat mehrere Bedeutungen, die untereinander ähnlich sind. Ein Polysem ist oft durch eine Metapher entstanden. In der finnischen Sprache gibt es zum Beispiel das Wort *selkä*, das entweder „Rücken“, „offene See“ oder „später Abend“ bedeuten kann. Austermühl (2001: 171) schreibt, dass in der MÜ Polyseme erst dann relevant werden, wenn die Bedeutung eines Wortes sich in der Übersetzung verändert.

Es ist nicht immer möglich, einen Unterschied zwischen Homonymie und Polysemie zu machen, und im Prinzip ist es für MÜ auch nicht wichtig, jedenfalls verursachen sie beim Übersetzen das gleiche Problem.

Der Aufbau der Sprache ist für MÜ-Systeme eine schwierige Frage. Wenn die Satzanalyse missglückt, muss das MÜ-System den Satz Wort-für-Wort übersetzen und die Struktur des Ausgangstexts wird in die Zielsprache übernommen. Austermühl (2001: 172) gibt als Beispiel den folgenden Satz und seine englische Übersetzung:

- C) „Schalten Sie einen ProPrint 100 am Schalter, der sich am PC befindet, aus.“
- D) „Switch off a ProPrint 100 at the switch, which is situated at the PC, from.“

Der Nebensatz beeinträchtigt die Satzanalyse, und die unfeste Zusammensetzung des Verbs *ausschalten* wird nicht erkannt. Wenn die Analyse misslingt, wird der Satz wortwörtlich übersetzt und das Resultat ist grammatikalisch falsch.



Auch Euphemisme, Metaphern, Idiome, Redewendungen und Metonymien verursachen Umstände beim Übersetzen.

Euphemismen werden oft verwendet, besonders wenn über Tabus die Rede ist oder ein empfindliches Thema angesprochen wird. Die Struktur und die Semantik der Sprache sind am weitesten voneinander entfernt, wenn die Sprache metaphorisch ist (Häkkinen 1998: 181).

Ein Idiom ist ein sinnbildlicher Ausdruck, der sich in die Sprache eingebürgert hat. MÜ-Systeme können besonders schwer mit Idiomen umgehen, und deswegen sollten sie eine eigenes Wörterbuch haben. Besser wäre es, wenn man schon bei der Ausgangstextsproduktion auf Idiome völlig verzichten könnte. (Austermühl 2001: 172.) Feste Redewendungen, wie „*Here you are*“ werden im Normalfall nicht direkt übersetzt, was die Maschine natürlich nicht wissen kann.

Eine Metonymie ist die Anwendung eines Konzepts anstelle der eigentlichen Bezeichnung (Korpela 2003). Man sagt zum Beispiel „*Er liest Shakespeare*“ wenn man eigentlich meint, dass er Shakespeares Werke liest. Auch dies ist für eine Übersetzungsmaschine nicht möglich zu wissen.

Das Verb ist das wichtigste Wort des Satzes, woneben auch die Bezugnahme der Pronomina Schwierigkeiten hervorruft. Besonders bei Sprachen wie dem Deutschen, wo ein Pronomen auch das Genus ausdrückt, wird es schwer, dies dem Computer zu schildern. (Austermühl 2001: 173.) Zum Beispiel bezieht sich im folgenden Satz das englische Pronomen „it“ auf verschiedene Objekte:

”The monkey ate the banana because it was hungry.”  
“The monkey ate the banana because it was ripe.”

Um eine korrekte Übersetzung machen zu können, müsste die Maschine linguistisches Wissen ausnutzen, das sie aber nicht hat. (Austermühl 2001: 173.)

## 6. Beispiele von MÜ-Programmen

Die Europäische Union (früher die Europäische Gemeinschaft) hat maschinelle Übersetzungen am längsten verwendet, wenn man die United States Air Force (USAF) einmal unbeachtet lässt. Die EU verwendet wahrscheinlich am meisten Maschinenübersetzungen in der Welt. Sie werden von Übersetzern gebraucht, die aus einer maschinellen Rohübersetzung eine publikationsreife Version machen. Nicht nur Übersetzer, sondern auch andere Beamte verwenden MÜ, wenn sie Dokumente in fremden Sprachen verstehen müssen. Basierend auf der Rohübersetzung treffen sie die Entscheidung, ob sie eine Humanübersetzung brauchen. (Hutchins 2005: 11.)

Zwei sogenannte ernsthafte MÜs sind EUROTRA und SYSTRAN. EUROTRA ist ein zwischen 1985 und 1993 entwickeltes Übersetzungsprogramm, das neun EU-Sprachen beherrscht (Carlson 1999). SYSTRAN stammt aus den 1950er Jahren, und wird von Übersetzern und Verwaltungsbehörden der EU verwendet (die Europäische Kommission, 2002). SYSTRAN übersetzt zwischen dem Englischen, Französischen, Italienischen, Deutschen, Holländischen, Spanischen, Griechischen und Portugiesischen (Hutchins 2005: 11).

Es gibt auch eine Menge kommerzielle, kostenlose Übersetzungsprogramme, zum Beispiel WebTranSmart, Sunda und Babel Fish. Die kommerziellen Programme können allerdings keine hohe Qualität garantieren.

## 7. Babel Fish und die Forschungsergebnisse

Als konkrete Beispiele eines Übersetzungsprogramms wählte ich drei technische Textstücke aus, die ich mit dem Babel Fish-Übersetzungsprogramm automatisch

übersetzen ließ. Babel Fish ist im Internet unter die Adresse <http://babelfish.altavista.com/> zu finden.

Die Ausgangstexte sind auf Englisch geschrieben, und als Vergleich hatte ich deutsche Humanübersetzungen. Das erste Textbeispiel ist ein Stück aus ein Bedienungsanleitung von Nokia 7600 Mobiltelefon, das zweite Beispiel ein Auszug aus ein Artikel von PR Newswire Association LLC und das dritte Beispiel ein Auszug aus Axons Unternehmensprospekt. Die drei bis zehn Sätze langen Textstücke wählte ich, weil sie vorwiegend standardsprachlich sind, aber trotzdem auch technische und kaufmännische Fachausdrücke enthalten. Die Texte und die Humanübersetzungen sowohl als auch die Maschinellen Übersetzungen sind im Anhang nachzulesen.

Die größte Schwäche des Forschungsergebnisses sind Wortstellungsfehler, falsche Übersetzungen und nicht übersetzte Wörter. Mit gutem Willen kann man sich ungefähr denken, was ursprünglich gemeint war. Diese Fehler waren jedoch zu erwarten. Ein Teil der Fehler hätte man vielleicht vermeiden können, indem man den Ausgangstext bearbeitet hätte, zum Beispiel durch Kürzen langer Sätze, Korrektur der Schreibfehler, Einsetzen von Kommas oder Vereinfachung der Satzstruktur. Solche Besserungen habe ich allerdings nicht vorgenommen, weil ich ein möglichst typisches Bild von MÜ bekommen wollte.

In den Beispielsübersetzungen gibt es deutliche Wortstellungsfehler und falsche Übersetzungen, die ein Humanübersetzer niemals machen würde. Auch technisch falsche Lösungen kommen vor, wie zum Beispiel beim Übersetzen von Namen: *Dr. W. A. Bashir* ist in der Maschinellen Übersetzung *Dr. W. A. erhielten Bashir* (Beispiel 2) und *Rolls-Royce Aero Engines* ist *Rolls-Royce-Royce Aero Maschinen* (Beispiel 3).

Wenn Babel Fish ein Wort nicht erkennt, bleibt es einfach ohne jede Fehlermeldung unübersetzt. Dies ist ein großer Defekt, denn wenn der Benutzer die Zielsprache gar nicht oder nur ein bisschen kennt, kann es passieren, dass er die fremdsprachlichen Wörter nicht bemerkt. Zum Beispiel gibt es in dem Beispiel 1 ein Fall, wo das englische Wort *Wildfire* nicht ins Deutsche übersetzt ist. Der englischsprachliche Benutzer könnte denken, dass es im Deutschen auch so heißt.

Ein weiteres Problem mit Babel Fish ist, dass keine Fachwörterbücher zu wählen sind. In Beispiel 1 sollte man etablierte Fachtermini für die Mobiltelefonindustrie verwenden, aber aus dem allgemeinen Wörterbuch kann Babel Fish sie nicht finden. Babel Fish benutzt zum Beispiel für *Operating time* die Übersetzung *Anzugszeit*, wobei die Humanübersetzung *Betriebszeit* ist. Und statt *Stand-by-Zeit* gibt Babel Fish *Bereitschaftszeit* als Übersetzung.

Falsche Wortstellungen kommen in den Textauszügen auch vor, zum Beispiel *Lassen Sie nicht die Batterie angeschlossen an eine Aufladeeinheit für länger, als eine Woche, seit dem Überladen, seine Lebenszeit verkürzen kann.*

In Beispiel 2 gibt es das Fachwort *lumbar spine*, das auf deutsch in der Humanübersetzung *Lendenwirbelsäule* ist, aber weil Babel Fish nur das allgemeine Wörterbuch hat, lautet die Maschinelle Übersetzung *lumbaler Dorn*.

Die automatische Übersetzung von Babel Fish kann man selbstverständlich nicht so kritisch analysieren wie Humanübersetzungen. Von einer Maschine wird keine hohe Qualität erwartet, nur eine Rohübersetzung. Die Forschungsergebnisse geben Hinweise darauf, dass die Maschinelle Rohübersetzungen in Ordnung sind, obwohl ich an einigen Stellen unsicher war, ob ein Leser der Zielkultur die Übersetzungen von Babel Fish überhaupt verstehen würde.

Die Nachbearbeitung des Textes in einen publikationsreifen Zustand ist sicher möglich, aber ich denke, dass es hier nicht der Mühe Wert ist. Es wäre besser, die ganzen Texte von Anfang an ‚menschlich‘ übersetzen zu lassen. Ein erfahrener Übersetzer, der mit der Terminologie bekannt ist und den Kontext kennt, würde sicherlich eine bessere Qualität herstellen, auch ohne eine maschinelle Rohübersetzung. Manchmal kann es sogar passieren, dass die maschinell erzeugte Rohübersetzung den Humanübersetzer verwirrt, so dass eine bessere Lösung unentdeckt bleibt, wenn das MÜ-System eine schlechter Lösung bietet.

Ein maschinelles Übersetzungsprogramm übertrifft an Schnelligkeit und preislicher Günstigkeit den Humanübersetzer: Babel Fish kann große Textmengen übersetzen und der Gebrauch ist für den Benutzer kostenlos.

## **8. Schlussfolgerungen**

Die Informationsgesellschaft hat die Arbeit des Übersetzers verändert. Informationen stehen viel leichter zur Verfügung als früher und für menschliche Übersetzer gibt es wertvolle technische Hilfsmittel. Elektronische Werkzeuge übernehmen die menschliche Arbeit nicht (völlig), sondern stützen sie. Es gibt heute keine Computerprogramme, die die menschliche Übersetzungsarbeit vollkommen ersetzen könnten, aber Übersetzungsprogramme von hoher Qualität trägt deutliche Vorteile bei. Der Übersetzer kann mit Hilfe von elektronischen Hilfsmitteln seine Arbeit schneller und besser machen.

In der Zukunft bleiben menschliche Eingriffe weiterhin erforderlich. Auch Nuutila (2005: 102) deutet darauf hin, dass der Mensch immer im Hintergrund bleibt, selbst wenn MÜ voll automatisiert wäre. Der Mensch hat die Maschinen tatsächlich gebaut und sie programmiert. Er gibt die Daten ein und erstellt schließlich die endgültige Version. Wenigstens so lange es keine Maschinen gibt, die so denken können wie der Mensch.

