

Katsaus

EI YHTÄ AINOAA POLKUA – SUOMA- LAISIA KOKEMUKSIA MATKALLA KIELI- TEKNOLOGISESTA TUTKIMUKSESTA LIIKETOIMINTAAN

Antti Arppe
Yleinen kielitiede,
Helsingin yliopisto
aarppe@ling.helsinki.fi

Tässä katsauksessa tarkastellaan kuinka lingvistikä ja kieliteknologista tutkimusta on kaupallistettu Suomessa. Katsauksessa havaitaan, että akateemisella tutkimuksella on ollut merkittävä taustarooli kieliteknologisen liiketoiminnan syntyemisessä Suomessa, vaikkakin yrityksiä on syntynyt myös muiden taustojen kautta. Katsauksessa käydään läpi aikajärjestyksessä kielitieteen ja kieliteknologian alan suomalaisten tutkijoiden ja tutkijaryhmien tutkimustuloksia sekä -projekteja, jotka ovat johtaneet kaupallistettuihin tuotteisiin tai yritysten perustamiseen. Tämän jälkeen tarkastellaan, minkälaisia strategioita nämä akateemisen taustan omaavat kieliteknologiayritykset ovat valinneet, minkälaisia eri vaiheita ne ovat kokeneet ja kuinka ne ovat ylipäänsä onnistuneet kehityksessään ja kasvussaan. Katsauksen havainto on, että nämä yritykset ovat menestyneet varsin vaihtelevasti ja vaihtelevin konseptein, eikä yhtään kansainvälistä läpimurtoa kansallisen tason menestystarinoista huolimatta ole vielä nähty. Jälkikäteen arvioituna vaikuttaa siltä, että ne yritykset jotka ovat nähneet itsensä ensisijaisesti ohjelmistoyrityksinä ovat pidemmällä tähtäimellä menestyneet parhaiten. Lopuksi tarkastellaan yritysten nykytilannetta ja arvioidaan minkälaisia tulevaisuuden mahdollisuuksia ja haasteita niillä on.

Kieliteknologisella liiketoiminnalla on Suomessa vahvat akateemiset juuret. Enemmistö suomalaisista ohjelmistoyrityksistä, joiden päätoimialana on kieliteknologisten ohjelmistotuotteiden kehitys ja myynti, voi johtaa alkunsa yksittäisiin tutkijoihin tai tutkijaryhmiin suomalaisissa yliopistoissa ja korkeakouluissa. Muuten nämä yritykset eroavatkin toisistaan melkoisesti. Eritoten historialtaan 'vanhempien' yritysten valitsemat strategiat ja polut akateemisista start-up-yrityksistä vakaiksi ja kannattaviksi, kaupallisiin periaattein toimiviksi yrityksiksi ovat vaihdelleet paljon. Toiset näistä yrityksistä ovat löytäneet itselleen selkeän ja kannattavan liiketoiminta-alueen, kun taas toisilla on vihreä oksa vielä löytämättä. Jälkikäteen arvioituna näyttää siltä, että ne yritykset jotka ovat osanneet nähdä itsensä ensi sijassa ohjelmistoyrityksinä ja vasta toissijaisesti kieliteknologiayrityksinä ovat menestyneet parhaiten. Kansallisen tason menestystarinoista huolimatta ei kuitenkaan vielä olla päästy näkemään suomalaisen kieliteknologiayrityksen kansainvälisen luokan läpimurtoa.

AKATEEMISET JUURET

Suomalainen kielitieteilijä tai kieliteknologi pitää äidinkieltään usein kallisarvoisena synnyinlahjana. Suomen rakenteellinen rikkaus, mikä näkyy taivutusmuotojen, johdosten ja yhdyssanojen runsaudessa, tarjoaa virkistävän erilaisen näkökulman kielentutkimukseen, missä rakenteeltaan yksinkertaisempi englanti on käytännössä ollut hallitsevassa asemassa niin pitkään kuin koko ala on ollut olemassa. Kun esimerkiksi englantia varten pystyy kehittämään yksinkertaisen kielenkäsittelyohjelmiston kuten oikolukijan käytännössä listaamalla ja kompressoimalla yleisimmät sata tuhatta sanaa, suomen kohdalla pitäisi samaa tekniikkaa noudattaen listata jos ei satoja niin vähintään kymmeniä miljoonia eri sanamuotoja, jotta vastaava oikolukija olisi yhtä kattava.

Yhteystiedot: Antti Arppe, Yleisen kielitieteen laitos
Helsingin Yliopisto
PL 9, Siltavuorenpenger 20 A
FIN-00014 Helsingin yliopisto

Teoriat ja ratkaisut, jotka toimivat englannin kieleen, eivät tyypillisesti sovellu suomen kieleen, ja tästä syystä suomalaisilla tutkijoilla oli 1970-luvulla motivaatio kehittää menetelmiä, joilla voitaisiin tehokkaasti mallintaa suomea tietokoneohjelmistoja varten. Näiden pioneerien eturintamassa olivat Kimmo Koskenniemi ja Fred Karlsson Helsingin yliopistosta ja Harri Arnola Teknillisestä korkeakoulusta.

Kimmo Koskenniemi esitti väitöskirjassaan vuonna 1983 ns. kaksitasomallin (Two-level Morphology, lyh. TWOL), jota voitiin käyttää periaatteessa minkä tahansa ihmiskielen morfologisen rakenteen mallintamiseen. Tämän mallin käytännön toteutuksena hän esitti sen mukaan tehdyn kuvauksen suomen kielestä, millä voitiin analysoida suomenkielisiä sanoja tietokoneella. Kansainväliseltä akateemiselta yhteisöltä kesti tämän jälkeen muutamia vuosia sulatella ja hyväksyä Koskenniemen innovaatio, mutta Suomessa IT-teollisuus otti sen vastaan innolla.

Tämä kiinnostus johti yhteistyöprojektien sarjaan Koskenniemen kollegan, Fred Karlssonin johtaman Helsingin yliopiston Yleisen kielitieteen laitoksen ja yritysten kuten Nokian ja Suomen IBM:n välillä. Lopulta tämä toiminta siirrettiin yritykseen nimeltä Lingsoft Oy, joka perustettiin vuonna 1986. Omana panoksenaan Karlsson kehitti kaksitasokuvauksen ruotsin kielestä 1980-luvun lopussa, ja vuonna 1990 hän esitti uuden kieliteknologisen formalismin, rajoitekieliopin, joka tunnetaan parhaiten englanninkielisellä nimellään Constraint Grammar (CG).

Sekä Karlssonin että Koskenniemen näkemyksenä on kieliteknologian tutkimuksessa ollut edetä askel kerrallaan, käsittelemällä perusteellisesti kielen yksinkertaisempi taso kuten morfologia ennen siirtymistä monimutkaisemmalle tasolle, kuten syntaksiin tai semantiikkaan. Samoin ajatteli myös Harri Arnola, joka oli tekoälyn tutkija ja opettaja Teknillisessä korkeakoulussa. Arnolan tavoitteena oli toteuttaa tämä tavoite suomen osalta niin morfologisen kuin syntaktisen analyysin suhteen. SITRAn rahoituksella Arnola käynnisti vuonna 1982 Kielikone-projektin, jonka käytännön tarkoituksena oli kehittää suomen-

kielinen tietokantojen hakurajapinta.

1980-luvun lopulla tästä alkuperäisestä päämäärästä kuitenkin luovuttiin, kun silloinen Telenokia ehdotti konekäännöstä suomesta englantiin projektin uudeksi käytännön sovellusalueeksi. Samoihin aikoihin projektin puitteissa olivat syntyneet omat mallit suomen morfologiasta ja syntaksista, joidenka kaupallistamista varten perustettiin vuonna 1987 Kielikone Oy.

Ensimmäisenä oman yrityksen perusti Helsingin yliopiston tutkija Olli Blåberg, joka oli sekä Koskenniemen että Karlssonin oppilas. Hänen 1987 perustamansa yritys Lanser Datan ensimmäinen projekti oli kehittää säätiedotteiden konekäännösjärjestelmä suomesta ruotsiin.

HY:n tutkijat Lauri Carlson ja Krister Lindén kehittivät vuosina 1988–1992 konekäännösjärjestelmää englannista suomeen yhteistyössä (Kimmo Koskenniemen veljen) Seppo Koskenniemen kanssa, joka oli jo 1970-luvulta kehitellyt kieliteknologisia ratkaisuja IBM:n Dictionaries and Linguistic Resources -yksikössä asemapaikkanaan Suomi. Tätä Mentor/F-nimistä järjestelmää ei kuitenkaan koskaan lopullisesti tuoteistettu. Yhtenä seurauksena oli kuitenkin myöhemmin, että Koskenniemi pysyvän professorin saatuaan pyysi Lindeniä vetämään Lingsoftia yrityksen ensimmäisenä kokopäiväisenä työntekijänä vuonna 1992. Konekäännöksen lisäksi Seppo Koskenniemi oli osallistunut lingvististen kuvausten laatimiseen morfologisesti rikkaita kieliä varten, kuten suomele, turkille, unkarille, tshekille ja puolalle. Näitä IBM on kaupallisesti hyödyntänyt tiedonhallintatuotteissaan.

Arto Anttila, Juha Heikkilä, Timo Järvinen, Pasi Tapanainen ja Atro Voutilainen osallistuivat ensiksi Karlssonin rajoitekieliopin soveltamiseen englantia varten, mitä teknologiaa ENGCG-nimisenä jakeli Lingsoft. 1990-loppupuolelle tultessa kolme näistä tutkijoista, Voutilainen, Tapanainen ja Järvinen jatkoivat yhteistyötään ja kehittivät uuden kieliopin kuvausformalismin, funktionaalisen dependenssi-kieliopin (Functional Dependency Grammar FDG), joka monessa suhteessa muistuttaa rajoitekielioppia mutta on tietojenkäsittelyteoreettisessa mielessä ilmaisuvoimaisempi. Tämän teknologian he päättivät

kaupallistaa omalla riskillään vuonna 1997 perustamassaan Conexor Oy:ssä.

Kielikone-projektiin osallistui lukuisia tutkijoita, joista pitänee mainita ainakin Timo Honkela ja Aarno Lehtola. Honkela kiinnostui 1990-luvulla itsejärjestäytyvistä kartoista, (englanniksi Self-Organising Maps eli SOM), jotka oli alunperin kehittänyt akateemikko Teuvo Kohonen. Omassa väitöskirjassaan Honkela tutki SOM-karttojen käyttöä luonnollisten kielten käsitteilyyn, ja yhdessä Kohosen, Samuel Kasken ja Krista Laguksen kanssa hän kehitti WEBSOM-projektissa menetelmiä ja työkaluja visuaaliseen tiedonhakuun. Tämän tutkimuksen pohjalta perustettiin vuonna 1997 Gurusoft, jonka tavoitteena on tarjota ratkaisuja tekstimuotoisen tiedon hallintaan. Aarno Lehtola puolestaan teki Kielikone-projektin jälkeen yhteistyötä Olli Blåbergin kanssa Lanser Datassa. Tämä jälkeen hän siirtyi VTT:een, missä on osallistunut tai vetänyt useita kieliteknologiaprojekteja, jotka ovat liittyneet mm. elektroniseen kaupankäyntiin (eBusiness).

Kielenkäsittelyn parissa toimivista yrityksistä löytyy myös sellaisia, joilla ei ole varsinaista akateemista taustaa. TimeHouse, joka on toteuttanut useita erilaisia ohjelmistoprojekteja vuodesta 1985 lähtien, kehitti jo vuonna 1991 yhteistyössä Teknillisen korkeakoulun Akustiikan ja äänenkäsittelytekniikan laboratorion kanssa suomenkielisen puhesyntetisaattorin nimeltä MikroPuhe. Samainen yritys on myös kehittänyt THText-nimisen elektronisen julkaisujärjestelmän, jota mm. WSOY on käyttänyt elektronisissa sanakirjoissaan. Vuonna 1987 perustettu Promentor Solutions on puolestaan pitkään ja menestyksekkäästi kehittänyt ja markkinoinut kielienopetusohjelmistoja ja kielikursseja. Sandstone.fi Oy, joka on perustettu 1997, tarjoaa kieliteknologiaratkaisuja langattomia verkkoja ja Internetiä varten. Tällaisia ovat esimerkiksi elektroniset sanakirjat, joita voidaan käyttää matkapuhelimessa. Vuonna 1999 perustettu Master's Innovations Oy on puolestaan erikoistunut tietokoneavusteisen käännösteknologian ja tiedonhakujärjestelmien kehittämiseen.

Lisäksi on muitakin yrityksiä, joiden kytkös kieliteknologian kaupallistamiseen on jäänyt löy-

hemmäksi, esimerkiksi siten että ne ovat vieneet läpi ainoastaan yhden kieliteknologiaprojektin muun toimintansa ohessa. Esimerkkinä tällaisesta yrityksestä käy vuonna 1996 perustettu Republica, joka on kehittänyt KONE Oyj:lle kontrolloituun kieleen perustuvan tuotedokumentoinnin prosessin. Tästä löytyy erillinen case-kertomus (<http://www.csc.fi/euromap/case/kone.phtml.fi>).

MENESTYKSEEN ON MONTA POLKUA

Varhaisimmat suomalaiset kieliteknologiayritykset lähtivät liikkeelle varsin verkkaisesti. Erittymisesti Lingsoft oli ja on ollut näihin päiviin asti klassinen esimerkki riskirahoitusvetoista liiketoimintaa edeltäneestä ajasta, sillä yritys on rahoittanut kasvunsa ja kehityksensä pelkästään tulorahoituksensa varassa. Kielikoneella on ollut ulkopuolisia rahoittajia, ensin SITRA ja myöhemmin TEKES, mutta tässä liikkuneet rahasummat ovat täysin eri kertaluokkaa kuin parin vuoden takaisissa IT-alan yritysten rehvakkaissa riskirahoituskierroksissa.

Aina vuoteen 1995 saakka Lingsoft toteutti yksittäisiä projekteja tai lisensoi teknologiaansa asiakkaille sekä Suomessa että ulkomailla. Vuoteen 1992 saakka yritys toimi itse asiassa täysin projektipohjalla, eikä sillä ollut pysyviä kokopäiväisiä työntekijöitä. Näin ollen Lingsoftin liikevaihto liikkui juuri sadan tuhannen euron alapuolella kyseisenä aikana. Merkittävä poikkeus tässä suhteessa tapahtui vuonna 1988, jolloin yritys lisensoi ensin suomen tavutusalgoritminsa ja sittemmin suomen oikolukijan WordPerfectille.

Lanser Data sai ensin lisensoitua suomen oikolukumoduulin Lotuksen AmiPro-tekstinkäsittelyohjelman yhteyteen vuonna 1990 ja teki samasta moduulista sopimuksen myös Microsoftin kanssa vuonna 1993.

Kielikone kaupallisti suomen kieliteknologiansa MORFO-nimisenä erillisenä oikolukuohjelmana ja VIRKKU-nimisenä kielioipintarkistusohjelmana. Vaikkakin VIRKKU oli ensimmäinen laatuaan millekään pohjoismaiselle kielelle, MORFO oli tuote josta yritys sai suurim-

man osan myyntituloistaan aina 1990-luvun alkupuolelle asti. Kielikone lisensoi suomen kielen kuvauksensa myös amerikkalaiselle Inso-yritykselle, joka lisensoi sen edelleen suomen oikolukijaksi kehitettynä Microsoftille 1990-luvun keskivaiheen paikkeilla.

Varovaisenpuoleisen alun jälkeen sekä Lingsoft että Kielikone saivat enemmän vauhtia ja volyymiä liiketoimiinsa 1990-luvun puoliväliin tultaessa. Kielikone oli koostanut keskeisen tuotekehitysprojektinsa, suomi–englanti–konekäännösjärjestelmän, ohessa elektronisen suomi–englanti–sanakirjan. Tästä resurssista tuli myöhemmin menestyksen siemen, sillä vuonna 1992 Käännöskone Oy, Kielikoneen tytäryhtiö, joka erikoistui kämmenkokoisiin sanakirjalaitteisiin, ryhtyi jakelemaan kyseistä sanakirjasisältöä laitteessaan. Käännöskone fuusioitiin myöhemmin takaisin emoyhtiöön.

Elektronisista sanakirjoista kehittyi Kielikoneelle tuotenimellä MOT nopeasti täysimittainen liiketoiminta-alue, joka ajan myötä tuotti valtaosan yrityksen myyntituloista. Kielikone ryhtyi määrätietoisesti sekä koostamaan itse kaksikielisiä sanakirjasisältöjä suomesta eurooppalaiseen valtakieliin ja päinvastoin että lisensoimaan näitä sisältöjä arvostetuilta kustantajilta. Lisäksi yritys kehitti tehokkaan tuotteistamisprosessin, mitä noudattaen sanakirjasisällöt saatiin nopeasti sisällytettyä valmiisiin loppuasiakastuotteisiin. Paljon merkitsi myös, että Kielikone pystytti hyvin toimivan markkinakanavan, jonka kautta se pystyi jakelemaan tuotteitaan erityisesti yritysasiakkaille.

Lingsoft onnistui vuonna 1995 sopimaan Microsoftin kanssa merkittävästä projektista, missä se lokalisoi Microsoftin Answer Wizard-apujärjestelmän muutamalle kielelle. Kun Lingsoft oli näin saanut jalan oven väliin, se pystyi 1990-luvun loppupuolella hankkiutumaan Microsoftin suhteen pohjoismaisten kielten ja saksan kielentarkistusmoduulien liki hovihankkijaksi. Lingsoft toimitti Microsoftille kaikille näille kielille kielentarkistusmoduulien lähes koko paletin – oikoluvun, tavutuksen, synonyymisanaston ja kieliopintarkistuksen – aloittaen suomen oikoluvusta ja tavutuksesta vuonna

1996. Avain tähän menestykseen löytyi kaikista niistä kielellisistä resursseista, joita Lingsoftissa oli kehitetty yrityksen projektivaiheen aikana, alkaen Koskenniemen suomen kaksitasokuvauksesta ja saksan mallista, joista jälkimmäisellä Lingsoft voitti vuonna 1994 järjestetyn Morpholympics-kilpailun. Molemmat kuvaukset, unohdamatta lukemattomia muita, toimivat pohjana niissä oikolukijoissa, tavuttimissa ja synonyymisanakirjoissa, jotka Lingsoft lisensoi Microsoftille - usein monia vuosia niiden alkuperäisen kehittämisen jälkeen.

Lingsoftin kannalta erittäin tervetullut kehityskulku liittyi rajoitekielioppiin (CG) ja sen sovelluksiin eri kielille, jotka eivät pitkälti akateemisesti profiloituneesta asiakaskunnasta johtuen olleet valitettavasti synnyttäneet suuria tuloja 1990-luvun aikana. Vuonna 1997 varmistui, että rajoitekieliopin formalismia käytännössäkin kannattaisi käyttää kieliopillisten virheiden havaitsemiseen, minkä seurauksena yritys ryhtyi kehittämään ruotsin kielen kieliopintarkistinta. Kun tämä oli lisensoitu Microsoftille vuonna 1998, siitä tuli ensimmäinen kaupallisesti saatavilla oleva tämän tyyppinen työkalua ruotsia varten. Tämä tuotekehitysprosessi ja lisensointijärjestely toistettiin suomen, tanskan ja norjan osalta vuosina 2000–2001.

Jälkikäteen on mielenkiintoista huomata, ettei kumpikaan menestystarina johtunut suoraan alkuvaiheessa valitusta strategiasta. Pikemminkin sekä Kielikone että Lingsoft sattuiivat olemaan oikeissa paikoissa oikeaan aikaan oikeiden resursien ja teknologioiden kanssa, ja lisäksi molemmat yritykset älyivät tarttua eteensä ilmestyneisiin mahdollisuuksiin.

MENESTYKSEN TIE ON VÄLILLÄ KUOPPAINEN

Tultaessa 1990-luvun lopulle Lingsoftilla oli Microsoftin eurooppalaisten alihankkijoiden joukossa kirkkaasti suurin lisensoitu kielentarkistus-työkalujen kielivalikoima. Ainoastaan yksi toinen alihankkija, unkarilainen MorphoLogic, lisensoi Microsoftille useamman kuin yhden kielen kie-

lentarkistusmoduuleja. Sitä mukaa kuin Lingsoftin tukemien kielten määrä oli kasvanut, yrityksen henkilökuntamäärä ja liikevaihto kasvoivat vastaavasti. Yrityksen liikevaihto kolminkertaistui vuosina 1995-1998 hieman yli kahteen miljoonaan euroon. Lingsoftin johdossa tajuttiin varhaisessa vaiheessa, että yhden suuren asiakkaan varaan nojaaminen olisi selvä riski pitemmällä aikavälillä, ja että yritys ei voisi turvata jatkuvuuttaan laajentamatta asiakaskuntaansa ja nostamatta omien tuotteidensa profilia.

Aluksi Lingsoftilla harkittiin ja kehiteltiin lingvististen työkalujen hyödyntämistä informaationhallinnassa, sen jälkeen puheteknologiassa ja lopuksi elektronisten sanakirjojen kehitysympäristössä. Lingsoft markkinoi myös kielentarkistustyökalujaan itsenäisinä ohjelmistoina tuotenimillä Orthografix ja Grammatifix joko sellaisenaan tai räätälöityinä versioina. Lingsoftin selkeä vahvuus oli lingvistisesti loppuun saakka mietittyjen ohjelmistojen kehittäminen. Yrityksellä oli pitkä kokemus tällaisten moduulien lisensoinnista toisille IT-yrityksille, jotka jakelevat niitä integroituna omiin ohjelmistotuotteisiinsa. Näin ollen Lingsoftin tarvitsi rakentaa oma markkinakanavansa tavoittaakseen eritoten lupaavat yritysmarkkinat. Yrityksen resurssit ja niin työnpanoksen kuin rahoituksen muodossa kului kuitenkin useaan eri kohteeseen, minkä johdosta Lingsoftin ensimmäisten, aidosti omien loppukäyttäjätuotteiden – Lingsoft Pointer ja Lingsoft Parrot – valmistuminen lykkääntyi vuoteen 2001. Kun samanaikaisesti rahoitusmarkkinat vielä kiristyivät, yritys joutui saneeraamaan toimintaansa reippaasti, mutta vuoden 2002 alussa pahin näyttää menneen ohi, missä on varmasti auttanut uusi merkittävä lisenssisopimus.

Kielikone puolestaan näyttää onnistuneen ottamaan irti mahdollisimman paljon laajasta tuotevalikoimastaan ja vahvasta elektronisten sanakirjojen markkina-asemastaan Suomessa. Tässä ei liene haitannut, että Kielikoneen kahdella toimitusjohtajalla vuoden 1997 jälkeen on ollut vahva kokemus ohjelmistoalan markkinoinnista. Vaikka vuosina 1995-1999 Kielikoneen liikevaihto kasvoi noin kahteen miljoonaan euroon,

yrityksen ei ole tarvinnut kasvattaa henkilökuntansa läheskään samassa suhteessa. Kautta koko 1990-luvun Kielikone on jatkanut TranSmart-konekäännösjärjestelmänsä kehittämistä yhteistyössä ensin Nokian ja myöhemmin mm. Rautaruukin kanssa. TranSmart tuntuu kuitenkin edelleen olevan pikemminkin osoitus Kielikoneen teknologisesta osaamisesta ja taustasta kuin merkittävä tulolähde. Itse asiassa Harri Arnola on siirtynyt kehittämään kieliriippumattomampaa konekäännösformalismia omaan Ganesa-nimiseen yritykseensä.

1990-luvun loppupuolella Lanser Data lisensoi ydinteknologiansa Xerox-yhtymän InXight-nimiselle tytäryhtiölle. Lanser Data oli kuitenkin jäänyt pienimmäksi kieliteknologian vanhimasta yrityskolmikosta, eikä sillä ole pariin vuoteen ollut aktiivista toimintaa.

UUDET TULOKKAAT OVAT LÄHTE-NEET LIIKKEELLE TOISIN KUVIOIN

Conexor ja Gurusoft ovat hyviä esimerkkejä yrityksistä, jotka ovat omaksuneet ohjelmistoliiketoiminnan uuden paradigman Suomessa. Päinvastoin kuin vanhemmat kieliteknologia-yritykset, ne ovat hyödyntäneet yksityistä tai riskirahoitusta heti käynnistymisvaiheesta lähtien kehittääkseen ensimmäiset tuotteensa ja saavuttaakseen ensimmäiset yhteydet potentiaaliin asiakkaisiinsa. Tämän lisäksi ne ovat tähtäämässä suoraan kansainvälisille markkinoille sen sijaan, että ne aloittaisivat varovaisesti kasvunsa Suomessa, laajentaen seuraavaksi Ruotsiin ja Saksaan ja sitten muuhun Eurooppaan, yrittäen lopuksi tunkeutua Yhdysvaltain markkinoille - polku joka oli liki pitäen suomalaisten yritysten kansainvälistymisen normi vielä vähän aikaa sitten.

Jälkikäteen ei voi olla pohdiskelematta, miksi vanhemmat suomalaiset kieliteknologia-yritykset ovat karttaneet ulkopuolista yksityistä tai riskirahoitusta, vaikka näillä yrityksillä on ollut heti perustamisvaiheestaan lähtien ja myöhemminkin vakuuttavaa teknologista osaamista omasta takaa. Tulokkaat muissa Pohjoismaissa, kuten Nordisk Språkteknologi Norjassa ja Hapax Ruotsissa, ovat

onnistuneet keräämään useita miljoonia euroja rahoitusta jo toimintansa käynnistämisvaiheessa. Olisiko niin, että nämä suomalaiset yritykset ovat vieroksuneet ulkoisia sijoituksia ja niihin liittyvää ulkoista kontrollia samoin kuin kunnianhimoisia tavoitteita, vai eivätkö kansalliset sijoittajat ole nähneet tai ymmärtäneet potentiaalia näiden yritysten strategioissa saati ylipäänsä kieliteknologiassa?

Conexorin tavoitteena on lisensoida suoraan kansainvälisesti toimiville IT-yrityksille paitsi lingvististä perusteknologiaansa Machines-tuotetimellä niin myös omia tuotteitaan, dokumenttien indeksointityökalua Navitermiä ja kielentarkistustyökalua TrueStyleria. Conexorista tekee kiinnostavan kansainvälisestä näkökulmasta se, että yrityksen teknologian ydin on verrattain tuore. Lisäksi Conexor on soveltanut teknologiaansa tärkeimmille länsieurooppalaisten kielten markkinoille, nimittäin englannille, ranskalle, espanjalle ja saksalle. Vaikkakin Conexor olikin aluksi varsin teknologiapainottunut yritys, se ei lopulta ole epäroinyt rekrytoida IT-markkinoinnin ja -johtamisen asiantuntemusta. Conexorilla on tällä hetkellä satakunta asiakasta ympäri maailmaa, mm. Toshiba Corporation, ja yritys on ollut tulorahoitteinen vuodesta 2000 lähtien.

Gurusoftware tähtää myös kansainvälisille markkinoille, missä pienikin siivu voi tarkoittaa suuria tuloja. Koska SOM-teknologia ei välttämättä edellytä mitään sääntöpohjaista esiprosessointia, Gurusoftwaren teknologia on aidosti kieliriippumaton, mikä on selkeä etu muihin tässä artikkelissa mainittuihin yrityksiin nähden, sillä kieliriippumaton periaate edellyttää kuitenkin aina jonkinasteista kielikohtaista kehitystyötä. Gurusoftwarella on kuitenkin vastassaan vakiintuneita kansainvälisiä kilpailijoita kuten Autonomy.

Näiden kahden yrityksen ilmiselvänä haasteena on miten pienet, tähän saakka käytännössä tuntemattomat yritykset pienestä pohjoiseurooppalaisesta maasta, tunnettiinpa Suomi Nokiasta tai ei, onnistuvat rajallisilla resursseilla saamaan jalan oven väliin vain voidakseen aloittaa kaupankäynnin.

TULEVAISUUDEN MAHDOLLISUUDET JA HAASTEET

Mitkä sitten ovat näiden suomalaisten kieliteknologiayritysten tulevaisuudennäkymät? Kaiken kaikkiaan ne työllistivät vuoden 2001 lopussa vajaat sata ihmistä, ja kunkin yrityksen liikevaihto liikkuu suurimmillaankin vain muutamassa miljoonassa eurossa. Itse asiassa huhu kertoo, että Nokialla on enemmän kieliteknologiaan tavalla tai toisella kytkettyneitä työntekijöitä, erityisesti puheteknologian parissa, kuin näillä yrityksillä yhteensä. Näin ollen kieliteknologiapohjainen liiketoiminta Suomessa näyttää toistaiseksi olevan korkeintaan lupaava nappu muiden muassa. Kun tarkastellaan kunkin yrityksen tämänhetkistä tilannetta, itse kullakin on edessään omat haasteensa, jos ne mielivät kehittyä eteenpäin nykyisestä asemastaan.

Lingsoftin haasteena on fokusoida toimintaansa sekä tuotteiden että kielien suhteen. Lingsoftilla on yllin kyllin lingvististä perusteknologiaa, erilaisia sisältöresursseja ja sovelluksia pohjoismaisia valtakieliä varten, jotka ovat kuitenkin valitettavasti osoittautuneet pienehköiksi, yksittäisiksi markkina-alueiksi, jotka edellyttävät kukin oman markkinoilletulostrategiansa ja investoinnit erityisesti loppuasiakastuotteiden osalta. Yrityksellä on laaja tuotevalikoima myös saksaa varten, joka Euroopan suurimpana markkina-alueena tarjoaa erittäin lupaavia mahdollisuuksia. Toisaalta suomalaisilla IT-yrityksillä on tunnetusti ollut vaikeuksia murtautua Saksan markkinoille. Lingsoftilla on myös mielenkiintoisia tuotenuppuja esimerkiksi puheteknologiaa ja elektronisia sanakirjoja varten, mutta yrityksen on tehtävä hartiavoimin työtä saadakseen markkinakanavat vetämään. Kun Kielikone dominoi kieliteknologisten tuotteiden vähittäis- ja yritysmarkkinoita Suomessa, ja samankaltaisia vahvoja yrityksiä löytyy muista Pohjoismaista, vaikuttaisi että Lingsoftin parhaat mahdollisuudet ovat teknologiansa lisensoinnissa kansainvälisille IT-yrityksille, räätälöitynä suurille yksittäisasiakkaille sekä ylipäänsä Saksassa.

Tästä huolimatta Lingsoft on puheteknologian suhteen ainutlaatuinen verrattuna muihin suo-

malaisiin kieliteknologiayrityksiin, mikä on selvästi mahdollisuus erityisesti yhdistettynä yrityksen pohjoismaisten kielten ja saksan kattavaan valikoimaan. Tässä ovat puolestaan vastassa potentiaalisina kilpailijoina muut toimijat kuten Philips ja Nordisk Språkteknologi Norjasta.

Kielikoneen haasteena on kansainvälistyminen, ellei yritys tyydy pysyvätyymään Suomessa. Kielikoneen vahvuuksia ovat tuotteistaminen ja loppuasiakkaat tavoittavan markkinakanavan hallinta, missä sillä on 2 500 yritysasiakasta. Tällä vahvuusalueella Kielikoneella on vastassaan muissa Pohjoismaissa vastaavanlaisia vakiintuneita yrityksiä kuten WordFinder Software Ruotsissa ja Clue Norjassa. Kieliteknologian suhteen, mikä voisi ehkä olla etu ulkomaille suunnattaessa, Kielikone on rajautunut ainoastaan suomeen ja Englantiin, mutta toisaalta yrityksen onnistuminen Suomessa elektronisissa sanakirjoissa ilman käytännössä mitään lingvistisiä ominaisuuksia tuntuisi viittaavan siihen, ettei tämä loppujen lopuksi ole rajoite ainakaan kyseisessä tuoteryhmässä.

Vaikka Kielikoneella on toimittamaansa konekäännösjärjestelmään ilmeisen tyytyväisiä yritysasiakkaita, Suomen kaltaisessa pienessä maassa sellaisten organisaatioiden lukumäärä, joilla olisi intoa ja resursseja ottaa käyttöön tämäntyyppinen järjestelmä, joka edellyttää parhaan toimivuuden saavuttamiseksi mittavaa räätelöintiä, on rajattu erityisesti kun järjestelmä kääntää tällä hetkellä ainoastaan suomesta Englantiin. Positiivista kehitystä on, että nykyiselle konekäännösjärjestelmälle on syntymässä LINGMACHINE-projektissa (osa MLIS-ohjelmaa) vastinkappale Englannista suomeen, mutta jää nähtäväksi tuleeko tästä uusi mittava tulonlähde yritykselle.

Ehkäpä kaikkein merkittäväintä LINGMACHINE-projektissa on, että Kielikone ja Conexor ovat valinneet yhteistyön omien polkujen seuraamisen sijasta. Tässä mielessä yleisesti arkaluontoinen ja potentiaalisesti ongelmallinen kysymys on, että merkittävä osuus kieliteknologiayritysten keskeisistä omistajista ja toimijoista ovat edelleen toimineet opettajina tai tutkijoina suomalaisissa yliopistoissa tai alan asiantuntijoina. Niin vanhat kuin uudet toimijat ovat voineet

kokea hankalaksi käynnistää uusia kieliteknologiaprojekteja törmäämättä muihin – ja mahdollisesti kilpaileviin – toimijoihin roolissa tai toisessa. Viimeaikainen kehitys, missä kieliteknologia-yrityksillä Kielikoneesta lähtien alkaa yhä useammilla olla IT-taustan omaava toimitusjohtaja, jolla ei ole henkilökohtaisia kytköksiä kieliteknologian akateemiseen puoleen, saattaa johtaa potentiaalisten ristiriitojen häviämiseen.

Jotta suomalaiset kieliteknologiayritykset onnistuvat kasvamaan hallitusti, niiden täytyy löytää tapoja murtautua nykyisistä liiketoiminnan lokeroistaan ja löytää ratkaisuja omiin haasteisiinsa. Kieliteknologisten ratkaisujen tarjoamisessa useille pienille kielialueille on erityishaasteena se, että näillä markkinoilla toimivien yritysten täytyy kehittää ja ylläpitää lingvistisiä malleja kaikille tukemilleen kielille, mikä voi olla potentiaaliin tuloihin nähden varsin kallista niinkin suurilla markkinoilla kuin Ruotsissa saati sitten Suomessa. Itse asiassa sekä Lingsoft, Lanser Data että Conexor ovat kaikki päätyneet kehittämään samankaltaiset lingvistiset mallit Englannille, ruotsille ja suomele sen lisäksi, että ne tukevat myös muita kieliä. Kun mukaan lasketaan vielä Kielikone, löytyy Suomesta neljä erilaista suomen kielen lingvististä mallia!

Mietittäessä kieliteknologista liiketoimintaa kokonaisvaltaisesti kansallisesta näkökulmasta käsin, tilanne ei vaikuta ideaalilta, erityisesti kun inhimillisiä resursseja on aina rajatusti. Olisiko yritysten kannattanut fokusoida täydentäen toisiaan tai tehdä yhteistyötä jo kauan ennen LINGMACHINE-projektia? Toisaalta voidaan spekuloida, olisivatko sentyyppiset kieliriippumattomat ratkaisut, joita Gurusoft kehittää, sittenkin niitä, joilla on suurin potentiaali globaaliin kasvuun – jos tällainen kasvu on sitä, mitä nämä yritykset haluavat? Tulisiko näiden suomalaisten yritysten olla aiempaa rohkeampia ulkopuolisen rahoituksen ja kansainvälisen kasvun tavoittelun suhteen? Jälkiviisuus lienee tulevaisuudessakin paras viisaus.

*

KIRJOITTAJASTA

Antti Arppe on koulutukseltaan diplomi-insinööri ja opiskellut tuotantotaloutta ja tietämysteknikkaa Teknillisessä korkeakoulussa sekä tietokoneingvistiikkaa Helsingin yliopistossa. 1990-luvun alusta lähtien hän on ollut tavalla tai toisella tekemisissä suomalaisten kieliteknologia-yritysten kanssa. Hän on selvittänyt Kielikoneelle konekäntämisen markkinoita 1993–1994 ja myöhemmin toiminut tuotepäällikkönä ja varatoimitusjohtajana Lingsoftilla 1994–1998. Tämän jälkeen hän on tehnyt yksittäisiä projektitoimeksiantoja Lingsoftille, missä hänellä on pieni omistusosuus. Hän on myös osallistunut EAGLES-projektiin kirjoittajana 1994. Tällä hetkellä hän on tutkijana TEKESin rahoittamassa USIX/GILTA-projektissa. Samalla hän on yleisen kielitieteen jatko-opiskelija Helsingin yliopiston Yleisen kielitieteen laitoksella, missä hän on vuosittain pitänyt kurssin Kaupallinen kieliteknologia.

KIITOKSET

Kirjoittaja kiittää seuraavia henkilöitä, jotka ovat keskeisesti edistäneet tämän artikkelin syntyä virkistämällä, täydentämällä ja tarkentamalla kirjoittajan muistia ja tietoja sekä kommentoimalla ja edistämällä kirjoitusta yleisellä tasolla: Harri Arnola (Kielikone, Ganesa), Olli Blåberg (Lanser Data), Lauri Carlson (Helsingin yliopisto), Jaakko Happonen (Lingsoft), Mika Herpiö (Kielikone), Timo Honkela (Gurusoft), Kaarina Hyvönen (Kielikone), Kimmo Koskenniemi (Helsingin yliopisto, Lingsoft), Seppo Koskenniemi (IBM), Katri Luostarinen (CSC), Jan Magnusson (Conexor), Harri Saarikoski (Republica), Pasi Tapanainen (Conexor), Juha Telkkinen (Promentor), Kristian Töyrä (Timehouse), Atro Voutilainen (Conexor), Hanna Westerlund (Helsingin yliopisto) ja Graham Wilcock (Helsingin yliopisto). Kyseinen artikkeli on kirjoitettu EUROMAP-hankkeelle.

NO SINGLE PATH – FINNISH LESSONS IN THE COMMERCIALISATION OF LANGUAGE ENGINEERING RESEARCH

Antti Arppe

Department of General Linguistics, University of Helsinki

This article provides an overview of how research in linguistics and language technology has been commercialized in Finland. It is observed that academic research has played an important role in the birth of the language technology business in Finland, though companies have also come to exist in the field with other types of background. This article goes through in chronological order the research results and projects of Finnish language technology researchers and research groups which have led to commercialized products or the founding of language technology companies. After this, the strategies which these companies with academic backgrounds have chosen are presented, followed by an overview of the subsequent development and growth of these companies and an assessment of how they have succeeded. A general observation is that the development of these companies has varied substantially, as have the strategies that the companies have pursued. Despite successes on the domestic level, a major international breakthrough for a Finnish language technology company is still in waiting. In retrospect, it seems that those that have managed to interpret themselves firstly as commercial IT companies have succeeded best. Finally, the present situation of these companies is evaluated as well as their future possibilities and challenges.”