

Linuxin lokalisointi

Marko Grönroos <magi@iki.fi>

Sisällys

- Johdanto
- Miten globalisointi toimii (xgettext)
- KDE
- OpenOffice
- Muut järjestelmät
- Suomenkielinen kirjallisuus
- Kääntämisen ongelmia

Mitä on lokalisointi? (I10n)

- Tietokoneohjelman spesialisointia tiettyyn kulttuuriseen, poliittiseen tai maantieteellinen kokonaisuuteen eli *lokaaliin*
- Lokaaleja voivat olla esimerkiksi kieli, maa tai tietty maantieteellinen sijainti
- Tavallisimmin lokalisointi tarkoittaa kielen kääntämistä

Johdanto

- Linuxin lokalisointi on tärkeää
 - Tavalliset ihmiset eivät muuten tykkää
 - Työpöytäkäyttö ei muuten yleisty
- Lokalisoitavat asiat:
 - Käyttöliittymät
 - Sovellusten sisäiset käyttöohjeet
 - Yleinen Linux-dokumentaatio

Lokalisoitavat asiat

- Sovellusten
 - Käyttöliittymät
 - Käyttöohjeet
- Sovellustyypit:
 - Ohjelmat, joissa graafinen käyttöliittymä
 - Komentoriviohjelmat
 - Palvelinohjelmat
- Yleiset Linux-käyttöohjeet
 - man- ja info-sivut
 - HOWTO:t
 - Muut käyttöohjeet

Internationalisointi (i18n)

- Tietokoneohjelman yleistäminen siten, että se on riippumaton kulttuurisesta, poliittisesta tai maantieteellisestä alueesta eli lokaalista
- Lokaalispesifinen tieto irroitetaan ohjelmasta erilliseksi resurssiksi, jota voi muokata ilman ohjelmointia
- Monia ohjelmia ei ole lainkaan internationalisoitu

Globalisointi (g11n = i18n + l10n)

- Koko prosessi joka sisältää
 - Internationalisoinnin (yleistys)
 - Lokalisoinnin (spesialisointi) eri lokaaleille

Internationalisointi...

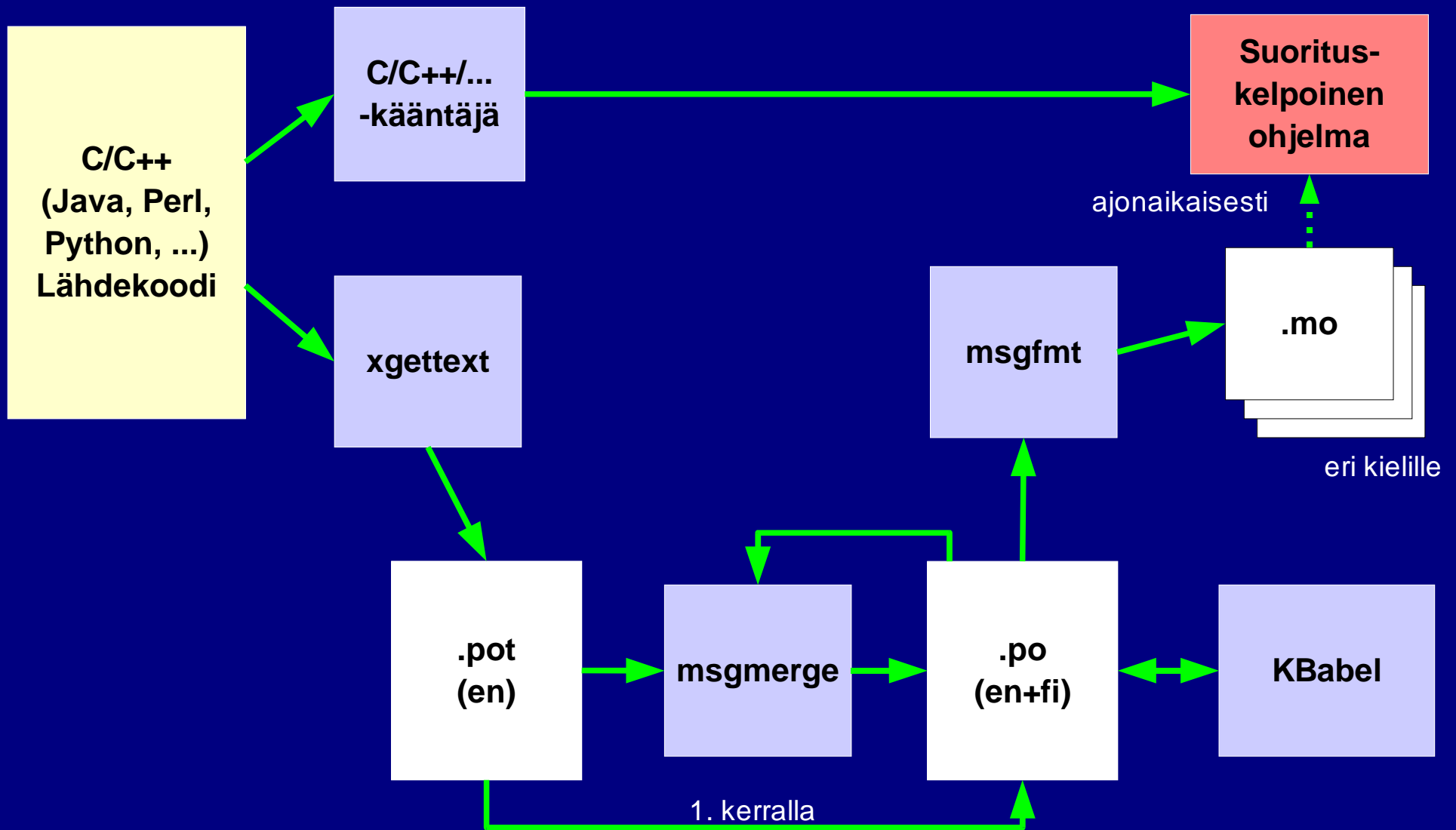
käytännössä

- Esim. `printf ("Hello world!\n");`
- `i18n` `xgettext`in tavalla:
 - `printf (gettext("Hello world!\n"));`
 - `gettext()`-funktio hakee käännöksen
 - Oletuskielen merkkijono toimii hakuavaimena
 - `xgettext` hakee lähdekoodista kaikki hakuavaimet
 - Tuottaa merkkijonoresurssitiedoston (kts. myöhemmin)
 - Nykyinen kieliasetus on globaalissa muuttujassa
 - Saadaan `LANG`-ympäristömuuttujasta
 - `ngettext` mahdollistaa yksikkö-/monikkomuodot
 - "1 tiedosto" -- "42 tiedostoa"

Globalisointi gettextillä

- i18n: xgettext-ohjelma
 - poimii C/C++ -lähdekoodista merkkijonot
 - tekee .pot-mallitiedostot
 - .pot -tiedosto mergetään/kopioidaan kielikohtaisiin .po-tiedostoihin
- l10n: Kielenkääntäjä kääntää .po-tiedoston
 - Käännetyistä .po-tiedostoista tehdään binääriset .mo-tiedostot
 - Ohjelma lukee merkkijonot .mo-tiedostosta ajonaikaisesti

...globalisointi gettextillä



Internationalisointi gettextillä

...yksinkertainen ohjelma

```
#include <stdio.h>

int main ()
{
    /* Tulosta merkkijono. */
    printf ("Hello, world!\n");
}
```

Internationalisointi gettextillä...

...internationalisoitu ohjelma

```
#include <stdio.h>
#include <libintl.h>
#include <locale.h>

int main ()
{
    /* Hae lokaali ympäristömuuttujista. */
    char* sl = setlocale (LC_ALL, "");

    /* Aseta lokaalitiedostojen polku. */
    bindtextdomain ("hello", "/usr/share");

    /* Aseta nykyinen teksti-domain. */
    textdomain ("hello");

    /* Hae ja tulosta merkkijono. */
    printf (gettext ("Hello, world!\n"));
}
```

Internationalisointi gettextillä...

...POT-tiedosto

```
# SOME DESCRIPTIVE TITLE.
# Copyright (C) YEAR THE PACKAGE'S COPYRIGHT HOLDER
# This file is distributed under the same license as the PACKAGE
package.
# FIRST AUTHOR <EMAIL@ADDRESS>, YEAR.
#
#, fuzzy
msgid ""
msgstr ""
"Project-Id-Version: PACKAGE VERSION\n"
"POT-Creation-Date: 2003-07-17 14:16+0300\n"
"PO-Revision-Date: YEAR-MO-DA HO:MI+ZONE\n"
"Last-Translator: FULL NAME <EMAIL@ADDRESS>\n"
"Language-Team: LANGUAGE <LL@li.org>\n"
"MIME-Version: 1.0\n"
"Content-Type: text/plain; charset=CHARSET\n"
"Content-Transfer-Encoding: 8bit\n"

#: hello.c:24
msgid "Hello, world!\n"
msgstr ""
```

Internationalisointi gettextillä...

...käännetty PO-tiedosto

```
...  
"Content-Type: text/plain; charset=UTF-8\n"  
...
```

```
#: hello.c:24  
msgid "Hello, world!\n"  
msgstr "Terve, met "  
"sä!\n"
```

Gettext...

...ngettext-internationalisointi

- Yksikkö- ja monikkomuotojen tulostus

```
/* Muuttujalla "i" on useita eri arvoja 1-4. */  
for (int luku=1; luku<=4; luku++)  
    /* Hae ja tulosta merkkijono. */  
    printf (ngettext ("%s file", /* Yksikkömuoto */  
                  "%s files", /* Monikkomuoto */  
                  i), /* Lukumäärä ngettextille */  
           i);          /* Tulostettava luku */
```

- Tulostaa:

```
1 file  
2 files  
3 files  
4 files
```

Gettext...

...nggettext-lokalisointi

- POT-tiedosto:

```
...  
#: hello.c:29  
#, c-format  
msgid "%d file\n"  
msgid_plural "%d files\n"  
msgstr[0] ""  
msgstr[1] ""
```

- PO-tiedosto:

```
...  
#: hello.c:29  
#, c-format  
msgid "%d file\n"  
msgid_plural "%d files\n"  
msgstr[0] "%d tiedosto"  
msgstr[1] "%d tiedostoa"
```

- Tulostaa:

```
1 file  
2 files  
3 files  
4 files
```

```
1 tiedosto  
2 tiedostoa  
3 tiedostoa  
4 tiedostoa
```


Avoimen lähdekoodin lokalisointiprojektit

- Työpöytäympäristöt
 - *KDE*
 - KOffice
 - Gnome
- Distrot
 - Debian
 - Mandrake
- Ohjelmistot
 - *OpenOffice*
 - Mozilla
- Järjestelmäohjelmat
 - Translation Project

KDE:n lokalisointi

- .pot ja .po -tiedostot CVS:ssä (kde-i18n)
- Lokalisointi tehdään KBabelilla
 - Kaikenlaista kivaa automatiikkaa
 - Käännöshallintaohjelma
- Dokumentaation lokalisointi
 - Tehdään myös KBabelilla!
 - Täytyy tehdä täydellisesti ohjekirja kerrallaan
 - yksikin kääntämätön tiedosto -> kirja kääntämätön
- Lokalisoinnin asennus ja testaus helppoa
- Käännökset ->
 - Kim Enkovaara <kim.enkovaara@iki.fi>

KDE:n gettext-erityispiirteitä

- Pari kivaa laajennusta:
 - Parametrien järjestys voi vaihdella
 - "% 1 lines of %2 different files match"
 - "42 lines of 6 different files match."
 - "%2:den tiedoston %1 riviä täsmää."
 - "6:den tiedoston %1 riviä täsmää"
 - Tapa hanskata yksikkö-/monikkomuodot
 - Eri tapa kuin ngettextillä
 - Osaa mielivaltaisen määrän monikkomuotoja
 - Lukumäärä on globaali kielikohtainen asetus
- Dokumentaatio on XML:ää
 - esim. "<quote>lainattua tekstiä</quote>"
 - merkkilyhenteet: "&kde; &etc;"

KDE:n lokalisointi...

Erityispiirteitä

- Tietyt käännökset yhteisiä
 - Määritelty kdelibs-moduulin käännöksissä
 - Valikkojen tekstit
 - Tiedosto, Muokkaa, Siirry, Ohje, ...
 - Dialogien painikkeet
 - Ok, Peruuta, Hylkää, Hyvähksy...
 - Monet perusdialogit kokonaan yhteisiä

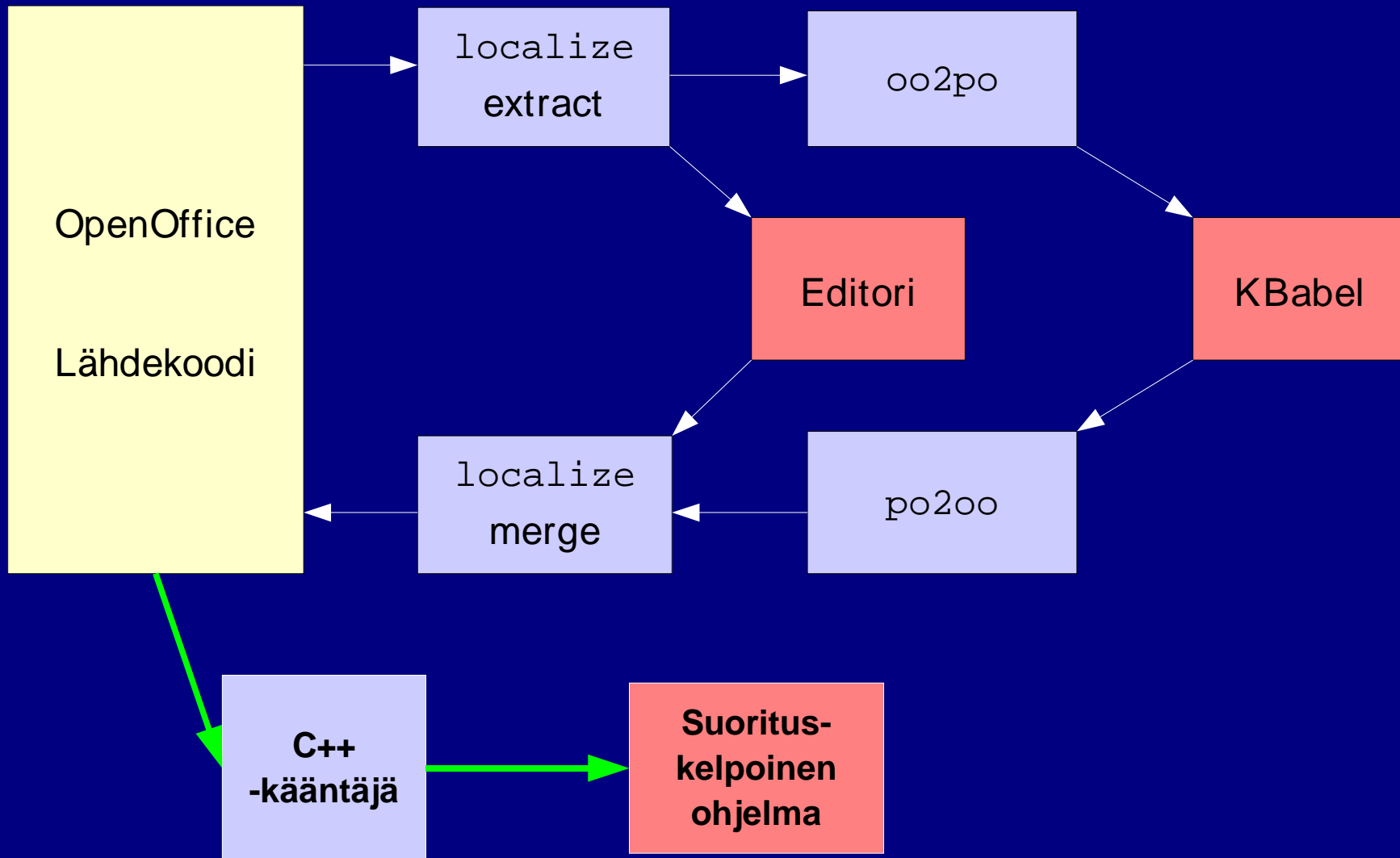
KDE:n lokalisoinnin tila

- Käyttöliittymistä suomennettu ~80%
 - Käännökset rappeutumassa vähitellen
 - Joitakuita satunnaisia kääntäjiä
- Dokumentaatiosta suomennettu <1%
 - Lokalisointi vasta aloitettu
 - Ei aktiivisia kääntäjiä
- Työskentely KDE_3_1_BRANCH:issä
 - Kohteena KDE 3.1.4
 - Runko kehittyä kohti KDE 3.2:ta
 - KDE_3_1_BRANCH mergetään ehkä lokakuussa runkoon

Gnomen lokalisointi

- Internationalisoitu gettextillä
 - Lokalisointiin suositellaan Emacsin .po-tilaa
- Status: ~40%
 - Tärkeimmät kokonaisuudet: ~81%
 - Työpöytäsovellukset: 78%
 - Peruskirjastot: 100%
 - ...useita uusia moduuleja...
 - Extras: ~24%
 - Esim. Evolution: 23%
- gnome-fi.sourceforge.net
- Sami Pesonen <sampeson@iki.fi>

OpenOfficen lokalisointi



OpenOfficen dokumentaatio

- XML-muodossa
 - Kielipaketissa mukana myös binääridataa
 - Ei aavistustakaan millä työkalulla XML-teksti on tuotettu
- SuomennoS Sunin tekemä
 - Tarkoitettu StarOffice 6.x:ään.
 - Toimii OpenOfficessa virittämällä
 - kts. Asmo Koskisen ohjeet

OpenOfficen lokalisoinnin tila

- Kuka tehnyt aiemmin?
 - Käyttöliittymä:
 - Kongo Group
 - Sun (ostettu Kongolta)
 - SOT (tehneet itse)
 - Dokumentaatio:
 - Sun (lähdekoodi kuulemma julkistettu elokuussa)
- Kuka tekee jatkossa?
 - ???
- OpenOffice native-lang -projekti
 - fi.openoffice.org (Asmo Koskinen)

Muut suomennokset

- Man-sivut
 - SOT organisoii aikoinaan suomennosprojektin
 - Käännösten kohtalosta ei tietoa
- Info-sivut
 - Ei käännetty tiettävästi lainkaan
- HOWTOt
 - Muutama käännetty, mutta ei löydy mistään
 - Finnish-HOWTO löytyy
- Weppisivustot
 - Debianin kotisivut suht hyvin

Linux-jakeluversioiden suomennokset

- Suomennosten laajuus
 - WWW-sivut
 - Asennus
 - Jakeluversiokohtaiset hallintaohjelmat
 - Muut sovellukset
 - Ohjekirjat
- Jakeluversiot:
 - Red Hat: ei asennuksessa
 - Debian: ei asennuksessa, www-sivut
 - Mandrake: myös asennuksessa, joten kuten
 - SuSE: ??

Suomenkielinen kirjallisuus

- IT Press
 - Linux Trainer Tehokurssi (1999)
 - Linux-ohjelmointi Trainer (1999)
 - Linux - käyttäjän käsikirja (Man-sivut ja järjestelmänhallinta) (2000)
 - KDE - käyttäjän käsikirja (2000)
 - Linux sertifikaatti (2002)
 - Inside Linux: Verkot (2002)
 - Linux - asentajan käsikirja (2002, IT Linux)
 - Linux - käyttäjän käsikirja (2003, Spectra Linux)
 - Inside Linux: Järjestelmänhaltijan käsikirja (2003)

Suomenkielinen kirjallisuus...

- Suomen ATK-kustannus
 - Red Hat Linux keltanokille (2000)
 - Linux ja Samba (2000)
 - Linuxin tarina (2000)
 - Linux palvelimena (2001)
- Pagina
 - Linux Internet-palvelimena (2001)
 - Linux -käsikirja (2.p. 2001)

Suomenkielinen kirjallisuus.....

- Docendo
 - Linux - Linux-käyttäjän peruskirja (1999)
 - StarOffice (SO 5.2, 2000)
- Edita
 - Linux: Peruskäyttäjän opas (1999)
 - Linux: Yrityksen avoin vaihtoehto (1999)

Kääntämisestä

- Mitä ongelmia kääntämisessä tulee?
 - Ohjelmakohtaiset ongelmat
 - Ohjelmissa erilaisia ohjelmakohtaisia koodeja ja sovittuja tapoja
 - Kaikkia koodeja ei saa kääntää
 - Ohjelmointikielten koodit
 - Esim. "Hakemistossa on %d tiedostoa.\n"

Kääntämisestä...

- Kääntäminen on vaikeaa
 - Amatöörimäisyys näkyy helposti
 - Suomen kieltä on vaikea kirjoittaa hyvin
 - Ajan käytöllä voi parantaa tulosta
- Oikoluku
 - Lue teksti aina läpi vielä pariin kertaan
 - Mieluiten parin viikon välein
 - Tai tai anna jonkun toisen tehdä oikoluku
- Kääntämistä oppii tekemällä

Kääntämisestä...

Kielellisiä ongelmia

- Sanat joilla useita eri merkityksiä
 - Esim. "column" on sekä taulukon sarake että sivun palsta
 - Internationalisointi on saattanut optimoida monimerkityksisen sanan yhdeksi merkkijonoksi
 - Lauserakenne ja sanajärjestys
 - Usein lähes samat, usein hyvinkin erilaiset
 - Suomessa suhteellisen vapaa sanajärjestys
 - Riippuu merkityksestä ja painotuksesta
 - On vaikeaa löytää hyvä sanajärjestys
- Hyvä sanajärjestys on vaikea löytää

Kääntämisestä...

Kielellisiä ongelmia

- Käännetään merkitys tai sanoma
 - Sanatasolla, lausetasolla, kappaletasolla
 - Kirjaimellisuus ei tarpeellista
 - Kuitenkin hyvä säilyttää jos toimii suomessakin
- Ilmaisun kulttuurierot
- Alkuperäistekstikin voi olla huonoa!
- Kaikkia lauseita ei ole pakko kääntää lainkaan
 - Erityisesti usein käytetty markkinointilöpinä ei aina sovi sellaisenaan suomalaiseseen kielenkäyttöön

Kääntämisestä...

Poistoja tekstistä

- Usein tekstistä täytyy poistaa tarpeettomuuksia
 - "*Cupsomatic is a Perl script that is needed to...*"
 - Mikä ihme on "Perl-skripti"? (Perl-komentosarja...?)
 - Onko relevanttia että ohjelma on Perl-skripti?
 - Suom. "*Cupsomatic-ohjelmaa tarvitaan...*"
 - Jos täsmennys *saattaa* olla relevantti jollekulle, voidaan se siirtää myöhemmäksi paremmin selitettynä
 - "*... Cupsomatic on toteutettu Perl-ohjelmointikielellä ja edistyneet käyttäjät voivat tehdä siihen helposti parannuksia.*"

Kääntämisestä...

Lisäyksiä tekstiin

- Usein tekstiin täytyy lisätä selityksiä
 - *"With the emerging CUPS concept, extending the use of PPDs even to non-PostScript printers..."*
 - *"...CUPS-tulostusjärjestelmän...PPD-kuvaustiedostojen..."*
- Lyhenteiden käyttö usein yli-innokasta
 - *"Ghostscript filter that can be used with any other program that needs a software-RIP."*
 - Suom. *"...joka tarvitsee ohjelmallista kuvanrasterointia (RIP)."*

Kääntämisestä...

Pikanäppäimet

- Merkitään esim. "*M&uokkaa*" -> "*Muokkaa*"
 - Muitakin merkintätapoja käytetään
 - OpenOfficessa tilde ~
- Eivät saisi koskaan olla ristiriidassa
 - Ristiriidattomuus täytyy tarkistaa ajamalla ohjelma
 - Joitain viestejä voi olla vaikea saada
 - Virheviestit
 - Vaikeaa jos internationalisointi on optimoinut usean eri kontekstin merkkijonot samaksi.
 - Esim. Ok, Ohje, Sulje, Hylkää ja Peruuta -painikkeet, jotka hyvin monissa dialogeissa.
 - KDE:ssä ristiriitojen tarkistusjuttu

Kääntämisestä...

Pituusongelmia

- Merkkijonot usein sovitettu englannin mukaisesti
 - Usein vaikea kääntää lyhyesti
 - Esim. "*is needed by*" (rpm)
 - Käännettiin joskus "*tarvitsee*" (huono)
- Lyhenteet, joita ei ole suomessa
 - Esim. N, S, E, W (nämä ok)
 - Mutta: NE, SE, SW, NW??
 - Esim. UT (Universal Time)
 - Maailmanaika, ei suomenkielistä lyhennettä
 - Käytetään engl. lyhennettä

Kääntämisestä...

Lauserakenteellisia ongelmia

- Katkaistut merkkijonot
 - Rajoittavat lauserakenteen muuttamista
 - Esim.
 - M1: *"Could not parse "*
 - *"Ei saatu jäsennettyä "* (tai tulkittua tai luettua)
 - *"Jäsentäminen epäonnistui"*
 - M2: *" as a valid month code."*
 - M3: *" as a valid day code."*
 - M1+M2: *"Could not parse <..> as a valid month code."*
 - M1+M3: *"Could not parse <..> as a valid day code."*
 - Ratkaisuja
 - *"Ei saatu jäsennettyä " + " kuukauden koodina."*
 - *"Arvo " + " ei ole kelvollinen kuukausi."*

Kääntämisestä...

Sanastot

- Käytettävä yhtenäistä sanastoa
- Yleinen ATK-sanakirja
- Gnomen käännöksistä sanakirja
 - Kts. gnome-fi.sourceforge.net
 - Usein vain yhden merkityksen käännös
- KDE käyttää pääsääntöisesti Gnomen sanakirjaa
 - Toissijaisesti Windows-sanastoa
 - Paitsi jos löytyy parempi sana
 - Pieni oma sanasto, ~40 sanaa
- Sanakirjoissa paljon typeriä käännöksiä
- Käännös riippuu aina merkityksestä!