

Apertium kaj Esperanto

Maŝintradukado en kaj el Esperanto
per malfermkoda platformo

Jacob Nordfalk
Hèctor Alòs i Font

KAEST
20an de novembro 2010

Skemo

- Kio estas Apertium?
- Historio de Apertium
- Historio de Esperanto en Apertium
- Stato de la Esperanto-tradukiloj
- Ebloj de novaj Esperanto-tradukiloj
- Pozitivaj aspektoj de Apertium

Kio estas Apertium?

- Apertium estas rimedaro (programaro kaj datumaro) por konstrui aŭtomatajn tradukilojn
- Ĝi estas bazita sur reguloj (ne sur statistiko)
- Ĝi uzas supraĵan (malprofundan) sintaksan analizon
- Ĉiu tradukilo inter du lingvoj estas sendependa de la aliaj (ne ekzistas pivota interlingvo), tamen eblas recikli materialojn de aliaj tradukiloj
- Evoluigantoj laboras super XML-dosieroj, tiel ke principe komputikaj scioj ne estas necesaj por evoluigi tradukilon (kvankam scipovo de skripto-lingvo ege utilas)
- La programaro estas libera kaj malfermfanta, preskaŭ ĉiam laŭ licenco GNU GPL, kiu igas, ke la aferoj disvolvita per Apertium devas plu esti libere redonataj al la komunumo

Historio de Apertium

- Apertium naskiĝis en 2005 kiel komuna laboro de tri laborgrupoj en Hispanio (el Valencilando, Galegio kaj Eŭskio) por krei liberan platformon por aŭtomata tradukado inter iberiaj latinidaj lingvoj
- En 2007 aperis la dua versio de Apertium, kiu donas pli da ebloj por traduki inter (iom) pli malproksimaj lingvoj.
- Samjare aperis la tria versio, kiu inkluzivis nur la uzadon de Unikodo

Apertium kaj malgrandaj lingvoj

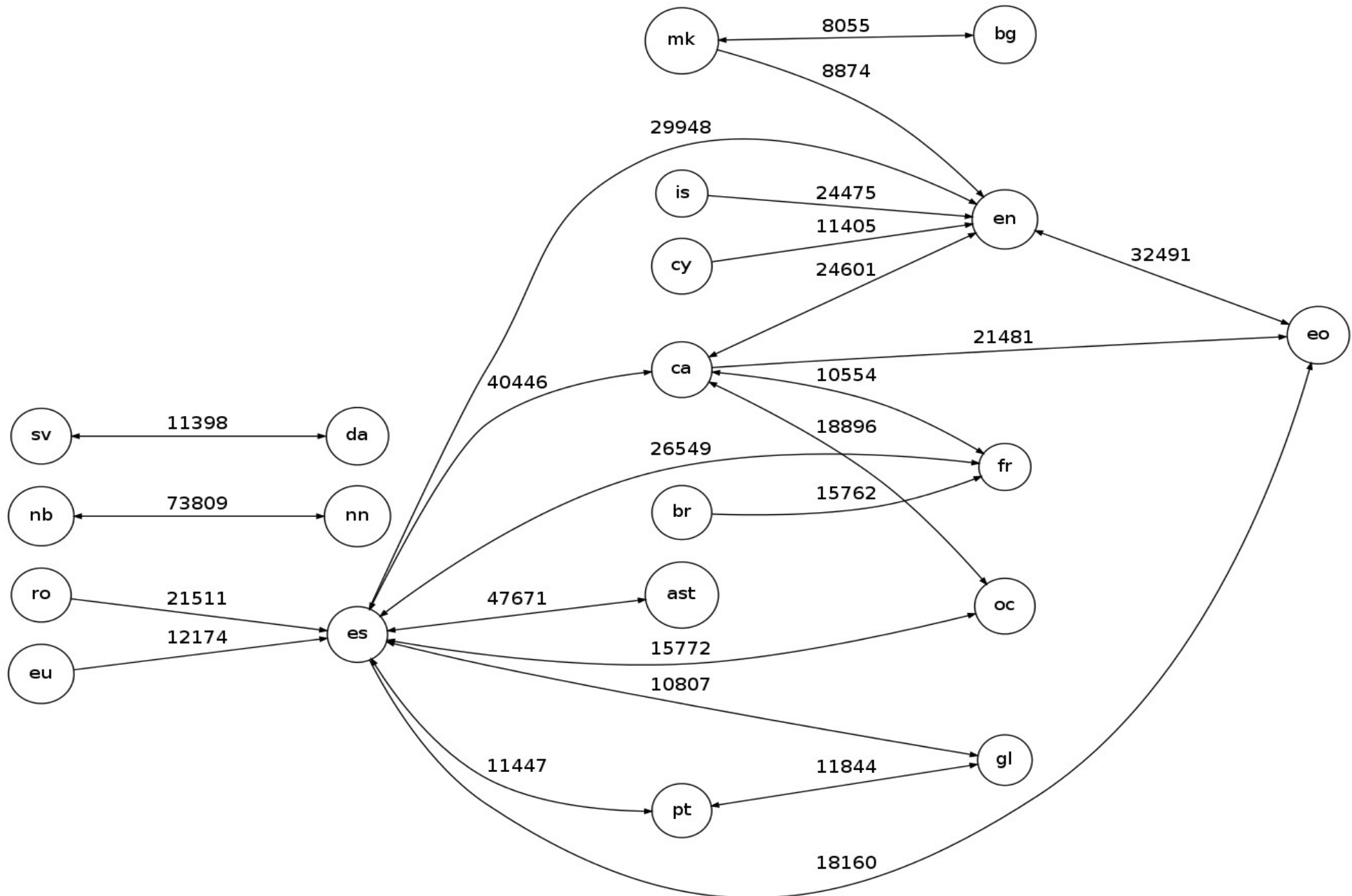
- Apertium estas aparte malfermita por t.n. "lingvoj kun malpli da rimedoj"
 - En ĝi estas funkciantaj tradukiloj por la kataluna, galega, eŭska, okcitana, asturia, bretona, kimra, makedona, nov-norvega, islanda, Esperanto... Oni evoluigas tradukilojn por la nord-samea, sud-samea, lule-samea, aragona, sarda, taĝika, afrikansa, haitia kreola k.a.
 - Fondintoj kaj ŝlosilaj evoluigantoj aktivas ankaŭ en la asocio SALT MIL (The ISCA Special Interest Group on Speech and Language Technology for Minority Languages)
 - Por tiaj ĉi lingvoj tradukiloj surbaze de reguloj ŝajnas pli promesplenaj ol la statistikaj, kiuj bezonas miriadon da dulingvaj tekstoj ĝuste ne ekzistantaj en multaj el tiuj lingvoj
 - En tia ĉi kunteksto Esperanto-tradukiloj estas aparte bone akceptataj en la Apertium-komunumo

Historio de Apertium

Evoluo de la tradukiloj

- 2005: 3 novaj tradukiloj
 - Inter latinidaj lingvoj
- 2006: 3 novaj tradukiloj
 - Unua tradukilo latinida ↔ angla
- 2007: 2 novaj tradukiloj
- 2008: 10 novaj tradukiloj
 - Unua tradukilo latinida ↔ eŭska
 - Unua tradukilo kun kelta lingvo (kimra ↔ angla)
 - Unuaj tradukiloj al Esperanto (el kataluna, hispana, angla)
- 2009: 5 novaj tradukiloj
 - Unuaj tradukiloj inter skandinavaj (nord-ĝermanaj) lingvoj
 - Unua tradukilo el Esperanto (Esperanto → angla)
- 2010: 3 novaj tradukiloj
 - Unuaj tradukiloj kun slavaj lingvoj

Publikigitaj tradukiloj de Apertium





Ĉu vi volas skribi malferman kodon?
Ĉu vi volas gajni monon?
Ĉu vi volas fari ambaŭ?



Historie, Google Summer of Code™ kunigis preskaŭ 3400 studentojn kun pli ol 200 projektoj kun malfermita kodo por krei milionojn da linioj da kodo. Ni serĉas mentorojn kaj studentojn por fari 2010 la plej bonan Summer of Code ĝis nun! Ni donos al sukcesaj studentaj kontribuantoj stipendion de 5000 USD (ĉ. 3630 eŭroj), tiel ke ili dum tri monatoj povu koncentriĝi je kodado de siaj projektoj.

La *Google Summer of Code*-programo kreiĝis por instigi studentojn partopreni en evoluado de malferma fontkodo (Open Source). Komenciĝinta en 2005, la programo havas plurajn celojn:

- Inspiri junajn programistojn ekpartopreni en evoluigo de malferma kodo
- Doni al studentoj de komputilscienco kaj rilataj kampoj la ŝancon fari laboron kiu rilatas al iliaj studoj dum la somero
- Doni al studentoj pli da kontakto kun la softvaro-evoluigaj situacioj en la vera mondo (ekz. distribuita evoluado, demandoj pri permesiloj, poŝtlista bontono ktp.)
- Igi pli da malferma kodo kreita kaj publikigata, por la utilo de ĉiu
- Helpi al projektoj kun malferma kodo trovi kaj envenigi novajn programistojn kaj kontribuantojn

Google akceptos kandidatojn de malfermfontaj projektoj inter la 8a kaj la 12a de marto 2010. Studentaj kandidatoj estos akceptataj inter la 29a de marto kaj la 9a de aprilo 2010.

Plenaj detaloj pri postuloj por partopreni, inkluzive kiel kandidatiĝi kaj pli pri la plej grava T-ĉemizo, ĉe <http://socghop.appspot.com>

Kiel funkcias Apertium?

Kelkaj gramatikaj etikedoj

- <n> - substantivo (nomo; o-vorto)
- <adj> - adjektivo (a-vorto)
- <vblex> - verbo
- <sg> - singularo, <pl> pluralo
- <nom> - nominativo, <acc> - akusativo
- Ekzemplo
 - kato<n><pl><nom> = katoj
 - kato<n><sg><acc> = katon

Apertium kaj Esperanto

- 2007

- La Universitato Pompeu Fabra el Barcelono (kreinta tri aliajn tradukilojn) evoluigas dum unu jaro tradukilojn kataluna → Esperanto hispana → Esperanto

- 2008

- Jacob Nordfalk evoluigas tradukilon angla → Esperanto

- 2009

- Jacob Nordfalk evoluigas tradukilon Esperanto → angla kaj novan version de la tradukilo angla → Esperanto (li krome okupiĝas pri la evoluigo de novaj rimedoj kaj de plibonigoj de la kerno)

- Hèctor Alòs evoluigas novajn versiojn de la tradukiloj kataluna → Esperanto hispana → Esperanto

- 2010

- Hèctor Alòs evoluigas tradukilon (ankoraŭ ne pretan) franca → Esperanto
- Jacob Nordfalk kunevoluigas la tradukilon sveda → dana
- Jacob Nordfalk estas elektita membro de la gvidorgano de Apertium

Stato de la Esperanto-tradukiloj

	vorto-paroj	t1x	t2x-2	t2x-1	t2x	t3x
kat-eo	21 481	187	-	-	72	1
hisp-eo	18 160	174	-	-	57	1
angl-eo	32 491	56	1	7	25	1
eo-angl	32 491	72	1	15	13	1
fr-eo	34 769	206	-	-	35	1

kat-angl	24 601	98	-	-	31	9
hisp-angl	29 948	87	-	-	27	8

La malfacilaĵoj de traduko el latinida lingvo

Ekzemploj el la kataluna

- Inversigo de la ordo de la adjektivoj en rilato al la nomo
- "Que" = ke, kio(n), kiu(j)(n), ol
- "seu" = lia, ŝia, ĝia, ilia, sia
- akuzativo
- Aldono de subjekta pronomo
 - Menjo -> Mi manĝas
 - Menja -> Li/Ŝi/Ĝi manĝas
- "es"
 - es menja = (li, ŝi, ĝi) manĝas, (li, ŝi, ĝi) manĝas sin, oni manĝas

Facile enkondukeblaj novaj lingvoj

- Pliaj latinidaj
 - Portugala (+revo); itala, rumana, galega, okcitana, asturia
- Pliaj ĝermanaj
 - Norvega(j), dana, sveda, islanda
 - germana?, nederlanda?
- Slavaj
 - Makedona, bulgara
 - Pola, serbokroata
- Keltaj
 - Kimra, bretona
- Aliaj: persa, kurda, hindia, bengala, nepala?

Pozitivaj aspektoj de Apertium (1)

- Temas pri malfermita platformo, universitate stirata, uzata de pluraj homoj/grupoj en diversaj landoj (pli ol 100 registritoj en sourcefourge)
- La platformo evoluadas kaj pliboniĝas sendepende de la fortoj de la Esperanto-komunumo; ankaŭ ĝia videbleco ne dependas de la esperantistoj
 - Tio ĉi rilatas ne nur al la reklama flanko por Esperanto pri la ekzisto de tiaj tradukiloj inter naciaj lingvoj kaj la internacia, sed ankaŭ ekz. pri la facileco publikigi artikolon en fakaj eldonaĵoj pri Esperanto-tradukilo aŭ ricevi stipendion por ĝia evoluigo
- La malfermeco de la kodo kaj de la lingvaj rimedoj permesas la kunlaboron de unuopaj evoluigantoj same kiel de malsamaj grupoj
 - Tio same validas por unuopaj Esperanto-evoluigantoj kiel inter tiuj ĉi kaj evoluigantoj de aliaj lingvoparoj

Pozitivaj aspektoj de Apertium (2)

- Ĝi estas relative facile uzebla kaj eblas partopreni en la evoluigo de Esperanto-tradukilo sen grandaj scioj pri komputiko kaj komputika lingvistiko (ĉefe, se staras apude pli fakaj gvidantoj)
- Kun la nuna sperto, la kreo de baza aŭtomata tradukilo el latinida lingvo al Esperanto (aŭ inversen) postulas 4-6 monato-homojn (600-1000 laborhorojn), depende de la sperto de la evoluiganto kaj de la preteco de la lingvorimedo; verŝajne tio estus ne multe pli por ĝermanaj kaj slavaj lingvoj, ĉefe por la ekzistantaj en Apertium
 - "baza" signifas por ni "nemirinda, sed tradukanta ne literaturajn tekstojn (ne solajn frazojn) al komprenebla lingvo sen eraroj en ĉiu frazo"
 - Estas evidente, ke la tradukilo de Eckhard Bick en->eo estas multe pli bona ol tiu de Apertium (kaj verŝajne eĉ ĝia kvalito estos malfacile atingebla de tiu ĉi), sed ĝi (tre bedaŭrinde) tute ne ŝajnas ĝeneraligebla; ĝi krome dependas nur de unusola homo kaj ne ebligas grupan plibonigon (pro kopirajteco kaj industria sekreto)

Kiel uzi Apertium?

- Ĉe <http://apertium.org/> eblas traduki
 - Platan tekston
 - Retejojn
 - Dokumentojn (OpenOffice, RTF, ...)
- De kaj al Esperanto