

# Maŝintradukado

Kiel ĝi funkcias, kion ĝi kapablas.  
Ilustrita per la malfermkoda sistemo de Apertium



Jacob Nordfalk  
Centro por plia edukado  
Kopenhaga Universitata Altlernejo de Inĝenieriko  
Center for Videreuddannelse  
Ingeniørhøjskolen i København  
[jacob.nordfalk@gmail.com](mailto:jacob.nordfalk@gmail.com)

Universala Kongreso  
29an de julio 2011



# Skemo

- Kio estas maŝintraduko (MT)?
- Ekzemploj de MT
- La du tipoj de MT plej uzataj
- Kialoj por uzi maŝintradukon
- Apertium
  - Apertium kaj Esperanto
  - Kiel kunlabori
  - Kiel funkcias Apertium
  - Fortoj kaj malfortoj de Apertium
- Kiel praktike uzi MT kaj Esperanto
- Elŝutejo por la lumbildoj  
<http://javabog.dk/esperanto/UK2011/>  
[bit.ly/rqqaoz](http://bit.ly/rqqaoz)



# Kio estas maŝintraduko?

- Aŭtomata tradukado de unu natura lingvo al alia pere de komputilo
- Unu el la grandaj defioj por komputilingvistiko
- "Sed.... Ĉu maŝintraduko povus malnecesigi Esperanton kiel pontan lingvon?"
  - Ĉu maŝino povus kompreni tekston?
  - Ĉu io/iu ajn povus traduki ion, sen kompreni la enhavon?





# La terura katastrofo en Norvegio



- Deadly terror attack in Norwegian capital - At least 91 youths killed in Utoya, May rise.....
  - Mortiga terorisma atako en norvega majuskla - Almenaŭ 91 junularoj mortigita en Utoya, maja altiĝo..... (Apertium)
  - Mortiga teroratako en norvega kapitalo - Almenaŭ 91 dekkelkjarulojn mortigita en Utoia, maja pliiĝo ... . . (Gramtrans)
  - Dødelige terror angreb i norske hovedstad - mindst 91 unge dræbt i Utoya, maj stiger ..... (Google)
  - Attaque terroriste meurtrière dans la capitale norvégienne - Au moins 91 jeunes tués dans Utoya mai élévation ..... (Google)
  - Tödlicher Terroranschlag in der norwegischen Hauptstadt - Mindestens 91 Jugendliche in Utoya getötet, Mai steigen ..... (Google)
  - Ataque terrorista mortífero en la capital de Noruega - Al menos 91 jóvenes muertos en Utoya, puede subir ..... (Google)

# ... nur eta eraro ŝanĝas tradukojn al sensaĵo

- Deadly terror attack in Norwegian capital - At least 91 youths killed in Utoya, **may** rise.....
  - Mortiga terorisma atako en norvega majuskla - Almenaŭ 91 junularoj mortigita en Utoya, povas altiĝi..... (Apertium)
  - Mortiga teroratako en norvega kapitalo - Almenaŭ 91 dekelkjarulojn mortigita en Utoia, povas pliiĝi ... (Gramtrans)
  - Dødelige terror angreb i norske hovedstad - mindst 91 unge dræbt i Utoya, kan stige ..... (Google)
  - Attaque terroriste meurtrière dans la capitale norvégienne - Au moins 91 jeunes tués dans Utoya, peut augmenter ..... (Google)
  - Tödlicher Terroranschlag in der norwegischen Hauptstadt - Mindestens 91 Jugendliche in Utoya getötet sind, können ..... (Google)
  - Ataque terrorista mortífero en la capital de Noruega - Al menos 91 jóvenes muertos en Utoya, puede subir ..... (Google)



# posta redakto estas necesa (vi helpu)

- Deadly terror attack in Norwegian capital - At least 91 youths killed in Utoya, May rise.....
  - Mortiga terorisma atako en norvega majuskla - Almenaŭ 91 junularoj mortigita en Utoya, povas altiĝi..... (Apertium)
  - Mortiga teroratako en norvega kapitalo - Almenaŭ 91 dekkelkjarulojn mortigita en Utoia, povas pliiĝi ... (Gramtrans)
  - Dødelige terror angreb i norske hovedstad - mindst 91 unge dræbt i Utoya, kan stige ..... (Google)
  - Attaque terroriste meurtrière dans la capitale norvégienne - Au moins 91 jeunes tués dans Utoya, peut augmenter ..... (Google)
  - Tödlicher Terroranschlag in der norwegischen Hauptstadt - Mindestens 91 Jugendliche in Utoya getötet sind, können ..... (Google)
  - Ataque terrorista mortífero en la capital de Noruega - Al menos 91 jóvenes muertos en Utoya, puede subir ..... (Google)



# postredakta versio

- Deadly terror attack in Norwegian capital - At least 91 youths killed in Utoya, May rise.....
  - Mortiga terorisma atako en norvega ĉefurbo - almenaŭ 91 junuloj mortis en Utoya, eble pli
  - Mortiga teroratako en norvega kapitalo - almenaŭ 91 dekkelkjarulojn mortigita en Utoia, povas pliiĝi...
  - Dødeligt terroristangreb i Oslo - mindst 91 unge dræbt i Utoya, måske flere...
  - Attaque terroriste meurtrière dans la capitale norvégienne - Au moins 91 jeunes tués dans Utoya, peut augmenter ..... (helpu!)
  - Tödlicher Terroranschlag in der norwegischen Hauptstadt - Mindestens 91 Jugendliche in Utoya getötet sind, können ..... (helpu!)
  - Ataque terrorista mortífero en la capital de Noruega - Al menos 91 jóvenes muertos en Utoya, puede subir ..... (helpu!)





# Kiel maŝintraduko funkcias

## La du tipoj de MT plej uzataj

- Regulbazitaj metodoj (Gramtrans, Apertium, ...)
  - Analizo, transfero, sintezo
    - Esence kiel lernanto tradukus per vortaro
  - Fidindaj, fidelaj al la originala teksto
    - Foje teda, ofte frazostrukturo kiel la originala teksto
- Statistikaj metodoj (Google, ...)
  - Teoremo de Bayes pri probablo
    - $P(A|B) P(B) = P(B|A) P(A)$
    - 'lernas' per amasego da eksemplaroj (paralelaj tekstaroj)
    - Pli viva lingvo, sed ofte kun eraroj



# Kialoj por uzi maŝintradukon

- Krei tekston por publikigo (disvastigo)
- Krei tekston por kompreni (asemilado)

## Kiaj tekstoj?

- Novaĵo, manlibro, poemo, faklibro, novelo, flugfolio, manifesto, hejmpaĝo, retmesaĝo, SMS, parola lingvo, subtekstoj de filmo, ...



# Publikigo (disvastigo)

- Postredakto de (profesia) tradukisto kiu konas ambaŭ lingvojn
  - La Grava Demando: **Kiom** oni devas postredakti por kontentiga rezulto ?
    - Ofta metriko estas la elcentaĵo de vortoj korektitaj dum la postredakto (WER = Word Error Rate)
- Fideleco al originalo vs. vivanta lingvo
  - Dependas de la tekstotipo kaj la celgrupo
    - Manlibroj? enorde... poemoj? forgesu!



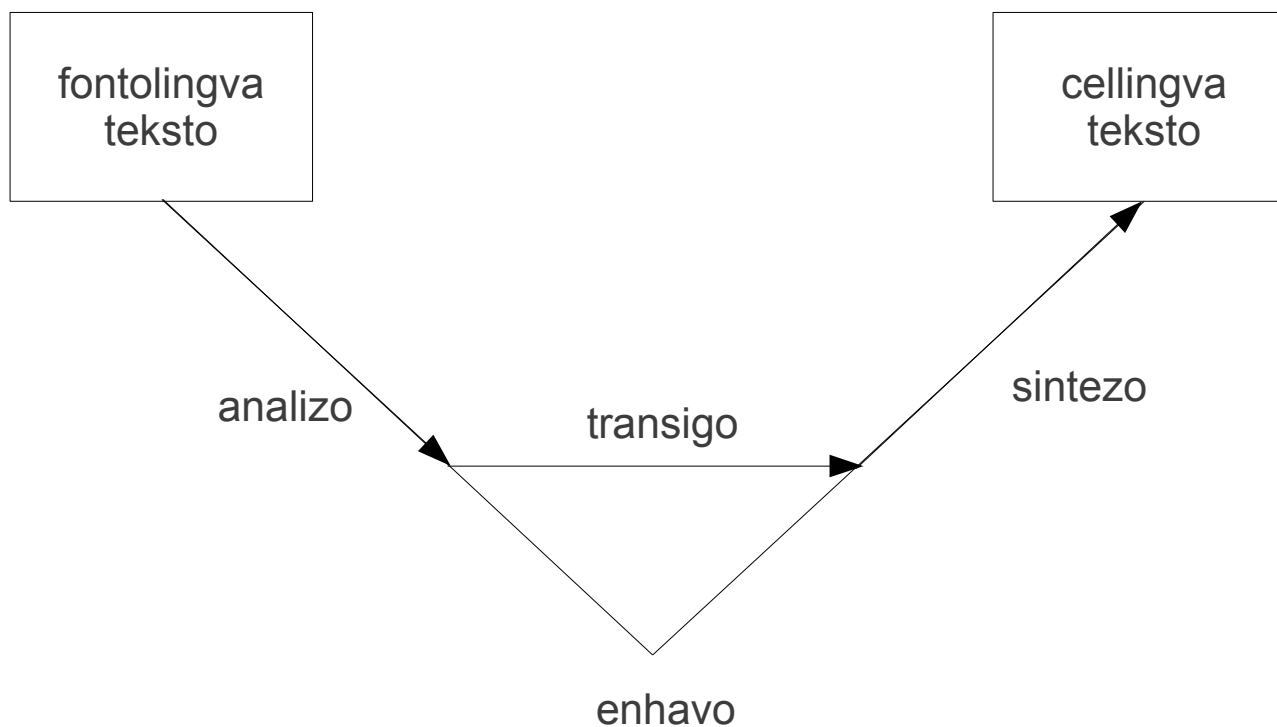
# Kompreni (asemilado)

- Krei tekston por kompreni (asemilado)
  - La Grava Demando: **Kiom** fidela la enhavo de la teksto estas al la originalo ?

... sed kiel oni mezuras tion?



# Regulbazita maŝintraduko



# 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 malfermfonta, preskaŭ ĉiam laŭ licenco GNU GPL, kiu igas, ke la aferoj disvolvitaĵ 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



# Historio de Apertium

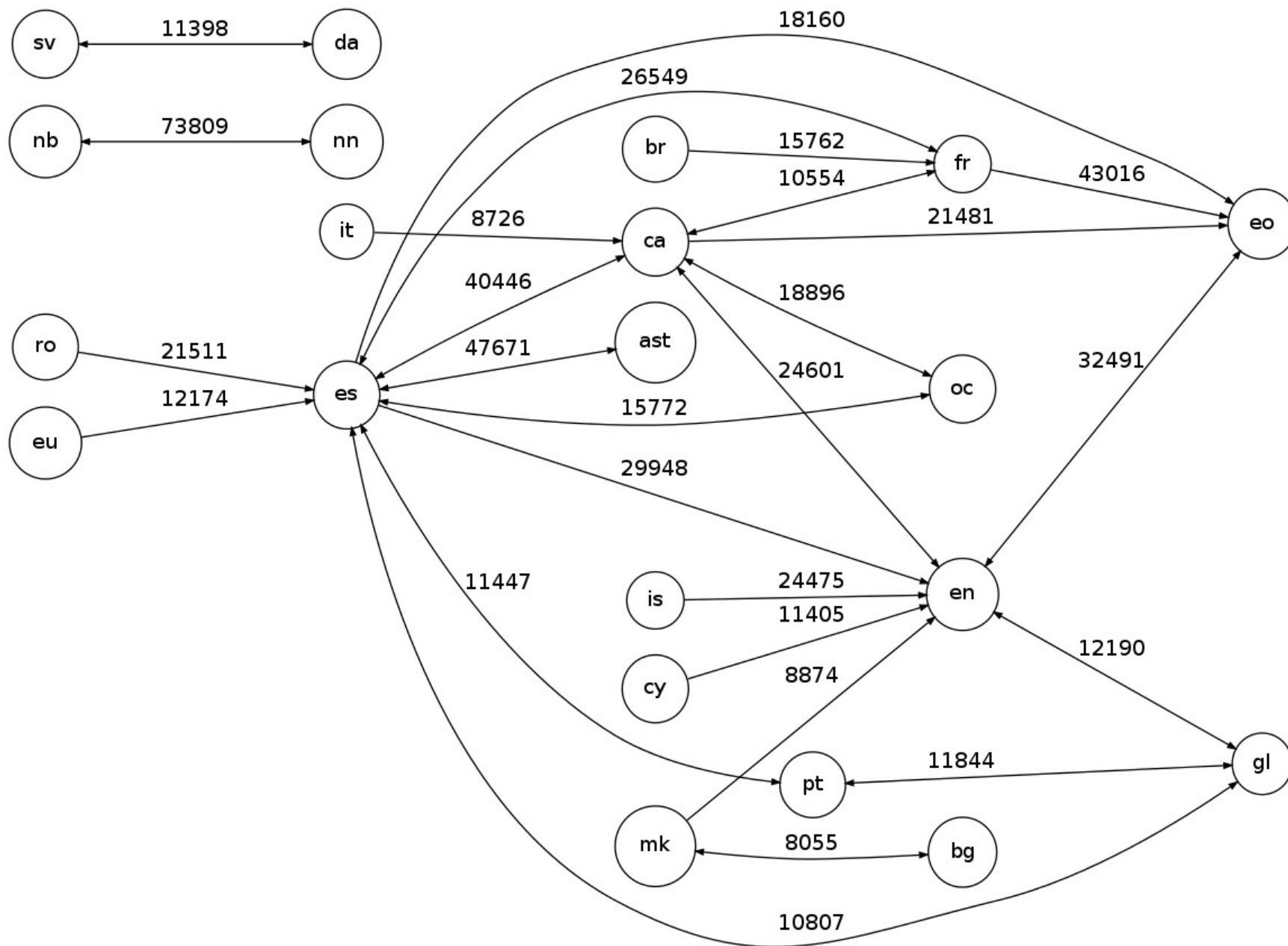
## Evoluo de la tradukiloj

- 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



# Apertium kaj malgrandaj lingvoj

- Apertium estas aparte malfermita por t.n. "lingvoj kun malpli da rimedoj"
  - En ĝi estas funkciaj 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 SALTMIL (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



# 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

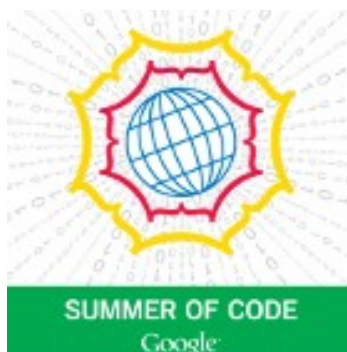
- Hèctor Alòs evoluigas tradukilon franca → Esperanto
- Jacob Nordfalk kunevoluigas la tradukilon sveda → dana
- Jacob Nordfalk estas elektita membro de la gvidorgano de Apertium



# Montrado

- Franca-Esperanto
  - <http://apertium.saluton.dk/common/browser.php?mark=>
- Angla-Esperanto
  - <http://apertium.saluton.dk/common/browser.php?mark=>
- Esperanto-angla
  - Ebla fenestro por ne-esperantistoj al Esperanto
  - Helpo bezonata!
  - <http://apertium.saluton.dk/common/browser.php?mark=>





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 kunlabori?

- Evoluigi novan lingvoparon  
(ekz. eo-de, eo-it, eo-sk)
- Evoluigi novan lingvodirekton  
(ekz. eo->es, eo->fr, eo->ca)
- Plibonigi ekzistantan tradukilon
  - Evoluigante novan version
    - Aldoni vortojn
  - Atentigante pri eraroj en retpaĝo

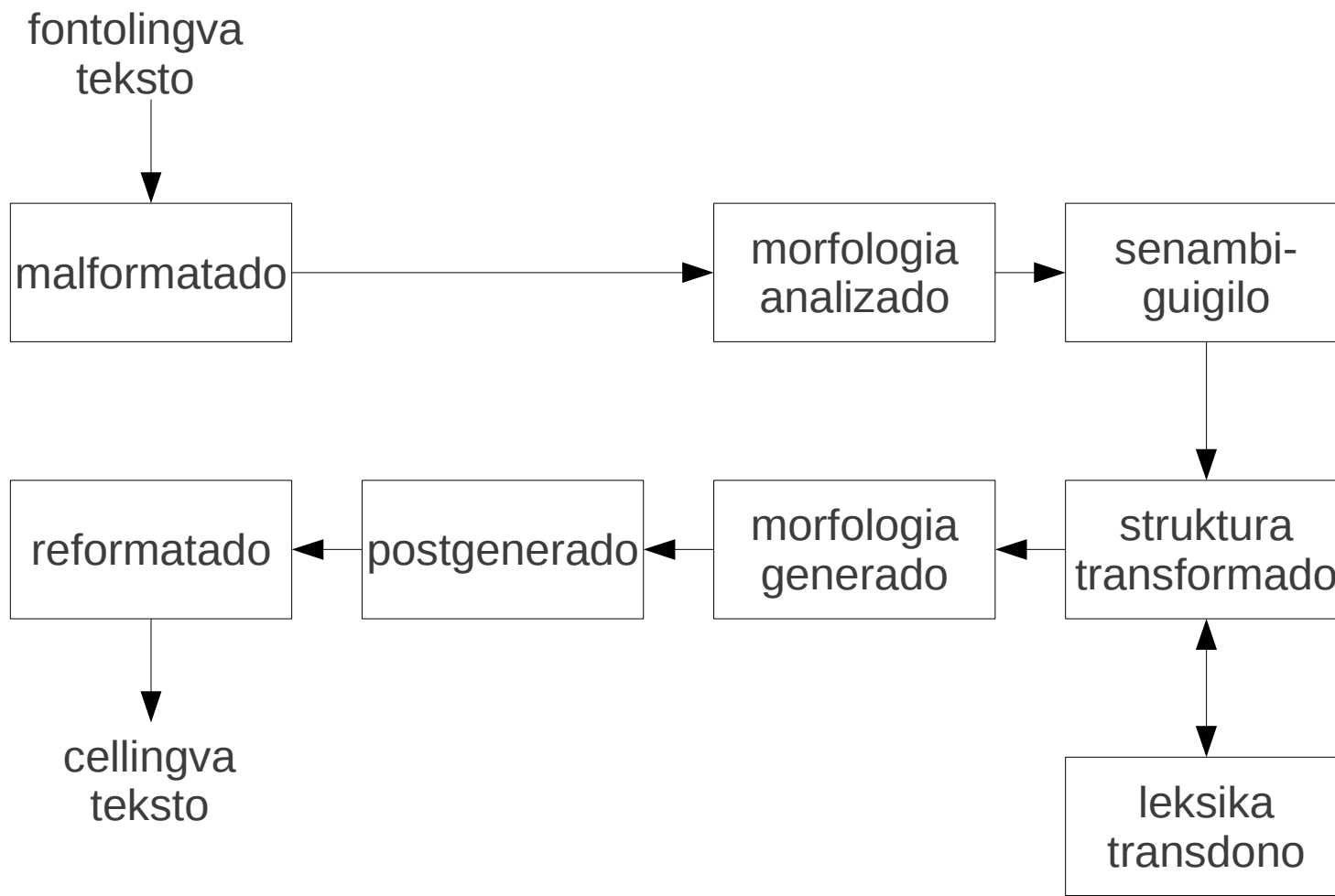


# Facile enkondukeblaj novaj lingvoj

- 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?



# Kiel funkcias Apertium? Ĉefaj elementoj





# Kiel funkcias Apertium?

## Programa arkitekturo

- Ĉiu etapo estas aparta programo
- Ĉiu etapo ricevas datumojn de la antaŭa etapo, transformas ilin kaj plusendas al la posta etapo
- La internaj interfacoj estas klare difinitaj por faciligi ŝanĝojn aŭ aldonojn
- La ĉeno estas facile manipulebla
  - Multaj lingvaj paroj ŝanĝas la ĉenerojn
    - alia aŭ asista senambiguilo
    - ekstraj transferaj niveloj
    - ortografiaj reguloj (ekz. *de + el* → *del*)



# Kiel funkcias Apertium? Transigaj reguloj

- Transigaj reguloj permesas ŝanĝi ecojn kaj strukturojn inter la fonta kaj cела lingvoj:
    - (es) *una señal inequívoca* -> (ca) *un senyal inequívoc*  
(ŝanĝo de genro de la tuta nom-sintagmo)
    - (en) *computer textbook* -> (eo) *komputila lernolibro*  
(ŝanĝo de gramatika kategorio: computer (nomo) -> komputila (adjektivo))
    - (en) *John's book* -> *libro de John*  
(ordoŝanĝo kaj aperigo de prepozicio)
    - (ca) *una casa gran i antiga* -> (eo) *granda kaj malnova domo*  
(ŝanĝo de la loko de la adjektivoj rilate al la nomo)
    - (fr) *j'aime ce livre* -> (eo) *mi ŝatas tiun libron*  
(aldono de akuzativo)
    - (es) *trabajo* -> (eo) *mi laboras*  
(aldono de persona pronomo)
    - (fr) *un ami n'aime pas ce livre* -> (eo) *amiko ne ŝatas tiun libron*  
(forigo de nedifinita artikolo kaj de la dua parto de neo)
- ktp.



# Kiel funkcias Apertium?

## Kelkaj gramatikaj etikedo

- <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



# Transigaj reguloj

- Etapo unu: Grupigo de frazelementoj
  - <det>, <adj>, <n> → grupo <SN>
    - ^a<det><ind><sg>\$ ^white<adj><sint>\$ ^cat<n><sg>\$
    - ^det\_adj\_nom<SN><sg><nom>{^blanka<adj><2><3>\$ ^kato<n><sg><3>\$}\$
  - <vblex> → grupo <SV>
- Etapo du: Apliko de reguloj sur la grupoj
  - <SN><nom>, <SV>, <SN><nom> → <SN><nom>, <SV>, <SN><acc>
- Etapo tri: Elpako de la grupoj
  - <SN><sg><acc>{^kato<n><2><3>\$} → kato<n><sg><acc>



# Kiel funkcias Apertium? Laŭpaŝa ekzemplo

- I saw the cat's tail
- `^I/I<num><sg>/prpers<prn><subj><p1><mf><sg>$`  
`^saw/saw<n><sg>/see<vblex><past>$ ^the/the<det><def><sp>$`  
`^cat/cat<n><sg>$ ^'s/'s<gen>$ ^tail/tail<n><sg>$`
- `^prpers<prn><subj><p1><mf><sg>$ ^see<vblex><past>$ ^the<det><def><sp>$`  
`^cat<n><sg>$ ^'s<gen>$ ^tail<n><sg>$`
- `^prnpers<SN><sg><nom>{^prpers<prn><p1><mf><sg><3>$}$`  
`^verb_all<SV><aliaj><past>{^vidi<vblex><3>$}$`  
`^det_nom_gen_det_nom<SN><sg><nom>{^la<det><def><sp>$ ^vosto<n><sg><3>$`  
`^de<gen>$ ^la<det><def><sg><nom>$ ^kato<n><sg><nom>$}$`
- `^prnpers<SN><sg><nom>{^prpers<prn><p1><mf><sg><3>$}$`  
`^verb_all<SV><aliaj><past>{^vidi<vblex><3>$}$`  
`^det_nom_gen_det_nom<SN><sg><acc>{^la<det><def><sp>$ ^vosto<n><sg><3>$`  
`^de<gen>$ ^la<det><def><sg><nom>$ ^kato<n><sg><nom>$}$`
- `^prpers<prn><p1><mf><sg><nom>$ ^vidi<vblex><past>$ ^la<det><def><sp>$`  
`^vosto<n><sg><acc>$ ^de<gen>$ ^la<det><def><sg><nom>$ ^kato<n><sg><nom>$`
- mi vidis la voston de la kato



# Kiel funkcias Apertium? Tri-ŝtupa transigo

- 1a ŝtupo

Surbaze de la morfologie etikedita fontolingva teksto,

- Kunigas lemojn en grupojn kaj aldonas al ĉi tiuj ĉapon kun sintaksaj informoj
- Tradukas laŭvorte
- Ŝanĝas morfologiajn informojn de la fontolingvo al la fontolingvaj, eventuale kun referencoj al la donitaĵoj de la grup-ĉapo (eventuale ŝanĝotaj en la dua ŝtupo)
- Eventuale, forigas elementojn (ĉi-okaze, la nedifinita artikolo)

- 2a ŝtupo

Surbaze de la sintakse etikedita cellingva teksto de la 1a ŝtupo,

- Modifas informojn de la ĉapo (ekz. kazoj)
- Eventuale, aldonas elementojn (ĉi-okaze, la subjekta pronomo)

- 3a ŝtupo

Surbaze de la sintakse etikedita cellingva teksto de la 2a ŝtupo,

- Heredigas la informojn de la ĉapo al la leksikaj elementoj
- Faras, se necese, lastajn strukturajn modifojn
- Forigas la vortgrupajn ĉapojn

- Aperis la eblo aldoni pli da ŝtupoj ol tiuj bazaj tri. La tradukilo en->eo uzas eĉ kvin ŝtupojn, kun du mezaj paŝoj inter la kutimaj 1a kaj 2a ŝtupo por malpliigi la kvanton de reguloj



# Kiel funkcias Apertium? Unulingvaj vortaroj

- Deklaroj de *lemoj* (=leksemoj kun ĉiuj surfacaj formoj de la sama gramatika kategorio)
  - `<e lm="ligo"> <i>lig</i><par n="o__n"/></e>`
  - `<e lm="ligi"> <i>lig</i><par n="i__vblex"/></e>`
- Deklaroj de *paradigmoj* (=ŝablonoj)
  - `<pardef n="o__n">`
  - `<e> <p><l>o</l>`  
`<r>o<s n="n"/><s n="sg"/><s n="nom"/></r></p></e>`
  - `<e> <p><l>oj</l>`  
`<r>o<s n="n"/><s n="pl"/><s n="nom"/></r></p></e>`
  - `<e> <p><l>on</l>`  
`<r>o<s n="n"/><s n="sg"/><s n="acc"/></r></p></e>`
  - `<e> <p><l>ojn</l>`  
`<r>o<s n="n"/><s n="pl"/><s n="acc"/></r></p></e>`
  - `</pardef>`
- Legado dekstren: analizo  
Legado maldekstren: generado



# Kiel funkcias Apertium? Dulingva vortaro

- `<e> <p><l>kato<s n="n" /></l>  
<r>cat<s n="n" /></r></p></e>`
- `<e> <p><l>slovaka<s n="adj" /></l>  
<r>Slovak<s n="adj" /></r></p></e>`
- `<e> <p><l>slovako<s n="n" /></l>  
<r>Slovak<s n="n" /></r></p></e>`
- `<e> <p><l>belorusa<s n="adj" /></l>  
<r>Belarusian<s n="adj" /></r></p></e>`
- `<e r="RL"><p><l>belorusa<s n="adj" /></l>  
<r>Byelorussian<s n="adj" /></r></p></e>`
- `<e r="LR"><p><l>bjelorusa<s n="adj" /></l>  
<r>Belarusian<s n="adj" /></r></p></e>`





# Kiel funkcias Apertium?

## Unulingvaj vortaroj: limigoj

- La uzado de paradigmoj principe povas funkcii nur per modifo en unu loko
  - Ne eblas ekzemple facile enkonduki en la vortaron la fleksion de plurvortaj esprimoj:
    - elementa lernejo
    - elementaj lernejoj
    - elementan lernejon
    - elementajn lernejojn



# Kiel funkcias Apertium?

## Dulingva vortaro: limigoj

- Unu vorton oni povas traduki principe nur al unu formo
  - *table* -> *tablo* aŭ *tabelo*
  - *accent* -> *akcento* aŭ *akĉento* aŭ *supersigno*
  - prepozicioj, konjunkcioj, subjunkcioj... !!!
- Tamen eblas
  - malsamaj tradukoj por malsamaj gramatikaj formoj
    - *Slovak* nomo (= *slovako*) kaj *Slovak* adjektivo (= *slovaka*)
    - *ĉiu* singurale (= *every*) kaj *ĉiu* plurale (= *all*)
  - plurvortaj leksikaj elementoj en vortaroj
    - *get in, get on, get off, get up...*
  - pritrakto de ŝablonoj en la transform-strukturaj reguloj
    - En la kataluna ofte estas egalaj konjugaciitaj formoj por la verboj *ŝajni* kaj *naski* (*parèixer, parir*) aŭ *voli* kaj *flugi* (*volar, volar*). Ĉar ne eblas "naski, ke" aŭ "flugi, ke", en tiaj kunteksto ni ĉiam elektas "ŝajni, ke" aŭ "voli, ke".



# Kiel funkcias Apertium?

## Transigaj reguloj: limigoj

- La maleblo uzi regulajn esprimojn en la ŝablonoj\* kaŭzas, ke amasiĝas la kvanto de kreendaj reguloj, ĉar montriĝas, ke ofte necesas plena analizo de la nom-grupoj
  - inter la latinidaj lingvoj kaj Esperanto, por inversigi la lokon de adjektivoj en rilato al la nomo
  - el la latinidaj lingvoj kaj el la angla al Esperanto, por trovi la kernon de la reprezentato de la rilativaj pronomoj tradukendajn al *kiu(j)(n)* kaj *kia(j)(n)*, por ekscii ties nombron
  - el Esperanto al la latinidaj lingvoj kaj al la angla, por kompreni la personon kaj nombron de la subjekto kaj konjugacii la verbon laŭe
  - Krome, el la franca al Esperanto necesas kapti la tutan verbon, kiu povas havi adverbojn, pronomojn kaj neajn adverbojn en si (inter la helpverbo kaj la ĉefverbo)
  - ktp.
- La skriptolingvo uzata en la transigaj reguloj estas iom malkutima kaj ne aparte afabla. Krome temas pri ordonema programlingvo, dum oni emus havi la suplecon kaj koncizecon donatajn alispecaj lingvoj por tiaj celoj.

\*Regulaj esprimoj ne estas akceptataj por ne malrapidigi la tradukilojn: rapideco estas unu el la dizajnaj ĉefpunktoj de Apertium



# Problemoj de Apertium: Senambiguigilo

- Granda parto de niaj problemoj originas de la senambiguigilo. Ekz:
  - He has the right to go. → Li havas la *ĝustan* iri.  
(Estas elektita right<adj> anstataŭ right<n><sg> = rajto)
- Kialoj
  - perfekta senambiguigilo ne eblas - eĉ ne per homo
  - la Apertium-senambiguiloj ne estas konstruataj surbaze de etikeditaj tekstaroj
  - La nuntempa senambiguigilo estas du-grama; provoj kun trigramo senambiguigilo montriĝis ne elstare pli bonaj
  - La senambiguigiloj de Apertium estas programeblaj per reguloj. Eble ni ne estis sufiĉe lertaj por uzi tiun eblon por solvi la renkontitaj problemojn

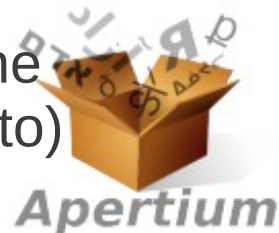
# Pozitivaj aspektoj de Apertium (1)

- 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 lingvorimedoj; 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 praktike uzi MT kaj Esperanto?

- Ĉe <http://apertium.org/> aŭ <http://saluton.apertium.org> eblas traduki
  - Platan tekston
  - Retejojn
  - Dokumentojn (OpenOffice, RTF, ...)
- De kaj al Esperanto
- ... sed ne forgesu <http://traduku.net>

