

Google Summer of Code: Get a summer job with
Apertium - open-source rule-based machine translation

Jacob Nordfalk

Center for Videreuddannelse

Ingeniørhøjskolen i København DTU Campus Ballerup

jacno@dtu.dk / jacob.nordfalk@gmail.com

CST 17. april kl. 15-17

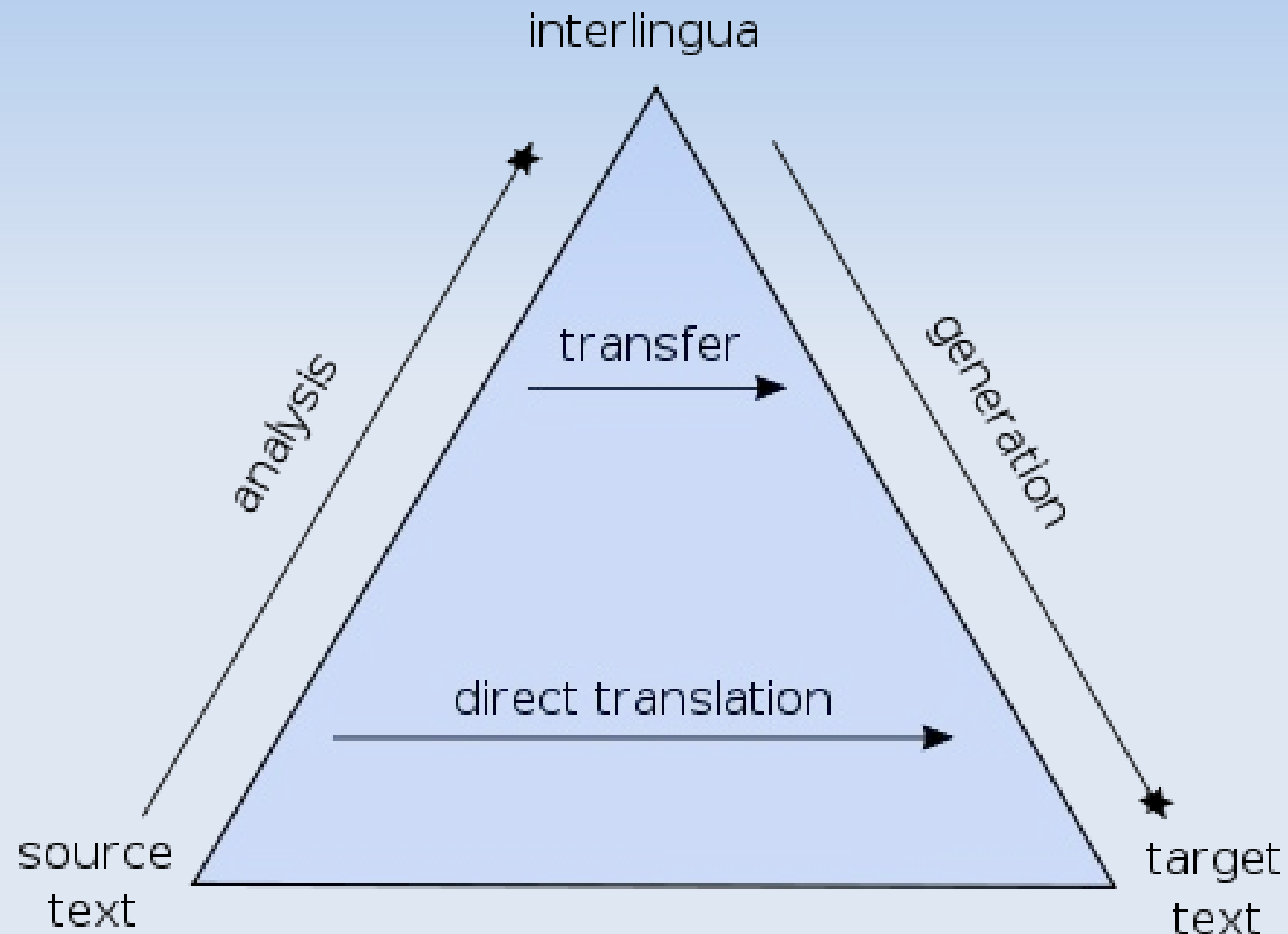
lokale 27.0.09.

Jacob Nordfalk

- Jacob Nordfalk
 - Assoc. professor (lektor) at Danish Technical University, Campus Ballerup
 - Java, Linux, Web, Android, advanced topics
 - Author of 3 Java programming books, <http://javabog.dk>
 - Active in the community of the International Language Esperanto
 - Self-chosen unemployed in Nepal 2006-2008
 - Working with i.a. Apertium and English-Esperanto (and mentoring Swedish-Danish during Google Summer of Code)
- Android
 - Mentor of ~100 Android projects
 - A few apps on Google Play, i.a. [Apertium offline translator](#)



Approaches to MT



Bernard Vauquois' pyramid

Angrebsvinkler på maskinoversættelse

- Regelbaseret MT (BabelFish, systrans, ...)
 - Sprogstrukturer "programmeres" med regler
 - Pålideligt, kedelig stil
 - "shallow-transfer":
 - Apertium - open source
 - dyb transfer (fuld grammattisk analyse):
 - Gramtrans, Babelfish
- Statistisk MT (Google translate, Giza++)
 - Gennem søger **store** mængder *parallel tekst* og søger v.hj.a statistik at stykke en tekst sammen på målsproget
 - Kvalitet afhænger meget af størrelse og kvalitet af corpus
 - Upålidelig, men mere levende stil, mange grammatiske fejl

Google Translate (2010)

En tilfældig side (fra svensk wikipedia):

Trakterna kring Fredriksberg räknas som bebodda sedan 1600-talet.

Google Oversæt giver (nok via engelsk):

Områderne omkring Fredriksberg tælles som har været besat siden 1600-tallet.

(er Google nu egentlig en hjælp?)

Se <http://translationparty.com>

Formål med maskinoversættelse

To brugssituationer:

Formidling - maskinen giver en råtekst til efterredigering. Målet er at få redigeringsdistancen (dvs. antal rettelser der skal til for at rette maskinoversættelsen så den er acceptabel som en rigtig oversættelse) så langt ned som muligt.

Forståelse - man kan ikke forstå kildesproget (f.eks. russisk) og bruger derfor en maskine til at lave en oversættelse som man kan forstå.

Swedish and Danish

- Standardised in the 12th to 15th centuries out of the Old Norse which was spoken across Scandinavia.
- Swedish on the speech around Stockholm,
- Danish on the speech around Copenhagen
- The languages are largely mutually intelligible
 - focus on production of text for dissemination (for post-editing)
 - production of text for assimilation (understanding) less important



Google Translate

Trakterna kring Fredriksberg räknas som bebodda sedan 1600-talet.

Områderne omkring Fredriksberg tælles som har været besat siden 1600-tallet.

havde det været hurtigere bare at oversætte originalteksten?

havde en ord-for-ord-oversættelse hjulpet?

hvad med en ord-for-ord-oversættelse + nogle regler?
== regelbaseret maskinoversættelse!

Exercises ;-)

- Download the presentations
 - <http://javabog.dk/filer/apertium-cst/> <http://bit.ly/Z29hbJ>
- Visit the wiki
 - <http://wiki.apertium.org>
http://wiki.apertium.org/wiki/Google_Summer_of_Code
- Try it out on your PC/Mac/Linux
 - <http://wiki.apertium.org/wiki/Apertium-viewer>
- Install the app
 - Search for 'Apertium' in Google Play, or open <https://play.google.com/store/apps/details?id=org.apertium.android>
- Instrukser på dansk
 - <http://bit.ly/5K45lb>
http://wiki.apertium.org/wiki/Dansk_introduktion

Apertium

- rule based
- open source (GPL - Gnu Public License)
 - the engine
 - all linguistic data
 - all supplementary tools
- host for an active and very helpfull community
 - Mailing list
 - Chat #apertium on Freenode
- relatively easy to learn, significant contributions can be made also by students without university degrees in linguistics or IT
- not that resource demanding
(a language pair takes ~4 months to make)

The Apertium project

Apertium is an open-source (GPL) machine translation platform. The platform provides

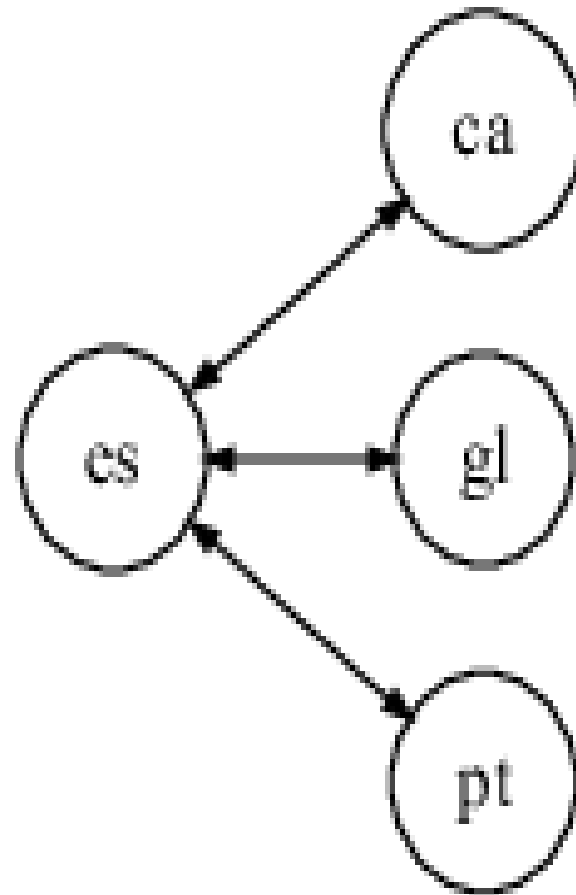
- a language-independent MT engine
- tools to manage linguistic data for language pairs
- linguistic data for >70 languages
 - Esperanto ↔ English Swedish ↔ Danish Catalan ↔ Romanian Welsh ↔ English English ↔ Afrikaans English ↔ Catalan English ↔ Spanish English ↔ Polish Esperanto ← Catalan Esperanto ← Spanish Esperanto ← Nepali Spanish ↔ Catalan Spanish ↔ Galician Spanish ↔ Italian Spanish ↔ Portuguese Spanish ← Romanian Basque ↔ Spanish French ↔ Catalan French ↔ Spanish Occitan ↔ Catalan Occitan ↔ Spanish Serbo-Croatian ↔ Macedonian Nynorsk ↔ Bokmål ...

The Apertium project

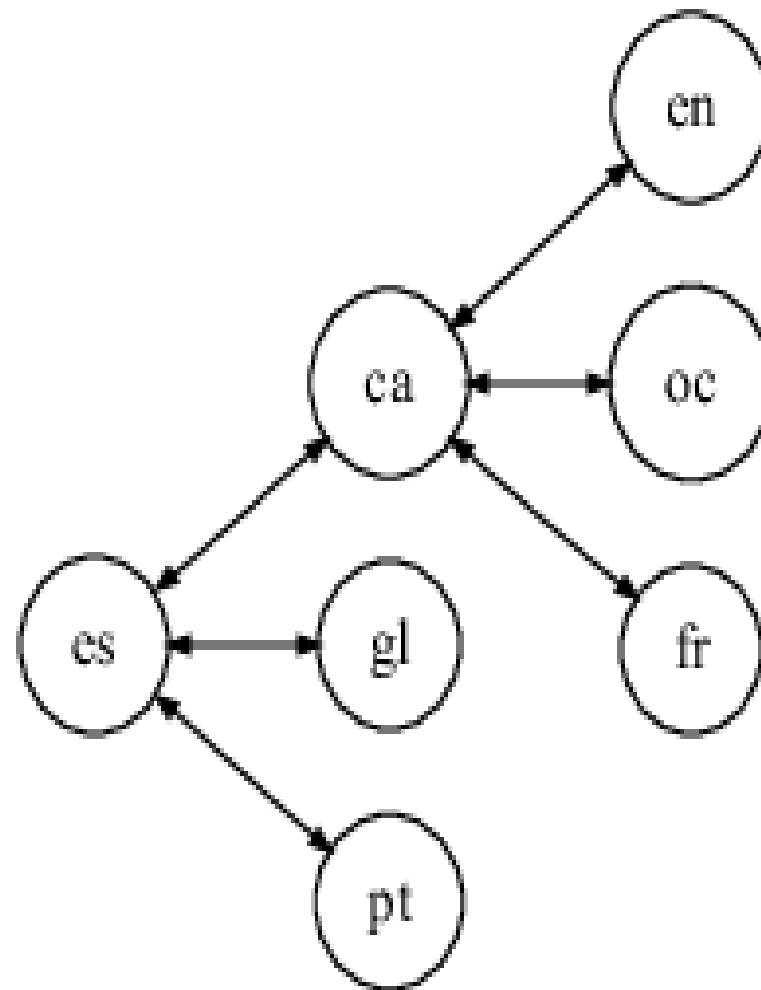
- uses a shallow-transfer MT
- processes in stages, as in an assembly line:
 - de-formatting, morphological analysis, part-of-speech disambiguation (tagging), shallow structural transfer, lexical transfer, morphological generation, and re-formatting.
- uses
 - finite-state transducers for all lexical processing operations
 - hidden Markov models for part-of-speech tagging
 - multi-stage finite-state based chunking for structural transfer.

2005 - Apertium 1.0

1 stage transfer - closely related languages

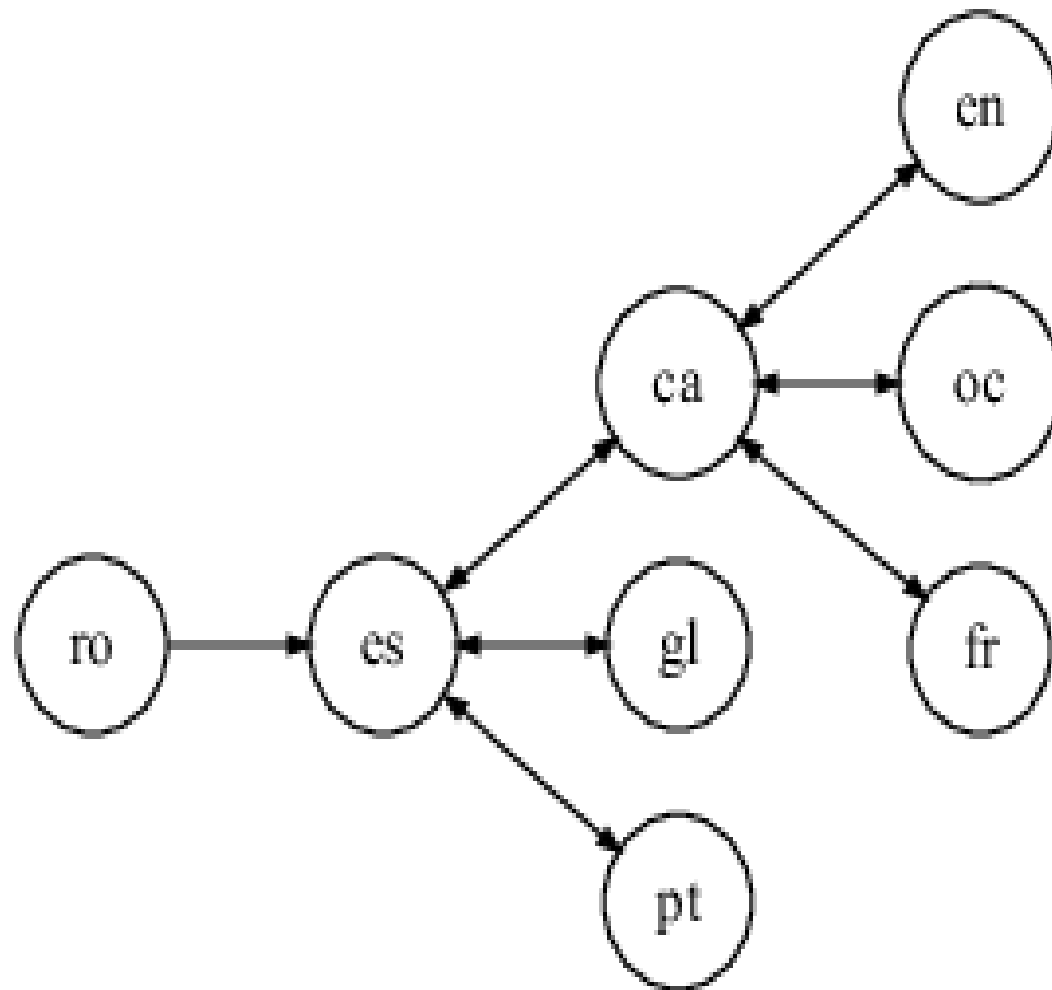


2006 - Apertium 2.0



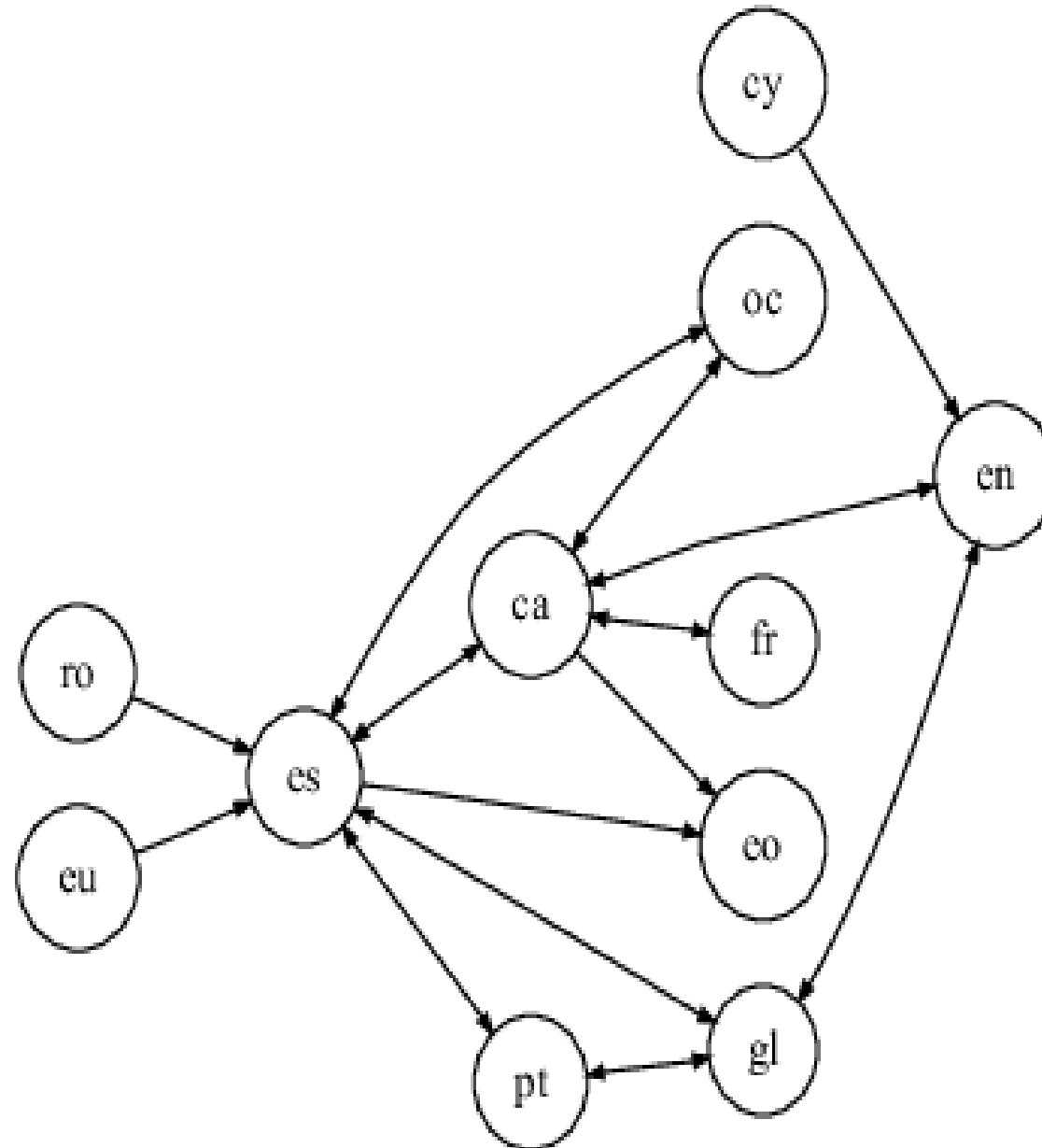
2007 - Apertium 2.0

3 stage structural transfer, chunking

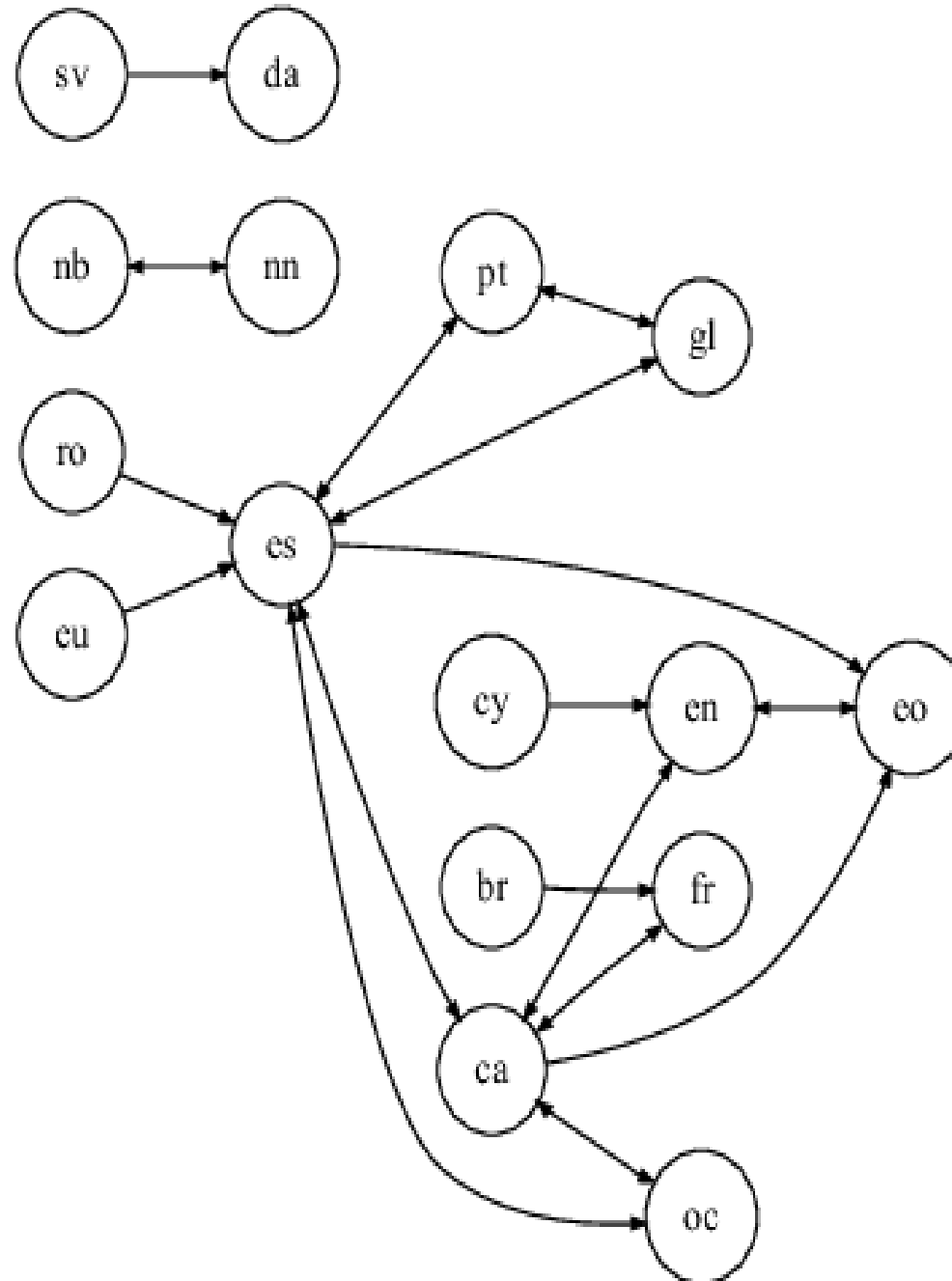


2008 - Apertium 3.0

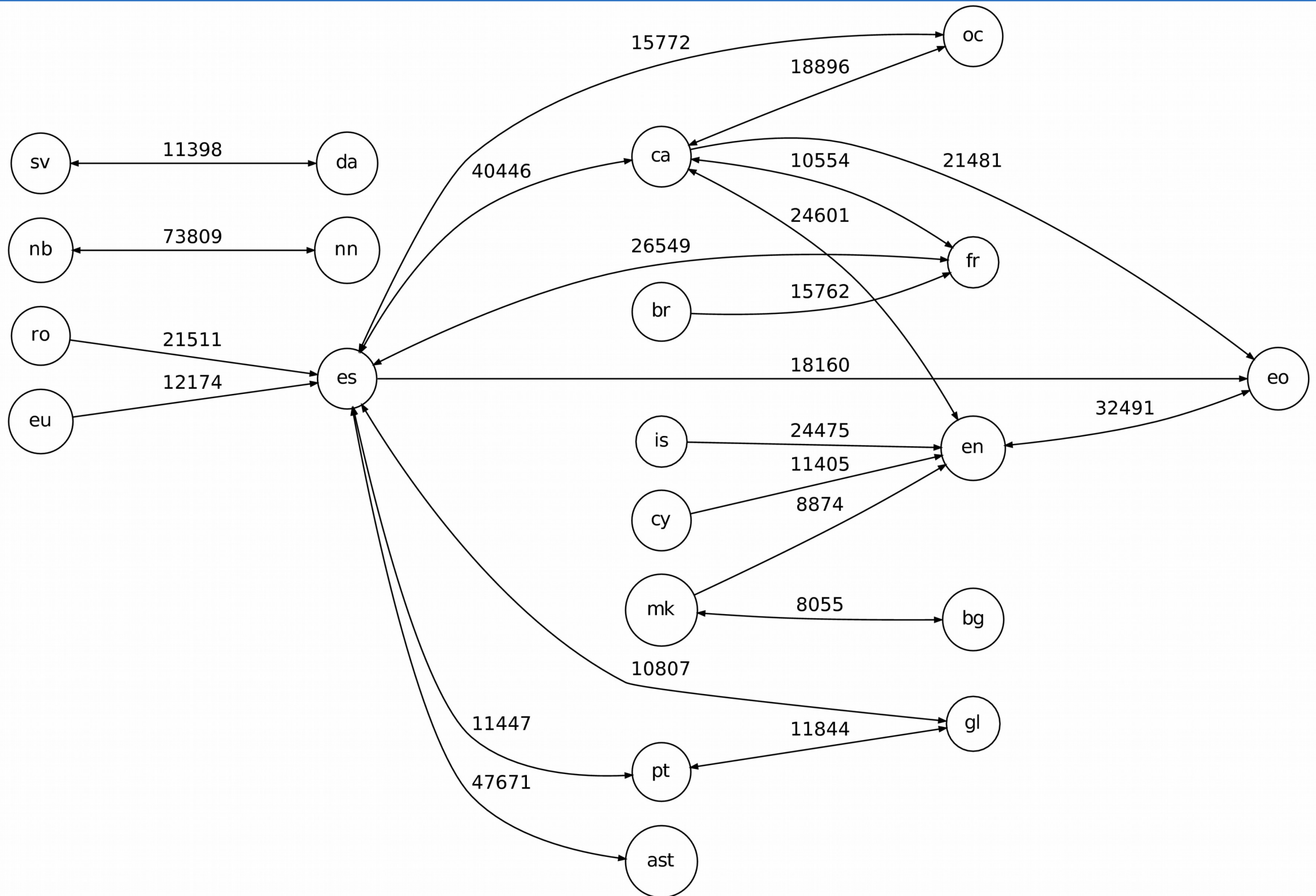
Unicode support



2009

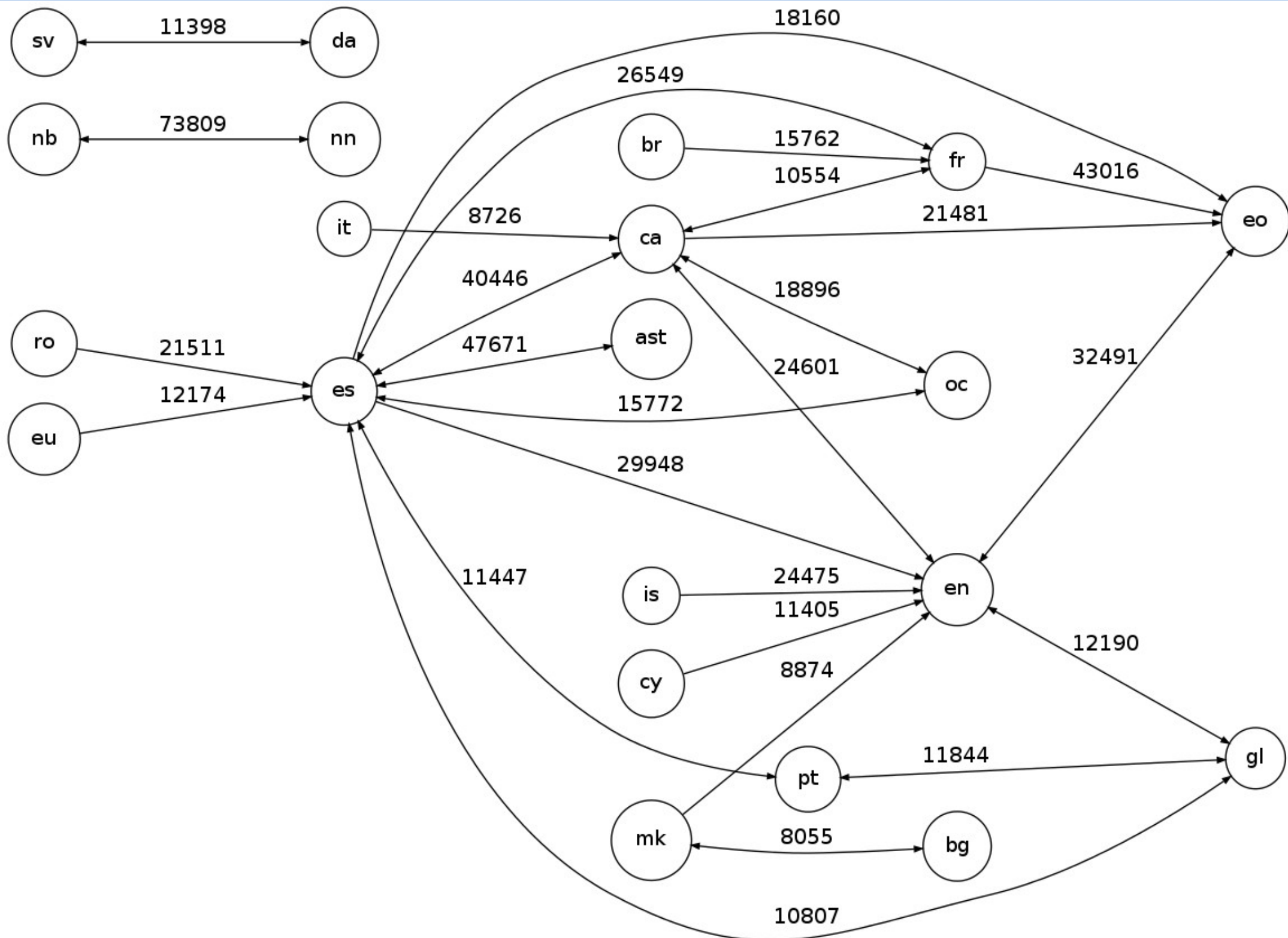


2010



2011

>3 stage structural stansfer



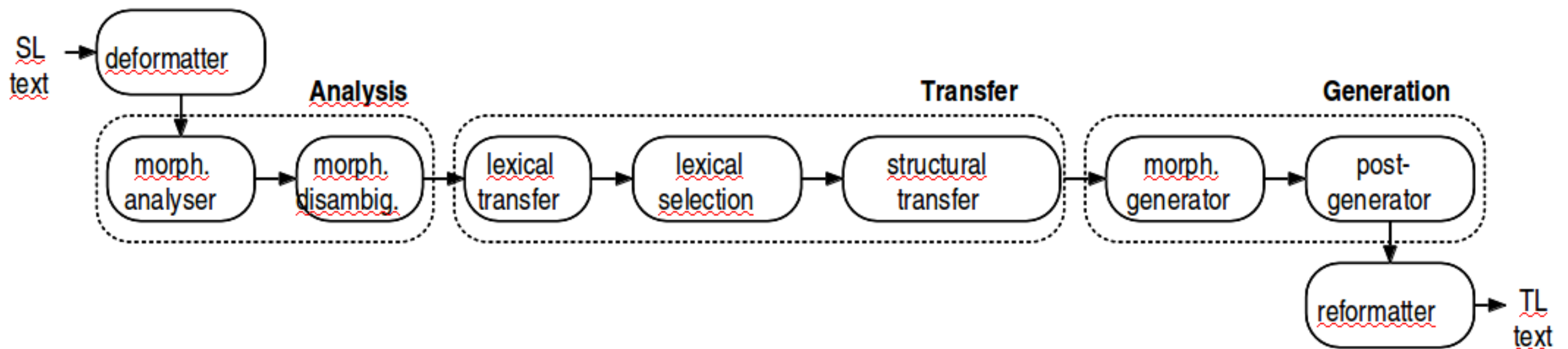
2013

- Released
 - 35 language pairs released - ready to use
 - 8 language pairs 'staging' - 'need a little more work'
 - 30 language pairs in nursery - 'needs more work'
 - data for 129 pairs waiting in incubator
- See http://wiki.apertium.org/wiki/List_of_language_pairs

Apertium project

- Rule based "shallow-transfer" MT
 - finite-state transducers for all lexical processing operations, hidden Markov models for part-of-speech tagging, and multi-stage finite-state based chunking for structural transfer.
- processing sker i trin, v.hj.a. UNIX pipes:
 - de-formattering (HTML, ODF, ...),
 - morfologisk analyse (v.hj.a. kildesprogs monodix),
 - part-of-speech disambiguation/tagging,
 - structurel transfer, leksikalsk transfer (her bruges bidix),
 - morfologisk generering (v.hj.a. målsprogs monodix),
 - re-formatting.

Architecture of Apertium MT



Morfological analysis

- English and Esperanto language pair
 - Marking accusative (n)
 - I saw a signal -> Mi vidis signalon
 - English words can be very ambiguous
 - ^saw/
 - saw<n><sg>/
 - saw<vblex><inf>/
 - saw<vblex><pres>/
 - saw<vblex><imp>/
 - see<vblex><past>\$

Apertium-viewer showing the pipeline

The screenshot shows the Apertium-viewer application window. The title bar reads "Apertium-viewer". The menu bar includes "File", "Tools", and "View". The toolbar contains several options: "Mark unknown words" (checked), "Show commands" (checked), "Fit", "Hide intermediate", "Copy", "Store", "Mode" (set to "English → Esperanto"), and "Local" (selected) / "Online" (unselected). The main text area displays the input sentence: "I saw a signal". Below this, the pipeline steps are shown, each with a "Freeze" button and a search icon:

- lt-proc** /usr/local/share/apertium/apertium-eo-en/en-eo.automorf.bin: Shows the initial tokenization and morphological analysis of the input words.
- apertium-tagger** -g \$2 /usr/local/share/apertium/apertium-eo-en/en-eo.prob: Shows the tagging process where parts of speech are assigned to the tokens.
- apertium-transfer** /usr/local/share/apertium/apertium-eo-en/apertium-eo-en.en-eo.t1x /usr/local/share/apertium/apertium-eo-en/en-eo.t1x.bin: Shows the transfer of the tagged English text into the Esperanto grammar rules.
- apertium-interchunk** /usr/local/share/apertium/apertium-eo-en/apertium-eo-en.en-eo.t2x /usr/local/share/apertium/apertium-eo-en/en-eo.t2x.b: Shows the chunking process where the Esperanto text is broken into syntactic chunks.
- apertium-postchunk** /usr/local/share/apertium/apertium-eo-en/apertium-eo-en.en-eo.t3x /usr/local/share/apertium/apertium-eo-en/en-eo.t3x.b: Shows the final post-chunking adjustments to the Esperanto text.
- lt-proc** \$1 /usr/local/share/apertium/apertium-eo-en/en-eo.autogen.bin: Shows the final output of the pipeline, which is the Esperanto translation: "mi vidis signalon".

Dictionaries and rules

- The three dictionaries

- English monodix

```
<e lm="see"><i>s</i><par n="s/ee__vblex"/></e>  
<e lm="saw"><i>saw</i><par n="house__n"/></e>  
<e lm="saw"><i>saw</i><par n="accept__vblex"/></e>
```

- English-Esperanto bidix

```
<e><p><l>vidi<s n="vblex"/></l><r>see<s n="vblex"/></r></p></e>  
<e><p><l>segilo<s n="n"/></l><r>saw<s n="n"/></r></p></e>  
<e><p><l>segi<s n="vblex"/></l><r>saw<s n="vblex"/></r></p></e>
```

- Esperanto monodix

```
<e lm="vidi"><i>vid</i><par n="verb__vblex"/></e>  
<e lm="segilo"><i>segilo</i><par n="nom__n"/></e>  
<e lm="segi"><i>seg</i><par n="verb__vblex"/></e>
```

- Transfer rules

- Also written in XML

Paradigm definitions

```
<pardef n="s/ee_vblex">
  <e>      <p><l>ee</l>          <r>ee<s n="vblex"/><s n="inf"/></r></p></e>
  <e>      <p><l>ee</l>          <r>ee<s n="vblex"/><s n="pres"/></r></p></e>
  <e>      <p><l>ees</l>         <r>ee<s n="vblex"/><s n="pres"/><s n="p3"/><s n="sg"/></r></p></e>
  <e>      <p><l>aw</l>          <r>ee<s n="vblex"/><s n="past"/></r></p></e>
  <e>      <p><l>een</l>         <r>ee<s n="vblex"/><s n="pp"/></r></p></e>
  <e>      <p><l>eeing</l>       <r>ee<s n="vblex"/><s n="ger"/></r></p></e>
</pardef>

<pardef n="house__n">
  <e>      <p><l></l>            <r><s n="n"/><s n="sg"/></r></p></e>
  <e>      <p><l>s</l>           <r><s n="n"/><s n="pl"/></r></p></e>
</pardef>

<pardef n="accept__vblex">
  <e>      <p><l></l>            <r><s n="vblex"/><s n="inf"/></r></p></e>
  <e>      <p><l></l>            <r><s n="vblex"/><s n="pres"/></r></p></e>
  <e>      <p><l>s</l>           <r><s n="vblex"/><s n="pres"/><s n="p3"/><s n="sg"/></r></p></e>
  <e>      <p><l>ed</l>         <r><s n="vblex"/><s n="past"/></r></p></e>
  <e>      <p><l>ed</l>         <r><s n="vblex"/><s n="pp"/></r></p></e>
  <e>      <p><l>ing</l>        <r><s n="vblex"/><s n="ger"/></r></p></e>
</pardef>

<e lm="see"><i>s</i><par n="s/ee_vblex"/></e>
<e lm="saw"><i>saw</i><par n="house__n"/></e>
<e lm="saw"><i>saw</i><par n="accept__vblex"/></e>
...
<e lm="adsorbate">      <i>adsorbate</i><par n="house__n"/></e>
<e lm="adsorbent">     <i>adsorbent</i><par n="house__n"/></e>
<e lm="adsorption">    <i>adsorption</i><par n="house__n"/></e>
<e lm="adulation">     <i>adulation</i><par n="house__n"/></e>
<e lm="adult">         <i>adult</i><par n="house__n"/></e>
...
<e lm="appeal">        <i>appeal</i><par n="house__n"/></e>
<e lm="appeal">        <i>appeal</i><par n="accept__vblex"/></e>
<e lm="appear">        <i>appear</i><par n="accept__vblex"/></e>
<e lm="appearance">    <i>appearance</i><par n="house__n"/></e>
<e lm="appease">       <i>appeas</i><par n="liv/e__vblex"/></e>
<e lm="append">       <i>append</i><par n="accept__vblex"/></e>

see: see<vblex><inf>
see: see<vblex><pres>
sees: see<vblex><pres><p3><sg>
saw: see<vblex><past>
seen: see<vblex><pp>
seeing: see<vblex><ger>

saw: saw<n><sg>
saws: saw<n><pl>

saw: saw<vblex><imp>
saw: saw<vblex><inf>
saw: saw<vblex><pres>
saws: saw<vblex><pres><p3><sg>
sawed: saw<vblex><past>
sawing: saw<vblex><ger>
```

Less related languages (chunking) / 3-stage transfer

I saw a signal

becomes after disambiguation

```
^prpers<prn><subj><p1><mf><sg>$  
^see<vblex><past>$  
^a<det><ind><sg>$  
^signal<n><sg>$.
```

which is transferred and chunked into

```
^prnpers<SN><p1><mf><sg>{^prpers<prn><subj><2><3><4>}$}$  
^verb<SV><past>{^vidi<vblex><past>}$}$  
^nom<SN><sg><nom>{^signalo<n><2><3><4>}$}$.
```

and transformed by rule SN SV SN -> SV SV SN<acc>

```
^prnpers<SN><p1><mf><sg>{^prpers<prn><subj><2><3><4>}$}$  
^verb<SV><past>{^vidi<vblex><past>}$}$  
^nom<SN><sg><acc>{^signalo<n><2><3><4>}$}$.
```

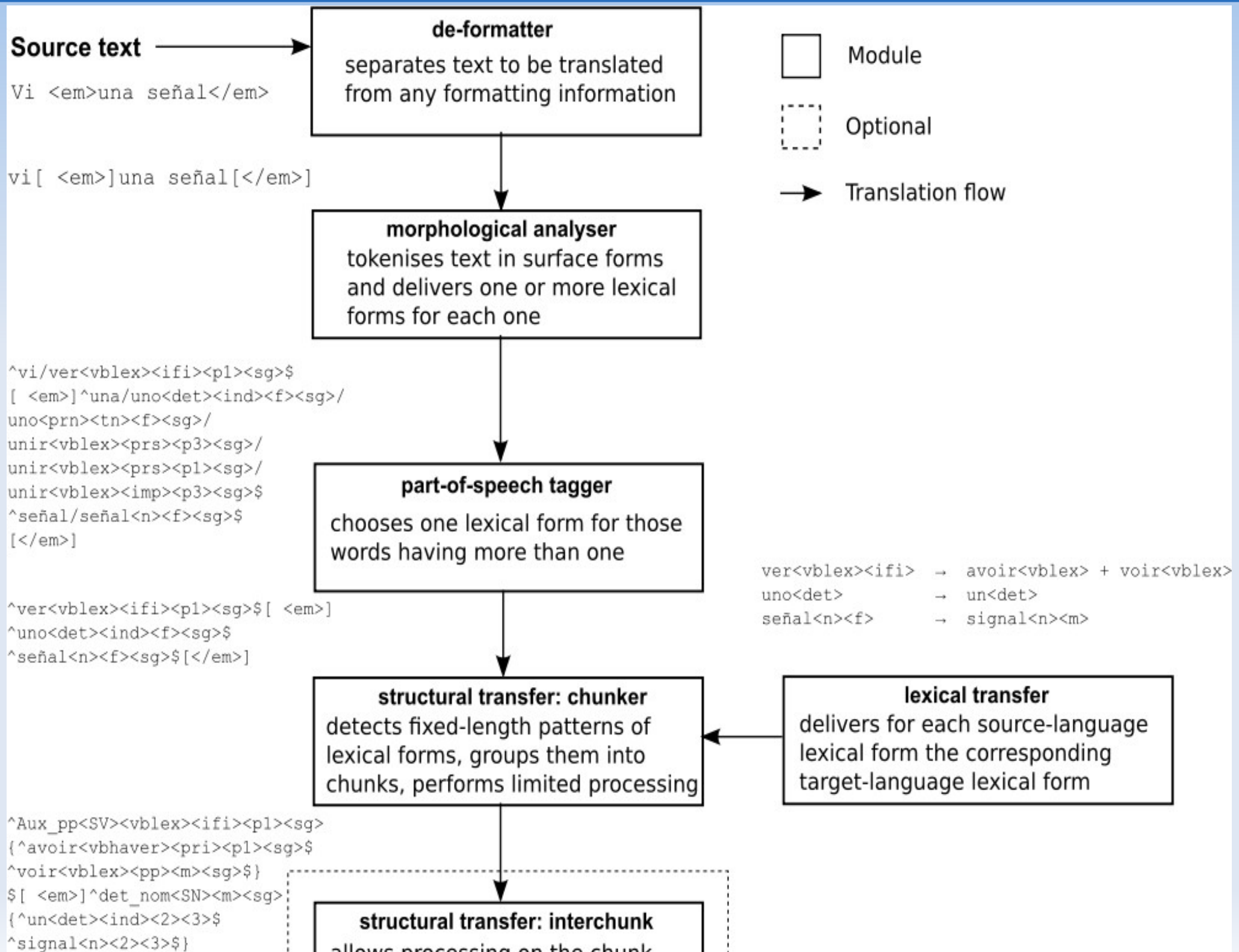
and then unchunked

```
^prpers<prn><subj><p1><mf><sg>$  
^vidi<vblex><past>$  
^signalo<n><sg><acc>$.
```

Less related languages (chunking) / 3-stage transfer

- all kinds of operations can be applied
 - word order can be changed (verb in end of sentence)
 - gender, tense and other tags can be replaced or moved
 - words and special features of source/destination language can be removed/added

Architecture of Apertium MT



Architecture of Apertium MT

```
{^un<det><ind><2><3>$
^signal<n><2><3>$}
$[</em>]
```

structural transfer: interchunk
allows processing on the chunk level of agreement, word order, and insertion or deletion of chunks

```
^Prnprerssubj<SN>{^je<prn><tn><pl><mf><sg>$}$
^aux_pp<SV><vblex><ifi><pl><sg>{
^avoir<vbhaver><pri><pl><sg>$
^voir<vblex><pp><m><sg>$}$[ <em>]
^det_nom<SN><m><sg>{
^un<det><ind><2><3>$
^signal<n><2><3>$}$[</em>]
```

structural transfer: postchunk
allows the modification of words inside the chunks, and de-chunks the output

```
^Je<prn><tn><pl><mf><sg>$
^avoir<vbhaver><pri><pl><sg>$
^voir<vblex><pp><m><sg>$[ <em>]
^un<det><ind><m><sg>$
^signal<n><m><sg>$[</em>]
```

morphological generator
delivers a target-language surface form for each target-language lexical form

```
~Je ai vu[ <em>]un signal[</em>]
```

postgenerator
performs orthographical operations such as contractions and apostrophations

```
J'ai vu[ <em>]un signal[</em>]
```

reformatter
restores the format information encapsulated by the de-formater

→ **Target text**

```
J'ai vu <em>un signal</em>
```

Exercises for interested parties

- Install the source code
 - <http://wiki.apertium.org/wiki/Installation>
 - Easiest on (Ubuntu) Linux
 - Java source is only for execution
 - C++ source is for development and execution
- Get in touch (IRC is best for getting help)
 - <http://wiki.apertium.org/wiki/Contact>
-

Swedish-Danish

Structural transfer sv-da

■ Double definiteness

Den stora utmaningen ('The big challenge')

^Den<det><def><ut><sg>\$ ^stor<adj><pst><un><pl><ind>\$ ^utmaning<n><ut><sg><def><nom>\$

^Den<det><def><ut><sg>\$ ^stor<adj><pst><un><pl><ind>\$ ^udfordring<n><ut><sg><ind><nom>\$

Den store udfordring

■ Swedish supine verb form

Han hade blivit trott ('He had been believed')

^Han<prn><subj><p3><m><sg>\$ ^ha<vbhaver><past><actv>\$ ^bli<vblex><supn><actv>\$

^tro<vblex><pp><nt><sg><ind>\$

^Han<prn><subj><p3><m><sg>\$ ^være<vbser><past><actv>\$ ^blive<vblex><pp>\$ ^tro<vblex><pp>\$

Han var blevet troet

(sometimes the auxiliary verb is omitted in Swedish - *Han blivit trott*. This is currently not supported)

■ Changes in auxiliary verbs

Två personer har börjat ('Two people has begun')

^Två<num><un><pl>\$ ^person<n><ut><pl><ind><nom>\$ ^ha<vbhaver><pres><actv>\$

^börja<vblex><supn><actv>\$

^To<num><un><pl>\$ ^person<n><ut><pl><ind><nom>\$ ^være<vbser><pres><actv>\$ ^begynde<vblex><pp>\$

To personer er begyndt ('Two people is begun')

Structural transfer sv-da

■ Changes in present passive formation

Det publiceras ('It is being published')

^Det<prn><subj><p3><nt><sg>\$ ^publicera<vblex><pres><pasv>\$

^Det<prn><subj><p3><nt><sg>\$ ^publicere<vblex><pres><pasv>\$

Det publiceres

Det upprepas ('It is being repeated')

^Det<prn><subj><p3><nt><sg>\$ ^upprepa<vblex><pres><pasv>\$

^Det<prn><subj><p3><nt><sg>\$ ^blive<vblex><pres><actv>\$ ^gentage<vblex><pp>\$

Det bliver gentaget

Changes in past passive formation

Det publicerades ('It was being published')

^Det<prn><subj><p3><nt><sg>\$ ^publicera<vblex><past><pasv>\$

^Det<prn><subj><p3><nt><sg>\$ ^blive<vblex><past><actv>\$ ^publicere<vblex><pp>\$

Det blev publiceret

Det upprepades ('It was being repeated')

^Det<prn><subj><p3><nt><sg>\$ ^upprepa<vblex><past><pasv>\$

^Det<prn><subj><p3><nt><sg>\$ ^blive<vblex><past><actv>\$ ^gentage<vblex><pp>\$

Det blev gentaget

Challenges in transfer

- Gender and number change in determiners, adjective, nouns
 - <nt> (Neuter), <ut> (Common) ⇔
<un> (Common/Neuter), <GD> (gender to be determined)
 - <sg>, <pl> ⇔
<sp>, <ND> (number to be determined)
 - Concordance: gender, number of determiner and adjectives follow must noun
 - Synthetic adjectives (better, best vs. more good, most good)

Bidix paradigms for simplicity

- <sp> words (singular and plural have same form)

^datum/datum<n><nt><sp><ind><nom>\$ →

^dato/dato<n><ut><sg><ind><nom>\$ or

^datoer/dato<n><ut><pl><ind><nom>\$

En atlas	^atlas<n><ut><sg><ind><nom>\$	^atlas<n><nt><sp><ind><nom>\$	Et atlas
Atlasen	^Atlas<n><ut><sg><def><nom>\$	^Atlas<n><nt><sg><def><nom>\$	Atlasset
Två atlaser →	^atlas<n><ut><pl><ind><nom>\$ →	^atlas<n><nt><sp><ind><nom>\$ →	To atlas
De två atlasen	^atlas<n><ut><pl><def><nom>\$	^atlas<n><nt><sp><ind><nom>\$	De to atlas

```
<pardef n="sgpl_sp__n">
```

```
<e r="RL"><p><l><s n="ND"/><s n="ind"/></l><r><s n="sp"/><s n="ind"/></r></p></e>
```

```
<e r="LR"><p><l><s n="sg"/><s n="ind"/></l><r><s n="sp"/><s n="ind"/></r></p></e>
```

```
<e r="LR"><p><l><s n="pl"/><s n="ind"/></l><r><s n="sp"/><s n="ind"/></r></p></e>
```

```
<e> <p><l><s n="sg"/><s n="def"/></l><r><s n="sg"/><s n="def"/></r></p></e>
```

```
<e> <p><l><s n="pl"/><s n="def"/></l><r><s n="pl"/><s n="def"/></r></p></e>
```

```
</pardef>
```

```
<e><p><l>atlas<s n="n"/><s n="ut"/></l><r>atlas<s n="n"/><s n="nt"/></r></p><par n="sgpl_sp__n"/></e>
```

```
<e><p><l>datum<s n="n"/><s n="nt"/></l><r>dato<s n="n"/><s n="ut"/></r></p><par n="sp_sgpl__n"/></e>
```

Dictionary entries for adjectives

- Swedish monodix

```
<pardef n="aktiv__adj">
<e><p><l></l> <r><s n="adj"/><s n="pst"/><s n="ut"/><s n="sg"/><s n="ind"/></r></p></e>
<e><p><l>t</l> <r><s n="adj"/><s n="pst"/><s n="nt"/><s n="sg"/><s n="ind"/></r></p></e>
<e><p><l>e</l> <r><s n="adj"/><s n="pst"/><s n="m"/><s n="sg"/><s n="def"/></r></p></e>
<e><p><l>a</l> <r><s n="adj"/><s n="pst"/><s n="un"/><s n="pl"/><s n="ind"/></r></p></e>
<e><p><l>a</l> <r><s n="adj"/><s n="pst"/><s n="un"/><s n="sp"/><s n="def"/></r></p></e>

<e><p><l>are</l> <r><s n="adj"/><s n="comp"/><s n="un"/><s n="sp"/></r></p></e>
<e><p><l>ast</l> <r><s n="adj"/><s n="sup"/><s n="un"/><s n="sp"/><s n="ind"/></r></p></e>
<e><p><l>aste</l><r><s n="adj"/><s n="sup"/><s n="un"/><s n="sp"/><s n="def"/></r></p></e>
</pardef>

<e lm="vit"> <i>vit</i><par n="aktiv__adj"/></e>
```

- Swedish-Danish bidix

```
<e><p><l>vit<s n="adj"/></l><r>hvid<s n="adj"/></r></p><par n="aktiv_aktiv__adj"/></e>
```

- Danish monodix

```
<pardef n="aktiv__adj">
<e><p><l></l> <r><s n="adj"/><s n="pst"/><s n="ut"/><s n="sg"/><s n="ind"/></r></p></e>
<e><p><l>t</l> <r><s n="adj"/><s n="pst"/><s n="nt"/><s n="sg"/><s n="ind"/></r></p></e>
<e><p><l>e</l> <r><s n="adj"/><s n="pst"/><s n="un"/><s n="pl"/><s n="ind"/></r></p></e>
<e><p><l>e</l> <r><s n="adj"/><s n="pst"/><s n="un"/><s n="sp"/><s n="def"/></r></p></e>
</pardef>

<e lm="hvid"> <i>hvid</i><par n="aktiv__adj"/></e>
```

Bidix paradigms... for simplicity (?)

- Adjective follows gender, number and can be synthetic

En vit atlas.	^vit<adj><pst><ut><sg><ind>\$	^hvid<adj><pst><nt><sg><ind>\$	Et hvidt atlas.
Atlasen		^mere<preadv>\$	Atlasset
Två vitare atlaser	→ ^vit<adj><comp><un><sp>\$	→ ^hvid<adj><pst><un><pl><ind>\$	→ To mere hvide atlas
De två vitaste atlaserna	^vit<adj><sup><un><sp><def>\$	^mest<preadv>\$	De to mest #hvid
		^hvid<adj><pst><sup><nt><pl><ind><def>\$	atlassene

```

<pardef n="aktiv_aktiv__adj">
<e> <p><l><s n="pst"/><s n="un"/><s n="sp"/><s n="def"/></l><r><s n="pst"/><s n="un"/><s n="sp"/><s n="def"/></r></p></e>
<e> <p><l><s n="pst"/><s n="un"/><s n="pl"/><s n="ind"/></l><r><s n="pst"/><s n="un"/><s n="pl"/><s n="ind"/></r></p></e>

<e r="LR"><p><l><s n="pst"/><s n="m"/><s n="sg"/><s n="def"/></l><r><s n="pst"/><s n="un"/><s n="sp"/><s n="def"/></r></p></e>
<e r="LR"><p><l><s n="pst"/><s n="ut"/></l><r><s n="pst"/><s n="ut"/></r></p></e>
<e r="LR"><p><l><s n="pst"/><s n="nt"/></l><r><s n="pst"/><s n="nt"/></r></p></e>

<e r="RL"><p><l><s n="pst"/><s n="GD"/></l><r><s n="pst"/><s n="ut"/></r></p></e>
<e r="RL"><p><l><s n="pst"/><s n="GD"/></l><r><s n="pst"/><s n="nt"/></r></p></e>

<e r="LR"><p><l><s n="comp"/><s n="un"/><s n="sp"/></l><r><s n="unsint"/><s n="comp"/><s n="GD"/><s n="ND"/></r></p></e>
<e r="RL"><p><l><s n="sint"/><s n="comp"/><s n="un"/></l><r><s n="comp"/><s n="ut"/></r></p></e>
<e r="RL"><p><l><s n="sint"/><s n="comp"/><s n="un"/></l><r><s n="comp"/><s n="nt"/></r></p></e>

<e r="LR"><p><l><s n="sup"/><s n="un"/><s n="sp"/></l><r><s n="unsint"/><s n="sup"/><s n="GD"/><s n="ND"/></r></p></e>
<e r="RL"><p><l><s n="sint"/><s n="sup"/><s n="un"/></l><r><s n="sup"/><s n="ut"/></r></p></e>
<e r="RL"><p><l><s n="sint"/><s n="sup"/><s n="un"/></l><r><s n="sup"/><s n="nt"/></r></p></e>
</pardef>

<e> <p><l>vit<s n="adj"/></l> <r>hvid<s n="adj"/></r></p><par n="aktiv_aktiv__adj"/></e>

```

Evaluation

Sv original: Historik.
Da postedit: Historik.
Apertium : Historik.
Gramtrans : Historik.
Googole SMT: Historie.

	Number entries
Monolingual dict. (sv)	5,230 lemmas
Bilingual dict.	6,854 lemmas
Monolingual dict. (da)	10,694 lemmas
Transfer rules (sv → da)	17 rules

System	Edit distance	WER	PWER
Apertium	350	30	28
Gramtrans	304	26	20
Google	415	35	22

Sv: Trakterna kring Fredriksberg räknas som bebodda sedan 1600-talet.
Da: Områderne omkring Fredriksberg regnes som beboede siden 1600-tallet.
Ap: *Trakterna omkring *Fredriksberg regnes som *bebodda siden 1600-talen.
Gr: Områderne omkring Fredriksberg regnes som beboede siden 1600-talet.
Go: Områderne omkring Fredriksberg tælles som har været besat siden 1600-tallet.

Sv: Området kring Fredriksberg utgjorde ursprungligen den södra delen av Nås finnmark,
Da: Området omkring Fredriksberg udgjorde oprindeligt den sydlige del af Nås finnmark,
Ap: Området omkring *Fredriksberg *utgjorde oprindeligt den *södra delen af Nås *finnmark,
Gr: Området omkring Fredriksberg udgjorde oprindeligt den sydlige del af Nås finnmark,
Go: Området omkring Frederiksberg var oprindeligt den sydlige del af Reachable Sverige,

Sv: och området räknas som en del av Västerdalarna
Da: og området regnes som en del af Västerdalarna
Ap: og området regnes som en del af *Västerdalarna
Gr: og området regnes som en del af Västerdalarna
Go: og området regnes som en del af den vestlige del af Dalarna

Corpus	Running tokens	Known tokens	Coverage
Wikipedia	30,662,861	22,030,690	71.84%
EuroParl	15,531,107	12,499,971	80.48%

Sv: (till skillnad från övriga Ludvika kommun, som räknas till Bergslagen).
Da: (til forskel fra øvrige Ludvika kommune, som regnes til Bergslagen).
Ap: (til forskel fra øvrige *Ludvika kommune, som regnes til *Bergslagen).
Gr: (til forskel fra den øvrige Ludvika kommune, som regnes til Bergslagen).
Go: (i modsætning til andre Ludvika Kommune, som rækker_Bergslagen).

Evaluation

System	Edit distance	WER	PWER
Apertium	350	30	28
Gramtrans	304	26	20
Google	415	35	22

	Translation	Gloss
Original	<i>Det finns en kort överfart vid det baltiska havet vid Helsingborg, på vilket ställe Själland kan ses från Skåne, ett vanligt tillhåll för vikingar.</i>	There exists a short passage by the Baltic Sea by Helsingborg, on which place Sjælland can be seen from Skåne, a common hangout for Vikings.
Apertium	<u>Det</u> findes en kort <u>överfart</u> ved det <i>baltiska</i> havet ved Helsingborg, på hvilket <u>ställe</u> <i>Sjælland</i> kan ses fra Skåne, et <u>vanligt</u> <u>tilhold</u> før vikinger.	<u>It</u> exists a short <u>överfart</u> by the <i>baltiska</i> Sea by Helsinborg, on which <u>ställe</u> <i>Sjælland</i> can be seen from Skåne, a <u>vanligt</u> <u>order</u> before Vikings.
Gramtrans	Der findes en kort overfart ved det baltiske hav ved Helsingborg, på hvilket sted <i>Sjælland</i> kan ses fra Skåne, et sædvanligt <u>tilhold</u> for vikinger.	There exists a short passage by the Baltic Sea by Helsingborg, on which place <i>Sjælland</i> can be seen from Skåne, a common <u>order</u> for Vikings.
Google	Der <u>er</u> en kort <u>passage</u> i Østersøen i Helsingborg, i hvilken <u>plads</u> <i>Zealand</i> kan ses fra <i>Scania</i> , en regelmæssig tilholdssted for vikingerne.	There <u>is</u> a short <u>passage</u> in the Baltic Sea <u>in</u> Helsingborg, <u>in</u> which <u>space/place/seat</u> <i>Zealand</i> can be seen from <i>Scania</i> , a <u>regular</u> hangout for <u>the</u> Vikings.

Table 4: Comparison of the three systems for a single sentence. Unknown words are marked with *emphasis* and incorrect translations are underlined.

Licence soup

- This presentation may be distributed under the terms of the GNU GPL, GNU FDL and CC-BY-SA licences.
 - GNU GPL v. 3.0
 - <http://www.gnu.org/licenses/gpl.html>
 - GNU FDL v. 1.2
 - <http://www.gnu.org/licenses/gfdl.html>
 - CC-BY-SA v. 3.0
 - <http://creativecommons.org/licenses/by-sa/3.0/>