

# What is machine translation?

7-10. desember 2015

## Before we start, we let the machine build our programs

- ▶ In the terminal, *compile* the source language
  - ▶ Open a new terminal, and write  
cd **main**/langs/sme  
./configure --with-hfst --enable-apertium  
time make
- ▶ Then compile the target language
  - ▶ Open a new terminal, and write  
cd main/langs/**sma**  
./configure --with-hfst --enable-apertium  
time make
- ▶ Then wait 20 minutes

# What is machine translation?

- ▶ Machine translation is translating texts from one language to another with the help of computer programs.

## The original idea: Background for the sme2smX project

- ▶ Language societies in the process of revitalisation have some basic needs
  - ▶ Small children need children's books and children's stories
  - ▶ School children need school books  
(many books, and then new editions)
  - ▶ Youngsters need books to widen their vocabulary
  - ▶ Official documents is often expected to be available in several Saami languages
- ▶ To move a language from the family sphere out in the public sphere, it must be present on public arenas
- ▶ **There are not enough writers and translators for all this**

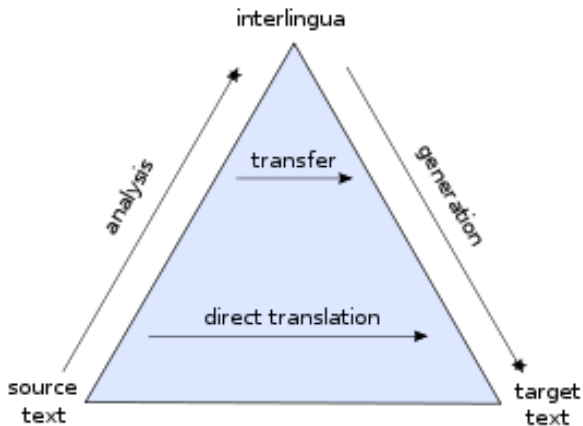
## The answer: MT via a relé language, and postediting

- ▶ Minority languages often come in groups
- ▶ Take North Saami as the starting point
  - ▶ Either a North Saami original
  - ▶ ... or a text already translated into North Saami
- ▶ **Do a machine translation from North Saami to the other Saami languages,  
and edit the result**

## Some things to keep in mind

- ▶ Machine translated text must *always* be corrected and postedited
- ▶ This does not prevent writers from making original texts in the other Saami languages
  - ▶ On the contrary, it will give writers more time to write
- ▶ There are some dangers, though (more on that later)

## Philosophy: Keep it close



## The two ways of using a MT program

- ▶ Understand or create?
  - ▶ You want to understand a text in an unknown language
  - ▶ You want to save time when producing text
- ▶ For a reader machine translated text is always problematic
  - ▶ The language may be of bad quality
  - ▶ There may be translation errors giving rise to misunderstanding
- ▶ We thus avoid MT for understanding, and prefer the original, whenever we can



## Is there a difference?

Text ...	Important	Not important
<b>understanding</b>	Understandability Fast translation	Syntactic <i>correctness</i> Lexical <i>correctness</i> Predictable errors Happy translators
<b>production</b>	Adequate syntactic transfer Predictable errors High accuracy (WER $\leq$ 15%) Happy translators	Understandability Fast translation

*With the binoculars the hat-having man sees the squirrel.*

## Is there a difference?

Text ...	Important	Not important
<b>understanding</b>	Understandability Fast translation	Syntactic <i>correctness</i> Lexical <i>correctness</i> Predictable errors Happy translators
<b>production</b>	Adequate syntactic transfer Predictable errors High accuracy (WER $\leq$ 15%) Happy translators	Understandability Fast translation

*With the binoculars the hat-having man sees the squirrel.*  
The man wearing a hat sees the squirrel with the binoculars.

## Is there a difference?

Text ...	Important	Not important
<b>understanding</b>	Understandability Fast translation	Syntactic <i>correctness</i> Lexical <i>correctness</i> Predictable errors Happy translators
<b>production</b>	Adequate syntactic transfer Predictable errors High accuracy (WER $\leq$ 15%) Happy translators	Understandability Fast translation

*The migration gave a great deal of criticism when it spoke out.*

## Is there a difference?

Text ...	Important	Not important
<b>understanding</b>	Understandability Fast translation	Syntactic <i>correctness</i> Lexical <i>correctness</i> Predictable errors Happy translators
<b>production</b>	Adequate syntactic transfer Predictable errors High accuracy (WER $\leq$ 15%) Happy translators	Understandability Fast translation

*The migration gave a great deal of criticism when it spoke out.*  
 The **organisation received** a great deal of criticism when it spoke  
 out.

## How is it used ?

```
De:      Trond Trosterud <...@uit.no>
Per a:   Hanna Martikainen <...@noos.fr>, Francis Tyers <...@ivixor.net>
Assumppte: Re: Vis1CG - POS Tagger for Finnish
Data:    Wed, 19 Jan 2011 12:46:05 +0100 (19/01/11 11:46:05)
```

Hanna Martikainen <...@noos.fr> kirjoitti 18.1.2011 kello 23.32:

```
> Dear Sirs,
> I got your contact information from Mr Tino Didriksen.
```

Kyllä, se onnistuu. (lopun englanniksi Franciksen takia)

## How is it used ?

```
De:      Trond Trosterud <...@uit.no>
Per a:   Hanna Martikainen <...@noos.fr>, Francis Tyers <...@ivixor.net>
Assumpte: Re: Vis1CG - POS Tagger for Finnish
Data:   Wed, 19 Jan 2011 12:46:05 +0100 (19/01/11 11:46:05)
```

Hanna Martikainen <...@noos.fr> kirjoitti 18.1.2011 kello 23.32:

```
> Dear Sirs,
> I got your contact information from Mr Tino Didriksen.
```

Kyllä, se onnistuu. (lopun englanniksi Franciksen takia)

Sí, lo consigo. (El resto está en Inglés porque Franciksen)

Máilmmi vuosttas eamiálbmogiid Pride-festivála ordnejuvvo Gironis | Yle Uutiset | yle.fi - Iisveassel

Eibeer Edita Yjsualitza Historial Adreces d'interés Eiges Ajuda

Máilmmi vuostt... Verdenens de fa... +

yle.fi/uutiset/mailmmi\_vuosttas\_eamialbmogiid\_pride-festivala\_ordnejuvvo\_gironis/7450625

Google

---

**Sápmi** 8.9.2014 klo 14:29 | beaivduvvon 8.9.2014 klo 14:45

## Máilmmi vuosttas eamiálbmogiid Pride-festivála ordnejuvvo Gironis

Sámi Nuorra ja Queering Sápmi -prošeakta ordnejit máilmmi vuosttas eamiálbmogiid Pride-festivála golggotmánus Gironis. Festivála sisttisdoallá earret eará ovtastallama, bargobájiid, logaldallamiid ja kulturdoaluid.



Sápmi Pride -leavga lea sáme- ja arvedáivgeleavvga seaguhus. Govva: Sápmi Pride

Máilmmi vuosttas eamiálbmogiid Pride-festivála ordnejuvvo Gironis golggotmánu 16.-19. beivviid. Festivála ordne Ruota sámenuoraid organisašuvdna Sámi Nuorra ja ruhtada Queering Sápmi -prošeakta.

Seksuála- ja sohkaabealvehádagaid Pride-festivála sisttisdoallá Gironis earret eará

---

### Váldoodđasat

**Sápmi** klo 8:50 0

#### Sovkina ii bala – giitá sámiid veahkis



Sii geat ná láhttejit min vuostá ja hehtejit min searvamis dákkár Čeahkkáni, sii dat ballet, Guotádagá sámi parlameantta ságadoalli Valentina Sovkina dajai marjebárga New Yorkkas.

---

**Sápmi** klo 14:56 0

#### Áile Javo ON:s: Árktalaš guovlu vuordá, ahte loahppadokumeantta boadusin ruvkeláгат rievdaduvvojit

Sámiráđi ságadoalli Áile Javo doalli árktalaš guovlu virggálaš sáhkavuoro máilmmikonferanssa eanarvuogitvuochtaságastallamis New Yorkkas marjebárgga.

---

**Sápmi** klo 12:04 0

#### Dutkamuš: Eanandoallit nagodedje ealihit eambo mánáid go sápmelaččat

Šibit- ja eanandoallu ealihedje 1700-1900-águin buorebut go árbevirolaš bivdu ja boazodoallu. Dan dihte suopmetaččat lassáneajo jodáneabbot go sápmelaččat, čájeha sámejuovlu girkojirjiid vuodii

Verdenens de første urfolkens Pride-festivalen ordnes i Kiruna | Yle Uutiset | yle.fi - Iceweasel

Eiäber Edita Visualizta Historial Adjecces d'interiès Eiges Ajuda

Mällimni vuostt... x Verdenens de fa... x

file:///tmp/test\_nob.html


Google

---

**Samen** 6.9.2014 klo 14:29 | beaivdávven 8.9.2014 klo de 14:45

## Verdenens første urfolks Pride-festivalen ordnes i Kiruna

Samens Ung og Queering Sameland -prosjektet ordner verdenens de første urfolkens Pride-festivalen i oktober i Kiruna. Festivalen inneholder blant annet samværet, bargebåtid, forelesningene og kulturarrangementet.




Sameland Pride -flagget er samens- og arvedávgeavvga en blanding. Et bilde: En same Pride


Verdenens de første urfolkens Pride-festivalen ordnes i Kiruna oktobers de 16.-19. dagene. Festivalen ordner Sveriges sameungdommer organisasjonen Samelands Ungdom og finansierer Queering Sameland -prosjektet.

---

### Váldoodđasat


**Samen** klo den 8:50 0. 

#### Sovkina frykter ikke – han takker på samenes hjelp



De som så oppfører de seg mot oss og de hindrer oss å slutte seg sammen til slikt møte, de frykter det, Kollas sames parlamentets talsmann Valentina Sovkina sa tirsdagen New Yorkkas.


---

**Sameland** klo 14:56 0. 

#### Áile Javo i ON: Den arktiske retningen venter, at sluttokumentet som resultat forandres ruvkelágat

Samerådets talsmann Áile Javo holdt det arktiske retningsens offisielle innleggs verdenskonferansen eananvuoiǵatvuhtasáǵastallamis New Yorkkas tirsdagens.

---

**Sameland** klo 12:04 0. 

#### Dutkamuš: Gårdbrukerne klarte livberge mye barn enn samer

Husdyr- og jordbruket livberget 1700-1900- med et tall bedre enn den tradisjonelle fisket og reindriften. for Det ákte finlenderne jodáneabbot nár samene, hun viser sameområdet på grunn av kirkebøkene gjort dutkamuš.



## MT for text production

- ▶ `http://no.wikipedia.org`

## The sme2smX project (Norway's Research Council)

- ▶ Make programs for letting all Saamis benefit from North Saami translations
- ▶ Gain insight in the differences between the Saami languages
- ▶ Make a foundation for intra-Saami lexicography
- ▶ ... and write scientific papers on this
- ▶ A side effect: The last year we have laid the foundation for all future work on Inari Saami language technology

## Goals for the week

- ▶ Learn to edit the bilingual dictionary *bidix*
- ▶ Learn to choose the correct word (*lexical selection*)
- ▶ Learn how the FST and MT components work together, and learn how to adjust them to each other
- ▶ Learn to write quasicode for the transfer component
- ▶ Learn some work principles

## Now, let us build the translation program

- ▶ Open a new terminal, and write  
cd **apertium**/nursery/apertium-sme-**sma**  
time make