

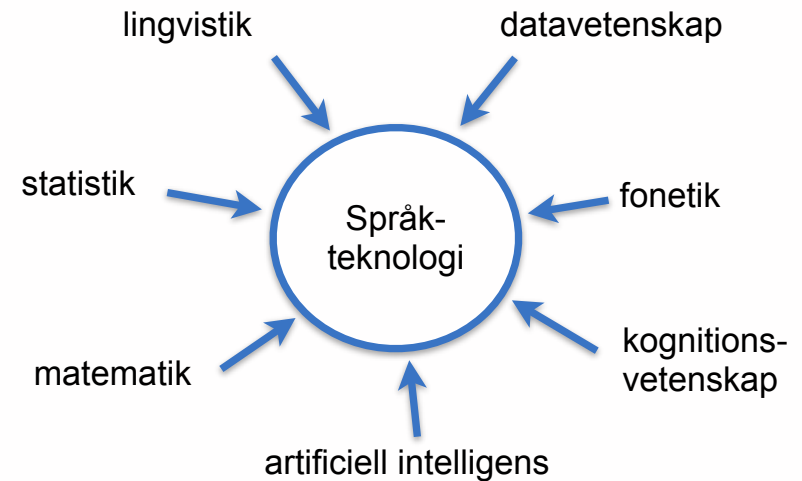
# Från frustration till dagens hype på väg till den lysande framtiden inom Maskinöversättning

Jörg Tiedemann

Språkteknologi  
Helsingfors Universitet  
jorg.tiedemann@helsinki.fi



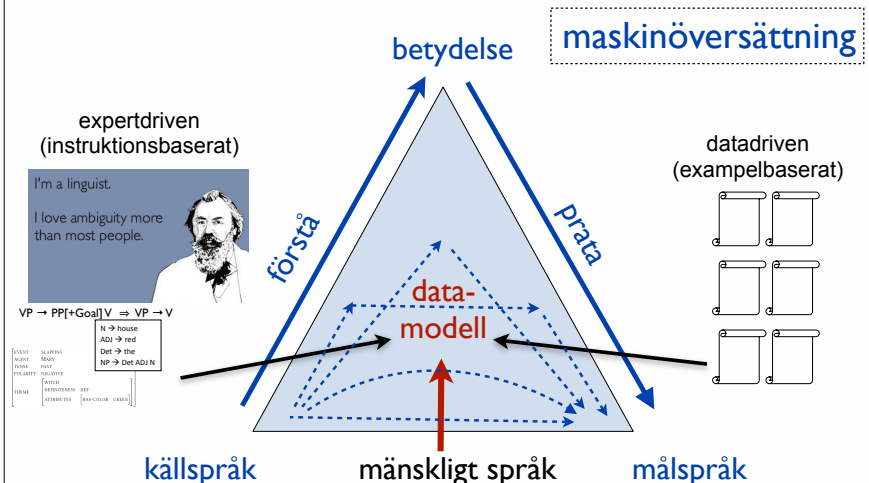
## Vad är språkteknologi?



## Språkteknologi i vardagslivet

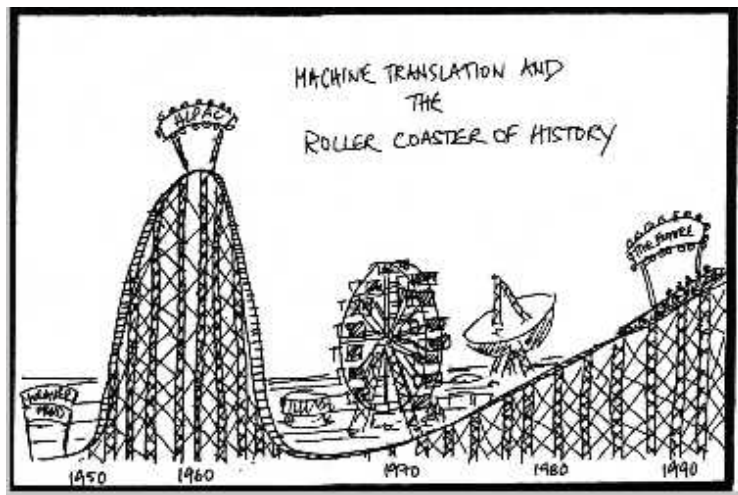


## Språkteknologi och översättning





## Maskinöversättning genom tiderna



(image from Arnold, 2008: "Machine Translation: An Introductory Guide")

HELSINGIN YLIOPISTO  
HELSINGFORS UNIVERSITET  
UNIVERSITY OF HELSINKI

5



## Automatisk översättning är inte lätt ...

- Hos en läkare: *Specialist in women and other diseases*
- På pubben: *Ladies are requested not to have children in the bar*
- På hotellet: *Please leave your values at the front desk*
- På en kinesisk restaurang: *Translation server error*



HELSINGIN YLIOPISTO  
HELSINGFORS UNIVERSITET  
UNIVERSITY OF HELSINKI

6



## Maskinöversättning som avkodning



inlärning av den  
okända koden



When I look at an article in Russian, I say:  
*This is really written in English,  
but it has been coded in some strange symbols.  
I will now proceed to decode.*

[Weaver, 1947, 1949]

HELSINGIN YLIOPISTO  
HELSINGFORS UNIVERSITET  
UNIVERSITY OF HELSINKI

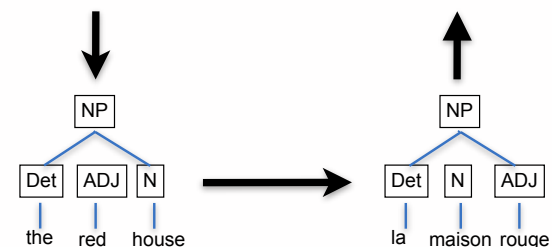
7



## Expertdrivna (regelbaserade) modeller

The red house .

La maison rouge.



N → house  
ADJ → red  
Det → the  
NP → Det ADJ N

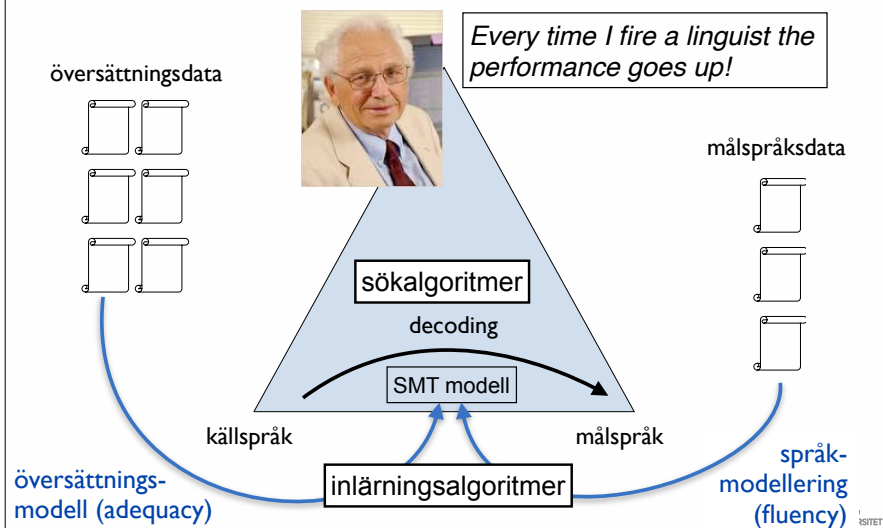
N → maison  
ADJ → rouge  
Det → la  
NP → Det N ADJ

HELSINGIN YLIOPISTO  
HELSINGFORS UNIVERSITET  
UNIVERSITY OF HELSINKI

8



## Datadriven maskinöversättning



9



## Statistisk maskinöversättning

- 1947: MT som avkodning (Warren Weaver)
- 1988: Ordbaserade modeller
- 1999: Öppna ordlänkingsverktyg (GIZA)
- 2003: **Frasbaserade SMT-modeller**
- 2004: Öppna SMT avkodare (Pharaoh)
- 2005: Hierarkiska SMT modeller
- 2006: **Google Translate**
- 2007: **Moses** (komplett och öppen SMT verktygslåda)

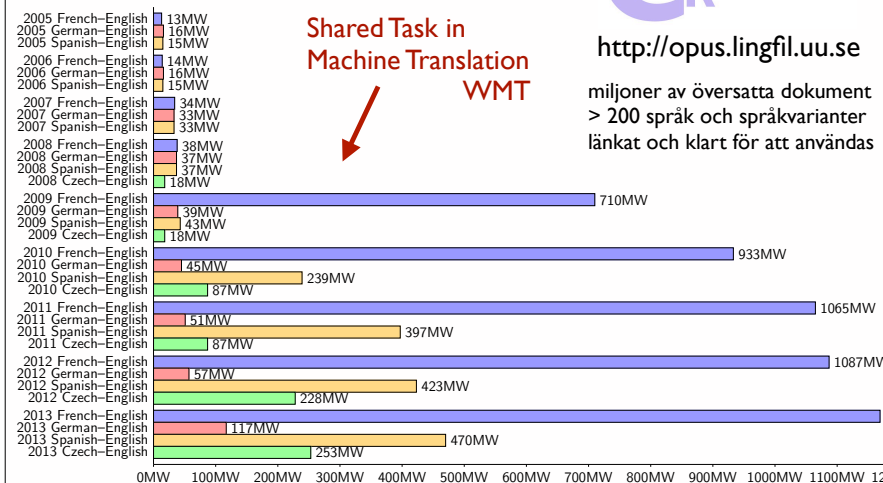
... och det var bara början ...

HELSINGIN YLIOPISTO  
HELSINGFORS UNIVERSITY  
UNIVERSITY OF HELSINKI

10



## Växande datamängder



Shared Task in  
Machine Translation  
WMT

**ORPUS**

<http://opus.lingfil.uu.se>

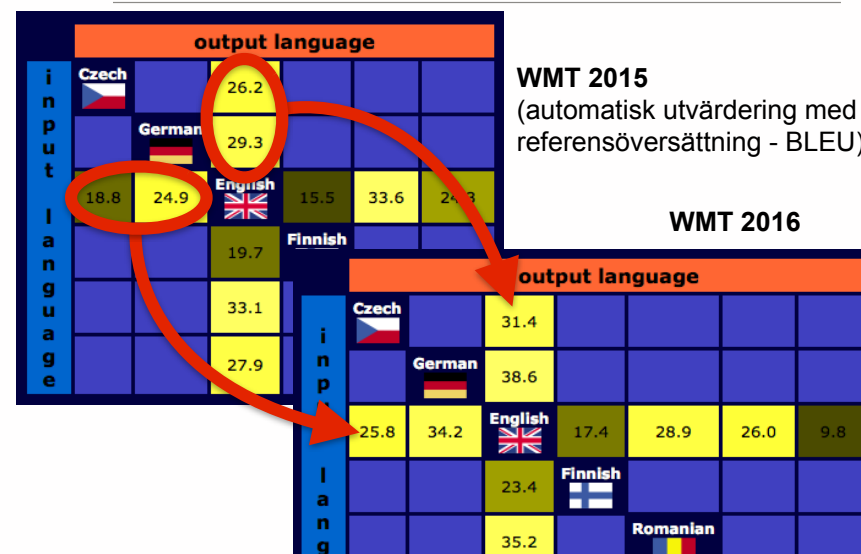
miljoner av översatta dokument  
> 200 språk och språkvarianter  
länkat och klart för att användas

HELSINGIN YLIOPISTO  
HELSINGFORS UNIVERSITY  
UNIVERSITY OF HELSINKI

11



## Var står vi nu?



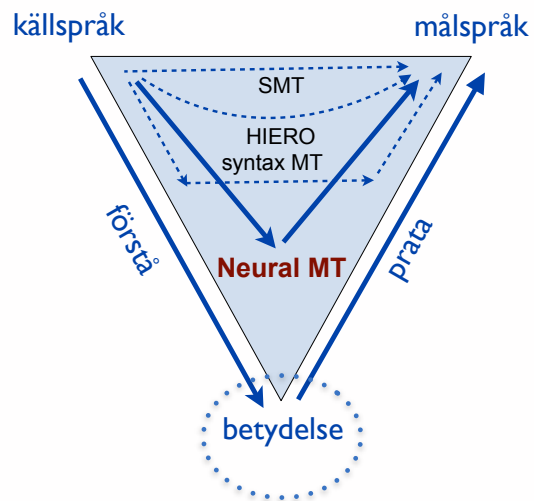
WMT 2015  
(automatisk utvärdering med  
referensöversättning - BLEU)

WMT 2016

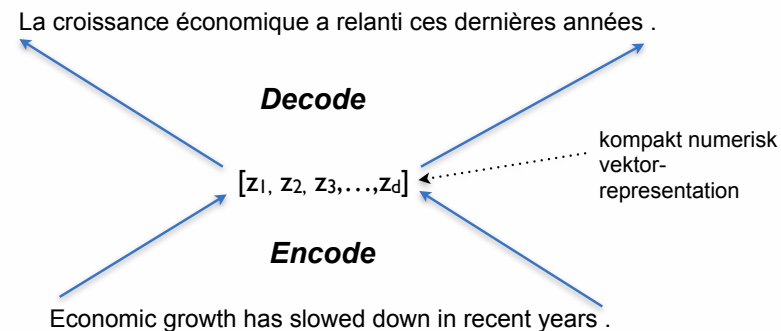
12



## Vad hände 2016?



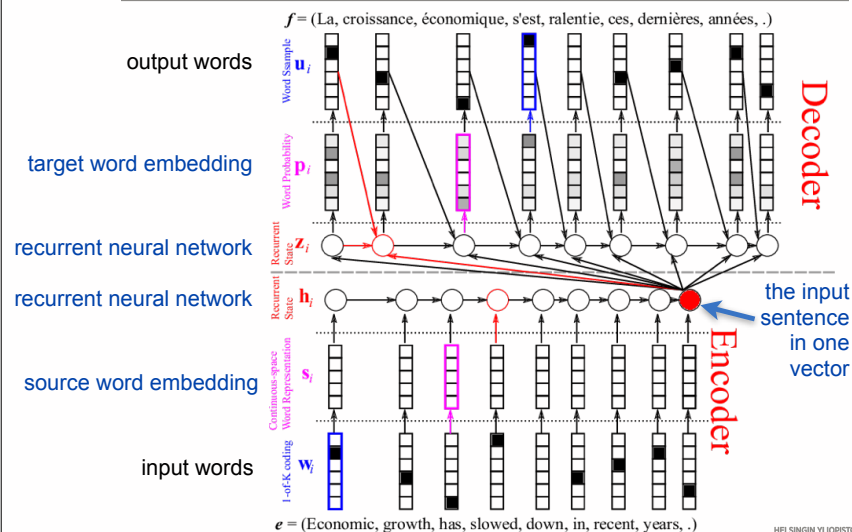
## Neurala “Encoder-Decoder” Modeller



(Example from Kyunghyun Cho et. al: “On the Properties of Neural Machine Translation: Encoder–Decoder Approaches”)



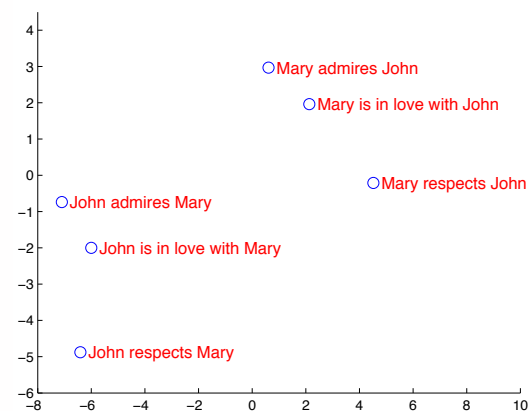
## Några detaljer av “Neural MT”



<https://devblogs.nvidia.com/parallelforall/introduction-neural-machine-translation-gpus-part-2/>



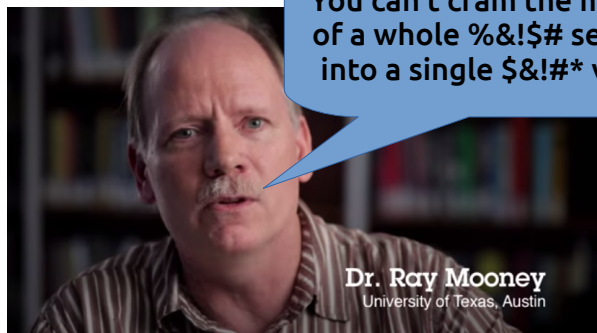
## Semantiska vektorrymdsmodeller



(From Sutskever et. al: “Sequence to Sequence Learning with Neural Networks”)



## Wait A Minute ...



You can't cram the meaning of a whole %&!\$# sentence into a single \$&!#\* vector!

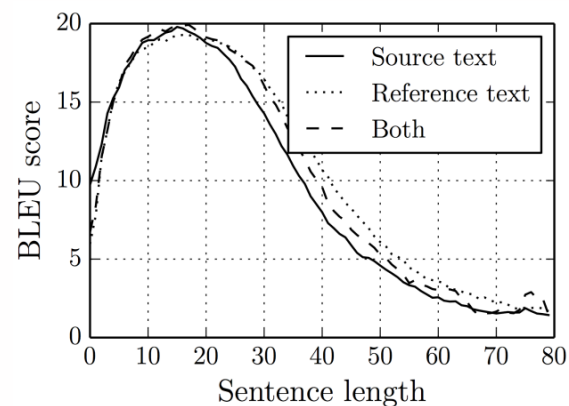
(From Ray Mooney: "Semantic Parsing: Past, Present, and Future", slide from Jacob Eisenstein)

HELSINGIN YLIOPISTO  
HELSINGFORS UNIVERSITET  
UNIVERSITY OF HELSINKI

17



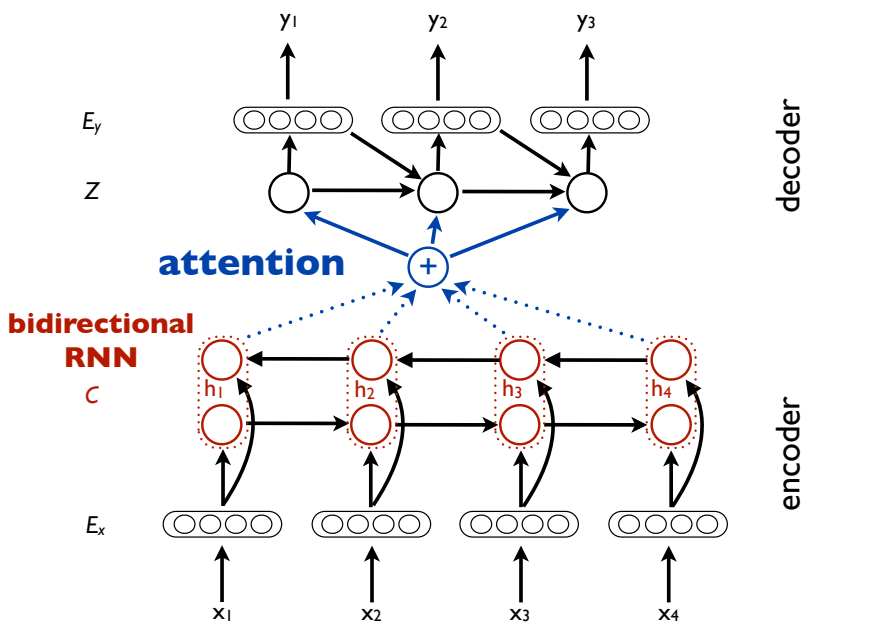
## Meningar med olika längder ...



<https://devblogs.nvidia.com/parallelfornall/introduction-neural-machine-translation-gpus-part-3/>

HELSINGIN YLIOPISTO  
HELSINGFORS UNIVERSITET  
UNIVERSITY OF HELSINKI

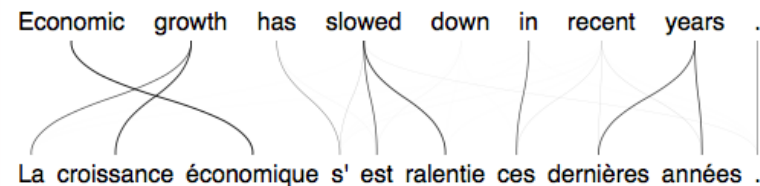
18



19



## Neural MT med "attention"



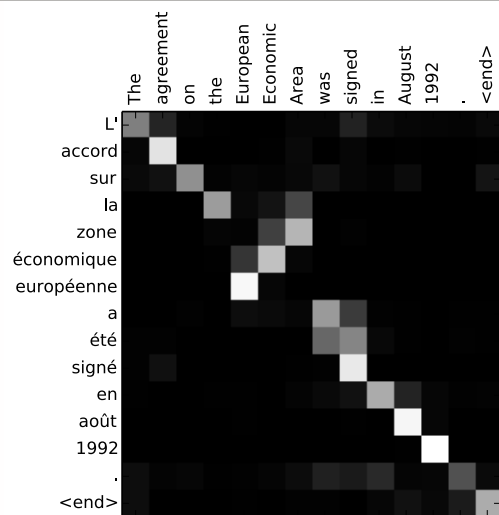
(From on-line system at <http://104.131.78.120/>)

HELSINGIN YLIOPISTO  
HELSINGFORS UNIVERSITET  
UNIVERSITY OF HELSINKI

20



## Neural MT med “attention”



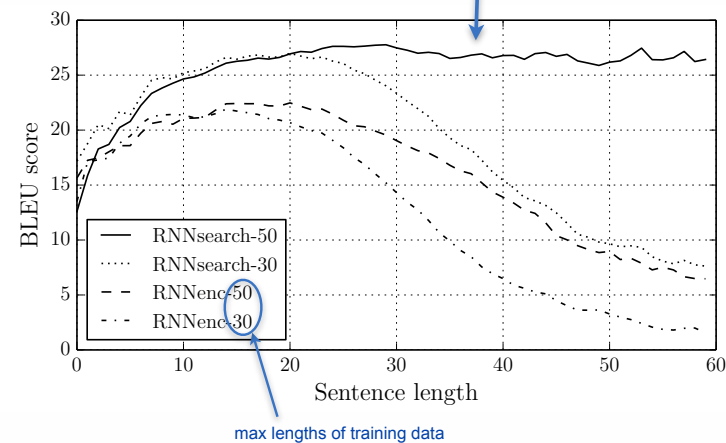
(Example from Bahdanau, Cho & Bengio: “Neural Machine Translation by Jointly Learning to Align and Translate”)

HELSINGIN YLIOPISTO  
HELSINGFORS UNIVERSITET  
UNIVERSITY OF HELSINKI

21



## Neural MT med “attention”



(From Bahdanau, Cho & Bengio: “Neural Machine Translation by Jointly Learning to Align and Translate”)

HELSINGIN YLIOPISTO  
HELSINGFORS UNIVERSITET  
UNIVERSITY OF HELSINKI

22



## Vad gör vi i Helsingfors?

NMT för  
svenska - finska

### HNMT: the Helsinki Neural Machine Translation system

This is a neural network-based machine translation system developed at the University of Helsinki.

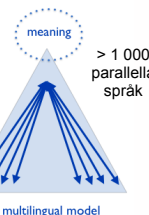
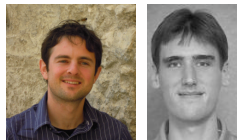
It is currently rather experimental, but the user interface and setup procedure should be simple enough for people to try out.

#### Features



WMT: engelska - finska

- biLSTM encoder which can be either character-based or hybrid word/character (Luong & Manning 2016)
- LSTM decoder which can be either character-based or word-based
- Variational dropout (Gal 2015) and Layer Normalization (Ba et al. 2016)
- Bayesian word alignment model for guiding attention mechanism (experimental)



NeuralMT / Home

### Neural Machine Translation Wiki

This is a Wiki to keep information and links about neural MT in 2015 and here is an on-line demo from the Montreal team short summary!

- News
- Background
- Tools
- Practical tutorials and code
- People and Labs
- Interesting translations

FINMT 2016 / 2017

Workshop on Machine Translation for Morphologically-Rich Languages

... and Finnish in particular

<http://opus.lingfil.uu.se>

23



## LT in MA *Linguistics in the Digital Age*

collaborations with linguistics, phonetics, digital humanities, data science, ...

Period	I	II	III	IV
1 (MA)	Theory and Method in Language Sciences Diversity and Variation Linguistics in the Digital Age		Computational Morphology Programming for Linguists	Computational Syntax Machine Learning for Linguists Mathematics for Linguists Speech synthesis and rec.
2 (MA)	Models and Algorithms in NLP Applications NLP for Low-Resource Lang. Intro to AI Intro to data science Teaching assistance	Research project Computational Semantics Computational creativity Intro to machine learning Intro big data management	Corpus clinic (5 cr) String processing alg. Current topics in LT Probabilistic models	Advanced course in ML Speech synthesis and rec.
MA Thesis work (including Seminar in LT) 30+5 cr				

if you missed the course earlier

HELSINGIN YLIOPISTO  
HELSINGFORS UNIVERSITET  
UNIVERSITY OF HELSINKI

24



**LANGUAGE TECHNOLOGY**  
 NLP Research at the University of Helsinki and Information about the Language Technology Programme

[http://blogs.helsinki.fi/language-technology/](#)

[NYHETER](#)
[FORSKNING](#)
[EVENEMANG](#)
[PERSONER](#)
[STUDIER](#)

## Språkteknologi i Helsingfors

 Svenska

Språkteknologi är ett tvärvetenskapligt forskningsområde. Ofta talar man också om datorlingvistik eller datalingvistik och NLP, som står för engelskans Natural Language Processing. Inom språkteknologin studerar vi metoder och utvecklar verktyg som gör det möjligt för datorer att hantera mänskligt språk (tal och skrift). Några kända tillämpningar är automatisk stav- och grammatikkontroll, maskinöversättning och automatisk taligenkänning.

Vid Helsingfors universitet har vi följande tyngdpunktsområden inom forskningen:

- [Språkteknologi för morfologiskt rika språk](#)
- [Tvåspråklig språkteknologi](#)
- [Språkteknologi inom humaniora](#)

Läroämnet datalingvistik grundades inom Institutionen för allmän språkvetenskap år 1987, och det nuvarande namnet språkteknologi togs i bruk 1999. Vid omorganiseringen 2010, blev språkteknologin en del av [Institutionen för moderna språk](#).

**CONTACT**  
 P.O. Box 24 (Unioninkatu 40 B)  
 FI-00014 University of Helsinki  
 Tel. +358 2 941 911  
 (switchboard/vaihde/växel)  
 Fax +358 2 941 28313  
[Telephone Directory](#)  
[E-mail Directory](#)

**KATEGORIER**  
[Okategoriserad](#)

25

# Tack! Frågor?



***Språkteknologi i Helsingfors***  
<http://blogs.helsinki.fi/language-technology/>  
[jorg.tiedemann@helsinki.fi](mailto:jorg.tiedemann@helsinki.fi)

HELSINGIN YLIOPISTO  
 HELSINGFORS UNIVERSITET  
 UNIVERSITY OF HELSINKI

26