



GÖTEBORGS  
UNIVERSITET

Språk  
BANKEN

# Språk och språkteknologi i Sverige

Lars Borin • Språkbanken/Swe-Clarin, Göteborgs universitet



European Language  
Resource Coordination  
Connecting Europe Faculty



# vad är språkteknologi?

En samling informationsteknologier som kommer ur forskning som syftar till att få datorer och datasystem att förstå och använda språk på ett sätt som vi föreställer oss att vi människor gör det.

Exempel:

- ▶ maskinöversättning
- ▶ informationsextraktion/mediabevakning
- ▶ epostsortering
- ▶ dokumentklassificering och -indexering
- ▶ dialogsystem/talsvarssystem
- ▶ automatiska dikteringsystem

# att utveckla språkteknologi

Språkteknologi bygger alltid på information om det språk eller de språk som ska hanteras.

Den dominerande tekniken idag använder **sjävlärande system**, som matas med **mycket stora mängder språkliga data** som innehåller det önskade resultatet ("träningsdata", "facit"), exempelvis:

- ▶ texter och deras översättningar
- ▶ texter där personer, platser och händelser är (manuellt) utmärkta som sådana
- ▶ epost sorterad i spam och icke-spam
- ▶ dokument sorterade efter ämnesinnehåll
- ▶ talinspelningar och deras transkriptioner

## att utveckla språkteknologi, 2

Om man inte har stora mängder språkliga data, finns (för en del av teknologierna) en alternativ väg, som dock innebär en betydligt större arbetsinsats, nämligen s.k. **regelbaserad språkteknologi**.

Den kräver dock fortfarande tillgång till språkliga data, så att man kan testa ut regler och överhuvudtaget få klart för sig vilka regelbundenheter man behöver kunna ta hand om.

Den kräver även betydligt djupare kunskap om språket/språken än vad som är fallet med de självlärande systemen.

(Naturligtvis finns även "hybridsystem", som försöker förena det bästa från de två approacherna.)

## Översätt



svenska engelska spanska Identifiera språk ▾



engelska danska ryska ▾

Översätt

Mittuniversitetet har erhållit flera anslag för att studera markvärdets relation till befolkningstillväxten.



My university has received several grants to study markvärdets relation to population growth.



Menade du: Mittuniversitetet har erhållit flera anslag för att studera **markvärde** relation till befolkningstillväxten.

## Översätt



svenska engelska polska Identifiera språk ▾



engelska danska svenska ▾

Översätt

Misje Centrum Archeologii Śródziemnomorskiej UW zostały nagrodzone w konkursie „Luxor Times Magazine” w kategoriach najważniejsze odkrycie i projekt konserwatorski. |



Uppdrag Centre of Mediterranean Archaeology UW tilldelades i "Luxor Times Magazine" i termer av den viktigaste upptäckten och bevarandeprojekt .



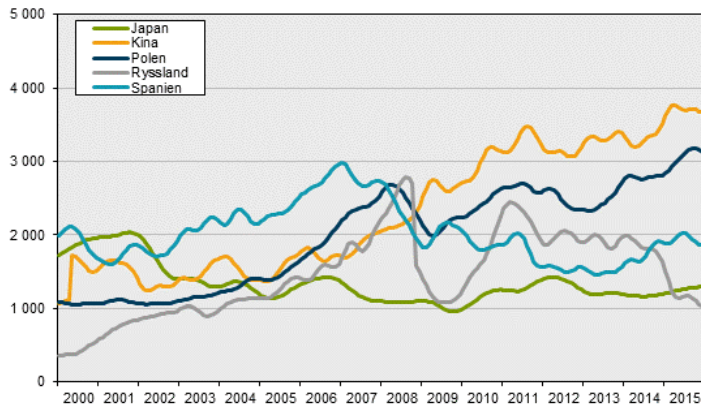
# språk i Sverige

- ▶ i Sverige talas omkring 150 språk
- ▶ svenska är huvudspråk i Sverige
- ▶ det finns fem officiella minoritetsspråk
- ▶ svenskt teckenspråk har en särställning i svensk lagstiftning
- ▶ de nordiska språken danska och norska får i många fall användas jämte svenska i Sverige
- ▶ engelska har en mycket stark ställning på alla områden i Sverige
- ▶ en handfull invandrarspråk har minst 50.000 talare vardera
- ▶ turismen är en viktig näring på många håll i Sverige

# dessutom:


## Export till stora handelspartner

Sveriges varuexport till fem stora mottagarländer; i miljoner kronor.  
Trend.



Källa: SCB

Data t.o.m december 2015

 Sveriges officiella statistik

# därför behöver vi språkteknologi för våra språk

- ▶ fullvärdiga samhällstjänster på lika villkor i den digitala tidsåldern
- ▶ "24-timmarsmyndigheten"
- ▶ kris- och samhällsbevakning
- ▶ samhällets och marknadens internationalisering (fri rörlighet för människor, varor och tjänster)



# den svenska META-NET-vitboken (2012)



<<http://www.meta-net.eu/whitepapers/overview>>

Här får vi bl.a. veta att svenska idag är större på internet än i den fysiska världen.

Hur länge kan det fortsätta utan språkteknologistöd?

# språkresursläget förr och nu

Högkvalitativa resurser	God resurstillgång	Mätlig resurstillgång	Fragmentariska resurser	Få eller inga resurser
	engelska	franska italienska nederländska polska spanska <b>svenska</b> tjeckiska ryska ungerska	bulgariska danska omniska finska galiciska grekiska katalanska kroatiska slovakiska portugisiska rumänska serbiska slovakiska slovenska	iriska isländska lettiska litauiska maltesiska

13: Språkresurser: Tillgång till tal- och textresurser för 30 europeiska språk

Högkvalitativa resurser	God resurstillgång	Mätlig resurstillgång	Fragmentariska resurser	Få eller inga resurser
	engelska	franska maltesiska nederländska spanska ryska	bulgariska danska finska galiciska grekiska katalanska norska polska portugisiska rumänska slovakiska <b>svenska</b> tjeckiska ungerska	omniska iriska isländska kroatiska lettiska litauiska maltesiska serbiska

12: Textanalyser: Tillgång till språkteknologi för 30 europeiska språk

Högkvalitativa resurser	God resurstillgång	Mätlig resurstillgång	Fragmentariska resurser	Få eller inga resurser
	engelska	franska spanska	italienska katalanska nederländska polska rumänska ryska ungerska	bulgariska bulgariska danska omniska finska galiciska grekiska iriska isländska kroatiska lettiska litauiska maltesiska norska portugisiska serbiska slovakiska slovenska <b>svenska</b> tjeckiska

11: Maskinöversättning: Tillgång till språkteknologi för 30 europeiska språk

Högkvalitativa resurser	God resurstillgång	Mätlig resurstillgång	Fragmentariska resurser	Få eller inga resurser
	engelska	finska franska italienska nederländska portugisiska spanska tjeckiska ryska	bulgariska bulgariska danska omniska galiciska grekiska iriska katalanska norska polska serbiska slovakiska slovenska <b>svenska</b> ungerska	isländska kroatiska lettiska litauiska maltesiska rumänska

10: Talteknologi: Tillgång till språkteknologi för 30 europeiska språk

För svenska är läget bättre – i några fall betydligt bättre – idag än då vitboken skrevs, men inte för flera av minoritetsspråken eller svenskt teckenspråk.

# ett par avslutande frågor och funderingar

- ▶ datorer / mobiltelefoner / bilar / operativsystem / etc.  
~ språkteknologi
- ▶ marknadskrafter ~ det allmänna ~ frivilligarbete

tack för uppmärksamheten!

