

Trädbanken SweLL

när andraspråkstalare annoterar andraspråkstexter

Arianna Masciolini
Språkbanken text

Andraspråkskorpusar



Swedish (SweLL)

```
<sentence> <w ref="1">"</w> <w ref="2" target_form="Det"  
correction_label="L-Ref">Den</w> <w ref="3">är</w>  
<w ref="4">en</w> <w ref="5">tredjedel</w>  
<w ref="6">av</w> <w ref="7">>din</w> <w ref="8">dag</w>  
<w ref="9">!</w> </sentence>
```

English (FCE)

I also suggest that more plays and films should

```
<ns type="RV"> <ns type="FV"><i>be taken</i><c>take</c>  
</ns> place</ns>.
```

Italian (VALICO)

Finse <MC><i>aveva paura</i><c>che aveva paura</c>
</MC> di un <DN><i>rapito</i><c>rapimento</c></DN>.

Begränsningar



- ▶ olika format som är inkompatibla med varandra
- ▶ manuell taggning
- ▶ grovkorniga felkategoritaggar (TODO: bättre översättning för “coarse-grained”?)
- ▶ snäv inriktning på språkliga fel

Universal Dependencies



Universal Dependencies är:

- ▶ ett ramverk för att skapa morfosyntaktiska analyser som är
 - ▶ trärspråkligt konsekventa
 - ▶ läsbara för både människor och maskiner
 - ▶ tillämpliga i både en- och flerspråkiga användningsscenarier
- ▶ en växande samling trädbanker för 180+ språk

UD för andraspråksdata



- ▶ UD:s **standardformat** (CoNLL-U) tillgängliggör etablerade sök- och bearbetningsverktyg
- ▶ befintliga **parsrar** underlättar annoteringen
- ▶ **detaljerade morfosyntaktiska analyser** kan användas för att studera L2-grammatik i stort
- ▶ att analyserna är **tvärspråkligt konsekventa*** gör det möjligt att jämföra:
 - ▶ andraspråksproduktioner med målspråket
 - ▶ andraspråket med modersmålet hos samma grupp
 - ▶ olika andraspråk hos samma grupp

* OBS!

TL;DR:

De officiella UD-riktlinjerna specificerar **inte** hur alla andraspråksfenomen ska analyseras; därför är andraspråksträdbanker **inte** 100% kompatibla med varandra

Läs mer:

Arianna Masciolini, Aleksandrs Berdicevskis, Maria Irena Szawerna, and Elena Volodina. *Annotating second language in Universal Dependencies: a review of current practices and directions for harmonized guidelines* (2025)

Andraspråksträdbanker



språk	namn	meningar	parallel	tillgänglig
kinesiska	CFL	451	(✓)	✓
engelska	ESL	5124	✓	(✓)
engelska	ESLSpok	2320		✓
grekiska	GLCII	670+	✓	snart!
italienska	Valico	398	✓	✓
koreanska	KSL	12977		✓
ryska		500	✓	
spanska		383		
svenska	SweLL	510+	✓	✓
tjeckiska		2000+	✓	

Källkorpus



SweLL (Swedish Learner Language) korpusen:

- ☒ består av elevuppsatser skrivna inom ramen för utbildning i svenska som andraspråk
- ☒ representerar vuxna inlärare med olika språkbakgrund och kunskapsnivåer
- ☒ är pseudonymiserad och annoterad med normaliseringar och felkategoritaggar

Annoterare



- ▶ första 510 meningar (med i UD 2.17):
 - ▶ Aleksandrs Berdicevskis (L1: ryska)
 - ▶ Maria Irena Szawerna (L1: polska)
 - ▶ Arianna Masciolini (L1: italienska)
- ▶ ytterligare ~140 meningar skrivna av elever med franska som modersmål eller skolspråk:
 - ▶ Caroline Grand-Clement (L1: franska)
 - ▶ Arianna Masciolini (L1: italienska)

Riktlinjer



1. analyserna försöker avspeglar elevens inre grammatik men samtidigt vara konsekventa med textens normalisering
2. medelväg mellan s.k. *ordagrann* och *distributionell* annotering:
 - ❖ LEMMA, UPOS och FEATS beskriver ordformen i sig
 - ❖ HEAD och DEPREL tilldelas baserat på hur ordet används
3. tydliga transferfenomenen annoteras enligt riktlinjerna för källspråket

Riktlinjer



1. analyserna försöker avspeglar elevens inre grammatik men samtidigt vara konsekventa med textens normalisering
2. medelväg mellan s.k. *ordagrann* och *distributionell* annotering:
 - ❖ LEMMA, UPOS och FEATS beskriver ordformen i sig
 - ❖ HEAD och DEPREL tilldelas baserat på hur ordet används
3. tydliga transferfenomenen annoteras enligt riktlinjerna för källspråket

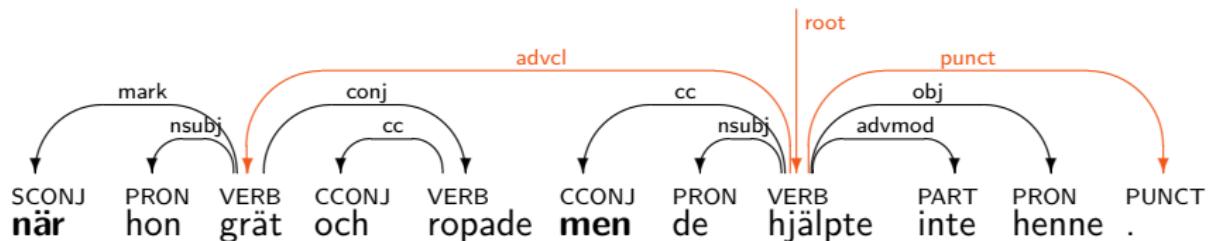
Exempel 1



SCONJ PRON VERB CCONJ VERB CCONJ PRON VERB PART PRON PUNCT
när hon grät och ropade **men** hjälpte de inte henne .

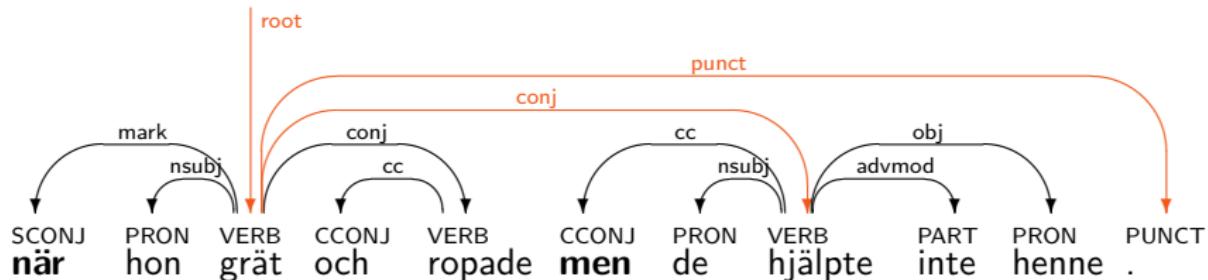
- ▶ när hon grät och ropade **men** hjälpte de inte henne.
- ▶ **när** hon grät och ropade, men de hjälpte inte henne.

Exempel 1



- ▶ **när hon grät och ropade men hjälpte de inte henne.**
- ▶ **när hon grät och ropade, men de hjälpte inte henne.**

Exempel 1



- ▶ nära hon grät och ropade men hjälpte de inte henne.
- ▶ nära hon grät och ropade, men de hjälpte inte henne.

Riktlinjer



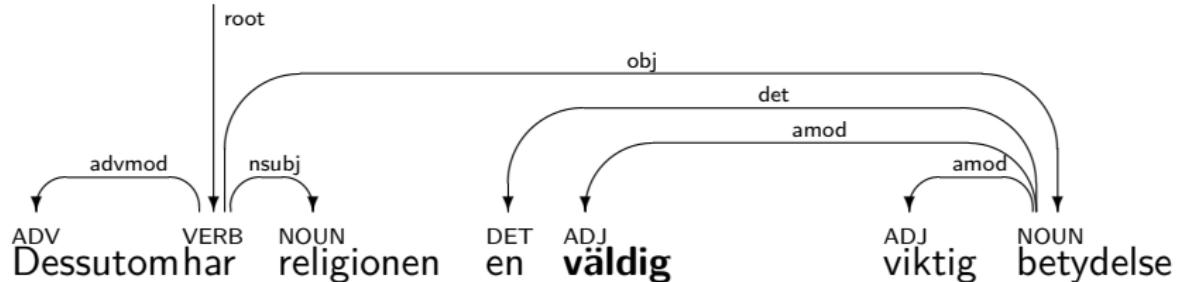
1. analyserna försöker avspeglar elevens inre grammatik men samtidigt vara konsekventa med textens normalisering
2. medelväg mellan s.k. *ordagrann* och *distributionell* annotering:
 - LEMMA, UPOS och FEATS beskriver ordformen i sig
 - HEAD och DEPREL tilldelas baserat på hur ordet används
3. tydliga transferfenomenen annoteras enligt riktlinjerna för källspråket

Exempel 2

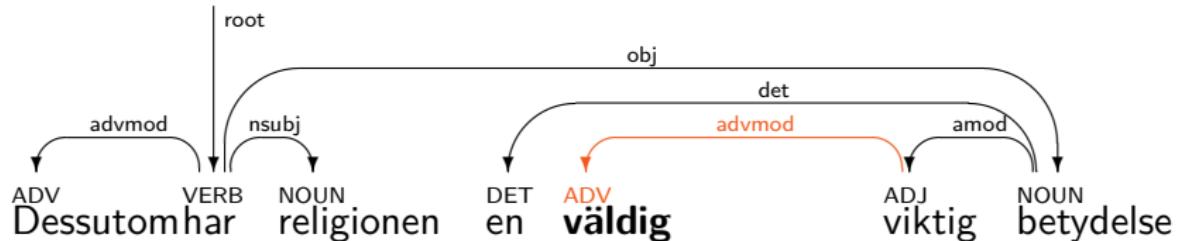


ADV VERB NOUN DET ADJ ADJ NOUN
Dessutom har religionen en väldig viktig betydelse

Exempel 2

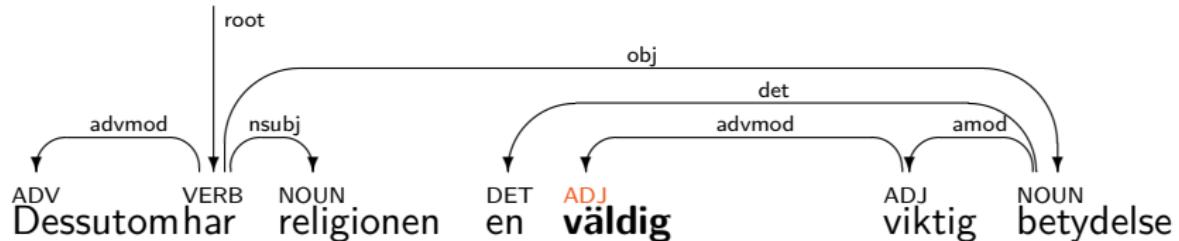


Exempel 2



- ➊ Dessutom har religionen en väldigt viktig betydelse

Exempel 2



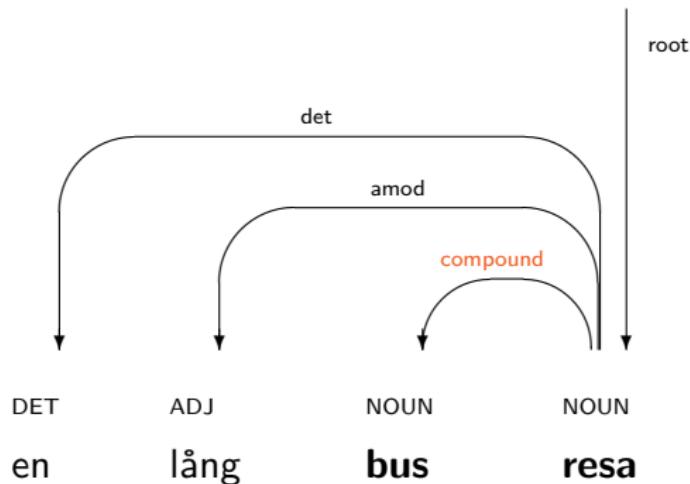
- Dessutom har religionen en väldigt viktig betydelse

Riktlinjer



1. analyserna försöker avspeglar elevens inre grammatik men samtidigt vara konsekventa med textens normalisering
2. medelväg mellan s.k. *ordagrann* och *distributionell* annotering:
 - ❖ LEMMA, UPOS och FEATS beskriver ordformen i sig
 - ❖ HEAD och DEPREL tilldelas baserat på hur ordet används
3. tydliga transferfenomenen annoteras enligt riktlinjerna för källspråket

Exempel 3



- ▶ en lång bussresa
- ▶ L1: engelska

Exempel 4



PRON AUX PRON/DET
det är **det**

ADJ
samma

PRON/ADP/... ADP
som i PROPN
 Sverige

Exempel 4



Exempel 4

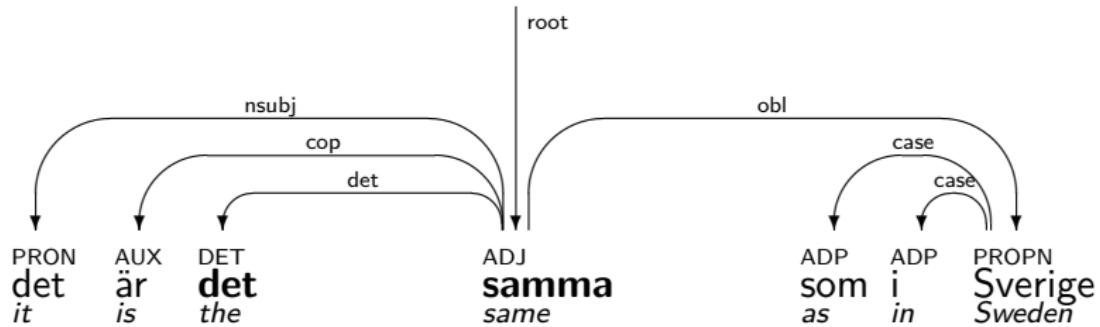


PRON AUX DET
det är **det**
it *is* *the*

ADJ
samma
same

ADP ADP PROPN
som i Sverige
as *in* *Sweden*

Exempel 4



Tillgänglig via



Tack!

