

# STUnD i användning

- presens perfekt i svenska och engelska -

**Arianna Masciolini & Márton A. Tóth**

Institutionen för svenska, flerspråkighet och språkteknologi  
Göteborgs universitet

Huminfra conference, Göteborgs universitet 2024.01.11

# Kort om STUnD

- Sökverktyg för Tvåspråkig Universal Dependencies
- Prototyp, tillgänglig på [github.com/harison/STUnD](https://github.com/harison/STUnD)
- Används med parallella (meningslänkade) UD-trädbanker
- Frågespråk:
  - UD-baserad → språkoberoende
  - Särskilt lämpligt för syntaktiska frågor
  - Tillåter tvåspråkiga sökningar

# Exempel frågor

/home/harisont/Data/ud-treebanks-v2.13/UD\_English-PUD/en\_pud-ud-test.conllu

/home/harisont/Data/ud-treebanks-v2.13/UD\_Swedish-PUD/sv\_pud-ud-test.conllu

FORM "har"

additional replacement rule (optional)

Mode:  text  CoNNL-U

[save as TSV](#)

That share has **been** rising steadily over the years — only 11 percent of the total vote was cast before Election Day in 1996 , according to the Census Bureau -- and seems likely to jump again this year .

Den andelen **har** ökat stadigt med åren – bara 11 procent av de samlade rösterna lades före valdagen 1996 enligt folkbokföringsbyrån – och det verkar troligt att den kommer öka ordentligt igen .

" We **'ve** requested other nations to help us populate the zoo with different species of animals , including a pig , " Saqib said .

" Vi **har** bett andra stater att hjälpa oss befolka djurparken med olika djurarter , inklusive en gris " , sade Saqib .

Several analysts **have** suggested Huawei is best placed to benefit from Samsung 's setback .

Flera analytiker **har** föreslagit att Huawei har bäst position för att tjäna på Samsungs tillbakagång .

Several analysts have suggested **Huawei is best placed to benefit from Samsung 's setback** .

Flera analytiker har föreslagit **att Huawei har bäst position för att tjäna på Samsungs tillbakagång** .

" But it may be that BA and IAG have cracked it and can offer something vaguely reliable . "

Men det kan vara så att BA och IAG **har** knäckt det och kan erbjuda något som kan verka pålitligt .

The 10 - week course **has** been " certified " by UK spy agency GCHQ .

10-veckorskursen **har** " certifierats " av brittiska spionmyndigheten GCHQ .

Throughout history , the international hair market **has** always had a political dimension , says Tarlo .

Genom historien **har** den internationella hårmarknaden alltid haft en politisk dimension , säger Tarlo .

Shenzhen 's traffic police **have** opted for unconventional penalties before .

Shenzhens trafikpolis **har** valt okonventionella straff förut .

Much of the debate , from the Democratic side this year , **has** been about white male identity .

Mycket av debatten , från Demokraternas sida i år , **har** handlat om vit manlig identitet .

Seagal , whose grandmother was from Vladivostok in Russia 's far east , **has** made frequent trips to Russia in recent years and visited Kamchatka and Sakhalin in September .

Seagal , vars farmor kom från Vladivostok i Rysslands östligaste delar **har** gjort flera resor till Ryssland de senaste åren och besökte Kamtjatka och Sakhalin i september .

/home/harisont/Data/ud-treebanks-v2.13/UD\_English-PUD/en\_pud-ud-test.conllu

/home/harisont/Data/ud-treebanks-v2.13/UD\_Swedish-PUD/sv\_pud-ud-test.conllu

AND [POS "AUX", LEMMA "ha", FEATS\_ "Tense=Pres"]

additional replacement rule (optional)

Mode:  text  CoNNL-U

[save as TSV](#)

That share has **been** rising steadily over the years — only 11 percent of the total vote was cast before Election Day in 1996 , according to the Census Bureau -- and seems likely to jump again this year .

" We **'ve** requested other nations to help us populate the zoo with different species of animals , including a pig , " Saqib said .

Several analysts **have** suggested Huawei is best placed to benefit from Samsung 's setback .

" But it may be that BA and IAG have cracked it and can offer something vaguely reliable . "

The 10 - week course **has** been " certified " by UK spy agency GCHQ .

Throughout history , the international hair market **has** always had a political dimension , says Tarlo .

Shenzhen 's traffic police **have** opted for unconventional penalties before .

Much of the debate , from the Democratic side this year , **has** been about white male identity .

Seagal , whose grandmother was from Vladivostok in Russia 's far east , **has** made frequent trips to Russia in recent years and visited Kamchatka and Sakhalin in September .

Researchers have **been** investigating potential for male hormonal contraceptives for around 20 years .

Den andelen **har** ökat stadigt med åren – bara 11 procent av de samlade rösterna lades före valdagen 1996 enligt folkbokföringsbyrån – och det verkar troligt att den kommer öka ordentligt igen .

" Vi **har** bett andra stater att hjälpa oss befolka djurparken med olika djurarter , inklusive en gris " , sade Saqib .

Flera analytiker **har** föreslagit att Huawei har bäst position för att tjäna på Samsungs tillbakagång .

Men det kan vara så att BA och IAG **har** knäckt det och kan erbjuda något som kan verka pålitligt .

10-veckorskursen **har** " certifierats " av brittiska spionmyndigheten GCHQ .

Genom historien **har** den internationella hårmarknaden alltid haft en politisk dimension , säger Tarlo .

Shenzhens trafikpolis **har** valt okonventionella straff förut .

Mycket av debatten , från Demokraternas sida i år , **har** handlat om vit manlig identitet .

Seagal , vars farmor kom från Vladivostok i Rysslands östligaste delar **har** gjort flera resor till Ryssland de senaste åren och besökte Kamtjatka och Sakhalin i september .

Forskare **har** undersökt potentialen för manliga hormonella preventivmedel i ungefär 20 år .

/home/harisont/Data/ud-treebanks-v2.13/UD\_English-PUD/en\_pud-ud-test.conllu

/home/harisont/Data/ud-treebanks-v2.13/UD\_Swedish-PUD/sv\_pud-ud-test.conllu

SEQUENCE [AND [POS "AUX", LEMMA "ha", FEATS\_ "Tense=Pres"], FEATS\_ "VerbForm=Sup"]

\_TER\_SUBTREES (OR [POS "VERB", DEPREL\_ "root", DEPREL\_ "csubj", DEPREL\_ "ccomp", DEPREL\_ "xcomp", DEPREL\_ "advcl", DEPREL\_ "acl"]) (OR [DEPREL\_ "aux", DEPREL\_ "cop"]), PRUNE (DEPREL\_ "aux") 0, PRUNE (DEPREL\_ "cop") 0]

Mode:  text  CoNNL-U

[save as TSV](#)

That share **has been rising** steadily over the years — only 11 percent of the total vote was cast before Election Day in 1996 , according to the Census Bureau -- and seems likely to jump again this year .

" We **'ve requested** other nations to help us populate the zoo with different species of animals , including a pig , " Saqib said .

Several analysts **have suggested** Huawei is best placed to benefit from Samsung 's setback .

" But it may be that BA and IAG have cracked it and can offer something vaguely reliable . "

Shenzhen 's traffic police **have opted** for unconventional penalties before .

Much of the debate , from the Democratic side this year , **has been** about white male **identity** .

Seagal , whose grandmother was from Vladivostok in Russia 's far east , **has made** frequent trips to Russia in recent years and visited Kamchatka and Sakhalin in September .

Researchers **have been investigating** potential for male hormonal contraceptives for around 20 years .

He added : " This has been one step in a long journey of finding the right combination for male hormonal contraception . "

Ms Pugh **has received** treatment at Papworth and Addenbrooke 's Hospitals in

Den andelen **har ökat** stadigt med åren – bara 11 procent av de samlade rösterna lades före valdagen 1996 enligt folkbokföringsbyrån – och det verkar troligt att den kommer öka ordentligt igen .

" Vi **har bett** andra stater att hjälpa oss befolka djurparken med olika djurarter , inklusive en gris " , sade Saqib .

Flera analytiker **har föreslagit** att Huawei har bäst position för att tjäna på Samsungs tillbakagång .

Men det kan vara så att BA och IAG **har knäckt** det och kan erbjuda något som kan verka pålitligt .

Shenzhens trafikpolis **har valt** okonventionella straff förut .

Mycket av debatten , från Demokraternas sida i år , **har handlat** om vit manlig identitet .

Seagal , vars farmor kom från Vladivostok i Rysslands östligaste delar **har gjort** flera resor till Ryssland de senaste åren och besökte Kamtjatka och Sakhalin i september .

Forskare **har undersökt** potentialen för manliga hormonella preventivmedel i ungefär 20 år .

Han tillade : " **Detta har varit ett steg i en lång resa** för att **hitta** den rätta kombinationen för manliga hormonella preventivmedel " .

Ms Pugh **har fått** behandling vid Papworths och Addenbrookes sjukhus i Cambridgeshire .

/home/harisont/Data/ud-treebanks-v2.13/UD\_English-PUD/en\_pud-ud-test.conllu

/home/harisont/Data/ud-treebanks-v2.13/UD\_Swedish-PUD/sv\_pud-ud-test.conllu

TREE\_ (FEATS\_ "VerbForm=Sup") [AND [POS "AUX", LEMMA "ha", FEATS\_ "Tense=Pres"]]

\_TER\_SUBTREES (OR [POS "VERB", DEPREL\_ "root", DEPREL\_ "csubj", DEPREL\_ "ccomp", DEPREL\_ "xcomp", DEPREL\_ "advcl", DEPREL\_ "acl"]) (OR [DEPREL\_ "aux", DEPREL\_ "cop"]), PRUNE (DEPREL\_ "aux") 0, PRUNE (DEPREL\_ "cop") 0]

Mode:  text  CoNNL-U

[save as TSV](#)

That share **has been rising** steadily over the years — only 11 percent of the total vote was cast before Election Day in 1996 , according to the Census Bureau -- and seems likely to jump again this year .

" We **'ve requested** other nations to help us populate the zoo with different species of animals , including a pig , " Saqib said .

Several analysts **have suggested** Huawei is best placed to benefit from Samsung 's setback .

" But it may be that BA and IAG have cracked it and can offer something vaguely reliable . "

The 10 - week course **has been " certified "** by UK spy agency GCHQ .

Throughout history , the international hair market **has** always **had** a political dimension , says Tarlo .

Shenzhen 's traffic police **have opted** for unconventional penalties before .

Much of the debate , from the Democratic side this year , **has been** about white male **identity** .

Seagal , whose grandmother was from Vladivostok in Russia 's far east , **has made** frequent trips to Russia in recent years and visited Kamchatka and Sakhalin in September .

Researchers **have been investigating** potential for male hormonal contraceptives for around 20 years .

Den andelen **har ökat** stadigt med åren – bara 11 procent av de samlade rösterna lades före valdagen 1996 enligt folkbokföringsbyrån – och det verkar troligt att den kommer öka ordentligt igen .

" Vi **har bett** andra stater att hjälpa oss befolka djurparken med olika djurarter , inklusive en gris " , sade Saqib .

Flera analytiker **har föreslagit** att Huawei har bäst position för att tjäna på Samsungs tillbakagång .

Men det kan vara så att BA och IAG **har knäckt** det och kan erbjuda något som kan verka pålitligt .

10-veckorskursen **har " certifierats "** av brittiska spionmyndigheten GCHQ .

Genom historien **har** den internationella hårmarknaden alltid **haft** en politisk dimension , säger Tarlo .

Shenzhens trafikpolis **har valt** okonventionella straff förut .

Mycket av debatten , från Demokraternas sida i år , **har handlat** om vit manlig identitet .

Seagal , vars farmor kom från Vladivostok i Rysslands östligaste delar **har gjort** flera resor till Ryssland de senaste åren och besökte Kamtjatka och Sakhalin i september .

Forskare **har undersökt** potentialen för manliga hormonella preventivmedel i ungefär 20 år .

/home/harisont/Data/ud-treebanks-v2.13/UD\_English-PUD/en\_pud-ud-test.conllu

/home/harisont/Data/ud-treebanks-v2.13/UD\_Swedish-PUD/sv\_pud-ud-test.conllu

TREE\_ (AND [FEATS\_ "VerbForm={Part->Sup}", {FEATS\_ "Tense=Pres"->TRUE}]) [AND [POS "AUX", LEMMA "{have->ha}", FEATS\_ "Tense=Pres"]]

\_TER\_SUBTREES (OR [POS "VERB", DEPREL\_ "root", DEPREL\_ "csubj", DEPREL\_ "ccomp", DEPREL\_ "xcomp", DEPREL\_ "advcl", DEPREL\_ "acl"]) (OR [DEPREL\_ "aux", DEPREL\_ "cop"]), PRUNE (DEPREL\_ "aux") 0, PRUNE (DEPREL\_ "cop") 0]

Mode:  text  CoNLL-U

[save as TSV](#)

That share **has been rising** steadily over the years — only 11 percent of the total vote was cast before Election Day in 1996 , according to the Census Bureau -- and seems likely to jump again this year .

Den andelen **har ökat** stadigt med åren – bara 11 procent av de samlade rösterna lades före valdagen 1996 enligt folkbokföringsbyrån – och det verkar troligt att den kommer öka ordentligt igen .

Researchers **have been investigating** potential for male hormonal contraceptives for around 20 years .

Forskare **har undersökt** potentialen för manliga hormonella preventivmedel i ungefär 20 år .

Students like Rai **have been meeting** with counsellors at the school to talk about what happened , but she said the biggest comfort has come from seeing her friends .

Studenter som Rai **har träffat** kuratorer på skolan för att prata om det som hände , men hon säger att den största trösten har kommit från att träffa sina vänner .

3 hits

→ "VerbForm={Part->Sup}"



/home/harisont/Data/ud-treebanks-v2.13/UD\_English-PUD/en\_pud-ud-test.conllu

/home/harisont/Data/ud-treebanks-v2.13/UD\_Swedish-PUD/sv\_pud-ud-test.conllu

TREE\_ (AND [FEATS\_ "VerbForm={Part->Sup}",{FEATS\_ "Tense=Pres"->TRUE}]) [AND [POS "AUX", LEMMA "{have->ha}", FEATS\_ "Tense=Pres"]]

\_TER\_SUBTREES (OR [POS "VERB", DEPREL\_ "root", DEPREL\_ "csubj", DEPREL\_ "ccomp", DEPREL\_ "xcomp", DEPREL\_ "advcl", DEPREL\_ "acl"]) (OR [DEPREL\_ "aux", DEPREL\_ "cop"]), PRUNE (DEPREL\_ "aux") 0, PRUNE (DEPREL\_ "cop") 0]

Mode:  text  CoNNL-U  [save L1](#) [save L2](#)

1 has have AUX VBZ  
 Mood=Ind|Number=Sing|Person=3|Tense=Pres|VerbForm=Fin 3 aux  
 5:aux ADJUSTED=True  
 2 been be AUX VBN Tense=Past|VerbForm=Part  
 3 aux 5:aux ADJUSTED=True  
 3 rising rise VERB VBG Tense=Pres|VerbForm=Part  
 0 root 0:root ADJUSTED=True|ORIG\_LABEL=root

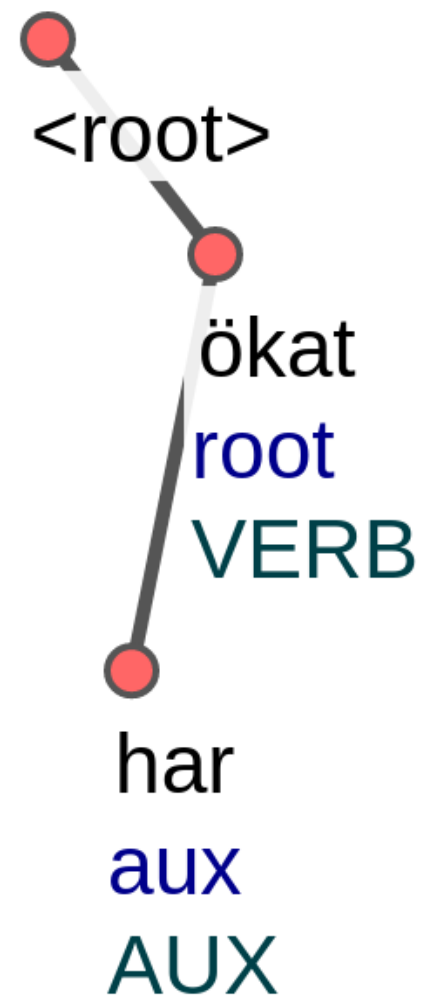
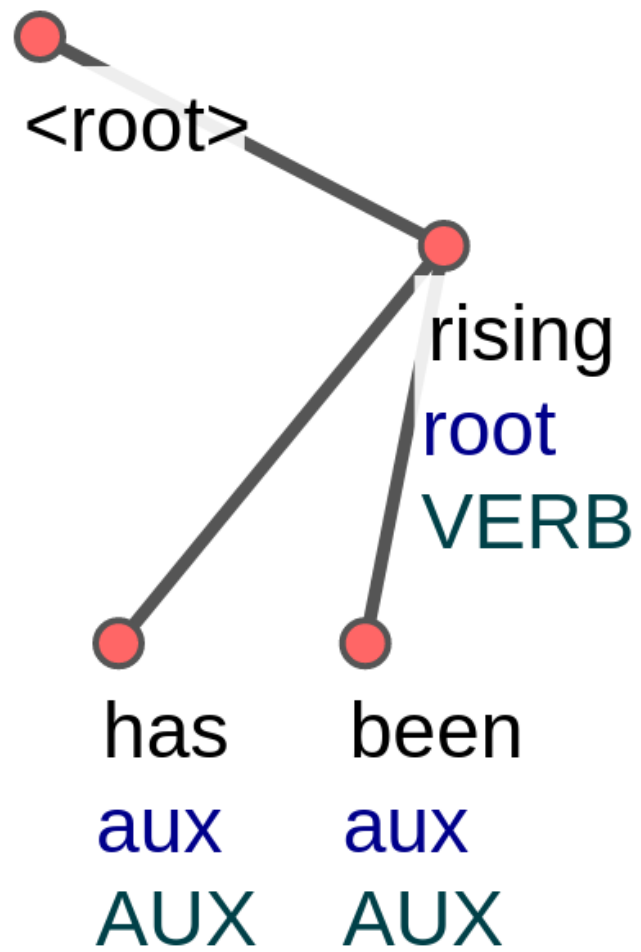
1 har ha AUX VB|PRS|AKT  
 Mood=Ind|Tense=Pres|VerbForm=Fin|Voice=Act 2 aux 4:aux  
 ADJUSTED=True  
 2 ökat öka VERB VB|SUP|AKT VerbForm=Sup|Voice=Act  
 0 root 0:root ADJUSTED=True|ORIG\_LABEL=root

1 have have AUX VBP  
 Mood=Ind|Tense=Pres|VerbForm=Fin 3 aux 4:aux  
 ADJUSTED=True  
 2 been be AUX VBN Tense=Past|VerbForm=Part  
 3 aux 4:aux ADJUSTED=True  
 3 investigating investigate VERB VBG  
 Tense=Pres|VerbForm=Part 0 root 0:root  
 ADJUSTED=True|ORIG\_LABEL=root

1 har ha AUX VB|PRS|AKT  
 Mood=Ind|Tense=Pres|VerbForm=Fin|Voice=Act 2 aux 3:aux  
 ADJUSTED=True  
 2 undersökt undersöka VERB VB|SUP|AKT  
 VerbForm=Sup|Voice=Act 0 root 0:root  
 ADJUSTED=True|ORIG\_LABEL=root

1 have have AUX VBP  
 Mood=Ind|Tense=Pres|VerbForm=Fin 3 aux 6:aux  
 ADJUSTED=True  
 2 been be AUX VBN Tense=Past|VerbForm=Part  
 3 aux 6:aux ADJUSTED=True  
 3 meeting meet VERB VBG Tense=Pres|VerbForm=Part  
 0 root 0:root ADJUSTED=True|ORIG\_LABEL=root

1 har ha AUX VB|PRS|AKT  
 Mood=Ind|Tense=Pres|VerbForm=Fin|Voice=Act 2 aux 5:aux  
 ADJUSTED=True  
 2 träffat träffa VERB VB|SUP|AKT VerbForm=Sup|Voice=Act  
 0 root 0:root ADJUSTED=True|ORIG\_LABEL=root



# Bakom kulisserna

1. Fras-och ordlänkning: verktyget hittar motsvarigheter mellan L1 och L2 delträd
2. Söksträngen transformeras till två separata L1 och L2 frågor
3. De två frågorna matchas mot steg 1:s L1-L2 motsvarande segment
4. (Ersättningsmönstret används för att förfina sökresultaten)

## **Fallstudie:**

presens perfekt i svenska och engelska

# Fallstudie - bakgrund

- Forskningsobjekt: hur används tempuset presens perfekt i svenska och engelska? Hur ser distributionen ut?
  - vilka motsvarigheter har engelskan till svenskans presens perfekt och vilka motsvarigheter har svenskan till engelskans presens perfekt?

- Presens perfekt

Svenska: t.ex. *har gjort*

Engelska: t.ex. *has done* men även *has been doing*

# Fallstudie - bakgrund

- Presens perfekt visar på stor tvärspråklig variation (se t.ex. de Swart 2016)
- Distribution: språk som har presens perfekt använder presens perfekt i olika utsträckning
  - (1) Har du sovit gott? (Larsson & Lyngfelt 2011:77)
  - (2) a. Did you sleep well?  
b. # Have you slept well?

# Fallstudie - material

- Material: parallellkorpusen Parallel Universal Dependencies (PUD)
  - 1000 meningar: wikipediatexter och tidningsartiklar
  - översatta från engelska i huvudsak, till svenska och andra språk

# Fallstudie - metod

- Söksträngar

Engelska:

```
TREE_ (FEATS_ "VerbForm=Part") [AND [LEMMA "have", FEATS_ "Tense=Pres"]]
```

Svenska:

```
TREE_ (FEATS_ "VerbForm=Sup") [AND [LEMMA "ha", FEATS_ "Tense=Pres"]]
```

- Söka på presens perfekt (i svenska) och present perfect (i engelska) och titta på motsvarande former - samt frekvens



# Fallstudie - resultat

**Tabell 1**

Svenska motsvarigheter till engelska present perfect

Form	Frekvens
Presens perfekt	77 (88,5%)
Supinum utan hjälpverb	6 (6,9%)
Enkelt presens	3 (3,4%)
Preteritum	1 (1,1%)
<b>Summa</b>	<b>87 (100%)</b>

**Tabell 2**

Engelska motsvarigheter till svenska presens perfekt

Form	Frekvens
Present perfect non-progressive	75 (94,9%)
Present perfect progressive	3 (3,8%)
Present tense	1 (1,3%)
<b>Summa</b>	<b>79 (100%)</b>

- Engelska present perfect användes i aningen större utsträckning än svenska presens perfekt
- Varierade vad som motsvarade presens perfekt/present perfect

# Fallstudie - resultat

Presens perfekt motsvarade present perfect non-progressive i de flesta fallen

Global warming **has caused** a change in the pattern of the rainy seasons .

Global uppvärmning **har orsakat** en förändring i regnperiodernas mönster .

Presens perfekt som motsvarade present perfect progressive

That share **has been rising** steadily over the years – only 11 percent of the total vote was cast before Election Day in 1996 , according to the Census Bureau -- and seems likely to jump again this year .

Den andelen **har ökat** stadigt med åren – bara 11 procent av de samlade rösterna lades före valdagen 1996 enligt folkbokföringsbyrån – och det verkar troligt att den kommer öka ordentligt igen .

Endast ett fall där presens perfekt motsvarade simple present i passiv

Outside Japan , beginning with Emperor Shōwa , the Emperors **are often referred** to by their given names , both whilst alive and posthumously .

Utanför Japan , med början med kejsar Shōwa , **har** kejsarna ofta **omtalats** med sina förnamn , både medan de lever och postumt .

# Fallstudie - resultat

Ha-utelämning: ensamt supinum som motsvarade present perfect

En taleskvinna för Harley-Davidson sade att de tidslinjer som **flutit** runt i medierna – 2017 eller 2020 – inte stämmer .

A Harley - Davidson spokeswoman said timelines that **have been floated** in the media -- 2017 or 2020 -- are n't accurate .

Enkelt presens som motsvarar present perfect

Efter en del genial kirurgi och mycket väldigt krävande rehabilitering **är** jag fullt **återställd** .

After some genius surgery and a lot of very tough rehab , I **have made** a full recovery .

Preteritum som motsvarar present perfect

Men en undersökning **visade** att tumören i Ms Pughs högra lunga växer och hon har varit tvungen att lämna försöket .

But a scan **has shown** the tumour in Ms Pugh 's right lung is growing , and she has had to leave the trial .

# Fallstudie - preliminära slutsatser och framtid

- Presens perfekt har liknande distribution i de två språken - men det finns smärre skillnader
  - Distributionen torde bero på olika faktorer: materialet, språkspecifika konventioner och översättningsrelaterade frågor
- Framöver: titta mer på
  - andra (förslagsvis större) material
  - kanske även fler språk, såsom japanska och italienska
  - semantik
  - fler faktorer än vi gjort hittills, t.ex. diates

# STUnD - framtida utveckling

- Optimering
- Integrering med automatisk UD-parser för att undersöka oannoterade texter
- Trädvisualiseringar direkt i verktyget
- Förtillverkade ersättningsmönster att välja mellan
- Omimplementering som webbapplikation?

# Litteratur

- de Swart, H. 2016. Perfect usage across languages. *Questions and Answers in Linguistics*, 3(2), 57-62.
- Larsson, I., & Lyngfelt, B. 2011. Tempus i svenskan. I C. Andersen, A. Granberg & I. Söhrman (redt.), *Tid och tidsförhållanden i olika språk*. 66-68. Institutionen för språk och litteraturer, Göteborgs universitet.
- Masciolini A. 2023. A query engine for L1-L2 parallel dependency treebanks. *Proceedings of the 24th Nordic Conference on Computational Linguistics (NoDaLiDa)*, 574–587. Tórshavn, Faroe Islands.
- Zeman D. et al. 2018. CoNLL 2018 shared task: Multilingual parsing from raw text to Universal Dependencies. *Proceedings of the CoNLL 2018 Shared Task: Multilingual Parsing from Raw Text to Universal Dependencies*, 1-21. Association for Computational Linguistics, Brussels, Belgium.