

Språk och Datorer (729G49)

Introduktion till Fördjupningsuppgiften

05/05/2025



Denna föreläsning

- Omfång och lärandemål
- Tema: Översättningssvenska
- Organisation och Tidsplan
- Genomförandet av Studien
- Rapportskrivning
- Betygskriterier

Omfång och lärandemål

Moduler och examination (1)

- 9 HP, fördelat på tre moduler:
 - 3 HP **Språkteknologi**: (5 labbar i par) — U, G
 - 3 HP **Lingvistik**: (tenta) — U, G, VG
 - 3 HP **Fördjupningsuppgift**: (individuell skriftlig rapport) — U, G, VG
- För G i kursen behövs G på alla tre moduler.
- För VG i kursen behövs VG på tentan och fördjupningsuppgiften.

Moduler och examination (1)

- 9 HP, fördelat på tre moduler:
 - 3 HP **Språkteknologi**: (5 labbar i par) — U, G
 - 3 HP **Lingvistik**: (tenta) — U, G, VG
 - 3 HP **Fördjupningsuppgift**: (individuell skriftlig rapport) — U, G, VG
- För G i kursen behövs G på alla tre moduler.
- För VG i kursen behövs VG på tentan och fördjupningsuppgiften.

Moduler och examination (2)

Språkteknologi

- 80 timmar
- 5 föreläsningar,
5x2 labbtimmar
- Börjar nu
- 5 labbar i par (U/G)

Lingvistik

- 80 timmar
- 6 föreläsningar,
5 seminarier
- Börjar nu
- Tenta (U/G/VG)

Fördjupningsuppgift

- 80 timmar
- 1 föreläsning,
3 seminarier
- Börjar v.19
- Individuell skriftlig
rapport (U/G/VG)

Moduler och examination (2)

Språkteknologi

- 80 timmar
- 5 föreläsningar,
5x2 labbtimmar
- Börjar nu
- 5 labbar i par (U/G)

Lingvistik

- 80 timmar
- 6 föreläsningar,
5 seminarier
- Börjar nu
- Tenta (U/G/VG)

Fördjupningsuppgift

- 80 timmar
- 1 föreläsning,
3 seminarier
- Börjar v.19
- Individuell skriftlig
rapport (U/G/VG)

Lärandemål

- Redogöra för grundläggande begrepp inom språkvetenskap, särskilt morfologi, syntax och semantik.
- Utföra grundläggande språkvetenskapliga analyser såsom ordklassbestämning och dependensanalys.
- Använda datorbaserade verktyg och programmering för att samla in, analysera och validera språkliga datamängder (korpusar).
- Planera och utföra enklare korpusundersökningar, samt redovisa och värdera resultaten.

Lärandemål

- Redogöra för grundläggande begrepp inom språkvetenskap, särskilt morfologi, syntax och semantik.
- Utföra grundläggande språkvetenskapliga analyser såsom ordklassbestämning och dependensanalys.
- Använda datorbaserade verktyg och programmering för att samla in, analysera och validera språkliga datamängder (korpusar).
- Planera och utföra enklare korpusundersökningar, samt redovisa och värdera resultaten.

Lärandemål

- Redogöra för grundläggande begrepp inom språkvetenskap, särskilt morfologi, syntax och semantik.
 - Utföra grundläggande språkvetenskapliga analyser såsom ordklassbestämning och dependensanalys.
 - Använda datorbaserade verktyg och programmering för att samla in, analysera och validera språkliga datamängder (korpusar).
- Planera och utföra enklare korpusundersökningar, samt redovisa och värdera resultaten.

Tema: Översättningssvenska

Översättningar

- Att översätta en text är att överföra den från ett språk till ett annat.
 - Mänsklig översättning: en person – en översättare – översätter texten. Professionella översättare har inte bara språkliga kompetens men även kulturella förståelse och kontextkännedom.
 - Maskinöversättning: görs av datorprogram, t.ex. Google Translate eller DeepL. Dessa system använder idag oftast neurala nätverk för att automatiskt översätta text.
- Man försöker bevara innebörden, tonen och stilen i det ursprungliga språket.

Översättningsspråk

- Översatta texter bevarar ofta egenskaper hos ursprungsspråket, samtidigt som de undviker språkliga konstruktioner på målspråket som ligger längre ifrån ursprungsspråket (Gellerstam 1986).
- Detta är bara ett fenomen som är känt inom både mänsklig och maskinell översättning som **översättningsspråk** (eng. *Translationese*, Gellerstam 1986).
- Översättarspråk är inte nödvändigtvis ett tecken på låg översättningskvalitet, även om termen kan uppfattas som nedsättande (Gellerstam 2005).

Egenskaper av Översättningsspråk

- Översatta texter är ofta mer explicita än sina originaltexter (Baker 1993).
- De är lexikalt och syntaktiskt enklare och följer en mer konventionell stil (Baker 1993).
- Översatta texter bevarar ofta egenskaper hos ursprungsspråket, samtidigt som de undviker språkliga konstruktioner på målspråket som ligger längre ifrån ursprungsspråket (Gellerstam 1986).

Översättningssvenska: Ord

Överrepresenterade Ord

- Definitivt
- Njuta
- Tillbringa
- Använda
- Vänligen
- Trevlig

Exempel från Katourgi (2022)

Underrepresenterade Ord

- Jobbig
- Klara
- Hinna
- Stå
- Ledig
- Sorry

Exempel från Katourgi (2022)

Översättningssvenska: Syntax

Med eller utan huvudord?

- Ett klassiskt exempel är att engelskan använder oftare huvudord med ett adjektiv där svenskan nöjer sig med endast ett adjektiv.
- I översättningar från engelska blir sådana huvudord därför vanligare än i svensk originaltext.
- *"A silly thing happened"*:
 - *"En fånig sak hände"* eller *"Något fånigt hände"*?
- *"He is a charming person"*:
 - *"Han är en charmig person"* eller *"Han är charmig"*?

Exempel från Gellerstam (2005)

Participformer

- Participformer är mycket mindre vanliga på svenska än på engelska:
 - *"Ta en bild på dig själv tittandes in i kameran"*
 - *"Ta en bild på dig själv när du tittar mot kameran"*.

Exempel från Katourgi (2022)

Artiklar efter *vara* och *bli*

- Artiklar efter *vara* och *bli* är mindre vanliga på svenska.
 - "*När jag var ett barn*" eller "*När jag var barn*"?
 - "*Jag är en översättare*" eller "*Jag är översättare*"?
- Men ibland behövs de?
 - "*Jag är en dålig översättare*".

Exempel från Katourgi (2022)

Possessivpronomen

- Possessivpronomen används mer sällan vid kroppsdelar:
 - ”*Jag måste träna mina armar*” eller ”*Jag måste träna armarna*”?

Exempel från Katourgi (2022)

Ordföljd i inledningsfraser

- På svenska placeras subjektet ofta tidigt i meningen:
 - *"Innan hon blev forskare jobbade mamma som lärare"* eller *"Innan mamma blev forskare jobbade hon som lärare"*?

Exempel från Katourgi (2022)

Uttryck av modalitet

- På engelska uttrycks modalitet ofta med frågor i slutet av meningar eller fraser som *"isn't it"?*, *"aren't they?"*, *"I suppose"*, eller *"I presume"*.
- Svenska använder istället korta adverbialord och modalpartiklar som *"väl"*, *"nog"*, *"dock"*, *"ju"* och *"visst"*, som saknar direkta motsvarigheter i engelskan.
- Detta gör att engelskinfluerade uttryck som *"du vet"*, *"eller hur?"*, *"antar jag"*, *"förmodligen"*, *"antagligen"*, *"säkert"* blir vanligare i översatt text.

Exempel från Katourgi (2022)

Referenser

Referenser (Utdrag finns i Lisam)

- Mona Baker. 1993. *Corpus linguistics and translation studies: Implications and applications*. In *Text and Technology*. John Benjamins.
- Martin Gellerstam. 1986. *Translationese in Swedish novels translated from English*. In L. Wollin and H. Lindquist, editors, *Translation studies in Scandinavia: Poceedings from the Scandinavian Symposium on Translation Theory (SSOTT) II*, number 75 in *Lund Studies in English*, page 88–95. CWK Gleerup, Lund.
- Martin Gellerstam. 2005. Chapter 13. *Fingerprints in Translation*, pages 201–213. *Multilingual Matters*, Bristol, Blue Ridge Summit.
- Alexander Katourgi. 2022. *Svenskan går bananer: En bok om översättningar som syns*. Lys förlag.

Mer Material

- **Populärvetenskaplig Podcast:** <https://www.sverigesradio.se/avsnitt/1518227>
- **Studie om LLM-genererade översättningar:** Vikas Raunak, Arul Menezes, Matt Post, and Hany Hassan. 2023. *Do GPTs produce less literal translations?* In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, pages 1041–1050, Toronto, Canada. Association for Computational Linguistics. <https://aclanthology.org/2023.acl-short.90/>
- **Evaluation av maskinöversättning för svenska:** Lars Ahrenberg. 2021. *Translation competence in machines: A study of adjectives in English-Swedish translation.* In *Proceedings for the First Workshop on Modelling Translation: Translatology in the Digital Age*, pages 57–65, online. Association for Computational Linguistics. <https://aclanthology.org/2021.motra-1.6/>

Intressanta Frågor

Frågor

- Vilka av dessa ord och konstruktioner kan vi hitta i text översatt med moderna översättningssystem?
 - För moderna översättningssystem finns det indikationer att stora språkmodeller producerar mindre ordagranna översättningar (Raunak et al., 2023).
- Dessa ord och konstruktioner är inte alltid exempel på översättningssvenska.
 - Hur kan vi avgöra detta?

Organisation och Tidsplan

Formell Tidsplan

- Introduktion 5/5
- Testdata tillgängligt: 20/5. 8:00.
- Inlämning: 5/6 17:00 (9.6. 8:00).
- Metodseminarium: 20/5 — Inlämning 2-3 slides i Lisam: 19/5 14:00.
- Resultatseminarium: 27/5 — Inlämning 2-3 slides i Lisam: 26/5 14:00.
- Rapportseminarium: 3/6 — Inlämning utkast i Lisam: 30/5 14:00.

Informell Tidsplan

Tid	Uppgifter	Skriva
5/5-11/5	Läsa material, egen litteratursökning, utveckla forskningsfrågorna.	Introduktion, Teori
12/5-18/5	Utveckla och implementera metoden för att besvara forskningsfrågorna.	Introduktion, Metod
19/5-25/5	Använd metoden på korpusen. Samla in resultaten. Tolka resultaten.	Data, Resultat
26/5-1/6	Tolka resultaten. Skriv klart rapporten.	Diskussion, Slutsatser
2/6-5/6	Finslipning av rapporten.	Inga nya!

Metodseminarium

- 20/5 10:15-12:00 i två grupper.
- Ni har nu utvecklat era forskningsfrågor och metoden för att besvara dem.
- 19/5 14:00: Lämna in 2-3 slides om dina forskningsfrågor och din metod.
- Bara dem som lämnar in något deltar i seminariet!
- ~3 minuters presentation, 3 minuters återkoppling från gruppen och läraren (beroende på antal deltagare).

Resultatseminarium

- 27/5 10:15-12:00 i två grupper.
- Ni har nu samlat in era resultat med hjälp av testdatan.
- 26/5 14:00: Lämna in 2-3 slides om dina resultat.
- Bara dem som lämnar in något deltar i seminariet!
- ~3 minuters presentation, 3 minuters återkoppling från gruppen och läraren (beroende på antal deltagare).

Rapportseminarium

- 3/6 10:15-12:00 i fyra grupper (en timme per grupp).
- Nu är rapporten nästan klar. I seminariet får och ger ni återkoppling.
- 30/5 14:00: Lämna in utkastet.
- Bara dem som lämnar in något deltar i seminariet!
- Ni förväntas läsa två andra deltagares rapporter innan seminariet och ge återkoppling till dem i seminariet.

Inlämning

- Deadline: 5/6 17:00 (9.6. 8:00).
- Lämna in:
 - Din rapport som ett PDF-dokument.
 - Din kod som Python-fil (.py) eller Jupyter Notebook (.ipynb).
 - Koderna måste producera alla resultat som du inkluderade i rapporten.
 - Vi måste kunna förstå koden. Inkludera gärna kommentarer eller textfält som förklarar vilken del av studien den tillhör.

Genomförandet av Studien

Kvantitativ och Kvalitativ

- Målet av fördjupningsuppgiften är att genomföra en korpusstudie med en kombination av:
 - **Kvantitativa metoder** (med hjälp av de programmeringstekniker ni lärt er i språkteknologidelen), och
 - **Kvalitativa analyser** (baserade på de lingvistiska kunskaper ni fick i lingvistikdelen).

Kvantitativ studie

- Välj fem ord som tycks vara vanligare i översättningssvenska. Notera antalet förekomster av varje ord i respektive korpus.
- Välj sedan fem olika syntaktiska konstruktioner och sök efter dem i korpusarna med hjälp av lingvistisk annotation. Notera antalet förekomster för varje konstruktion.
- Observera att det är fullt möjligt att vissa konstruktioner eller ord inte förekommer i korpusen. I sådana fall anger du 0 som frekvens.
 - Dubbelkolla dock att koden fungerar korrekt genom att testa t.ex. med ett manuellt konstruerat exempel som innehåller den aktuella konstruktionen.

Kvalitativ Studie

- Titta nu manuellt på de exempel du har hittat.
 - Känns de naturliga i svenskan?
 - Kan du komma på ett alternativ som känns mer naturligt?
- Skulle du säga att de är exempel på översättningssvenska?
- Om du tycker att de låter naturliga, är det möjligt att äldre personer skulle inte tycka likadant?
- Argumentera med hjälp av lingvistiska begrepp och din kunnskap om det svenska språket.

Datamaterial

- Datan kommer att publiceras den 20/5.
 - Ni ska utveckla hypoteser och metoden utan att känna till den exakta korpusen!
 - Ni kan dock redan testa koden med korpusen från labb 4.
- Det kommer vara två korpusdelar: *Dialog* och *Böcker*.
- Både är automatiskt översatt från svenska till engelska och automatiskt taggade.
- Obs att den automatiska taggningen är långt från perfekt!

Rapportskrivning

Formella kriterier

- Mellan 2 000 och 4 000 ord, motsvarande ca. 4–8 sidors text (exkl. titel, bilder, tabeller, formler, referenser).
- Akademiskt språk, väl genomarbetad och korrekturläst.
- Referenser enligt vedertaget format.

Tilltänkt läsare

- När jag och Katriin bedömer din rapport kommer vi läsa den som om vi vore personer som inte alls vet vad uppgiften handlar om.
- Detta innebär att du behöver förklara utförligt vad du har gjort, varför du har gjort det och hur du har gjort.
- Skriv för den som du var första dagen på programmet!
- Detta innebär t.ex. att du inte ska förutsätta att läsaren kan något om korpusar eller programmering.

Struktur

- **Inledning.** Förklara syftet med studien med egna ord.
- **Teori.** Definiera viktiga koncept med egna ord.
- **Data.** Beskriv datamaterialet med egna ord.
- **Metod.** Beskriv hur du genomfört ditt arbete.
- **Resultat.** Presentera dina resultat på ett objektivet sätt.
- **Diskussion.** Ge din tolkning av resultaten gentemot frågeställningen.
- **Slutsatser.** Dra tydliga slutsatser. Vad blev svaret på frågan?

Inledning

- Beskriv studien översiktligt och sätt den i ett sammanhang.
- Försök att göra läsaren intresserad av arbetet: Varför är studien intressant och viktig?
- Ge en översikt över problemområdet och beskriv syftet med din studie.
- Avsluta med en tydligt formulerad forskningsfråga som du besvarar i rapporten.
- Det ska framgå redan i introduktionen vad du har undersökt och hur.

Teori

- Introducera konceptet av en översättning, och vilka problem kan uppstå.
- Definiera översättningssvenska med egna ord.
 - Hur visar fenomenet sig i en text?
 - Illustrera med exempel.
- Referera till arbeten som länkas till från kurshemsidan och ev. till annat material du hittar om översättningssvenska.
 - Hur relaterar de till din studie?

Data

- Beskriv datamaterialet med egna ord.
 - Vilken domän är det?
 - Vilka begränsningar har datan? Hur skulle man kunna utöka korpusen för att förbättra studien?
- Inkludera deskriptiv statistik, t.ex. antal meningar och ord.
- Ju mer du vet om hur datan kom till, desto bättre!

Metod

- Beskriva hur du faktiskt genomfört din korpusstudie.
 - Hur har du valt ut ord och konstruktioner?
 - Hur du har implementerat dem?
- **Replikerbarhet:** Någon som har läst din metodbeskrivning ska kunna göra om samma studie och få samma resultat.
- Kod hör **inte** hemma i metoddelen. Den som läser din metoddel ska inte behöva kunna Python.

Resultat

- Presentera dina resultat på ett objektivet sätt. Tolka inte!
- Använd tabeller eller grafer för att presentera dina siffror från den kvantitativa delen av studien men glöm inte att även sammanfatta resultaten i en kort text.
- Det kvalitativa resultatet ska innehålla konkreta exempel från korpusen, förklarat med lingvistiska begrepp.
- Dina resultat är underlaget till din diskussion.

Diskussion

- Resultatdiskussion
 - Vad blev resultaten? Hur tolkar du dem?
 - Tycker du att exemplen du hittade känns översatta eller inte? Varför?
- Metoddiskussion
 - Var din metod lämplig? Vilka begränsningar har den? Vad skulle kunna förbättras — till exempel val av korpus, analysmetod, kod, eller urval av exempel?
- Relatera dina resultat till tidigare forskning från teoridelen.

Slutsatser

- Sammanfatta de viktigaste resultaten.
- Besvara forskningsfrågan tydligt och kortfattat.
- Lyft gärna fram eventuella förslag på vidare studier.

Betygskriterier

Kriterier för Godkänd

- För att bli godkänd ska du:
- Genomföra en korpusstudie med hjälp av datorlingvistiska analyser.
 - Leta efter minst 5 ord och 5 olika syntaktiska mönster som är kända tecken på översättningssvenska.
 - Analysera exemplen du hittade kritiskt med korrekt lingvistisk terminologi.
- Redovisa din metod, dina resultat, din diskussion och dina slutsatser i en skriftlig rapport. Strukturen nämnd i avsnitt *Rapportskrivning* följs.
- Lämna in kod som är korrekt och begriplig. Koden producerar alla resultat som beskrivs i rapporten.

Dimensioner för bedömning

- Hur mycket material är rapporten baserad på? Diskuteras relevant litteratur i relation till den egna studien på ett meningsfullt sätt?
- Hur komplexa är argumenten i rapporten?
- I vilken utsträckning problematiserar rapporten? Visar studenten god förståelse för studiens begränsningar?
- I vilken utsträckning används konkreta exempel?
- Hur väl genomarbetad är rapporten (struktur, språk)?
- Används lingvistiska begrepp korrekt och integreras de konsekvent i analysen?

Kriterier för Väl Godkänd

- För betyget VG krävs en omfattande diskussion. En sådan diskussion kräver en utökning av din studie med minst två av dessa tre föreslag:
 - Undersök fler konstruktioner och ord — inte bara de som föreslås i instruktionerna men gärna fler som är inspirerade av litteraturen eller egna idéer.
 - Högkvalitativ kod som visar på språkteknologisk och lingvistisk förståelse. Det kan till exempel handla om att använda lemma i stället för graford när det är mer lämpligt, hantera specialfall, och anpassa sökningen efter konstruktionens variationer.
 - Gör en mer djupgående kvalitativ analys där du även jämför med de engelska originaltexterna.

Användning av Generativ AI

- Allmän kursinformation: *"För fördjupningsuppgiften (UPG1) får dessa verktyg endast användas för att hjälpa dig att förbättra eller korrigera text som du redan har skrivit; de får inte användas för att generera helt ny text till din rapport, oavsett om du bara kopierar texten eller omformulerar den."*
- Ni får använda generativ AI för att generera och förbättra kod.
 - Men ni måste ha **full kontroll** att den verkligen gör vad ni vill!
 - Vi måste kunna köra koden utan att installera nya bibliotek. Importera alltså inga biblioteker som inte finns i Pythons standardbibliotek!

Tack för idag!