



Kursplaneöversättaren

Lina Stadell

lina.stadell@convertus.se

2017-11-13



Innehåll

- Allmänt
- Språkliga resurser
- Översättningsprocessen
- Översättningsproblem
- Stavningskontroll

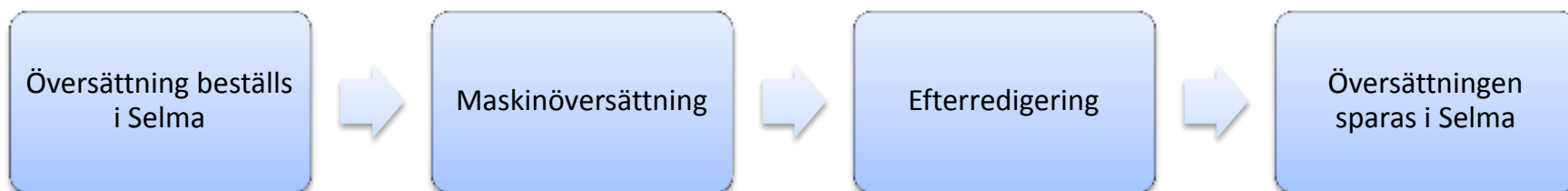


Allmänt

- Bygger på egenutvecklad översättningsteknologi som arbetats fram under många år
- Teknologin är baserad på grammatisk analys och manuella översättningsregler
- Specialanpassat till språkbruket i akademiska kursplaner
- Förutsätter manuell granskning, redigering och godkännande före publicering



Användning





Språkliga resurser

- Lexikon
- Grammatik
- Översättningsminne



Lexikon

1. Ämnesspecifika lexikon
2. UHR:s svensk-engelska ordbok
3. Allmän utbildningsterminologi
4. Allmänspråkliga lexikon

Lexikonets olika delar är hierarkiskt ordnade. Ämnesspecifika lexikon har högst prioritet och det allmänspråkliga lexikonet har lägst.

Totalt cirka 60 000 lexikonenheter



Språkdomäner

- Varje kursplan tilldelas en språkdomän vid översättning (t.ex. kemi, medfarm)
- Språkdomänen avgör i vilken ordning ämnesspecifika lexikon används
- Språkdomän väljs automatiskt baserat på institutionstillhörighet eller kurskod

Grammatik

- Handskrivna regler i ett speciellt format
- Svensk grammatik (ca 1000 regler)
 - satsanalys av den svenska meningen
- Översättningsgrammatik (ca 4000 regler)
 - överför den grammatiska strukturen till målspråket
- Engelsk grammatik (ca 1000 regler)
 - genererar målspråksmeningen



Översättningsminne

- I översättningsminnet sparas alla meningar som granskats och godkänts av översättarna.
- De lagras som par av källspråksmening med motsvarande målspråksmening.
- Minnet är uppdelat i språkdomäner.
- Om en ny redigering görs av en översättning, skriver den över den tidigare varianten i minnet.



Översättningsprocessen

Meningssegmentering



Minnessökning



Maskinöversättning



Meningssegmentering

- Kursplanen delas upp i mindre delar, s.k. översättningssegment
- Ett översättningssegment kan vara t.ex. en mening, en rubrik, ett element i en punktlista
- Uppdelningen sker utifrån formella kriterier som skiljetecken samt stor/liten bokstav
- En lista med förkortningar förhindrar felaktig segmentering



Minnessökning

- För varje översättningssegment görs en sökning i översättningsminnet.
- Om sökningen leder till en *exakt* träff, så görs ingen maskinöversättning – översättningen hämtas direkt från minnet.
- Minnessökningen görs i första hand i den aktuella språkdomänen, i andra hand söks resten av minnet igenom.



Maskinöversättning

- Lexikonuppslagning
 - Ord som inte hittas återges på svenska
- Grammatisk analys
 - Identifierar ordklasser och böjningsinformation
 - Identifierar satsled och deras funktion
 - Om den fullständiga analysen misslyckas görs delanalyser i stället.
- Översättning
 - Kontextuella regler kan ändra översättningen



Översättningsproblem

- Flertydighet i källspråket
- Val av preposition
- Böjning och ordföljd

Flertydighet

- genom -> genome/through
 - Lösning: grammatisk analys
- tal -> number/speech
 - Lösning: lexikonhierarki, kontextuella grammatikregler
- tal -> number/numbers
 - Lösning: grammatisk analys



Val av preposition

- Val av preposition
 - på -> on/of/in/to/at...
- En översättning i lexikonet
 - på -> on
- Grammatikregler som ändrar översättningen i vissa kontexter
 - exempel på -> example of



Böjning och ordföljd

- Den grammatiska analysen ger rätt böjning och ordföljd i översättningen.
- Om systemet inte lyckas göra en fullständig analys av meningen kan satsled hamna på fel plats och böjas felaktigt.

Efter bearbetning baserad på återkopplingen
examineras slutversionen av uppsatsen.

After processing based on the feedback
be examined the final version of the thesis.



Stavningskontroll

- En välskriven svensk text ger bättre översättning.
- Kontrollera stavning innan översättning
 - Stavfel, ihopskrivningar, särskrivningar
 - Stavfel gör att ord inte hittas i lexikon, vilket kan leda till sämre översättning av hela meningen.