

Hur svårt är det att analysera bulgariska?

Från skolgrammatiken känner man igen ett relativt enkelt sätt att analysera meningar i språket. Det gäller t.ex. att hitta vem som är subjekt (utförare av handlingen) och vem som är objekt (föremål för handlingen). Generellt sett försöker man hitta relationer mellan orden i en mening och ge dessa relationer olika etiketter. Svetoslav Marinovs doktorsavhandling handlar om hur man kan skapa dataprogram gällande bulgariska som genererar avancerade grammatiska analyser.

En sk dependensanalys av meningar kan man göra automatisk genom att man skapar en *parser*. En parser är ett dataprogram där man kan mata in en eller kanske flera tusen meningar och få ut varje mening med dess grammatiska analys. Vanligtvis behöver parsern en grammatik eller explicita regler på hur orden kan kombineras med varandra i tillägg till en instruktion, som anger hur programmet ska gå till väga för att skapa den önskade analysen. I stället för att använda explicita regler finns det möjlighet att parsern lär sig grammatiken från en samling meningar som redan är analyserade av lingvister, den så kallade trädbanken. En sådan metod för inlärning och parsning kallas för datadriven.

Marinovs doktorsavhandling handlar om hur man kan skapa en parser för bulgariska som returnerar dependensanalyser. För att göra detta används datadrivna metoder i samband med en trädbank, där meningarnas analyser är dependensbaserade. Ursprungligen har analyserna i trädbanken ett helt annat, icke-dependensbaserat format.

Avhandlingen tar också upp hur man automatiskt kan skapa det nödvändiga formatet från ett annat. Slutligen testas och utvärderas parsern. Marinov har tittat på konstruktioner och fenomen i språket som gör det mycket svårt och ibland omöjligt för parsern att korrekt analysera vissa meningar.

En sådan studie av dependensparsning för bulgariska ökar inte bara förståelsen av själva språket utan ger också insikt i hur man kan förbättra metoder för att automatisk analysera uttryck i språk med friare ordföljd.

Avhandlingens titel: **Dependency-Based Syntactic Analysis of Bulgarian.**

Disputationen äger rum fredagen den 9 januari 2009 kl. 10.00

Plats: Sal T 307, Arkeologen, Olof Wijksgatan 6

Opponent: Dr Adam Przepiorkowski, Warszawa

För ytterligare information kontakta Svetoslav Marinov, mobiltel. 0733-24 86 83,

e-post: sedalti99@yahoo.com

Hemsida: www.gslt.hum.gu.se/~svet

Avhandlingen kan beställas från Institutionen för filosofi, lingvistik och vetenskapsteori, tel. 031-786 19 24 och finns även tillgänglig digitalt:

www.gslt.hum.gu.se/dissertation/marinov.pdf