

# Instruktioner för attityduppmärkning av svensk text med WebAnno

Kulturomikprojektet  
(Lars Borin, Jacobo Rouces, Nina Tahmasebi, Stian Rødven Eide)  
Språkbanken, Inst. för svenska språket  
Göteborgs universitet

## 1 Uppgiften: Attityduppmärkning av svensk text

### 1.1 Bakgrund

Attityduppmärkning av text (eng. *sentiment analysis* eller *opinion mining*) är en textteknologi med många tillämpningar inom språkteknologi och andra områden. Dock finns inte verktyg och resurser för den typen av uppmärkning för mer än en handfull språk. I Språkbanken utarbetar vi nu verktyg för attityduppmärkning av svensk text, ett arbete som förutsätter tillgång till en relativt stor textmängd med pålitlig sådan uppmärkning, vilket i praktiken innebär att uppmärkningen måste göras manuellt.

Den arbetsuppgift som beskrivs i det här dokumentet är just sådan manuell uppmärkning av ett antal texter hämtade ur tidningar och diskussionsforum på internet. Eftersom 2018 är ett valår har texturvalet gjorts med tanke på en av de dominerande politiska frågorna inför detta val, nämligen invandring/migration, och de inställningar till den frågan som kommer fram i texterna. Därför ingår en deluppgift där vi särskilt ber dig att markera textförfattarens attityd i denna fråga, ifall den tar sig uttryck i texten.

### 1.2 Attityd och hur man känner igen den i text

#### 1.2.1 Attityduttryck

De uttryck vi är intresserade av förmedlar någons – positiva eller negativa – *värdering* av något eller *känsla* inför något. I det enklaste fallet handlar det om explicita värdeomdömen (här och nedan markerar fetstil värdeomdömen i exemplen):

- (1) När hon föddes **var allt underbart**.
- (2) Att ha "working poor" **är skamligt**, Löfven.
- (3) Nej, **tyvärr kan de inte** heller denna gång **operera** Celin

Oftast är värdeomdömena dock mer subtilt uttryckta och det kräver en god träning i språklig analys för att man ska se dem.

Det är vanligt att ett val mellan denotativa synonymer kan användas för att signalera (olika grader av) attityd genom att synonymerna har olika konnotationer. Några exempel:

<b>Adjektiv(fraser):</b>	<b>Substantiv(fraser):</b>	<b>Verb(uttryck):</b>
<i>försiktig med pengar</i>	<i>intern</i>	<i>X är ofta här</i>
<i>sparsam</i>	<b><i>fängelsekund</i></b>	<b><i>X hänger här hela tiden</i></b>
<i>ekonomisk</i>		
<b><i>snål</i></b>	<i>läkare</i>	<i>anklaga</i>
	<b><i>kvacksalvare</i></b>	<b><i>skuldbelägga</i></b>
<i>varnande</i>		
<b><i>alarmistisk</i></b>	<i>ort</i>	<i>röra (vid)</i>
	<b><i>håla</i></b>	<b><i>tafsa (på)</i></b>
<i>stark/stor/kraftig</i>		
<b><i>brutal</i></b>	<i>vovve</i>	<i>försenas/bli försenad</i>
	<i>hund</i>	<b><i>söla</i></b>
	<b><i>byracka</i></b>	
		<i>spilla</i>
	<i>särart</i>	<b><i>söla</i></b>
	<b><i>utanförskap</i></b>	
		<i>säkerställa</i>
		<b><i>värna</i></b>

I den här annoteringsuppgiften ska ni markera/ange tre sorters information:

- 1: själva attityduttrycket, dess polaritet (positivt eller negativt) och dess intensitet (positivt/negativt, mycket positivt/negativt eller – i undantagsfall – neutralt)
- 2: (före)målet för attityden
- 3: vem som har attityden (källan)

### 1.2.2 Hur avgör man att något är ett attityduttryck?

Ett av de bästa testen på att man har att göra med ett språkligt uttryck för attityd i vår mening är ifall någon av följdfrågorna

*Så X tycker (bra) om (vad) Y (gör)?*

*Så X tycker inte/illa om (vad) Y (gör)?*

naturligt följer av det uttryck som man funderar över. Exempel:

(4) *Just nu är jag rasande på svensk sjukvård.*

⇒ *Så du tycker illa om svensk sjukvård (just nu)?*

⇒ *\*Så du tycker bra om svensk sjukvård (just nu)?*

- (5) *alla berörda samarbetar nära och prestigelöst*  
⇒ \*Så du tycker illa om (vad) alla berörda (gör)?  
⇒ Så du tycker bra om (vad) alla berörda (gör)?

Detta test är inte alltid naturligt utan kan bli krystat ibland, och då kan man istället undersöka om en parafraas eller definition av det uttryck som man ska bedöma naturligt innehåller en klart värderande betydelse, som *god/bra/bättre/bäst, dålig/illa/sämre/sämst, (o)behaglig, rätt, fel*, etc. Exempel:

- (6) *Men det är ändå en sak som grämer mig djupt.* ⇒ 'få att känna sig illa till mods'  
(7) *Vi är många som ser hur intoleransen ökar i vårt samhälle.* ⇒ 'det att tänka illa om andra'  
(8) *Det blir extra genant i sammanhanget* ⇒ 'skapa en obehaglig känsla'  
(9) *SFI-eleverna hamnar som grupp i underläge* ⇒ 'hamna i en sämre situation'

Ytterligare några användbara test nämns av Hedquist (1978, 25ff).

### 1.3 Hur identifierar man (före)målet för attityden?

När man ska identifiera föremålet för attityden finns det två huvudfall:

- 1: **Explicita attityduttryck**, där attityduttryckets semantik avgör vad som ska uppfattas som föremålet (och även källan). I exempel (4) är *svensk sjukvård* (ett oblikt objekt eller prepositionsobjekt) föremålet och *jag* (subjektet) är källan, medan det i exempel (6) är *mig* (objektet) som är källan och *en sak* (subjektet) är föremålet.
- 2: **Attityd uttryckt genom konnotation**, där det typiskt är ett subjekt (eller agent) som blir föremålet i verbala attityduttryck (inklusive predikativer) – som i exempel (1–2) och (5) – och det substantiviska huvudordet om attityduttrycket är ett adjektivattribut (*en hal försäljare*). Ifall attityduttrycket är ett substantiv eller en substantivfras som inte används predikativt, kan föremålet sammanfalla med attityduttrycket (*Deras byracka har grävt upp hela gräsmattan*).

I de fall där texten har ett **attitydbärande satsadverbial i en neutral sats**, där föremålet skulle kunna sägas vara hela den sats där satsadverbialen står (se exempel 3), ber vi dig att när det går ange ett attityduttryck som omfattar attitydbärande satsadverbial, negation och hjälp- och huvudverb, som den fetstilta markeringen i exempel (3), och att ange attitydvärde för hela uttrycket. Detsamma gäller negeterade attityduttryck: Inkludera negationen i uttrycket och ange attitydvärde för helheten (som i undantagsfall skulle kunna bli neutralt p.g.a. negationen).

## 2 Annoteringsinstruktioner

Här följer nu instruktioner för hur attitydannoteringen ska utföras med hjälp av WebAnno, ett webbaserat annoteringsgränssnitt för språkliga material.

Om du har frågor om annoteringen – om gränssnittet eller den språkliga analysen – kan du använda den epostlista som vi har satt upp för annoterarna, med adressen `absa-annotation@listserv.gu.se`.

Du bör ha fått ett meddelande om att du har blivit medlem av listan, men hör av dig ifall detta inte skulle ha hänt.<sup>1</sup>

Allmän information om WebAnno finns publicerade på nätet (se WebAnno, 2016), men i princip bör detta dokument och epostlistan räcka för att du ska kunna utföra annoteringsarbetet.

## 3 WebAnnos gränssnitt

I WebAnnos gränssnitt är all information på engelska.

Du hittar inloggningssidan här: <https://spraakbanken.gu.se/tools/webanno>

Logga in med det användar-id och lösenord som du har fått från oss per epost.

På den sida du då kommer till, klicka på ”Annotation”. I fönstret med titeln ”Open document” kan du välja bland ett antal dokument som du ska annotera.

När du öppnar ett dokument hamnar du i annoteringsgränssnittet som har tre paneler:

- Den vänstra panelen visar dokumentets text.
- Den högra panelen visar den annotation du just arbetar med samt menyer för att välja annotationstyp, etc.
- Överst finns en navigations- och inställningspanel. Under ”Document” kan du navigera mellan olika dokument. Under ”Page” navigerar du fram och tillbaka i det öppna dokumentet. Under ”Settings” kan du ställa in sidstorleken, alltså hur många textrader en sida innehåller. När du är klar med ett dokument trycker du på ”Finish”. Pågående annoteringsarbete sparas automatiskt så att du kan stänga ett dokument och sedan öppna det igen för att fortsätta med annoteringen.

Texten i dokumentpanelen har rader som inleds med ”#” och som fungerar som ”platshållrar” eller avskiljare, representanter för det textparti som börjar under dem, så att du kan lägga till annoteringar för större delar av ett dokument. Platshållrarna har information om vilken typ av textparti som följer:

- ”#D” för hela dokumentet
- ”#T” för dokumentets titel, som kommer omedelbart nedanför ”#T”. Somliga dokument kan sakna titel.
- ”#P” för det stycke som följer omedelbart nedanför.

Varje platshållarrad avslutas med ”IMM”, som ska användas vid annoteringen av författarens (eventuella) attityd till invandring, se avsnitt 4.2 nedan.

---

<sup>1</sup>Du kan även själv hantera ditt medlemskap i listan på URL:en <https://listserv.gu.se/sympa/arc/absa-annotation>. Om inget verkar fungera kan du också skriva till [lars.borin@svenska.gu.se](mailto:lars.borin@svenska.gu.se). **Observera att det du skriver till epostlistan skickas till alla listans medlemmar.**

## 4 Annotationer

Annotationerna som du ska lägga till dokumenten är av två slag, *textannotationer* och *textpartiannotationer*.

### 4.1 Textannotationer

Textannotationer knyts som namnet antyder direkt till dokumentets text. Vid textannoteringen spelar platshållarraderna (de rader som inleds med "#") ingen roll.

Det finns två strukturella typer av annotationer:

- **(Text)spann**<sup>2</sup> spänner över ett sammanhängande stycke text. Spann markeras på det förväntade sättet, genom att du klickar med musen och, muspekaren med knappen nedtryckt och släpper upp knappen när det önskade spannet har markerats. Ett spann kan förses med vissa särdrag/egenskaper, exakt vilka beror på typen av spann.
- **Relationer** visualiseras som pilar mellan två spann. De skapas genom att du klickar på utgångspunkten (ett spann), håller ner musknappen och drar pekaren till målet (ett annat spann) och släpper upp knappen.

Avsnitt 4.1.1 beskriver de olika typer av spann som ska annoteras. Avsnitt 4.1.2 beskriver relationerna. Avsnitt 4.1.3 behandlar ytterligare aspekter av annoteringen.

#### 4.1.1 Textspann och spanntyper

För denna annoteringsuppgift har vi definierat två typer av textspann ("Layer"): SENTIMENT EXPRESSION och SOMETHING.

**SENTIMENT EXPRESSION** Ett spann för ett eller flera ord som uttrycker (direkt eller indirekt) en attityd. Denna typ kan tilldelas följande egenskaper:

- **sentiment**: Ett siffervärde mellan 1 och 5, som anger hur positiv eller negativ den uttryckta attityden är. Denna annotation är obligatorisk och måste fyllas i först. Värdena betyder:
  - 1 : mycket negativ
  - 2 : negativ
  - 3 : neutral
  - 4 : positiv
  - 5 : mycket positiv

(Det neutrala värdet kommer i praktiken inte att användas så mycket, eftersom uppgiften i huvudsak är att annotera uttryck med positivt eller negativt attitydvärde.)

- **source**: Attitydens källa, den som har attityden, t.ex. *Många kommuner* i exemplet i figur 1. Denna kan anges först efter det att ett attitydspann har definierats och dess "sentiment"-värde har fyllts i. Då dyker en ruta som heter "arguments" upp automatiskt, med ett fält för "source" och ett för "target" (se nästa punkt). Klicka på knappen <Click to activate> och markera sen ett spann i texten, vilket automatiskt blir av typen SOMETHING och kopplas till attityduttrycket som dess "source".

---

<sup>2</sup>I vårt fall är textspannen ytterligare begränsade till att bara kunna omfatta hela textord (eller skiljetecken).

- **target:** Objektet/föremålet för attityden, t.ex. *fyrkantiga regelverk och en statlig arbetsmarknadspolitik som förhindrar en individuell anpassning och lokala behov* i figur 1. Det definieras på samma sätt som ”source”.
- **ironi:** Om du anser att attityduttrycket används ironiskt, bocka då i rutan ”irony” (precis ovanför ”sentiment”-menyn).

**SOMETHING** Denna typ av spann används för att markera tre sorters språkliga uttryck:

- 1: det aktuella attityduttryckets källa
- 2: det aktuella attityduttryckets objekt/föremål
- 3: länkar i koreferenskedjor



Figur 1

Arbetsgången vid annoteringen blir att du först markerar ett SENTIMENT EXPRESSION-spann och därefter dess källa och föremål. Observera att du först kan behöva välja spanntyp (”Layer”) i menyn överst i högerpanelen.

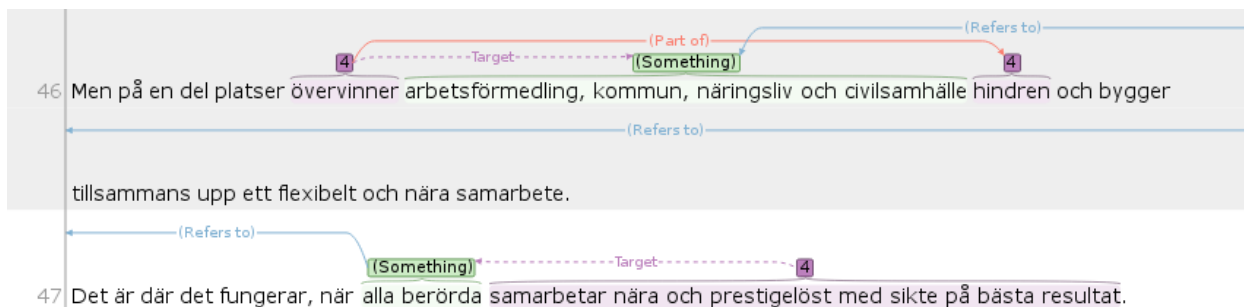
#### 4.1.2 Relationer

Ibland är det textspann som fungerar som direkt källa eller föremål för ett attityduttryck ett anaforiskt (eller kataforiskt) uttryck, typiskt ett pronomen, men ibland även t.ex. ett överordnat begrepp. I såna fall innehåller texten även ett uttryck som syftar direkt på referenten för källan eller föremålet. Då behöver du lägga till en koreferenslänk – en pil med etiketten ”Refers to” – mellan det direkta föremålet för attityden och dess antecedent. Detta görs på följande sätt:

- Annotera först antecedenten (det som syftar direkt på referenten) med ett spann av typen SOMETHING.
- Klicka sedan på pronomenspannet, håll nere musknappen, dra markören till antecedentspannet och släpp musknappen. Nu ska en pil dyka upp med etiketten ”Refers to” som går mellan de två spannen.
- För att fortsätta attitydannoteringen, byt spanntyp i rullgardinsmenyn (”Layer”) högst upp i den högra panelen.

Ett exempel finns i figur 2.

En annan typ av relation används för att hantera diskontinuerliga attityduttryck. Ett exempel syns i figur 3, där attityduttrycket i mening 23 är *leder till stora ... påfrestningar*, vars delar är förenade av en pil med etiketten ”Part of”. Att skapa en sådan annotering är enkelt: Du definierar först de två (eller flera) spannen och klickar och drar sen en pil från ett av dem till ett annat. Pilen får automatiskt etiketten ”Part of”.

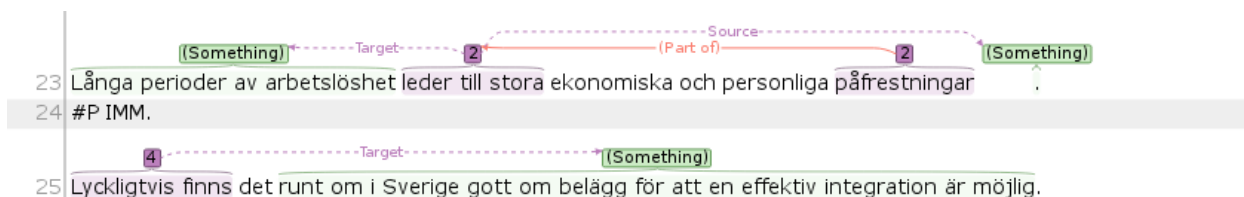


Figur 2

### 4.1.3 Diverse

I tillägg till de enkla typerna av attitydannotering som beskrivs ovan, finns en rad mer komplicerade fall samt information som är nyttig att känna till:

- Om det är klart att texten uttrycker ett föremål för en attityd, men det inte går att klart ange vilket uttrycket för själva attityden är i texten, markera då skiljetecknet i slutet av den aktuella meningen som ett spann av typen SENTIMENT EXPRESSION.
- När attitydkällan är textens författare, men detta inte är explicit uttryckt i texten (med pronomenet "jag" eller dylikt), så ska fältet "source" lämnas tomt. Om källan däremot är någon annan än artikelförfattaren, men den inte är uttryckt, så ska "source"-spannet annoteras på det avslutande skiljetecknet i den aktuella meningen. Det är viktigt att dessa två typer av icke-uttryckt källa skiljs åt konsekvent. Se figur 3, där mening 23 har en icke-uttryckt källa som inte är artikelförfattaren, medan källan i mening 25 är författaren.



Figur 3

- Föremålet för attityden kan lämnas ospecificerat ifall det inte uttrycks på något sätt i meningen.
- Om källan eller föremålet består av flera icke-sammanhängande delar, kan dessa läggas till med hjälp av menyn "Add" i rollformuläret för SENTIMENT EXPRESSION.
- Försök att göra både källor och föremål maximala, t.ex. hela nominalfraser (inklusive attribut och relativsatser; se figur 1 och 2).
- Om det inte känns onaturligt, försök att inkludera satsens verb i attityduttrycket (flera exempel i figurerna).
- På grund av att vi behöver annotera så stor mängd texter samlas texterna in helt automatiskt, utan manuellt efterarbete, och på grund av uppgiftens natur hämtas texterna från källor där det är sannolikt att vi kommer att finna attitydladdat språk (inklusive åsikter om invandring), nämligen debattartiklar och ledare i tidningar samt inlägg i online-diskussionsforum. Därför kan det hända att enstaka texter kan innehålla åsikter och/eller språk som är stötande. Det

är viktigt att de verktyg vi utarbetar klarar av att identifiera även sådan text, och vi önskar därför helst att du annoterar även dessa fall. Om du inte skulle vilja det, accepterar vi givetvis din önskan, men ber dig då att meddela detta till `lars.borin@svenska.gu.se`.

## 4.2 Textpartiannotationer

I tillägg till textannotationerna som beskrevs ovan, vill vi även att du i förekommande fall annoterar hela textpartier med avseende på författarens attityd till invandring. En delmängd av texterna har sökts fram med hjälp av söktermer avsedda att hitta dokument på temat immigration.

Denna annoteringsuppgift har två aspekter:

- 1: Även om en del av dokumenten uttryckligen har sökts fram så att de ska handla om invandring, så kan sökningen ändå ha gett ifrån sig några dokument som inte har det temat. Omvänt kan det gömma sig någon text om invandring bland de dokument som i första hand har valts ut för att representera andra teman. Därför ska du i varje dokument som handlar om invandring – **och endast i dokument som handlar om invandring** – annotera ”IMM” på första raden (“#D IMM”) som ett spann av typen SENTIMENT EXPRESSION och ange dess attitydvärde (inklusive neutralt, om texten som helhet håller sig neutral till detta tema).
- 2: Oberoende av detta, ber vi dig på samma sätt annotera varje ”IMM” där det textparti som det inleder handlar om immigration. I princip kan ett sånt textparti dyka upp även i ett dokument som inte (huvudsakligen) handlar om invandring och omvänt kan en text om invandring innehålla partier med något annat tema. Åter ska hela attitydskalan användas, alltså inklusive ’neutral’.

I denna uppgift behöver du inte ange ”source” (som ju alltid är författaren) och heller inte ”target” (som ju per definition måste vara invandring). Observera att en text kan innehålla partier som *beskriver* en negativ attityd till något fast författaren uppenbarligen inte delar den attityden. Textpartiuppmärkningen ska alltid ange *författarens* attityd.

## Referenser

Hedquist, Börje. 1978. *Emotivt språk: En studie i dagstidningars ledare*. Umeå: Inst. för nordiska språk, Umeå universitet. <http://umu.diva-portal.org/smash/get/diva2:631135/FULLTEXT01.pdf>.

WebAnno. 2016. <https://webanno.github.io/webanno/releases/3.3.5/docs/user-guide.html>.