

Synonymutvärdering:
Synonymlexikon slår ordvektormetod

Anna Westin
Augusti 2018

1 Abstrakt

En synonymutvärdering utfördes genom en onlinenkät, och resultatet blev att det nuvarande synonymlexikonet vi använder idag presterade bäst i testet, men inte tillräckligt bra och vidare arbete kommer behövas på att förbättra det.

2 Introduktion

Ett tillvägagångssätt för att göra en text mer lättläst är att ge synonymförslag. Idag på TeCST-hemsida används detta redan, men det finns ett missnöje över synonymerna som genereras och därför har man sökt efter en ny källa för synonymer.

Jag har gjort en online enkät för att utvärdera folks åsikter om synonymer från tre synonymgeneratorer. Deltagarna blev presenterade med fyra påståenden om synonymerna de genererat, och fick sedan svara ”Stämmer” eller ”Stämmer inte”. Den första synonymgeneratorn tar sina synonymer från Folkets synonymlexikon, och de två andra synonymgeneratorerna bygger på Goran Glavaš och Sanja Štajners metod som har beskrivits i sin artikel (Glavaš and Štajner 2015)

3 Metod

3.1 Deltagare

Studien bestod av 56 deltagare. Deltagarna bestod av facebook-användare i åldern 23-66 år. Samtliga hade inga läs-och skrivsvårigheter. 83.9% svarade att de hade en avslutad utbildning på eftergymnasiellnivå 3 år eller mer, och 16.1% svarade att de hade en avslutad gymnasieutbildning. 66.1% arbetar heltid eller deltid, 25% studerar, och 8.9% svarade ”inget av de ovanstående” på frågan om vad de gör idag.

3.2 Material

3.2.1 Light-LS

Goran Glavaš och Sanja Štajners metod bygger på textförenkling av en hel text genom att byta ut ord mot enklare synonymer. Metoden utför regelbaserade lexikala förenkling och i deras artikel evaluerade de Light-LS genom tre olika metoder och kom fram till att Light-LS presterade bäst inom alla tre.

Sanja Štajners är känd för att alltid ha open source på all kod från sina artiklar, och koden hon använde

i sin artikel fanns därför lätt tillgänglig på Github. URLen till biblioteket finns här: <https://github.com/codogogo/lightls>

Kodens output ger ut två filer där ena filen innehåller den förenklade texten och den andra filen innehåller alla ord som blivit utbytta samt deras originalord. Det är denna andra fil som skulle användas till synonymförslag på hemsidan om den visat sig vara det bättre alternativet.

Metoden avgör om ett ord ska bytas ut till ett enklare ord beroende på:

- Komplexiteten utav ordet från inputen.
- Komplexiteten utav ersättningsordet (dvs om det är enklare).
- Likheten mellan inputordet och ersättningsordet.
- Likheten mellan ersättningsordet i kontexten med inputordet.

Metoden kräver två filer för att den ska fungera. Den första filen ska innehålla unigram ordfrekvenser man måste skapa själv. Den andra filen ska innehålla förtränade ordvektorer. Beroende på vilken ordvektor man använde skiljde sig synonymerna avsevärt mycket. Därför fick denna metod i enkäten utvärderas som två olika synonymgeneratorer, där den ena använde sig av ordvektor kallat Swectors som är framtagen på Linköpings Universitet. Den andra använde sig av en ordvektor tränad på Wikipedia genom att de använt FastText. Vektorerna är framtagna genom att använda en skip-gram model beskriven i Bojanowski et al. (2016) med default parametrar. Korpusen kommer i en fil som heter Wiki.sv.vec, men för enkelhetens skull kommer den därför kallas Wiki i resten av denna rapport.

Det var även valbart om man ville lägga till en fil för stoppord. Denna fil kom också från Evelina Rennes efter att vi diskuterat hur många dåliga stoppordslistor det finns där ute, men hon hade hittat en som vi båda ansåg var bra.

Metoden innehåller även fem stycken parametrar man kan ändra på. Dessa parametrar ska anpassas till ordvektorn som används och parametrarna

som användes till de två ordvektorna står under utförande. Parametrarna var följande:

- Tc: Minsta komplexiteten på ordet för att det ska övervägas till att bytas till ett enklare ord. Värdet kan vara mellan 0.0 (vilket innebär att alla ord övervägs för förenkling) och 1.0 (inga ord övervägs för förenkling).
- Nc: Antal kandidater av ersättningsord som ska övervägas till förenklingen. Alla kandidater kommer från vad ordvektorn anser är liknande ord till inputordet. Värdet kan vara mellan 1-10 stycken.
- St: Minst cosinus likheten i ordvektorn mellan inputordet och ersättningsordet. Värdet kan vara mellan 0.0 (den bästa kandidaten kommer alltid ersätta ordet) och 1.0 (den bästa kandidaten kommer aldrig ersätta inputordet). Det logiska här är förstås att alltid ta 0.0, men det kan leda till att orden är semantiskt relaterade i ordvektorn men stämmer inte överrens i betydelsen (dvs huvud och ansikte kom bland annat upp som synonymer när man hade 0.0 för swectors).
- Cd: Den minsta komplexiteten för ersättningsordet. Värdet kan vara mellan 0.0 (all komplexitetsreduktion är bra) och 1 (komplexitetsreduktionen är 100% enklare).
- W (window): Storleken av kontexten, dvs rankingen av kandidaterna av ersättningsorden i förhållande till kontexten av texten.

3.2.2 Synonymlexikon

Den tredje synonymgeneratorn tar sina synonymer från Folkets synonymslexikon. Synonymerna är sammanställda i en XML-fil. Detta lexikon är det som används på TeCST-hemsida idag.

Synonymlexikonet ger tre synonymalternativ för läsaren att välja mellan. Då Light-LS metoden endast ger en synonym i nuläget valdes därför synonymer med högst siffra på variabeln level för att synonymlexikonet också endast skulle erbjuda ett synonymalternativ.

3.2.3 Textkällor

På grund av den stora skillnaden mellan synonymkvaliteten i olika textkategorier testade onlineenkäten synonymgeneratorerna på meningar från fyra olika källor;

- Dagstidningar (Webbnyheter 2012, Österbottens tidning 2012, GP 2002, 2011)
- Romaner (bonnierromaner I och II, August Strindbergs romaner)
- Texter med fackspråk (läkartidningen 2004, 2005)
- Akademiska texter (Akademiska texter – Humaniora, Akademiska texter – Samhällsvetenskap)

Samtliga korpusar kom från Språkbankens bibliotek.

För att träna unigram frekvensen av ord i Light-LS metoden användes korpusar lättlästa korpusar framtagna av Evelina Rennes, samt korpus kallat HungarianAlignment och MaximumAlignment.

3.3 Pilotstudie

Två pilotstudier gjordes med samma 4 stycken deltagare i båda. Den första pilotstudien upplevde deltagarna var för kognitivt belastat. Den bestod av 10 meningar från de fyra kategorierna nämnda i avsnitten ovan, samt 10 meningar från ytterligare en kategori kallat Populärvetenskapliga artiklar. Detta kategori bestod av meningar från tidningen Forskning och Framsteg. Deltagarna fick originalmeningen som en fråga, och svaren bestod av tre liknande meningar, men med ersatta synonymer tjockmarkerade från vardera synonymgenerator. Studien bestod av 50 sådana mening-frågor. Resultatet från denna studie blev att den var för lång och för kognitivt belastande för läsarna.

För att göra studien mindre kognitivt belastad gjordes hela konceptet om. Istället utvärderades endast en mening i taget, och deltagarna fick kryssa i två kolumner av ”Stämmer” och ”Stämmer inte” på fyra påståenden om meningen. En kategori togs bort

för att korta ner studien, samt alla längre meningar togs bort i samband med att meningsantalet för varje kategori kortades ner till 4 meningar per kategori. Denna studie gav ett bra resultat, och därför användes denna studie till huvudstudien.

Pilotstudiens resultat användes inte i huvudstudien då deltagarna endast fick prova på en del av kategorierna.

3.4 Utförande

3.4.1 Ta fram parametrar

Första steget var att anpassa parametrerna för de båda ordvektorerna. Det var utmanande att hitta rätt parametrar för Light-LS då parametrarna påverkade varandra. Även kategorin av texter på inputen påverkade då vissa parametrar funkade perfekt till en vanlig wikipediatext, men när skriptet med samma parametrar användes till exempelvis en artikeltext presterade den inte alls lika bra. Parametrarna bestämdes genom att testas fram på olika texter. De blev aldrig optimala för alla texter då studien fastnade i denna testfas länge och behövdes tas vidare. Onlineenkäten får avgöra om de togs till ett "bra nog" stadiet för att kunna användas på hemsidan.

Parametrarna anpassade för vardera ordvektor blev följande:

Table 1: Parametrar i Swectors

Tc	0.2
Nc	10
St	0.55
Cd	0.03
W (window)	5

Table 2: Parametrar i Wiki

Tc	0.0
Nc	3
St	0.0
Cd	0.03
W (window)	3

Folkets lexikon hade inga parametrar man kunde ställa in.

3.4.2 Samla ihop meningar

De tre synonymgeneratorerna var så pass olika varandra att det var svårt att hitta meningar där alla tre synonymgeneratorer genererade synonymer till meningen. Det var också nästintill omöjligt att hitta meningar där alla tre synonymgeneratorer hade synonymer till samma ord. För att kunna jämföra de olika synonymgeneratorerna med varandra behövde synonymgeneratorerna åtminstone generera synonymer inom samma mening, och det blev därför fokuset när meningar togs fram.

Meningarna slumpades fram, testades på alla synonymgeneratorer, och valdes sedan om alla synonymgeneratorer fick fram synonymer i alla meningar. Långa meningar och meningar som ansågs vara för komplicerade av deltagare i pilotstudierna togs även bort.

3.4.3 Onlinenkäten

Onlinenkäten gjordes i Google forms.

Fyra påståenden gjordes om synonymerna i meningarna. Påståendena byggde på fyra egenskaper som kommit fram i olika diskussioner som synonymer borde ha. Påståendena tjockmarkerades med vissa ord för att lätta på belastningen kognitivt för läsaren om vad vi var ute efter. Dessa fyra påståenden var:

- Alla synonymer gör **orden** i meningen **lättare att förstå**
- Meningen har inte ändrat **betydelsen** av orden som bytts ut
- Alla ord jag anser är lite **svårare att förstå** i meningen hade ett synonymförslag till sig
- De utbytta orden är ersatta av en **synonym**

Det första påståendet var för att mäta om läsarna upplevde att orden hjälpte till i förståelsen av ordet. Det andra påståendet var för att mäta om synonymerna i meningen hade ändrat kontexten eller

sammanhanget i meningen. Det tredje påståendet var för att mäta om synonymgeneratorerna lyckades fånga upp alla ord som kunde tolkas som svårare att förstå. Det fjärde och sista påståendet var för att mäta om synonymgeneratorerna hade lyckats generera vad läsarna ansåg som faktiska synonymer till orden.

En förklaring till vad alla dessa påståenden var ute efter gavs i introduktionstexten i enkäten.

Efter pilotstudierna delades onlineenkäten ut på facebook samt flera relevanta facebookgrupper.

3.4.4 Analys

Inga statistiska analyser har gjorts på datan, och den kommer bedömmas utefter om antalet svar på "Stämmer" är fler än "Stämmer inte" inom varje påstående. Detta blev resultatet efter att Google Forms excel-fil gav sitt resultatet i fel form, och alla svar hamnade i en och samma cell istället för olika celler. Datan samlades istället in genom att räkna ihop summan av alla svar, både inom kategorierna och totalen.

4 Resultat

Antalet svar på alla påståenden i alla kategorier för alla synonymgeneratorer finns nedan i Table 3, Table 4 och Table 5. Folkets synonymlexikon var det enda synonymgeneratorn folk ansåg lyckades generera faktiska synonymer. Den var även nära på att generera synonymer man ansåg inte ändrade betydelsen av orden i sig. Däremot lyckades Folkets synonymlexikon inte generera tillräckligt med synonymer i meningarna som fångade alla ord de ansåg var svårare att förstå, och inte heller generera synonymer som blev enklare för läsare att förstå. Synonymgeneratorerna Wiki och Swectors misslyckades också med att uppnå hög standard nog för att kunna få tillräckligt med positiva svar inom alla kategorier.

Table 3: Resultat för Folkets synonymlexikons prestation inom alla kategorier

Påståenden	Stämmer	Stämmer inte
Alla synonymer gör orden i meningen lättare att förstå	279	617
Meningen har inte ändrat betydelsen av orden som bytts ut	445	451
Alla ord jag anser är lite svårare att förstå i meningen hade ett synonymförslag till sig	355	541
De utbytta orden är ersatta av en synonym	493	403

Table 4: Resultat för Wikis prestation inom alla kategorier

Påståenden	Stämmer	Stämmer inte
Alla synonymer gör orden i meningen lättare att förstå	205	691
Meningen har inte ändrat betydelsen av orden som bytts ut	250	646
Alla ord jag anser är lite svårare att förstå i meningen hade ett synonymförslag till sig	277	619
De utbytta orden är ersatta av en synonym	330	566

Table 5: Resultat för Swectors prestation inom alla kategorier

Påståenden	Stämmer	Stämmer inte
Alla synonymer gör orden i meningen lättare att förstå	290	582
Meningen har inte ändrat betydelsen av orden som bytts ut	266	630
Alla ord jag anser är lite svårare att förstå i meningen hade ett synonymförslag till sig	389	507
De utbytta orden är ersatta av en synonym	347	549

Table 6: Resultat för Folkets synonymlexikon

Påståenden	Stämmer	Stämmer inte
Alla synonymer gör orden i meningen lättare att förstå	56	168
Meningen har inte ändrat betydelsen av orden som bytts ut	112	112
Alla ord jag anser är lite svårare att förstå i meningen hade ett synonymförslag till sig	67	157
De utbytta orden är ersatta av en synonym	130	94

Adademiska texter

I kategorin Akademiska texter var deltagarnas åsikter negativa för alla synonymgeneratorers prestation inom att generera lättare synonymer samt att generera tillräckligt med synonymer i meningarna för att fånga alla ord som kunde tolkas som svårare att förstå. För Folkets synonymlexikon fanns det en delad mening om synonymerna hade ändrat betydelsen av orden, medan det ansågs på de andra synonymgeneratorerna att synonymerna ändrade betydelsen. För synonymgeneratorerna Folkets synonymlexikon och Swectors var folk däremot positiva till att det var faktiskt synonymer som hade genererats, medan i synonymgeneratoren Wiki ansåg de inte att den hade lyckats generera faktiskt synonymer alltid.

Table 7: Resultat för Wiki

Påståenden	Stämmer	Stämmer inte
Alla synonymer gör orden i meningen lättare att förstå	67	157
Meningen har inte ändrat betydelsen av orden som bytts ut	56	168
Alla ord jag anser är lite svårare att förstå i meningen hade ett synonymförslag till sig	76	148
De utbytta orden är ersatta av en synonym	89	135

Table 8: Resultat för Swectors

Påståenden	Stämmer	Stämmer inte
Alla synonymer gör orden i meningen lättare att förstå	91	133
Meningen har inte ändrat betydelsen av orden som bytts ut	74	150
Alla ord jag anser är lite svårare att förstå i meningen hade ett synonymförslag till sig	95	129
De utbytta orden är ersatta av en synonym	124	100

I Table 6 ser vi att deltagarna upplevde synonymer från Folkets synonymlexikon som (1) Inte lättare att förstå, (2) Det fanns en delad mening om ifall betydelsen hade ändrats av den nya synonymen, (3) Synonymgeneratorn gav inte tillräckligt med synonymer i meningarna, och (4) Synonymgeneratorn genererade faktiska synonymer till orden i meningarna.

I Table 7 ser vi att deltagarna upplevde synonymer från synonymgeneratorn Wiki som (1) Inte lättare att förstå, (2) Synonymerna som gavs ändrade betydelsen av orden, (3) Synonymgeneratorn gav inte tillräckligt med synonymer i meningarna, och (4) Synonymgeneratorn genererade inte alltid faktiska synonymer till ordet.

I Table 8 ser vi att deltagarna upplevde synonymer från synonymgeneratorn Swectors som (1) Inte lättare att förstå, (2) Synonymerna som gavs ändrade betydelsen av orden, (3) Synonymgeneratorn gav inte tillräckligt med synonymer i meningarna, och (4) Synonymgeneratorn genererade faktiska synonymer till orden i meningarna.

Romaner

I kategorin Romaner var deltagarnas åsikter nega-

tiva för alla synonymgeneratorers prestation inom att generera lättare synonymer samt att generera tillräckligt med synonymer i meningarna för att fånga alla ord som kunde tolkas som svårare att förstå. Med Folkets synonymlexikon ansåg man att synonymerna inte hade ändrat betydelsen av orden, medan för de andra synonymgeneratorerna hade synonymerna ändrade betydelsen. För synonymgeneratorerna Folkets synonymlexikon var folk däremot positiva till att det var faktiskt synonymer som hade genererats, medan i synonymgeneratorn Wiki och Swectors ansåg de inte att den hade lyckats generera faktiskt synonymer alltid.

Table 9: Resultat för Folkets synonymlexikon

Påståenden	Stämmer	Stämmer inte
Alla synonymer gör orden i meningen lättare att förstå	100	124
Meningen har inte ändrat betydelsen av orden som bytts ut	135	89
Alla ord jag anser är lite svårare att förstå i meningen hade ett synonymförslag till sig	109	115
De utbytta orden är ersatta av en synonym	139	85

I Table 9 ser vi att deltagarna upplevde synonymer från Folkets synonymlexikon som (1) Inte lättare att förstå, (2) Synonymerna som gavs ändrade inte betydelsen av orden, (3) Synonymgeneratorn gav inte tillräckligt med synonymer i meningarna, och (4) Synonymgeneratorn genererade faktiska synonymer till orden i meningarna.

Table 10: Resultat för wiki

Påståenden	Stämmer	Stämmer inte
Alla synonymer gör orden i meningen lättare att förstå	34	190
Meningen har inte ändrat betydelsen av orden som bytts ut	34	190
Alla ord jag anser är lite svårare att förstå i meningen hade ett synonymförslag till sig	63	161
De utbytta orden är ersatta av en synonym	57	167

tydelsen av orden, (3) Synonymgeneratoren gav inte tillräckligt med synonymer i meningarna, och (4) Synonymgeneratoren genererade inte alltid faktiska synonymer till orden i meningarna.

I Table 11 ser vi att deltagarna upplevde synonymer från synonymgeneratoren Swectors som (1) Inte lättare att förstå, (2) Synonymerna som gavs ändrade betydelsen av orden, (3) Synonymgeneratoren gav inte tillräckligt med synonymer i meningarna, och (4) Synonymgeneratoren genererade inte alltid faktiska synonymer till orden i meningarna.

Table 11: Resultat för Swectors

Påståenden	Stämmer	Stämmer inte
Alla synonymer gör orden i meningen lättare att förstå	26	198
Meningen har inte ändrat betydelsen av orden som bytts ut	27	197
Alla ord jag anser är lite svårare att förstå i meningen hade ett synonymförslag till sig	56	168
De utbytta orden är ersatta av en synonym	36	188

I Table 10 ser vi att deltagarna upplevde synonymer från synonymgeneratoren Wiki som (1) Inte lättare att förstå, (2) Synonymerna som gavs ändrade be-

Dagstidningar

I kategorin Dagstidningar var deltagarnas åsikter negativa för alla synonymgeneratorernas prestation inom att generera lättare synonymer. Swectors lyckades med att generera tillräckligt med synonymer i meningarna för att fånga alla ord som kunde tolkas som svårare att förstå, medan Folkets synonymlexikon och Wiki inte var lika lyckade inom detta. Folkets synonymlexikon lyckades med att inte ändra betydelsen av orden med sina synonymer, men Wiki och Swectors synonymer ändrade betydelsen. Folkets synonymlexikon var också den enda som lyckades med att generera faktiska synonymer till orden, medan samma åsikt inte fanns för Swectors och Wiki.

Table 12: Resultat för Folkets synonymlexikon

Påståenden	Stämmer	Stämmer inte
Alla synonymer gör orden i meningen lättare att förstå	96	128
Meningen har inte ändrat betydelsen av orden som bytts ut	130	94
Alla ord jag anser är lite svårare att förstå i meningen hade ett synonymförslag till sig	84	140
De utbytta orden är ersatta av en synonym	140	84

Table 13: Resultat för wiki

Påståenden	Stämmer	Stämmer inte
Alla synonymer gör orden i meningen lättare att förstå	50	174
Meningen har inte ändrat betydelsen av orden som bytts ut	74	150
Alla ord jag anser är lite svårare att förstå i meningen hade ett synonymförslag till sig	56	168
De utbytta orden är ersatta av en synonym	84	140

I Table 12 ser vi att deltagarna upplevde synonymer från Folkets synonymlexikon som (1) Inte lättare att förstå, (2) Synonymerna som gavs ändrade inte be-

tydelsen av orden, (3) Synonymgeneratorn gav inte tillräckligt med synonymer i meningarna, och (4) Synonymgeneratorn genererade faktiska synonymer till orden i meningarna.

I Table 13 ser vi att deltagarna upplevde synonymer från synonymgeneratorn Wiki som (1) Inte lättare att förstå, (2) Synonymerna som gavs ändrade betydelsen av orden, (3) Synonymgeneratorn gav inte tillräckligt med synonymer i meningarna, och (4) Synonymgeneratorn genererade inte alltid faktiska synonymer till orden i meningarna.

Table 14: Resultat för Swectors

Påståenden	Stämmer	Stämmer inte
Alla synonymer gör orden i meningen lättare att förstå	95	129
Meningen har inte ändrat betydelsen av orden som bytts ut	65	159
Alla ord jag anser är lite svårare att förstå i meningen hade ett synonymförslag till sig	152	72
De utbytta orden är ersatta av en synonym	82	142

I Table 14 ser vi att deltagarna upplevde synonymer från synonymgeneratorn Swectors som (1) Inte lättare att förstå, (2) Synonymerna som gavs ändrade betydelsen av orden, (3) Synonymgeneratorn gav tillräckligt med synonymer i meningarna, och (4) Synonymgeneratorn genererade inte alltid faktiska synonymer till orden i meningarna.

Texter med fackspråk

I kategorin Texter med fackspråk var deltagarnas åsikter negativa för alla synonymgeneratorers prestation inom alla påståenden.

Table 15: Resultat för Folkets synonymlexikon

Påståenden	Stämmer	Stämmer inte
Alla synonymer gör orden i meningen lättare att förstå	27	197
Meningen har inte ändrat betydelsen av orden som bytts ut	68	156
Alla ord jag anser är lite svårare att förstå i meningen hade ett synonymförslag till sig	95	129
De utbytta orden är ersatta av en synonym	84	140

Table 16: Resultat för Wiki

Påståenden	Stämmer	Stämmer inte
Alla synonymer gör orden i meningen lättare att förstå	54	170
Meningen har inte ändrat betydelsen av orden som bytts ut	86	138
Alla ord jag anser är lite svårare att förstå i meningen hade ett synonymförslag till sig	82	142
De utbytta orden är ersatta av en synonym	100	124

I Table 15 ser vi att deltagarna upplevde synonymer från Folkets synonymlexikon som (1) Inte lättare att förstå, (2) Synonymerna som gavs ändrade be-

tydelsen av orden, (3) Synonymgeneratorn gav inte tillräckligt med synonymer i meningarna, och (4) Synonymgeneratorn genererade inte alltid faktiska synonymer till orden i meningarna.

I Table 16 ser vi att deltagarna upplevde synonymer från synonymgeneratorn Wiki som (1) Inte lättare att förstå, (2) Synonymerna som gavs ändrade betydelsen av orden, (3) Synonymgeneratorn gav inte tillräckligt med synonymer i meningarna, och (4) Synonymgeneratorn genererade inte alltid faktiska synonymer till orden i meningarna.

Table 17: Resultat för Swectors

Påståenden	Stämmer	Stämmer inte
Alla synonymer gör orden i meningen lättare att förstå	102	122
Meningen har inte ändrat betydelsen av orden som bytts ut	100	124
Alla ord jag anser är lite svårare att förstå i meningen hade ett synonymförslag till sig	86	138
De utbytta orden är ersatta av en synonym	105	119

I Table 17 ser vi att deltagarna upplevde synonymer från synonymgeneratorn Swectors som (1) Inte lättare att förstå, (2) Synonymerna som gavs ändrade betydelsen av orden, (3) Synonymgeneratorn gav inte tillräckligt med synonymer i meningarna, och (4) Synonymgeneratorn genererade inte alltid faktiska synonymer till orden i meningarna.

5 Diskussion

Något som upptäcktes när man försökte samla in tillräckligt med meningar för studien var att de tre synonymgeneratorerna fångade olika ord i meningarna, samt att bara runt 1/5 av meningarna faktiskt kunde användas då de var de enda meningarna som alla synonymgeneratorer fann synonymer till. Ett förslag skulle därför vara att använda åtminstone två av synonymgeneratorerna för att fler ord i texter ska få synonymförslag till sig på hemsidan och göra det enklare för läsare att förstå ordet.

Nedan presenteras sammanfattning

5.1 Folkets synonymlexikon

Folkets synonymlexikon var den synonymgenerator som presterade bäst i denna utvärdering. Table 3 visar att folks totala åsikt om synonymgeneratorn var att den lyckades med att generera faktiska synonym till orden i fråga. Synonymgeneratorn var även några röster ifrån att lyckas helt med att generera synonymer som inte ändrade betydelsen av orden. Däremot misslyckades Folkets synonymlexikon med att generera synonymer som gjorde det enklare för läsaren att förstå orden samt att fånga alla ord de ansåg var svårare att förstå i meningarna.

Folkets synonymlexikon presterade bäst inom kategorierna Akademiska texter (se Table 6), Romaner (se Table 9) och Dagstidningar (se Table 12), men mindre bra i kategorin Texter med fackspråk (se Table 15).

Dessa resultat får det att framstå som att Folkets synonymlexikon är det bästa alternativet för hemsidan. Däremot var det missnöjet över just detta lexikon som startade en sökning efter andra alternativ. Folkets synonymlexikon skulle behöva förbättras en hel del, och förslag på detta finns under vidare forskning.

5.2 Wiki

Table 4 visar Wikis totala prestation inom alla kategorier. Synonymgeneratorns synonymer var inte om-

tyckta av deltagarna inom några påståenden, eller inom någon av kategorierna. Detta medför att Wiki kan anses ha presterat sämst av de tre synonymgeneratorerna. Anledningen till dess dåliga presterande kan bero på inställningarna av parametrarna och/eller korpusen för träningen av unigram frekvenserna kan ha varit för liten.

5.3 Swectors

Table 5 visar Swectors totala prestation inom alla kategorier. Synonymgeneratorn lyckades inte uppnå tillräckligt med positiva svar inom någon av påståenden. Däremot i Table 8) kan man se att i kategorin Akademiska texter ansåg folk att synonymgeneratorn lyckades generera faktiska synonymer till orden. I kategorin Dagstidningar lyckades synonymgeneratorns fånga alla ord deltagarna ansåg var svårare att förstå (se Table 14). Detta innebär att Swectors kan anses ha presterat nästbäst av de tre synonymgeneratorerna. Anledningarna till att synonymgeneratorn inte lyckades bättre kan bero på inställningarna av parametrarna och/eller korpusen för träningen av unigram frekvenserna borde ha varit större.

5.4 Vidareforskning

Light-LS går bara att göra bättre, till skillnad från Folkets synonymlexikon som är statisk med sina synonymalternativ. För att hitta optimala variabler för Light-LS metoden skulle man kunna träna dem i ett neuralt nätverk. Metoden skulle även kunna bli bättre om man skapade ett större korpus av lättlästa texter för den att träna unigrams frekvensen till att ha högre frekvenser av alla ord.

Det skulle också behövas en vidareforskning på hur synonymerna i Folkets lexikon kan bli bättre. Exempelvis hur den ska generera synonymer som inte ändrade betydelsen av orden, synonymer som gör det enklare för läsaren att förstå orden, och att fånga alla ord som kan anses som svårare att förstå i meningar.

5.5 Slutsats

Folkets synonymlexikon presterade bäst, Swectors näsbäst, och sämst var Wiki. Detta innebär däremot inte att Folkets synonymlexikon ska få fortsätta vara huvudkällan till synonymer på hemsidan. Det enda i sin helhet den lyckades med var att producera vad läsarna ansåg var faktiska synonymer till orden. Den lyckades även helt okej i tre av fyra textkategorier. Denna studie påbörjades av ett missnöje över just detta lexikon, och vidare forskning skulle behövas göras för att förbättra den. Light-LS är inte heller helt uteslutet som ett alternativ. I Goran Glavaš och Sanja Štajners utvärdering lyckades Light-LS mycket bra, och det kan därför snarare ha varit fel inställningar av parametrarna som skapade det stora missnöjet över de två synonymgeneratorerna hos användarna. Även när unigram frekvenserna tränades för orden kan korpusen för detta ha varit för liten. Detta är dock endast spekulationer.

Ett förslag från min sida är även att man ska försöka använda åtminstone två av synonymgeneratorerna till hemsidan, efter att mycket arbete har lagts på att förbättra dem förstås. Detta skulle medföra att fler ord fångades upp och gavs synonymförslag till.

References

- Bojanowski, Piotr, Edouard Grave, Armand Joulin, and Tomas Mikolov (2016). “Enriching Word Vectors with Subword Information”. In: *arXiv preprint arXiv:1607.04606*.
- Glavaš, Goran and Sanja Štajner (2015). “Simplifying lexical simplification: do we need simplified corpora?” In: *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 2: Short Papers)*. Vol. 2, pp. 63–68.

6 Appendix

6.1 Välkomsttext

TeCST är ett delprojekt i Diginclud och syftar till skapa ett kostnadsfritt verktyg som underlättar för webbredaktörer och skribenter att skriva texter som är lätta att förstå. Synonymer används bland annat inom detta verktyg för att underlätta ords betydelser, och vi i TeCST-projektet behöver nu hjälp med att utvärdera tre synonymgeneratorer. Du kommer bli presenterad med totalt 48 meningar som har fyra påståenden till sig, och din uppgift är att kryssa i om de påståendena stämmer eller inte.

Deltagandet är helt frivilligt och du kan avbryta din medverkan när du vill. Utvärderingen tar ungefär 20 minuter att avsluta.

Utvärderingen är indelad i fem kategorier; akademiska texter, romaner, dagstidningar, texter med fackspråk, och populärvetenskapliga artiklar. Du kommer få kolla på 12 meningar inom varje kategori.

6.2 Kontrollfrågor

Har du läs -och/eller skrivsvårigheter?

- Ja
- Nej
- Vet ej

Hur gammal är du?

Vad är din högst avslutade utbildningsnivå?

- Folk- och grundskoleutbildning
- Gymnasieutbildning
- Eftergymnasiellnivå 3 år eller mer
- Forskarutbildning
- Inget av de ovanstående

Vad sysslar du med idag? (Studerar du och arbetar samtidigt kan du kryssa i det du gör mest av, eller välja en av de två du tycker passar dig bäst)

- Arbetar heltid eller deltid
- Studerar
- Inget av de ovanstående

6.3 Introduktionstext i frågorna

Genom att svara på dessa frågor hjälper du oss utvärdera tre synonymgeneratorer. Du kommer i detta avsnitt bli presenterad med 12 stycken meningar som gått igenom de tre olika synonymgeneratorerna. Varje mening har samma fyra påståenden till sig som du ska kryssa i om de stämmer eller inte. Det finns inget

”Vet ej”-alternativ då det är vanligt att man kryssar i det istället för att ta ett beslut, vilket inte ger oss stark data.

De fyra påståenden är:

Alla synonymer gör **orden** i meningen **lättare att förstå** - Här frågar vi om ordet har bytts ut till vad du upplever som ett enklare ord.

Meningen har inte ändrat **betydelsen** av orden som bytts ut - Här frågar vi om kontexten/sammanhanget i meningen är detsamma eller om den ändrat betydelse av meningen.

Alla ord jag anser är lite **svårare att förstå** i meningen hade ett synonymförslag till sig - Här frågar vi om synonymgeneratören för meningen har lyckats fånga alla svårare ord eller om den missat (vad du upplever som) svårare ord.

De utbytta orden är ersatta av en **synonym** - Här frågar vi om ordet har bytts ut till synonymer, och inte vad som verkar som slumpmässiga ord.

Ord med synonymförslag är överstrukna ~~såhär~~, och den föreslagna synonymen till ordet är understruket och kommer direkt efter det överstrukna ordet. Dvs: ~~Exempelord~~ Synonym

6.4 Svartalernativ

Alla synonymer gör **orden** i meningen **lättare att förstå**

Meningen har inte ändrat **betydelsen** av orden som bytts ut

Alla ord jag anser är lite **svårare att förstå** i meningen hade ett synonymförslag till sig

De utbytta orden är ersatta av en **synonym**

6.5 Meningarna som gjordes bedömning på

Mening med synonymer genererad av Folkets synonymlexikon:

Magnus von Platen berättar hur deltagarna i en kongress 1972 kring temat ”modern skandinavisk lyrik ”fick i uppdrag mission att tolka Göran Palms dikt ”Havet ”.

Mening med synonymer genererad av Wiki:

Magnus von Platen berättar hur deltagarna i en kongress 1972 kring temat ”modern skandinavisk ~~lyrik~~ diktning ” fick i uppdrag att tolka Göran Palms dikt ” Havet ”.

Mening med synonymer genererad av Swectors:

Magnus von Platen berättar hur deltagarna i en kongress årsmöte 1972 kring temat " modern skandinavisk nordisk lyrik litteratur " fick i uppdrag att tolka Göran Palms dikt text " Havet " .

Mening med synonymer genererad av Folkets synonymlexikon:

Dels för att det analytiskt blev nästan omöjligt att skilja på när en individs utsatthet berodde på till exempel prov rasism och när det berodde på sexism.

Mening med synonymer genererad av Wiki:

Dels för att det analytiskt matematiskt blev nästan hela omöjligt att skilja på när en individs utsatthet berodde på till exempel rasism och när det berodde på sexism.

Mening med synonymer genererad av Swectors:

Dels för att det analytiskt blev nästan omöjligt att skilja på när en individs elevs utsatthet berodde på till exempel rasism och när det berodde på sexism.

Mening med synonymer genererad av Folkets synonymlexikon:

Tabellen ovan avser dem som svarat på frågorna och avseende gällande läsning av brott, endast de som läst om brott i dagspressen minst någon gång i veckan.

Mening med synonymer genererad av Wiki:

Tabellen ovan nedan avser dem som svarat på frågorna och avseende läsning av brott, endast de som läst om brott i dagspressen minst någon gång i veckan.

Mening med synonymer genererad av Swectors:

Tabellen ovan avser dem som ~~svarat~~ röstat på frågorna och avseende läsning av brott, endast de som läst om brott i ~~dagspressen~~ tidningarna minst någon gång i veckan

Mening med synonymer genererad av Folkets synonymlexikon:

Ombuden menar att barn lämnas utan ~~tillräcklig~~ nog tillsyn skötsel och att vuxna inte visar den ~~omtanke~~ bry sig mot de egna barnen som de förtjänar.

Mening med synonymer genererad av Wiki:

Ombuden menar att barn lämnas utan tillräcklig tillsyn och att ~~vuxna barn~~ vuxna barn inte visar den ~~omtanke~~ respekt mot de egna barnen som de förtjänar.

Mening med synonymer genererad av Swectors:

Ombuden menar att barn lämnas utan tillräcklig tillsyn och att vuxna inte visar den ~~omtanke~~ omsorg mot de egna barnen som de förtjänar.

Mening med synonymer genererad av Folkets synonymlexikon:

Konungens makt att i brottmål giva nåd var ännu ~~inskränkt~~ uppgörelse av målsägandens rätt till hämnd eller ~~förlikning~~ enkelspårig.

Mening med synonymer genererad av Wiki:

Konungens makt att i ~~brottmål~~ domstolar giva nåd var ännu inskränkt av målsägandens rätt till ~~hämnd~~ eller förlikning.

Mening med synonymer genererad av Swectors:

Konungens makt att i brottmål giva nåd var ännu inskränkt av ~~målsägandens~~ kvinnans rätt till hämnd eller överenskommelse.

Mening med synonymer genererad av Folkets synonymlexikon:

Publiken blev ~~nämligen~~ alltså utäcklad, vände ryggen och spottade bakut.

Mening med synonymer genererad av Wiki:

Publiken blev nämligen eftersom utäcklad, vända ryggen huvudet och spottade bakut.

Mening med synonymer genererad av Swectors:

Publiken blev nämligen utäcklad, vände ryggen och ~~spottade~~ slängde bakut.

Mening med synonymer genererad av Folkets synonymlexikon:

Men ~~eftersom~~ på grund av Sloane förlorade ganska friskt, ljög han ~~naturligtvis~~ självfallet för sin fru.

Mening med synonymer genererad av Wiki:

Men eftersom Sloane förlorade ~~ganska~~ runt friskt, ~~ljög~~ visste han naturligtvis för sin fru.

Mening med synonymer genererad av Swectors:

Men eftersom Sloane förlorade ganska friskt, ~~ljög~~ ljugit han naturligtvis för sin ~~fru~~ mamma.

Mening med synonymer genererad av Folkets synonymlexikon:

Jag går fram mot skynket, böjer mig ner och lyfter ~~långsamt~~ sakta upp det.

Mening med synonymer genererad av Wiki:

Jag går fram mot skynket, ~~böjer~~ vänder mig ner och lyfter långsamt upp det.

Mening med synonymer genererad av Swectors:

Jag går fram mot skynket, ~~böjer~~ böjde mig ner och lyfter långsamt upp det.

Mening med synonymer genererad av Folkets synonymlexikon:

~~Samtidigt~~ Medan slog en och annan fast att texten nog var ett försök till satir, eller att Sannfinländarna i alla fall skulle ~~försöka~~ testa bortförklara inlägget som ett skämt.

Mening med synonymer genererad av Wiki:

Samtidigt slog en och annan fast att texten nog mindre var ett försök till satir, eller att Sannfinländarna i alla fall skulle försöka hjälpa bortförklara inlägget som ett skämt.

Mening med synonymer genererad av Swectors:

Samtidigt slog en och annan fast att texten nog var ett försök till ~~satir~~ humor, eller att Sannfinländarna i alla fall skulle försöka ~~bortförklara~~ dölja inlägget som ett skämt.

Mening med synonymer genererad av Folkets synonymlexikon:

Hon talar om den egenartade stillheten i bilden, det bleka ljuset som faller in i rummet, känslan av en ~~närvaro~~ deltagande som inte tränger sig på, som finns där men håller sig i skymundan.

Mening med synonymer genererad av Wiki:

Hon talar om den egenartade stillheten i bilden, det bleka ljuset som ~~faller~~ slår in i rummet, känslan av en närvaro som inte tränger sig på, som finns där men håller sig i skymundan.

Mening med synonymer genererad av Swectors:

Hon talar om den ~~egenartade~~ poetiska ~~stillheten~~ tystnaden i bilden, det bleka ljuset som faller in i rummet, känslan av en närvaro som inte ~~tränger~~ drar sig på, som finns där men håller sig i ~~skymundan~~ skuggan.

Mening med synonymer genererad av Folkets synonymlexikon:

Försvarsminister Ben-Eliezer betonade att Israel också kommer att ~~fortsätta~~ gå vidare gripa palestinier efterlysta för terrorism som tagit sin ~~tillflykt~~ fristad till Arafats ~~residens~~ säte och den palestinska säkerhetstjänstens högkvarter i Ramallah.

Mening med synonymer genererad av Wiki:

Försvarsminister Ben-Eliezer betonade att Israel också kommer att fortsätta gripa palestinier efterlysta för terrorism som tagit sin tillflykt till Arafats residens och den palestinska säkerhetstjänstens högkvarter huvudkontor i Ramallah.

Mening med synonymer genererad av Swectors:

Försvarsminister Ben-Eliezer ~~betonade~~ försäkrade att Israel också kommer att fortsätta ~~gripa~~ stoppa palestinier ~~efterlysta~~ misstänkta för terrorism som tagit sin tillflykt till Arafats residens och den palestinska ~~säkerhetstjänstens~~ arméns högkvarter kontor i Ramallah.

Mening med synonymer genererad av Folkets synonymlexikon:

Beskedet om att en ~~kollega~~ medarbetare misshandlats till döds under ett arbetspass i Flemingsberg kom som en chock.

Mening med synonymer genererad av Wiki:

Beskedet om att en ~~kollega~~ misshandlats till ~~döds~~ dödade under ett arbetspass i Flemingsberg kom som en chock.

Mening med synonymer genererad av Swectors:

Beskedet om att en ~~kollega~~ vän ~~misshandlats~~ mördats till ~~döds~~ ihjäl under ett arbetspass i Flemingsberg kom som en ~~chock~~ lättnad.

Mening med synonymer genererad av Folkets synonymlexikon:

Om dessa världar kunde mötas lite oftare skulle ~~möjligen~~ kanske också användningen av ~~medicinsk teknologi~~ teknik bli mer ~~balanserad~~ sansad.

Mening med synonymer genererad av Wiki:

Om dessa världar kunde ~~mötas~~ möts lite oftare skulle möjligen också användningen av medicinsk teknologi bli mer balanserad.

Mening med synonymer genererad av Swectors:

Om dessa världar kunde mötas lite oftare skulle möjligen kanske också användningen av medicinsk ~~teknologi~~ forskning bli mer balanserad.

Mening med synonymer genererad av Folkets synonymlexikon:

Läkemedelsindustrins ~~intresse~~ hobby är ~~förvisso~~ otvivelaktigt mycket stort, inte minst i ~~riktning~~ håll mot att ~~försöka~~ prova hitta någon form av vaccin mot demenssjukdomar

Mening med synonymer genererad av Wiki:

Läkemedelsindustrins ~~intresse~~ är ~~förvisso~~ ändå mycket stort, inte minst i riktning mot att ~~försöka~~ hjälpa hitta någon form av vaccin mot demenssjukdomar.

Mening med synonymer genererad av Swectors:

Läkemedelsindustrins ~~intresse~~ är ~~förvisso~~ ändå mycket stort, inte minst i riktning mot att ~~försöka~~ hitta någon form av ~~vaccin~~ läkemedel mot demenssjukdomar.

Mening med synonymer genererad av Folkets synonymlexikon:

Ett par föreningar har ~~emellertid~~ dock inkomster från sidoverksamhet, i till ~~exempel~~ prov Göteborg producerar föreningen en ~~katalog~~ register över privatläkare, och Stockholm har inkomster från ~~kapital~~ pengar.

Mening med synonymer genererad av Wiki:

Ett ~~par~~ flera föreningar har ~~emellertid~~ dock inkomster från sidoverksamhet, i till exempel Göteborg producerar föreningen en katalog över privatläkare, och Stockholm har inkomster från kapital.

Mening med synonymer genererad av Swectors:

Ett par föreningar har emellertid inkomster från sidoverksamhet, i till exempel Göteborg producerar föreningen en katalog förteckning över privatläkare vårdgivare, och Stockholm har inkomster från kapital.

Mening med synonymer genererad av Folkets synonymlexikon:

Det mest ~~använda~~ nyttja testet, blödningstid, är en invasiv metod som dras med en viss ~~osäkerhet~~ osjälvständighet, bl a ~~beroende~~ ambivalens av vem som utfört testet.

Mening med synonymer genererad av Wiki:

Det mest använda testet, blödningstid, är en ~~invasiv~~ smittspridning metod som ~~dras~~ dras med en viss osäkerhet, bl a beroende av vem som utfört testet.

Mening med synonymer genererad av Swectors:

Det mest använda ~~testet~~ undersökningen, blödningstid, är en invasiv metod som dras med en viss ~~osäkerhet~~ oro, bl a beroende av vem som utfört ~~testet~~ undersökningen.