



LUNDS
UNIVERSITET

Språk- och litteraturcentrum
Översättarutbildningen

EXAMENSARBETE VT 2014
MASTER I ÖVERSÄTTNING
Specialisering i engelska

Översättningsuniversalier i svenska översatta texter

En undersökning av tecken på förenkling i svenska översatta
texter och svenska icke-översatta texter

Författare:

Malin Kärnebro

malin.karnebro@gmail.com

Handledare:

Lisa Holm

Sammandrag:

I denna uppsats har översättningsuniversaliers förekomst i svenska översatta texter undersökts. Fokus ligger på förenkling, och metoden för undersökningen bygger på tidigare undersökningar utförda av Laviosa (1996, 1997, 1998). Laviosas resultat tydde på att vissa tecken på förenkling alltid existerar i översatta engelska tidningstexter och skönlitterära texter, och därför har hennes metoder använts för att undersöka om detsamma gäller för svenska översatta skönlitterära och icke-skönlitterära texter. Materialet till undersökningen som har genomförts i denna uppsats kommer från the English-Swedish Parallel Corpus (ESPC), en korpus som bland annat innehåller ett stort antal svenska icke-översatta texter och svenska översatta texter. Med hjälp av dessa texter har frekvenslistor över de 108 vanligaste typorden för varje delkorpus skapats. Utifrån dessa frekvenslistor har undersökningar av bland annat lexikal densitet, andel vanliga lexikonord, upprepning av vanligaste korpusord och antal lemman genomförts. Det visade sig att översatta texter i de allra flesta fall visar upp fler tecken på förenkling än icke-översatta texter. Resultaten är statistiskt signifikanta och tyder på att översättningsuniversalien förenkling existerar i svenska översatta texter.

Nyckelord: översättning, översättningsuniversalier, translation universals, förenkling, simplification, ESPC, korpusbaserad undersökning, lexikal densitet

Engelsk titel: Translation Universals in Swedish Translated Texts

Innehåll

| | |
|---|----|
| 1. Inledning..... | 1 |
| 2. Översättningsuniversalier..... | 2 |
| 2.1 Förenkling | 3 |
| 2.2 Tidigare forskning av översättningstrender i svenska..... | 6 |
| 3. Material | 7 |
| 3.1 STfict..... | 8 |
| 3.2 SOfict | 9 |
| 3.3 STnfict..... | 9 |
| 3.4 SONfict | 10 |
| 4. Metod | 11 |
| 4.1 Lexikal densitet | 12 |
| 4.2 Andel vanliga lexikonord | 14 |
| 4.3 Upprepning av de vanligaste korpusorden | 15 |
| 4.4 Lemmaberäkning..... | 15 |
| 4.5 Hjälpverben <i>ha</i> och <i>vara</i> | 16 |
| 4.6 Sammanfattning | 16 |
| 4.7 Trender i ordförekomster | 16 |
| 4.8 Statistisk signifikans..... | 17 |
| 5. Resultat..... | 18 |
| 5.1 Lexikal densitet | 18 |
| 5.1.1 Skönlitterära texter | 18 |
| 5.1.2 Icke-skönlitterära texter..... | 19 |
| 5.2 Andel vanliga lexikonord | 19 |
| 5.2.1 Skönlitterära texter | 19 |
| 5.2.2 Icke-skönlitterära texter..... | 20 |
| 5.3 Upprepning av de vanligaste korpusorden | 20 |
| 5.3.1 Skönlitterära texter | 21 |
| 5.3.2 Icke-skönlitterära texter..... | 21 |
| 5.4 Antal olika lemman | 21 |
| 5.4.1 Skönlitterära texter | 22 |
| 5.4.2 Icke-skönlitterära texter..... | 22 |
| 5.5 Hjälpverben <i>ha</i> och <i>vara</i> | 22 |
| 5.5.1 Skönlitterära texter | 22 |
| 5.5.2 Icke-skönlitterära texter..... | 23 |
| 5.6 Resultatsammanfattning för Laviosas fem mönster | 23 |
| 5.7 Trender i ordförekomster | 25 |
| 6. Avslutande diskussion..... | 27 |
| Referenser..... | 29 |
| Bilaga 1. Material..... | 32 |
| Bilaga 2. Frekvensordlistor | 33 |
| Bilaga 3. Lemmatiserade listor..... | 42 |
| Bilaga 4. Konfidensintervall | 51 |

1. Inledning

Översättningsuniversalier (eng. 'translation universals') kan beskrivas som lingvistiska särdrag som typiskt förekommer i översatta texter snarare än icke-översatta texter och som tros förekomma oberoende av vilka språkpar som är involverade i översättningsprocessen (Baker 1993:243). Översättningsuniversalier är alltså lingvistiska särdrag som alla översatta texter har gemensamt. De mest etablerade översättningsuniversalierna är enligt Laviosa (2009:307) **förenkling** (eng. 'simplification'), **explicitering** och **normalisering**, vilka kommer att förklaras mer ingående i uppsatsens bakgrundsavsnitt (avsnitt två).

I denna uppsats kommer jag att undersöka översättningsuniversaliers förekomst i svenska översatta texter. Jag kommer att fokusera på förekomsten av tecken på förenkling i svenska översatta texter, och för att ta reda på om översatta texter uppvisar spår av förenkling kommer jag att jämföra de översatta texterna med icke-översatta texter. Min hypotes är att jag kommer att finna regelbundenheter som tyder på att vissa av de existerande hypoteserna kring förenkling i översättning stämmer. Denna undersökning kommer att vara deskriptiv och kommer inte söka förklara varför översättningsuniversalier (eventuellt) existerar i svenska texter. Jag kommer att fokusera på produkten snarare än processen. Undersökningen hör alltså till den produktorienterade deskriptiva översättningsvetenskapen (Malmkjær, 2005:18f).

Syftet med undersökningen är att utforska eventuella förekomster av översättningsuniversalier i svenska översatta texter. Översättningsuniversaliers existens i översättningar på olika språk har undersökts många gånger, men förekomsten av dem i svenska översättningar är ännu inte utforskad i någon nämnvärd omfattning. Därför hoppas jag att min undersökning ska kunna belysa vilka exempel på förenkling man eventuellt kan förvänta sig att finna i svenska översatta texter och möjligtvis även bana väg för fler undersökningar av både förenkling och andra sorters översättningsuniversalier i svenska texter.

2. Översättningsuniversalier

Enligt Chesterman (2004:39) finns det två sorters översättningsuniversalier: S-universalier (S för ”source”) och T-universalier (T för ”target”). S-universalier är universella skillnader mellan översättningar och deras källtexter, och T-universalier är universella skillnader mellan översättningar och icke-översatta texter på samma språk. I denna uppsats kommer jag att fokusera på T-universalier.

Potentiella T-universalier är enligt Baker (1996:176ff) **förenkling**, **explicitering**, **utjämning** (eng. ’levelling out’) och **normalisering**. Förenkling innebär att det finns en tendens att språkbruket förenklas i översättningar, t.ex. genom att de i jämförelse med icke-översatt text har lägre genomsnittlig meningslängd eller genom att den så kallade lexikala densiteten (förhållandet mellan innehållsord och funktionsord) är lägre. Explicitering innebär att det finns en tendens att saker och ting uttrycks explicit i översättningar i högre grad än i icke-översatt text. Utjämning innebär att det finns en tendens till att översättningar är mer homogena än vad icke-översatta texter är. Normalisering innebär att det finns en tendens att målspråkstypiska konstruktioner överanvänds i översättningar. Det kan innebära att översättaren ersätter en relativt ogrammatisk mening i källtexten med ett målspråkstypiskt uttryck i måltexten. Baker (1996:184) exemplifierar detta med hjälp av den hebreiska översättningen av Faulkners roman *Sanctuary*:

a. Engelskt original:

He couldn't walk, even.

b. Hebreisk översättning (en målspråkstypisk konstruktion, återöversatt):

He could hardly stand on his feet.

Normalisering kan också innebära att översättaren tenderer att undvika att föra över exempelvis experimentell interpunktion eller talspråkliga drag från en källtext till en måltext.

Bland forskare i översättningsvetenskap pågår en debatt om huruvida översättningsuniversalier finns eller inte. Självklart finns det logiska skäl att vara tveksam till översättningsuniversaliers existens – till exempel det faktum att man omöjligt kan undersöka alla översatta texter som någonsin producerats. Det närmsta man kan komma är en tendens, och den tendensen är enligt exempelvis Toury (2004) och Chesterman (2004) att översättningsuniversalier existerar.

Toury uttrycker att han accepterar existensen av regelbundenheter (eng. ’regularities’) men skriver också:

that, while universals do presuppose regularities, the reverse does not necessarily hold: It is one thing to say that certain regularities were found *in* translation, and something quite different to claim that the observed regularities are there *because* it is translation. (originalets kursiv)

Toury 2004:17

Toury menar dock att det är viktigt att se till likheterna i översatta texter snarare än skillnaderna och att ju mer forskning som utövas inom området, desto större är möjligheten att ta reda på exakt vilka mönster som finns för olika fenomen.

Chesterman (2004:33) verkar se det som en självklarhet att forskningen letar efter likheter mellan olika fall för att kunna generalisera utifrån dessa. Det är den enda möjligheten för vetenskap att kunna förutspå vad som ska ske i ännu ostuderade fall, och det är också den enda möjligheten för att man såsmåningom ska kunna förstå de grundläggande lagarna inom en disciplin. Chesterman (ibid.) säger också att en disciplin som översättningsvetenskap skulle stagnera om man bortsåg från de likheter som översättningar faktiskt uppvisar.

En forskare som är tveksam till översättningsuniversaliers existens är House. Enligt henne (2008:11) ”är det fruktlöst att söka efter översättningsuniversalier eftersom det inte finns och inte skulle kunna finnas några universalier inom översättning” (min översättning). Hon skriver att översättningen alltid beror av språkparen som är inblandade samt att termer som ’explicitation’, ’simplification’ och ’conventionalization’ (numera ’normalization’) är för generella och inte bör användas om man inte är helt på det klara med deras definition och hur de ska användas. Vidare uttrycker House att något annat som talar emot att översättningsuniversalier skulle finnas är att de till viss del verkar vara genrespecifika.

Som utgångspunkt för min undersökning väljer jag att gå in i översättningsuniversaliernas värld med ett öppet sinne. Jag utgår varken ifrån att universalier inte existerar eller att de gör det. Ämnet är som sagt ännu relativt utforskat, framförallt inom svenskan, och därför ser jag min undersökning som ett steg på vägen för att kunna ta reda på om det möjligtvis finns universella drag som alltid går att finna i svenska översättningar.

2.1 Förenkling

Förenkling innebär alltså att översättningar tenderar att vara enklare än icke-översatta texter. Det är viktigt att skilja på förenkling och explicitering eftersom de inte nödvändigtvis innebär samma sak, även om förenklingsprocessen kan innebära att texten blir mer explicit eftersom översättaren till exempel kan välja att undvika tvetydighet genom att förmedla en specifik tolkning till mottagaren (Baker 1996:182).

Enligt Sara Laviosa-Braithwaite (1996:115) har minst tre sorters förenkling identifierats – lexikalisk, syntaktisk och stilistisk. Lexikalisk förenkling definieras av Blum-Kulka och Levenstone (1983:119) som ”the process and/or result of making do with *less* words” (Blum-Kulka och Levenstons kursiv). Detta kan ta sig uttryck i att man i måltexten använder sig av en överordnad term när det inte finns någon motsvarighet till en av källspråkets hyponymer.

Syntaktisk förenkling kan enligt Vanderauwera (1985, citerad i Laviosa-Braithwaite 1996:116) innebära att svår syntax görs enklare genom att man till exempel ersätter infinita satser med finita satser i översättningsprocessen.

Vanderauwera ger också exempel på stilistisk förenkling, vilket kan vara att man bryter upp långa meningar, ersätter invecklade kollokationer med enklare sådana eller att man utelämnar eventuell upprepning i källtexten.

Laviosa har undersökt förekomsten av förenkling i olika texttyper (1996, 1997 och 1998). I sin doktorsavhandling (1996) undersöker hon bl.a. förenkling i engelska översatta tidningstexter och engelska översatta skönlitterära texter. Resultaten från dessa undersökningar presenteras sedan mer kortfattat i två artiklar (1997 och 1998). De är också dessa artiklar som ligger till grund för min undersökning. I den första av de två artiklarna (1997) vill Laviosa visa på hur förenkling tar sig uttryck i översatta tidningsartiklar. Hon använder sig av en så kallad ’comparable corpus’ som enligt Baker (1995:234) är ”two separate collections of texts in the same language: one corpus consists of original texts in the language in question and the other consists of translations in that language from a given source language or languages”. Laviosas korpus innehåller texter och textutdrag från europeiska och amerikanska tidningar, och de översatta texterna är översatta från många olika europeiska språk, men de flesta texterna är översatta från franska, tyska, spanska och ryska. Med hjälp av korpustexterna kommer hon fram till att dessa sju mönster för förenkling går att finna i översatt tidningstext (1997:538):

1. Translated articles have a relatively lower percentage of content words versus grammatical words (i.e. their lexical density is lower)
2. The value of lexical density is unaffected by the source language
3. Translated articles have a lower mean sentence length
4. The proportion of high frequency words [de 200 vanligaste engelska lexikonorden ur the Cobuild Bank of English] versus less frequent words is relatively higher in translated texts
5. The list heads of a corpus of translated texts [de 108 vanligast förekommande typorden för varje delkorpus] account for a larger area of the corpus
6. The list heads of translated texts contain fewer lemmas
7. Translated articles use the present tenses of the auxiliary verbs *to be* and *to have* more frequently

Laviosa 1997:538

Laviosa presenterar i en senare artikel (1998) sina resultat från samma typ av undersökning gjord på översatta narrativa texter. Hon använder sig även här av en 'comparable corpus' och kommer fram till att det bara är vissa av de sju mönstren som också stämmer för översatt skönlitteratur. Jag har försökt att göra en sammanställning av resultaten från Laviosas undersökning av skönlitteratur. Det har inte varit så enkelt att göra denna sammanställning eftersom hon i sin artikel (1998) inte redovisar alla resultat helt tydligt, så för vissa resultat har jag gått tillbaka till hennes avhandling (1996). Jag reserverar mig för eventuella misstolkningar av resultaten. Jag kommer härnäst att benämna de typord som finns i ett visst språks lexikon för "lexikonord" och de typord eller löpord som finns i mina korpusar för "korpusord" eller "korpusförekomster".

| Mönster | Skönlitterära översatta texter | Skönlitterära icke-översatta texter |
|--|---|--|
| 1. Lexikal densitet | 52,87 | 54,95 |
| 2. Källspråkets inverkan på lexikal densitet | Går inte att testa eftersom antalet olika texter är för få | Går inte att testa eftersom antalet olika texter är för få |
| 3. Genomsnittlig meningslängd | 24,09 ord | 15,62 ord |
| 4. Andel vanliga lexikonord | 59,74 % | 58,51 % |
| 5. Samtliga korpusförekomster av de 108 vanligaste typordens andel av korpusen | 56,2 % | 51,6 % |
| 6. Antal lemman i frekvenslistan | 82 | 87 |
| 7. Position för olika tempus av hjälpverben <i>to have</i> och <i>to be</i> i frekvenslistorna | IS 12 ARE 48 WAS 9 WERE 35 HAS 63 HAVE 31 HAD 21 | IS 25 ARE 61 WAS 8 WERE 31 HAS - HAVE 33 HAD 13 |

Tabell 1. Laviosas resultat, skönlitterära texter

Jag valde att inte göra en likadan sammanställning av resultaten från undersökning av tidningstexterna av tre skäl. Dels eftersom korpusarna är mycket mindre, dels eftersom min undersökning huvudsakligen bygger på resultaten från Laviosas skönlitterära korpusar och dels eftersom det är ännu svårare att uttyda resultaten från hennes undersökning av tidningstexterna.

För att återgå till tabell 1 så visar det sig att översatta skönlitterära texter inte har en lägre genomsnittlig meningslängd än icke-översatta skönlitterära texter. De översatta texterna har inte heller en högre förekomst av presensformen av verbet *to be*. Mönster 3 stämmer alltså inte för skönlitterära texter och 7 stämmer endast delvis. Laviosa kan heller inte undersöka om den lexikala densiteten påverkas av källspråket (mönster 2) eftersom materialet är för litet. Mönster 1, 4, 5 och 6 stämmer däremot och Laviosa fastställer därför att dessa fyra mönster för förenkling stämmer för både engelsk översatt tidningstext och engelsk översatt skönlitteratur:

1. Translated texts have a relatively lower percentage of content words versus grammatical words (i.e. their lexical density is lower)
2. The proportion of high frequency words versus low frequency words is relatively higher in translated texts
3. The list head of a corpus of translated texts accounts for a larger area of the corpus (i.e. the most frequent words are repeated more often)
4. The list head of translated texts contains fewer lemmas

Laviosa 1998:565

Laviosas skönlitterära översatta texter är översatta från flera olika källspråk, bland vilka tyska, franska, spanska och italienska är de vanligaste.

En undersökning som bygger på Laviosas mönster för förenkling är Grabowskis (2013). Han använder sig av delar av Laviosas metod för att genomföra en undersökning av förenkling i polsk översatt skönlitterär text. Grabowski använder sig av två korpusar, en med polska översatta skönlitterära texter och en med polska icke-översatta skönlitterära texter. Grabowski undersöker endast två av de mönster för förenkling som stämmer för både tidningstext och skönlitterär text – andelen vanliga lexikonord (mönster nummer 2) och upprepning av de vanligaste korpusorden (mönster nummer 3). Men han undersöker också den genomsnittliga meningslängden för texterna i de olika korpusarna. Resultaten av hans undersökning är att den genomsnittliga meningslängden är lägre i de översatta texterna än i de icke-översatta texterna vilket bekräftar förenklingshypotesen. Däremot upprepas inte de vanligaste orden oftare i de översatta texterna än i de icke-översatta texterna, och andelen vanliga ord är högre för icke-översatta texter än översatta texter. De två mönster för förenkling som Grabowski undersöker visar sig alltså inte stämma för hans material.

2.2 Tidigare forskning av översättningstrender i svenska

Jag har funnit två tidigare undersökningar av trender i svenska översättningar. Det ska dock understrykas att dessa undersökningar snarare behandlar ämnet översättningssvenska (jmf.

eng. 'translationese') än översättningsuniversalier. Översättningssvenska innebär att målspråket influeras av källspråket vid översättning (Gellerstam 1986).

Gellerstam (1986) undersöker källspråkets influens i romaner som översatts från engelska. Hans undersökning fokuserar på skillnader i vokabulären i översatta romaner och icke-översatta romaner. Det han kommer fram till är bland annat att engelska lånord är vanligare i översatt text, talspråksuttryck är vanligare i originalproducerad text och att lånord från latin (exempelvis *massiv*, *drastisk* och *lokal*) används i översättningarna med en ny betydelse som stammar från det engelska lexemet. Gellerstam betonar dock att hans undersökning endast gäller språkparet engelska och svenska; därför kan denna undersökning inte fullt ut räknas som en undersökning av översättningsuniversalier.

Nilsson (2004) undersöker förekomsten av det svenska ordet *av* i svensk översatt skönlitteratur och icke-översatt skönlitteratur. Hans material kommer från parallellkorpusen ESPC (English-Swedish Parallel Corpus) och de översatta texterna är översatta från engelska. Hans undersökning är inte bara deskriptiv utan också till viss del komparativ; därför går han tillbaka till källtexterna för att se vad som skapar en viss användning av *av*. Resultaten av undersökningen är att *av* är överrepresenterat i de översatta texterna och att detta beror på att vissa ordförbindelsemönster är vanligare i de översatta texterna än i de icke-översatta texterna (huvudsakligen PREPOSITION + x + *av* [t.ex. ”i början av”] och LOKATIVT SUBSTANTIV + *av* [t.ex. ”foten av”]). Nilssons undersökning behandlar även den endast språkparet engelska och svenska.

3. Material

Mitt material består av texter från the English-Swedish Parallel Corpus (ESPC). ESPC är en korpus som startade som ett forskningsprojekt 1997 på Lunds universitet, men sedan dess har korpusen utvecklats både i Lund och på Göteborgs universitet. ESPC innehåller originaltexter och deras översättningar (engelska-svenska och svenska-engelska). Korpusen innehåller både skönlitterära texter och vad korpuskaparna kallar för ”icke-skönlitterära” texter (”non-fiction”). (ESPC:s webbplats). Flertalet texter förekommer inte i sin helhet i korpusen, utan består av utdrag på ca 10 000 – 15 000 ord. De skönlitterära texterna består av både barnböcker och vuxenlitteratur, och bland de icke-skönlitterära texterna finns bland annat årsredovisningar, historieböcker, tal från Europaparlamentet, lagar och biografier. De fyra

delkorpusarna kommer jag hädanefter att benämna **STfict** (översatt skönlitteratur), **SOfict** (icke-översatt skönlitteratur), **STnfict** (översatt icke-skönlitteratur) och **SONfict** (icke-översatt icke-skönlitteratur). ”S” står i dessa förkortningar för svenska, ”T” står för translated och ”O” står för original. ”Fict” står för fiction och ”nfict” står för non-fiction. Varje delkorpus innehåller ca 300 000 till 350 000 löpord. Laviosas skönlitterära korpusar innehåller fler löpord än mina korpusar – mellan 700 000 och 1 000 000 löpord, men korpusarna med tidningstexter innehåller endast mellan 10 000 och 65 000 löpord.

Jag kommer i min undersökning endast använda mig av svenska översatta texter och svenska icke-översatta texter. Laviosa har, som tidigare nämnts, funnit mönster för engelska översatta skönlitterära texter och tidningstexter. Jag har dessvärre ingen tillgång till tidningstexter, men jag kommer att undersöka huruvida hennes mönster stämmer för svenska översatta skönlitterära texter och icke-skönlitterära översatta texter. Till skillnad från Laviosas undersökning är de översatta texter jag använder till min undersökning endast översatta från ett annat språk – engelska. Detta kan uppfattas som ett problem eftersom man kan hävda att de resultat jag får från min undersökning endast stämmer för svensk text som är översatt från engelska. Det ska dock nämnas att Grabowskis undersökning (2013) även den endast behandlar två språkpar – engelska och polska.

Eftersom materialet ur ESPC är upphovsrättsskyddat har jag inte tillåtelse att exempelvis citera från materialet i min uppsats. Jag har dock fått lov att göra frekvenslistor över de 108 vanligaste typorden för varje delkorpus. Dessa frekvenslistor har gjorts med hjälp av webbplatsen lix.se, där jag har kopierat in samtliga texter ur en delkorpus och på så vis fått fram en frekvenslista för just den delkorpusen. De texter jag fått ta del av från ESPC har varit fyllda av taggar och tecken som inte hör till själva texten. Jag beskriver mer ingående hur jag löst detta problem i metodavsnittet. Efter att jag rensat upp filerna anser jag att de frekvenslistor jag fått ut kan räknas som pålitliga.

Här följer en beskrivning av varje delkorpus uppbyggnad. För en detaljerad lista över texternas titlar och författarnas och översättarnas namn, se bilaga 1.

3.1 STfict

Det översatta skönlitterära materialet består av utdrag från 25 romaner. Den totala ordmängden uppgår till 346 649 ord. Vissa av verken är böcker för ungdomar eller barn, men de flesta kan sägas vara vuxenlitteratur. Både författarna till originalverken och översättarna är i de flesta fall välkända. Ingen författare förekommer mer än en gång, men vissa översättare

har översatt fler än ett verk ur korpusen. Totalt är det 21 stycken olika översättare/översättarpar som har översatt de 25 romanerna. Det faktum att fyra översättare/översättarpar har översatt två verk vardera borde dock inte ha någon inverkan på frekvenslistan som gjorts på STfict-korpusen. Alla översättare utom en har översatt till sitt modersmål.

| Översatt skönlitteratur | | | |
|-------------------------|--------------|-----------|------------------|
| | Antal texter | Antal ord | % av underkorpus |
| Till modersmål | 24 | 333 161 | 96,11 |
| Från modersmål | 1 | 13 488 | 3,89 |

Tabell 2. Översättningsriktning, skönlitteratur

3.2 SOfict

I min undersökning har jag använt mig av utdrag från 25 svenska icke-översatta romaner. Den totala ordmängden uppgår till 306 935 ord. Verken är skrivna av relativt välrenommerade svenska författare. Även för SOfict-korpusen är en del av böckerna skrivna för ungdomar eller barn, men de flesta är dock vuxenlitteratur. Två författare förekommer med två av sina verk. En av dessa författare är Henning Mankell, som har med två av sina böcker i serien om poliskommissarien Wallander. Detta har gett utslag i delkorpusens frekvenslista eftersom namnet Wallander finns med bland de 108 vanligaste typorden för SOfict-korpusen.

| Parallellkorpus skönlitterära texter | | |
|--------------------------------------|-----------------------|------------------|
| | Icke-översatta texter | Översatta texter |
| TIDSSPANN | 1968 – 1995 | 1981 – 1996 |
| FÖRFATTARSKAP/ ÖVERSÄTTARE | | |
| - Individ man | 64 % | 44 % |
| - Individ kvinna | 36 % | 44 % |
| - Grupp | 0 % | 12 % |
| TOTALT ANTAL ORD | 306 935 | 346 649 |

Tabell 3. Jämförelse, skönlitterära texter

3.3 STnfict

Materialet som innehåller översatta icke-skönlitterära texter består av utdrag av texter samt, i vissa fall, hela texter. 27 texter har använts och det totala antalet ord är 291 153. I STnfict-

korpusen finns bland annat årsredovisningar, EU- och EG-dokument (direktiv och avtal), historieskildringar och biografier.

För samtliga fyra årsredovisningar är översättarna okända, liksom för vissa EU- och EG-dokument. I övrigt är det två översättare som har översatt tre respektive två av texterna i korpusen. Utöver dessa två har varje översättare/översättarpär endast översatt en text var ur korpusen. De flesta översättare har översatt till sitt modersmål. I ett av översättarpären har en person engelska som modersmål och en person svenska som modersmål.

| Översatt icke-skönlitteratur | | | |
|------------------------------|--------------|-----------|------------------|
| | Antal texter | Antal ord | % av underkorpus |
| Till modersmål | 18 | 194 604 | 66,84 |
| Blandat | 1 | 12 544 | 4,31 |
| Okänt | 8 | 84 005 | 28,85 |

Tabell 4. Översättningsriktning, icke-skönlitteratur

3.4 SONfict

De icke-översatta icke-skönlitterära texterna består av både utdrag och hela texter. 47 texter har använts, och totalt antal ord är 353 303. Några exempel på texttyper från SONfict-korpusen är lagar, historieskildringar, årsredovisningar, biografier och tal från Europaparlamentet.

Det faktum att det förekommer rätt många tal i korpusen skulle kunna vålla problem för min undersökning eftersom de eventuellt är mer talspråkliga än de andra texterna i korpusen. Men med tanke på att de är officiella tal och förekommer i skriftlig form i korpusen har de troligtvis inte alltför många talspråkliga drag. Och även om de skulle vara mer talspråkliga än andra texter i korpusen, utgör de endast en sjundedel (ca 50 000 löpord) av det totala antalet ord för SONfict-korpusen.

| Parallellkorpus icke-skönlitterära texter | | |
|---|-----------------------|------------------|
| | Icke-översatta texter | Översatta texter |
| TIDSSPANN | 1985 – 2000 | 1980 – 2000 |
| FÖRFATTARSKAP/ ÖVERSÄTTARE | | |
| - Individ man | 43 % | 52 % |
| - Individ kvinna | 26 % | 15 % |
| - Grupp | 31 % | 33 % |
| TOTALT ANTAL ORD | 353 303 | 291 153 |

Tabell 5. Jämförelse, icke-skönlitterära texter

4. Metod

Jag kommer att undersöka om Laviosas mönster för förenkling stämmer för svensk översatt text. Jag väljer att i huvudsak fokusera på de fyra mönster som stämde både för tidningstext och för skönlitterär text eftersom jag, på grund av begränsningarna i mitt material, helt enkelt inte kan undersöka vissa av Laviosas sju mönster för förenkling i tidningstext. Det finns dock ett mönster av de som endast stämde för tidningstext som jag kan undersöka, nämligen huruvida hjälpverben *ha* och *vara* är vanligare i presens i översatta texter och vanligare i preteritum i icke-översatta texter. Jag kommer att undersöka presens- och preteritumformerna av hjälpverben *ha* och *vara* och också de fyra mönster som stämmer för både tidningstext och skönlitterär text, alltså dessa:

1. Translated texts have a relatively lower percentage of content words versus grammatical words (i.e. their lexical density is lower)
2. The proportion of high frequency words versus low frequency words is relatively higher in translated texts
3. The list head of a corpus of translated texts accounts for a larger area of the corpus (i.e. the most frequent words are repeated more often)
4. The list head of translated texts contains fewer lemmas

(Laviosa 1998:565)

Min undersökning kommer att genomföras med hjälp av frekvenslistor som har gjorts på de delkorpusar som beskrevs i avsnitt 3. Uttrycket 'list heads' som Laviosa använder sig av kommer från Piemontese (1995:480) och står för de 108 vanligaste typorden för varje samling

texter. Varför det är just 108 ord vet jag inte, men jag väljer att, liksom Laviosa, genomföra min undersökning på de 108 vanligaste typorden för varje delkorpus, för att mina resultat ska kunna jämföras med hennes.

Att göra frekvenslistor utifrån mitt material visade sig kräva mer arbete än vid första anblicken. Texterna jag fick från ESPC var, som tidigare nämnts, fyllda av taggar och andra tecken som försvårade för mig när jag skapade frekvenslistorna. För att få fram textutdrag som skulle ge tillförlitliga frekvenslistor behövde jag göra mellanslag mellan taggar och ord samt mellan skiljetecken och ord. Till en början gjorde jag bara mellanslag mellan taggar och ord, och då visade det sig att jag fick två olika inlägg av exempelvis ordet *han* bland de 108 vanligaste typorden, eftersom vissa förekomster av *han* följdes av ett skiljetecken. Vanligtvis har skiljetecken inte någon betydelse på *lix.se* (två förekomster av *han* räknas som just två förekomster av ordet, oavsett om det en gång följs av ett mellanslag och en gång av en punkt), men i detta fall fungerade det inte så. Men efter att jag fört in mellanslag mellan ord och skiljetecken var problemet löst och jag fick fram mina frekvenslistor. Frekvensordlistorna går att finna i bilaga 2.

Ett annat problem med materialet jag fick var att bokstäverna å, ä och ö inte fanns med. Dessa var istället utbytta mot Σ (ä), σ (å) och \div (ö). Jag var alltså tvungen att byta ut dessa mot å, ä och ö i samtliga fall de förekom. För vissa ord, t.ex. *f÷r* (*för*) och *før* (*får*), gällde det att få rätt vokal på rätt plats.

Jag vill understryka att samtliga resultat från min undersökning baseras på de frekvenslistor jag har skapat. Det är bara från dem jag hämtar information om de texter som finns i delkorpusarna.

Jag har som sagt använt mig av Laviosas artiklar från 1997 och 1998 som grund för min undersökning. Ofta är hon dock rätt kortfattad när det kommer till metoddelen, så ibland har jag fått försöka lista ut hur hon genomfört vissa delar av sin undersökning och varför hon har gjort på ett visst sätt. Inte heller hennes doktorsavhandling från 1996 beskriver metoden för undersökningen av förenkling i översatta texter särskilt ingående.

Nedan följer en detaljerad beskrivning av hur varje enskilt mönster kommer att undersökas.

4.1 Lexikal densitet

För att räkna ut den lexikala densiteten för varje enskild delkorpus kommer jag att räkna ut hur stor del av samtliga korpusförekomster av de 108 vanligaste typorden i varje enskild

delkorpus som är innehållsord respektive formord. Anledningen till att jag endast gör denna undersökning på samtliga korpusförekomster av de 108 vanligaste typorden är dels att jag inte har tillgång till rensade filer av texterna i helhet, dels att jag inte har den tid som krävs för att undersöka samtliga ord i varje text i alla delkorpusar.

Här följer en lista över hur jag delat in de olika typorden som funnits bland frekvenslistorna för de fyra delkorpusarna i funktionsord och innehållsord:

Funktionsord

Prepositioner: i, av, på, till, med, från, under, vid, genom, inom, över, än, efter, mot, utan, mellan, hos, enligt, bland, åt, ur

Artiklar: en, ett, den, det

Pronomen: de, han, jag, vi, sig, detta, sin, andra, man, denna, dessa, alla, hans, sina, dem, deras, honom, oss, mig, något, sitt, allt, dess, någon, varje, hon, vår, några, vilket, henne, hennes, min, ni, dig, du, ingen, själv, mitt, dom

Subjunktioner: att, som, om

Konjunktioner: och, för, men, eller, samt

Hjälperb: är, var, har, hade, kan, skulle, skall, vara, måste, ha, kunde, blev, bli, vill, ska, kunna, varit, ville

Adverb: inte, så, när, där, vad, också, bara, ut, upp, hur, in, här, fram, även, därför, dock, ju, aldrig, kanske, ner, alltid, igen

Räkneord: två, första

Innehållsord

Substantiv: år, artikel, sätt, del, kommissionen, miljoner, Sverige, svenska, tid, gång, Wallander

Verb: kommer, finns, göra, se, ta, få, får, gäller, kom, fick, sa, såg, gick, sade, tog, gjorde, fanns, vet, stod, gå, visste, satt

Adjektiv: mycket, mer, många, nya, stora, hela, olika, stor, större

Adverb: då, nu, sen, sedan

Att dela upp orden som finns på frekvenslistorna i funktionsord och innehållsord har, i vissa fall, varit en relativt svår uppgift eftersom jag inte har haft direkt tillgång till kontexten. Därför har vissa ord som exempelvis homonymer och ord som kanske förekommer i texterna som talspråkliga uttryck varit svåra att bedöma. Exempel på sådana ord har varit *få*, *får*, *åt*, *dom* och *sen*. Jag har i de flesta fall valt att bedöma vilken homonym som rimligen är den mest vanliga. Detta gäller dock inte för ordet *dom* som förekommer i SOFict-korpusen. Där har jag inte valt att tolka ordet som ”ett beslut i domstol” utan jag har valt att bedöma det som den talspråkliga formen för *de* och *dem*, eftersom det är troligt att det har förekommit som just det i texterna. Hade *dom* däremot förekommit i en av korpusarna med icke-skönlitterära texter eller i STFict-korpusen hade jag nog bedömt det annorlunda.

En annan problemgrupp har varit hjälpverb. Hjälpverb ska bedömas som funktionsord, men även vid bedömningen av denna grupp har avsaknaden av kontext gjort det problematiskt. Ett verb som *kommer* kan ju mycket väl vara ett hjälpverb, men det kan också vara huvudverb. Gällande kategoriseringen av verb har jag också försökt att bedöma om de oftast används som hjälpverb eller huvudverb.

Jag har bedömt fyra adverb som innehållsord (*nu, då, sen* och *sedan*) i samspråk med min handledare eftersom de uttrycker tid och därför ligger rätt tätt på verbens tempusformer. Spatiala adverb som *upp, ner, här* och *där* ansågs inte vara tillräckligt betydelsebärande för att räknas som innehållsord i just den indelning jag har gjort.

Jag har inte hittat en svensk indelning av funktions- och innehållsord som har varit tillförlitlig. Det som går att finna är riktlinjer (t.ex. på NE.se) men bedömningen sker ju i stort sett alltid med hjälp av kontexten, och utan kontexten blir det svårare att bedöma vilka ord som hör hemma var. Därför har det oftast varit en bedömningsfråga. Det positiva med de olika frekvenslistorna är att många av svenskans vanligaste ord ofta förekommer i samtliga fyra korpusar. Eftersom alla listor har behandlats lika behöver det därför inte vara ett stort problem att jag inte har kunnat bedöma orden med hjälp av kontexten.

Själva uträkningen av andel innehållsord för varje delkorpus har jag gjort på samma sätt som Laviosa (1996), bortsett från att hon, utifrån hur jag förstätt hennes metod, undersöker hela korpusen, medan jag endast undersöker samtliga korpusförekomster av de 108 vanligaste typorden. Jag förväntar mig en lägre andel innehållsord än vad Laviosa fick fram för sitt material eftersom de vanligaste orden i svenskan oftast är funktionsord (Allén 1972), och därför borde rimligen huvuddelen av de 108 vanligaste typorden för mina korpusar vara funktionsord. Hon räknar ut andelen innehållsord genom att dividera antalet innehållsord med det totala antalet löpord. Jag kommer att addera samtliga korpusförekomster av de innehållsord som finns bland mina 108 vanligaste typord och dividera denna siffra med det totala antalet korpusförekomster av de 108 vanligaste typorden för att få fram andelen innehållsord.

4.2 Andel vanliga lexikonord

Att undersöka andelen vanliga lexikonord i översatta och icke-översatta texter fungerar snarare som ett mått på normalisering än förenkling, eftersom det ger ett mått på hur målspråkstypiska de olika texterna är. Men eftersom Laviosa ser andelen vanliga lexikonord som ett tecken på förenkling, väljer även jag att göra det i denna undersökning.

Enligt Laviosa finns det en högre andel vanliga lexikonord i översatta texter än icke-översatta texter, och detta mönster är kanske det som är svårast att undersöka med tanke på mina förutsättningar. Laviosa (1997:536) beskriver hur hon har undersökt hur stor del av hela delkorpusen som upptas av engelskans 200 vanligaste lexikonord. Jag har ju, som tidigare nämnts, inte tillgång till hela texterna på samma sätt som Laviosa hade, så jag har undersökt andelen vanliga lexikonord genom att kontrollera hur många av de 108 vanligaste typorden för varje delkorpus som finns med bland de tvåhundra vanligaste lexikonorden i Alléns *Tiotusen i topp* (1972). Sedan har jag räknat alla korpusförekomster av dessa ord och räknat ut hur stor del av varje delkorpus frekvenslista som förekommer bland svenskans 200 vanligaste lexikonord.

Det ska dock understrykas att Alléns material utgjordes av tidningstexter. Det innebär att de texter som är icke-skönlitteratur troligtvis kommer att ha en högre andel vanliga lexikonord än de skönlitterära texterna, eftersom Alléns material var icke-skönlitterärt. Att Alléns bok idag är över 30 år gammal är också värt att nämna. Men eftersom svenskans vanligaste lexikonord till stor del utgörs av funktionsord – en grupp ord som inte förändras över tid – anser jag att bokens ålder inte har någon betydelse för min undersökning. *Tiotusen i topp* har dessutom varit den enda befintliga källan för frekvenslistor över de vanligaste orden i svenskan.

4.3 Upprepning av de vanligaste korpusorden

Vid undersökningen av Laviosas tredje mönster, att vanliga korpusord upprepas oftare i översatt text än icke-översatt text, har jag, liksom Laviosa, räknat ut hur stor del av korpusen de 108 vanligaste typordens korpusförekomster upptar. Beräkningen blir alltså alla antal korpusförekomster av de 108 vanligaste orden för en delkorpus dividerat med totalt antal löpord för delkorpusen.

4.4 Lemmaberäkning

Ett lemma är en ”ordform betraktad som grundform eller uppslagsform i en ordbok eller lexikalisk teori” (NE.se). I min undersökning kommer jag att använda mig av Svenska Akademiens ordlista (SAOL) för att ta reda på hur många olika lemman som förekommer bland de 108 vanligaste typorden för varje delkorpus. Se bilaga 3 för de lemmatiserade listorna för respektive frekvenslista.

4.5 Hjälpverben *ha* och *vara*

För att undersöka om presensformerna av *ha* och *vara* är vanligare i översatt text än i icke-översatt text kommer jag att se till vilken position orden har i frekvenslistorna. Ju lägre siffra desto vanligare är ordet i den delkorpusen. Jag kommer också att undersöka orden *hade* och *var* på samma sätt för att se om de är vanligare i icke-översatt text än i översatt text, vilket deras engelska motsvarigheter var i Laviosas undersökning av tidningstexter.

4.6 Sammanfattning

Jag kommer att använda de frekvenslistor jag har skapat för att ta reda på om:

1. översatt text har lägre lexikal densitet än originaltext
2. översättningar har en högre andel ”vanliga lexikonord” (enligt vad som listas som vanliga lexikonord i Alléns *Tiotusen i topp* [1972]) än icke-översatta texter
3. de vanligaste korpusorden upprepas oftare i översatt text än icke-översatt text
4. det förekommer färre lemman bland de vanligaste korpusorden i översatt text än i icke-översatt text
5. presensformerna av hjälpverben *ha* och *vara* är vanligare i översatt text än i icke-översatt text och om preteritumformerna av samma verb är vanligare i icke-översatt text än i översatt text

4.7 Trender i ordförekomster

Som en sidoundersökning kommer jag också att observera om det finns korpusord som är vanligare i en viss korpus än den andra för samma genre. Exempelvis kan vissa korpusord vara vanligare i SOfict än i STfict. Jag kommer alltid att undersöka en genre för sig, så jag jämför exempelvis inte STnfict med SOfict. För att ett ord ska ha varit ”vanligare” i en korpus har jag sett till dess position i frekvenslistorna. Om skillnaden har varit mer än 6 positioner (exempelvis ett typord som har position 12 i SOfict och position 19 i STfict) har jag ansett att ordet är vanligare i den korpus där det har en högre position i frekvenslistan. Jag kommer inte att behandla alla ord som är vanligare i en viss korpus eftersom det, framförallt mot slutet av frekvenslistorna, är många ord som finns med i en korpus men inte i den andra inom samma genre. Korpusförekomsterna är nästintill lika många mot slutet av frekvenslistan (mellan två

positioner i frekvenslistan kan det skilja en eller två korpusförekomster av orden) vilket gör att det är svårt att yttra sig om huruvida ett ord är vanligare eller inte.

De fall som blir mest intressanta är när ett ord exempelvis är vanligare i båda de översatta korpusarna om man jämför med de icke-översatta korpusarna. Det skulle eventuellt kunna vara ett tecken på översättningssvenska, d.v.s. att målspråket har influerats av källspråket vid översättningen.

4.8 Statistisk signifikans

Jag kommer att räkna ut statistisk signifikans för resultaten för mönster nummer 1–4. Detta eftersom jag vill undersöka om resultaten beror på slumpen eller inte. Det visade sig vara svårt att ta reda på exakt hur jag skulle göra för att räkna ut signifikansen för just min undersökning eftersom det är en ja-/nej-fråga som undersöks. Men med hjälp av teknologie doktor Stefan Hellström (privat mejlkonversation 2014) lyckades jag få tag på de relevanta formlerna för att räkna ut signifikans för mitt material.

För varje värde för respektive delkorpus kommer jag först att räkna ut standardavvikelse med hjälp av denna formel (Mathematics Stack Exchange 2014):

$$\sqrt{\frac{p(1-p)}{n}} = S$$

I formeln står p för värdet jag räknat fram för respektive mönster och delkorpus, och n står för populationen (antal undersökta löpor för respektive mönster och delkorpus).

Därefter kommer jag att räkna ut de så kallade konfidensintervallen för varje mönster och delkorpus. Konfidensintervall är ett slags osäkerhetsintervall, och för att räkna ut det väljer man en konfidensgrad i förväg (exempelvis 95 % eller 99 %) enligt denna tabell:

| Prob = r | z-score = z_r |
|-----------------|-----------------------------------|
| 0,90 | 1,645 |
| 0,95 | 1,96 |
| 0,98 | 2,326 |
| 0,99 | 2,576 |

För min undersökning väljer jag en mycket hög konfidensgrad – 99 % eller 0,99 vilket gav mig ett z_r på 2,576. Detta värde, tillsammans med värdet för standardavvikelsen,

populationen och grundvärdet från undersökningen av mönstret kommer sedan sättas in i denna formel (Western Kentucky University 2013):

$$\bar{x} \pm \frac{z_r \times S}{\sqrt{n}} = \mu$$

\bar{x} står i denna formel för grundvärdet som jag fått fram för respektive korpus och mönster. S är standardavvikelsen och n är populationen. μ blir i detta fall ett konfidensintervall som visar mellan vilka värden resultatet från ett mönster för en delkorpus hamnar med 99 % säkerhet. Ett konfidensintervall skulle exempelvis kunna bli $42 \pm 1,4$. Värdena för detta exempel ligger då mellan 40,6 och 43,4. Jag kommer att jämföra konfidensintervallen för exempelvis STfict och SOfict för mönster 1, och om dessa två intervall inte överlappar har jag bevisat att mina resultat är statistiskt signifikanta. För att undvika att skriva så mycket siffror i löpande text kommer jag bara skriva om resultaten är statistiskt signifikanta eller ej. För att se de exakta konfidensintervallen se bilaga 4.

5. Resultat

I denna analysdel kommer jag att först gå igenom Laviosas fem mönster, ett efter ett, för att se vilka resultat min undersökning har gett för respektive mönster. Därefter kommer jag att gå in närmre på vissa specifika trender i ordförekomster.

5.1 Lexikal densitet

En text med låg andel innehållsord uppfattas ofta som enklare än en text med hög andel innehållsord. Innehållsorden bär informationen och ju högre andel innehållsord desto mer information blir det att ta in för mottagaren. Den lexikala densiteten kan alltså fungera som ett mått på hur komplex en text är (Halliday 1985:62). Enligt Laviosas första mönster för förenkling ska översatta texter ha en lägre andel innehållsord än icke-översatta texter (och alltså lägre lexikal densitet).

5.1.1 Skönlitterära texter

Efter att ha räknat antal olika innehållsord bland de 108 vanligaste typorden för SOfict-korpusen och STfict-korpusen stod det klart att det faktiskt var fler ord som var innehållsord

bland dessa 108 typord i STfict-korpusen än i SOfict-korpusen (STfict – 22, SOfict – 21). Men efter att ha adderat ihop alla korpusförekomster av dessa innehållsord stod det klart att andelen innehållsord för samtliga korpusförekomster av de 108 vanligaste typorden var 6,92 % för SOfict och 6,21 % för STfict. Laviosas första mönster för förenkling verkar alltså stämma för svenska översatta skönlitterära texter, och trots att skillnaden mellan delkorpusarna är liten gör det stora materialet att skillnaden är statistiskt signifikant.

| Mönster | STfict | SOfict |
|--|---------------|--------|
| 1. Andel innehållsord/lexikal densitet | 6,21 % | 6,92 % |

Tabell 6. Lexikal densitet, skönlitterära texter

5.1.2 Icke-skönlitterära texter

Av de 108 vanligaste typorden för SONfict-korpusen var 24 av dem innehållsord. För STNfict-korpusen fanns det 22 innehållsord bland de 108 vanligaste typorden. Efter en beräkning av samtliga korpusförekomster av de 24 respektive 22 innehållsorden för varje delkorpus stod det klart att andelen innehållsord för de 108 vanligaste orden i SONfict var 7,47 % och i STNfict 5,89 %. Laviosas första mönster för förenkling stämmer alltså även för de svenska översatta icke-skönlitterära, och för dessa texter var dessutom skillnaden mellan korpusarna något större. Skillnaden mellan delkorpusarna är statistisk signifikant.

| Mönster | STNfict | SONfict |
|--|---------------|---------|
| 1. Andel innehållsord/lexikal densitet | 5,89 % | 7,47 % |

Tabell 7. Lexikal densitet, icke skönlitterära texter

5.2 Andel vanliga lexikonord

Enligt Laviosas mönster ska översatta texter vara mer målspråkstypiska, vilket innebär att andelen ord som finns med bland svenskans 200 vanligaste ord borde vara högre i de översatta texternas frekvenslistor än de icke-översatta texternas frekvenslistor.

5.2.1 Skönlitterära texter

Av de 108 vanligaste typorden för SOfict fanns 91 av dem med bland Alléns (1972) 200 vanligaste lexikonord i svenskan. Frekvenslistan för STfict innehöll 94 typord som fanns med

bland svenskans 200 vanligaste lexikonord. Då löporden summerades för dessa ”vanliga lexikonord” för respektive korpus blev resultatet att STfict innehöll 94,8 % vanliga lexikonord baserat på frekvenslistan och SOfict 93,24 %. Den översatta korpusen har alltså en högre andel vanliga lexikonord, sett till frekvenslistorna, än de icke-översatta texterna, vilket följer Laviosas andra mönster för förenkling. Även för detta mönster är dock skillnaden mellan korpusarna mycket liten, men skillnaden är statistiskt signifikant.

| Mönster | STfict | SOfict |
|-----------------------------|----------------|---------|
| 2. Andel vanliga lexikonord | 94,82 % | 93,24 % |

Tabell 8. Andel vanliga lexikonord, skönlitterära texter

5.2.2 Icke-skönlitterära texter

För de icke-skönlitterära texterna var resultatet ett annat. Av de 108 vanligaste typorden för SONfict-korpusen fanns 106 av dem med bland svenskans 200 vanligaste lexikonord. För de STnfict-korpusen fanns 105 av de 108 vanligaste typorden med bland svenskans 200 vanligaste lexikonord. Efter uträkningen som genomfördes på samtliga korpusförekomster av dessa ”vanliga ord” blev resultatet att andelen vanliga lexikonord för frekvenslistan gjord på STnfict-korpusen var 99,34 %, och andelen vanliga lexikonord i SONfict-korpusen var 99,52 %. Gällande icke-skönlitterära texter var alltså andelen vanliga ord mindre i översatta texter, sett till frekvenslistorna, än för icke-översatta texter, och dessa resultat följer därmed inte Laviosas andra mönster för förenkling. Trots att skillnaden är ytterst liten är resultatet statistiskt signifikant.

I denna undersökning ser vi också att andelen vanliga lexikonord var markant högre i de icke-skönlitterära texterna än i de skönlitterära texterna. Detta resultat beror som tidigare nämnts troligtvis på att Alléns (1972) material bestod av icke-skönlitterära texter (tidningstexter).

| Mönster | STnfict | SONfict |
|-----------------------------|---------|----------------|
| 2. Andel vanliga lexikonord | 99,34 % | 99,52 % |

Tabell 9. Andel vanliga lexikonord, icke-skönlitterära texter

5.3 Upprepning av de vanligaste korpusorden

Att räkna ut hur stor del av delkorpusen de 108 vanligaste typordens korpusförekomster för varje delkorpus tar upp ger ett slags mått på hur ordvariationen ser ut i de olika texterna. Enligt Laviosas mönster ska de 108 vanligast typordens korpusförekomster i översatt text uppta större del av delkorpusen än icke-översatt text.

5.3.1 Skönlitterära texter

De skönlitterära texterna, icke-översatta och översatta, ligger väldigt nära varandra i denna undersökning. De 108 vanligaste typordens korpusförekomster i den icke-översatta delkorpusen upptar totalt 51,82 % av korpusen, medan de 108 vanligaste typordens korpusförekomster i den översatta delkorpusen upptar 52,04 % av delkorpusen. Laviosas tredje mönster verkar stämma för svenska skönlitterära texter, och skillnaden är statistiskt signifikant.

| Mönster | STfict | SOfict |
|--|----------------|---------|
| 3. Samtliga korpusförekomster av de 108 vanligaste typordens andel av korpusen | 52,04 % | 51,82 % |

Tabell 10. Upprepning av vanliga korpusord, skönlitterära texter

5.3.2 Icke-skönlitterära texter

Gällande de icke-skönlitterära texterna finns en tydligare trend när det kommer till upprepningen av vanliga korpusord. De översatta texternas 108 vanligaste typordens korpusförekomster upptar 53,12 % av delkorpusen, medan de 108 vanligaste typordens korpusförekomster för de icke-översatta texterna endast upptar 43,65 % av delkorpusen. Det är alltså markant skillnad mellan de två icke-skönlitterära delkorpuserna när det kommer till upprepning av de vanligaste korpusorden, och resultatet stödjer Laviosas tredje mönster för förenkling i översatta texter. Resultatet är statistiskt signifikant.

| Mönster | STnfict | SONfict |
|--|----------------|---------|
| 3. Samtliga korpusförekomster av de 108 vanligaste typordens andel av korpusen | 53,12 % | 43,65 % |

Tabell 11. Upprepning av vanliga korpusord, icke-skönlitterära texter

5.4 Antal olika lemman

Om en frekvenslista för en korpus innehåller få olika lemman kan detta ses som en indikator för att ordvariationen i texterna är låg och vice versa. Enligt Laviosas mönster för antal olika

lemman ska översatt text ha färre lemman bland de 108 vanligaste typorden än icke-översatt text har.

5.4.1 Skönlitterära texter

De 108 vanligaste typorden för den icke-översatta delkorporusen innehöll 79 olika lemman. De 108 vanligaste typorden för den översatta delkorporusen innehöll 76 olika lemman. Detta resultat tyder på att Laviosas fjärde mönster för förenkling stämmer för svensk översatt skönlitterär text. Skillnaden är statistiskt signifikant.

| Mönster | STfict | SOfict |
|----------------------------------|-----------|--------|
| 4. Antal lemman i frekvenslistan | 76 | 79 |

Tabell 12. Antal lemman, skönlitterära texter

5.4.2 Icke-skönlitterära texter

De 108 vanligaste typorden för den icke-översatta delkorporusen innehöll 84 olika lemman. De 108 vanligaste typorden för den översatta delkorporusen innehöll 82 olika lemman. Liksom i fallet för skönlitterära texter tycks Laviosas fjärde mönster stämma även för svensk översatt icke-skönlitterär text. Detta resultat är statistiskt signifikant, men inte till 99 % som för de andra mönstren, utan till 95 %. Det är dock fortfarande en mycket hög signifikansnivå.

| Mönster | STnfict | SONfict |
|----------------------------------|-----------|---------|
| 4. Antal lemman i frekvenslistan | 82 | 84 |

Tabell 13. Antal lemman, icke-skönlitterära texter

5.5 Hjälpverben *ha* och *vara*

Enligt Laviosa är ett mönster för förenkling att hjälpverben *ha* och *vara* förekommer oftare i presens i översatt text än i icke-översatt text. Samtidigt förekommer samma verb oftare i preteritum i icke-översatt text än i översatt text. Laviosa kom dock fram till att detta mönster endast stämde för de tidningstexter hon undersökte och inte för skönlitterära texter.

5.5.1 Skönlitterära texter

För de skönlitterära texterna är orden *är* och *hade* vanligare i SOfict-korporusen medan *har* är vanligare i STfict-korporusen. Ordet *var* ligger på samma plats i båda korpusarnas

frekvenslistor. Skillnaderna mellan korpusarna är dock mycket små, men detta mönster för förenkling verkar inte stämma för svenska översatta skönlitterära texter.

| Verb | STfict, position | SOfict, position |
|------|------------------|------------------|
| ÄR | 21 | 19 |
| HAR | 27 | 28 |
| VAR | 10 | 10 |
| HADE | 14 | 13 |

Tabell 14. Hjälppverb, skönlitterära texter

5.5.2 Icke-skönlitterära texter

För icke-skönlitterära texter blev resultatet det motsatta mot hypotesen. Hjälppverben *är* och *har* var vanligare i de icke-översatta texterna, medan *var* och *hade* var vanligare i de översatta texterna. Skillnaderna mellan positionerna var i de flesta fall större i de icke-skönlitterära korpusarna än i de skönlitterära. Detta mönster för förenkling stämmer alltså inte för svensk översatt icke-skönlitterär text.

| Verb | STnfict, position | SONfict, position |
|------|-------------------|-------------------|
| ÄR | 14 | 9 |
| HAR | 20 | 15 |
| VAR | 18 | 19 |
| HADE | 26 | 37 |

Tabell 15. Hjälppverb, icke-översatta texter

5.6 Resultatsammanfattning för Laviosas fem mönster

Resultaten av undersökningen får lov att betraktas som tendenser, snarare än fakta. Eftersom den korpus jag använt mig av bara behandlar språkparet engelska-svenska skulle man också behöva ta in översättningar från fler olika språk för att kunna säga att det inte är språkspecifika mönster som mina resultat visar på.

Materialet jag har använt mig av är relativt stort (mellan 300 000 och 350 000 löpord per delkorpus). Eftersom skillnaderna i resultaten mellan delkorpusarna är så små krävs det ett stort material för att kunna slå fast att de är statistiskt signifikanta.

| Mönster | STfict | SOfict | STnfict | SONfict |
|---|--------------------------------------|--------------------------------------|--------------------------------------|-------------------------------------|
| 1. Andel innehållsord/ lexikal densitet | 6,21 % | 6,92 % | 5,89 % | 7,47 % |
| 2. Andel vanliga lexikonord | 94,82 % | 93,24 % | 99,34 % | 99,52 % |
| 3. Samtliga korpusförekomster av de 108 vanligaste typordens andel av korpusen | 52,04 % | 51,82 % | 53,12 % | 43,65 % |
| 4. Antal lemman i frekvenslistan | 76 | 79 | 82 | 84 |
| 5. Position för olika tempus av hjälpverben <i>ha</i> och <i>vara</i> i frekvenslistorna | ÄR 21 HAR 27 VAR 10 HADE 14 | ÄR 19 HAR 28 VAR 10 HADE 13 | ÄR 14 HAR 20 VAR 18 HADE 26 | ÄR 9 HAR 15 VAR 19 HADE 37 |

Tabell 16. Resultat för samtliga korpusar

Om man jämför mina resultat med Laviosas ser man att hennes korpusar inte heller uppvisar så stora skillnader för vissa av de mönster jag har undersökt.

| Mönster | Skönlitterära översatta texter | Skönlitterära icke-översatta texter |
|--|--|---|
| 1. Lexikal densitet/ andel innehållsord | 52,87 % | 54,95 % |
| 2. Andel vanliga lexikonord | 59,74 % | 58,51 % |
| 3. Samtliga korpusförekomster av de 108 vanligaste typordens andel av korpusen | 56,2 % | 51,6 % |
| 4. Antal lemman i frekvenslistan | 82 | 87 |
| 5. Position för olika tempus av hjälpverben <i>to have</i> och <i>to be</i> i frekvenslistorna | IS 12 ARE 48 WAS 9 WERE 35 HAS 63 HAVE 31 HAD 21 | IS 25 ARE 61 WAS 8 WERE 31 HAS - HAVE 33 HAD 13 |

Tabell 17. Laviosas resultat, skönlitteratur

Laviosas olika korpusar, tidningskorpusarna och de skönlitterära korpusarna, skilde sig mycket åt i storlek. Tidningskorpusarna innehöll bara runt 50 000 löpord per delkorpus, och de skönlitterära korpusarna innehöll mellan 700 000 och 1 000 000 löpord. Hon räknade inte

ut statistisk signifikans för resultaten från undersökningen av tidningskorpusarna. Det gjorde hon däremot för de skönlitterära korpusarna där det stort klart att hennes resultat var statistiskt signifikanta.

Men för att återgå till resultaten från min undersökning, så verkar de alltså stämma för tre av de fyra mönstren för förenkling som enligt Laviosa går att finna i både översatt tidningstext och översatt skönlitterär text. Resultaten visar inte att det finns en större andel ”vanliga lexikonord” i svensk översatt icke-skönlitterär text än i svensk icke-översatt icke-skönlitterär text. Och gällande det mönster som endast stämde för tidningstext finns ingenting som tyder på att hjälpverben *vara* och *ha* skulle vara vanligare i presens i svenska översatta texter än i svenska icke-översatta texter.

5.7 Trender i ordförekomster

Frekvenslistorna visade att det fanns många ord som var vanligare i olika korpusar, men här tänkte jag fokusera på ett par av dessa – huvudsakligen sådana ord som var vanligare i antingen översatta texter eller i icke-översatta texter. Till att börja med observerade jag att ordet *ju* var vanligare i de icke-översatta texterna. I SOfict var det på position 71 och i SONfict position 105. *Ju* fanns inte alls med i frekvenslistorna för de översatta texterna, vilket skulle kunna ses som ett stöd för hypotesen att översatta texter innehåller färre unika element som är målspråkstypiska. Det är ett tecken på translationese (Tirkkonen-Condit 2002), och det faktum att *ju* inte finns med i frekvenslistorna för de översatta texterna kan vara ett tecken på översättningssvenska.

| Ord | STfict, position | SOfict, position | STnfict, position | SONfict, position |
|-----|------------------|------------------|-------------------|-------------------|
| JU | ej med på listan | 71 | ej med på listan | 105 |

Tabell 18. Positioner för ordet *ju*, samtliga korpusar

En annan trend som fanns i korpusarna var att vissa tidsadverb var vanligare i icke-översatta texter än i översatta texter. Detta gällde framförallt för orden *nu* och *då* som var mellan 15 och 30 positioner högre upp i de icke-översatta korpusarnas frekvenslistor än i de översatta korpusarnas frekvenslistor. En möjlig förklaring till detta skulle kunna vara att engelska texter kanske uttrycker tid med andra medel än *now* och *then* (som är direktöversättningarna av *nu* och *då*), och att översättarna har influerats av detta. I sådana fall är detta också ett tecken på översättningssvenska.

| Ord | STfict, position | SOfict, position | STnfict, position | SOnfict, position |
|-----|------------------|------------------|-------------------|-------------------|
| NU | 54 | 38 | 81 | 66 |
| DÅ | 70 | 37 | 77 | 59 |

Tabell 19. Positioner för orden nu och då, samtliga korpusar

Slutligen ville jag också undersöka om Nilssons (2004) hypotes att *av* är överrepresenterat i skönlitterära översatta texter stämmer även för icke-skönlitterära texter. Det visade det sig att det inte verkade göra vid observationen av positionerna för typordet *av* då det hade position 4 i SOnfict-korpusen och position 5 i STnfict-korpusen. Men om man däremot ser på frekvensernas procentandel av respektive delkorpus ser man att korpusförekomsterna för *av* upptar 1,93 % av SOnfict-korpusen och 2,42 % av STnfict-korpusen. Ur det hänseendet var alltså *av* vanligare i den översatta korpusen. Om man ska lita på procentsiffran eller positionen är svårt att säga eftersom tidigare undersökningar har visat att de vanligaste typorden upprepas oftare i den översatta korpusen än i den icke-översatta korpusen. Därför är det inte oväntat att korpusförekomsterna för ett typord på position 5 i STnfict-korpusen utgör en större andel av delkorpusen än vad korpusförekomsterna för ett typord på position 4 i SOnfict-korpusen gör. Det står dock klart att metoden att endast se på typordets position för att undersöka i vilken korpus ett specifikt ord är vanligare kan ha sina brister.

Nilsson genomförde sin undersökning av skönlitterära texter på samma material som jag har gjort min undersökning på. Mina skönlitterära frekvenslistor visar därför, precis som Nilssons resultat visar, att *av* är vanligare i de översatta texterna (position 13 i STfict och position 21 i SOfict).

| Ord | STfict, position | SOfict, position | STnfict, position | SOnfict, position |
|------------------------------------|------------------|------------------|-------------------|-------------------|
| AV, position | 13 | 21 | 5 | 4 |
| AV, andel av delkorpusen i procent | 1,16 % | 0,80 % | 2,42 % | 1,93 % |

Tabell 20. Positioner samt andel av korpusarna för ordet *av*, samtliga korpusar

Dessa fall är de enda relevanta iakttagelser som har gjorts. Resultaten av denna undersökning visar att inslagen av översättningssvenska inte verkar vara särskilt stora i det material som har undersökts.

6. Avslutande diskussion

Syftet med denna uppsats har varit att undersöka om svenska översatta texter generellt uppvisar fler tecken på förenkling än svenska icke-översatta texter. I den stora majoriteten av fallen är svaret ja, översättningar innehåller fler tecken på förenkling än icke-översatta texter. Svenska översatta texter har en lägre lexikal densitet än svenska icke-översatta texter. I översatta texter upprepas de vanligaste korpusorden oftare än i icke-översatta texter. Och de vanligaste 108 typorden i översatta texter innehåller färre lemman än de 108 vanligaste typorden i icke-översatta texter. Detta stämmer för både skönlitterära texter och icke-skönlitterära texter. De icke-skönlitterära korpusarna visar dock upp större skillnader sinsemellan för vissa mönster än vad de skönlitterära korpusarna gör.

Detta skulle kunna bero på att översättningsnormerna är olika i de olika genrerna. Lundquist (2007:37) förklarar att en funktionell översättningsstrategi, med vilken man fokuserar på att textens funktion blir det samma vid översättningen, är vanligast för översättning av fackprosa medan en imitativ översättningsstrategi, med vilken man fokuserar på källtextens ton och stil, är vanligast för översättning av skönlitteratur. Möjligtvis förenklas en text mer när man översätter med en funktionell strategi än med en imitativ strategi, eftersom man med en funktionell strategi inte behöver förhålla sig lika troget till källtextens ordalydelse. Om man översätter en skönlitterär text vill man, enligt dagens norm, ofta göra originalförfattaren rättvisa. Därför är det kanske ovanligt att man väljer att förenkla texten i översättningen. Översätter man däremot fackprosa är förenkling möjligtvis en naturlig del i översättningsprocessen eftersom översättaren strävar efter att förmedla informationen i texten till den nya mottagaren på ett så smidigt sätt som möjligt. I framtiden skulle det vara mycket intressant med en undersökning som fokuserar på översättningsuniversalier för olika genrer, speciellt eftersom House (2008) misstänker att översättningsuniversalier är genrespecifika.

Med tanke på det material ESPC har att erbjuda är det underligt att översättningsuniversalier i svenska översatta texter inte har undersökts mer. Det finns stor potential för framtida forskning om översättningsuniversalier i svenskan, både fler undersökningar om förenkling, men även om andra översättningsuniversalier. Det behövs stora korpusar med översatt och icke-översatt text för att tecken på översättningsuniversalier ska kunna undersökas, och ESPC är för närvarande den enda svenska korpusen med tillräckligt stort material för att forskare ska kunna genomföra sådana undersökningar. Dessvärre är copyright ett stort problem för både korpuskapare och forskare, men det vore ett

stort steg framåt inom översättningsvetenskapen om det fanns en korpus med översatta och icke-översatta texter där man som forskare kunde få tillgång till texterna i sin helhet för att ytterligare utvidga möjligheterna att undersöka översättningsuniversalier i svenska översatta texter.

Referenser

- Allén, S. (1972). *Tiotusen i topp. Ordfrekvenser i tidningstext*. Stockholm: Almqvist & Wiksell Förlag AB.
- Baker, M. (1993). "Corpus Linguistics and Translation Studies. Implications and Applications". I: Baker, M., Francis, G. och Tognini-Bonelli, E. (red.) *Text and Technology: In Honour of John Sinclair*. Amsterdam: John Benjamins, s.233-250.
- Baker, M. (1995). "Corpora in Translation Studies. An Overview and Suggestions for Future Research", *Target* 7 (2) s. 223-243.
- Baker, M. (1996). "Corpus-based translation studies: The challenges that lie ahead". I: Somers, H. (red.) *Terminology, LSP and Translation. Studies in language engineering in honour of Juan C. Sager*. Amsterdam och Philadelphia: John Benjamins Publishing Company, s.175-186.
- Blum-Kulka, S. och Levenston, E. A. (1983). "Universals of Lexical Simplification". I: Faerch, C. och Kasper, G. (red.) *Strategies in Interlanguage communication*. London: Longman, s119-139.
- Chesterman, A. (2004). "Beyond the particular". I: Mauranen, A. och Kujamäki, P. (red.) *Translation Universals: Do they exist?*. Amsterdam och Philadelphia: John Benjamins Publishing Company, s. 34-49.
- Gellerstam, M. (1986). "Translationese in Swedish novels translated from English" I: Wollin, L. och Lindquist, H. (red.) *Translation studies in Scandinavia: Proceedings from the Scandinavian Symposium on Translation Theory (SSOTT) II Lund 14-15 June, 1985*. Lund: CWK Gleerup, s. 88-95.
- Grabowski, Ł. (2013). "Interfacing corpus linguistics and computational stylistics: Translation universals in translational literary Polish". *International Journal of Corpus Linguistics*, 18 (2), s. 254-280
- Halliday, M. A. K. (1985). *Spoken and written language*. Victoria: Deakin University.
- Hellström, S. (2014). Privat mejlkonversation.
- House, J. (2008). "Beyond Intervention: Universals in Translation?". *Trans-kom*, 1 (1), s. 6-19.
- Laviosa-Braithwaite, S. (1996). *The English Comparable Corpus (ECC): a Resource and a Methodology for the Empirical Study of Translation*. PhD Thesis, Manchester, UMIST.
- Laviosa-Braithwaite, S. (1997). "Investigating Simplification in an English Comparable

- Corpus of Newspaper Articles”. I: Klaudy, K. och Kohn, J. (red.) *Transfere Necesses Est, Proceedings of the Second International Conference on Current Trends in Studies of Translation and Interpreting 5-7 September, 1996, Budapest, Hungary*. Budapest: Scholastica, s. 531-540.
- Laviosa, S. (1998). ”Core patterns of lexical use in a comparable corpus of English narrative prose”. *Meta*, 43 (4), s. 557–570.
- Laviosa, S. (2009). ”Universals”. I: Baker, M. och Saldanha, G. (red.), *Routledge Encyclopedia of Translation Studies*. London och New York: Routledge, s.306-310.
- Lundquist, L. (2007). *Oversættelse*. Fredriksberg: Samfundslitteratur.
- Malmkjær, K. (2005). *Linguistics and the Language of Translation*. Edinburgh: Edinburgh University Press.
- Mathematics Stack Exchange. (2014).
<http://math.stackexchange.com/questions/495346/expected-value-and-standard-deviation-for-sample>, hämtad 2014-05-16.
- Nationalencyklopedins nätutgåva. (2014). NE.se, sökord *lemma*, hämtad 2014-03-02, sökord *funktionsord*, hämtad 2014-04-22.
- Nilsson, P-O. (2004). ”Translation-specific lexicogrammar? Characteristic lexical and collocational patterning in Swedish texts translated from English”. I: Mauranen, A. och Kujamäki, P. (red.) *Translation Universals: Do they exist?* Amsterdam och Philadelphia: John Benjamins Publishing Company, s. 129-141.
- Piemontese, M. E. (1995). ”Il LIP: uno strumento per la didattica della lingua italiana in Italia e all'estero”. *Italica*, 72 (4), s.474-487.
- SAOL:s nätutgåva. (2011).
http://www.svenskaakademien.se/svenska_spraket/svenska_akademiens_ordlista/saol_paatet/ordlista, hämtad mellan 2014-02-20 och 2014-03-02.
- Tirkkonen-Condit, S. (2002). ”Translationese – a myth or an empirical fact?: A study into the linguistic identifiability of translated language”. *Target*, 14 (2), s. 207-220.
- Toury, G. (2004). ”Probabilistic explanations in translation studies: Welcome as they are, would they qualify as universals?” I: Mauranen, A. och Kujamäki, P. (red.) *Translation Universals: Do they exist?* Amsterdam och Philadelphia: John Benjamins Publishing Company, s. 15-32.
- Vanderauwera, R. (1985). *Dutch Novels Translated into English: The Transformation of a "Minority" Literature*. Amsterdam: Rodopi. Citerad i Laviosa-Braithwaite, S. (1997). ”Investigating Simplification in an English Comparable Corpus of Newspaper Articles”. I:

Klaudy, K. och Kohn, J. (red.) *Transfere Necesses Est, Proceedings of the Second International Conference on Current Trends in Studies of Translation and Interpreting 5-7 September, 1996, Budapest, Hungary*. Budapest: Scholastica, s. 531-540.

Vidén, G. (2014) ESPC:s webbplats

<http://www.sprak.gu.se/forskning/korpuslingvistik/korpusar-vid-spl/espc/> (hämtad 2014-03-03)

Western Kentucky University. (2013).

<http://people.wku.edu/david.neal/109/Unit4/ConfInt.pdf>, hämtad 2014-05-16.

Bilaga 1. Material

##Denna bilaga endast i pappersarkivexemplaret.##

Bilaga 2. Frekvensordlistor

Här följer frekvensordlistor för samtliga delkorpus. Först kommer en siffra som hänvisar till ordets position i listan, sedan kommer själva typordet, därefter visas frekvensen och slutligen visas en procentsiffra som beskriver hur stor del av delkorpusen förekomsterna av det specifika typordet upptar.

Frekvensordlista SOfict

| | | |
|----|--------|----------------|
| 1 | och | 11918 (3,88 %) |
| 2 | det | 7742 (2,42 %) |
| 3 | att | 6607 (2,15 %) |
| 4 | i | 6549 (2,13 %) |
| 5 | han | 6365 (2,01 %) |
| 6 | en | 5407 (1,76 %) |
| 7 | som | 5190 (1,69 %) |
| 8 | jag | 5108 (1,66 %) |
| 9 | på | 4861 (1,58 %) |
| 10 | var | 4835 (1,57 %) |
| 11 | inte | 3899 (1,27 %) |
| 12 | hon | 3895 (1,27 %) |
| 13 | hade | 3603 (1,17 %) |
| 14 | med | 3529 (1,15 %) |
| 15 | den | 3027 (0,99 %) |
| 16 | för | 2811 (0,92 %) |
| 17 | till | 2697 (0,88 %) |
| 18 | om | 2589 (0,84 %) |
| 19 | är | 2556 (0,83 %) |
| 20 | de | 2514 (0,82 %) |
| 21 | av | 2460 (0,80 %) |
| 22 | sig | 2434 (0,79 %) |
| 23 | men | 2380 (0,78 %) |
| 24 | ett | 2011 (0,66 %) |
| 25 | så | 1960 (0,64 %) |
| 26 | skulle | 1572 (0,51 %) |
| 27 | när | 1459 (0,48 %) |
| 28 | har | 1352 (0,44 %) |
| 29 | sa | 1261 (0,41 %) |
| 30 | du | 1243 (0,40 %) |
| 31 | mig | 1238 (0,40 %) |
| 32 | där | 1155 (0,38 %) |
| 33 | honom | 1129 (0,37 %) |
| 34 | man | 1124 (0,37 %) |
| 35 | vi | 1012 (0,33 %) |
| 36 | ut | 988 (0,32 %) |
| 37 | då | 971 (0,32 %) |
| 38 | nu | 948 (0,31 %) |
| 39 | vad | 892 (0,29 %) |
| 40 | kunde | 872 (0,28 %) |
| 41 | från | 871 (0,28 %) |

| | | |
|----|--------|--------------|
| 42 | henne | 843 (0,27 %) |
| 43 | hans | 827 (0,27 %) |
| 44 | bara | 822 (0,27 %) |
| 45 | över | 805 (0,26 %) |
| 46 | upp | 796 (0,26 %) |
| 47 | eller | 794 (0,26 %) |
| 48 | sin | 787 (0,26 %) |
| 49 | in | 779 (0,25 %) |
| 50 | här | 729 (0,24 %) |
| 51 | efter | 697 (0,23 %) |
| 52 | dem | 677 (0,22 %) |
| 53 | mot | 674 (0,22 %) |
| 54 | såg | 658 (0,21 %) |
| 55 | kom | 658 (0,21 %) |
| 56 | kan | 634 (0,21 %) |
| 57 | vid | 616 (0,20 %) |
| 58 | gick | 615 (0,20 %) |
| 59 | alla | 593 (0,19 %) |
| 60 | min | 572 (0,19 %) |
| 61 | hur | 571 (0,19 %) |
| 62 | ha | 570 (0,19 %) |
| 63 | något | 566 (0,18 %) |
| 64 | mycket | 551 (0,18 %) |
| 65 | vara | 542 (0,18 %) |
| 66 | fick | 531 (0,17 %) |
| 67 | än | 509 (0,17 %) |
| 68 | blev | 505 (0,16 %) |
| 69 | ner | 502 (0,16 %) |
| 70 | hennes | 492 (0,16 %) |
| 71 | ju | 484 (0,16 %) |
| 72 | utan | 476 (0,16 %) |
| 73 | åt | 473 (0,15 %) |
| 74 | någon | 463 (0,15 %) |
| 75 | måste | 454 (0,15 %) |
| 76 | aldrig | 452 (0,15 %) |
| 77 | sedan | 444 (0,14 %) |
| 78 | också | 440 (0,14 %) |
| 79 | allt | 435 (0,14 %) |
| 80 | ingen | 433 (0,14 %) |
| 81 | andra | 432 (0,14 %) |
| 82 | varit | 430 (0,14 %) |
| 83 | fram | 429 (0,14 %) |
| 84 | under | 413 (0,13 %) |
| 85 | själv | 410 (0,13 %) |
| 86 | tog | 406 (0,13 %) |
| 87 | stod | 404 (0,13 %) |
| 88 | ville | 402 (0,13 %) |
| 89 | igen | 395 (0,13 %) |
| 90 | ska | 388 (0,13 %) |
| 91 | vet | 383 (0,12 %) |
| 92 | sina | 383 (0,12 %) |
| 93 | få | 382 (0,12 %) |
| 94 | kanske | 381 (0,12 %) |

| | | |
|-----|-----------|--------------|
| 95 | två | 373 (0,12 %) |
| 96 | sitt | 372 (0,12 %) |
| 97 | gång | 369 (0,12 %) |
| 98 | hela | 365 (0,12 %) |
| 99 | ur | 365 (0,12 %) |
| 100 | sen | 359 (0,12 %) |
| 101 | mitt | 348 (0,11 %) |
| 102 | wallander | 348 (0,11 %) |
| 103 | se | 344 (0,11 %) |
| 104 | gå | 338 (0,11 %) |
| 105 | visste | 336 (0,11 %) |
| 106 | dom | 335 (0,11 %) |
| 107 | några | 332 (0,11 %) |
| 108 | satt | 329 (0,11 %) |

Frekvensordlista STfict

| | | |
|----|--------|----------------|
| 1 | och | 12077 (3,48 %) |
| 2 | att | 8528 (2,46 %) |
| 3 | det | 8462 (2,44 %) |
| 4 | i | 7121 (2,01 %) |
| 5 | en | 6505 (1,88 %) |
| 6 | som | 6494 (1,87 %) |
| 7 | på | 6282 (1,81 %) |
| 8 | han | 6281 (1,81 %) |
| 9 | jag | 5819 (1,68 %) |
| 10 | var | 5144 (1,48 %) |
| 11 | inte | 4330 (1,25 %) |
| 12 | med | 4256 (1,23 %) |
| 13 | av | 4028 (1,16 %) |
| 14 | hade | 3926 (1,13 %) |
| 15 | för | 3763 (1,10 %) |
| 16 | den | 3732 (1,08 %) |
| 17 | hon | 3691 (1,06 %) |
| 18 | till | 3305 (0,95 %) |
| 19 | om | 3147 (0,91 %) |
| 20 | de | 3003 (0,87 %) |
| 21 | är | 2894 (0,83 %) |
| 22 | sig | 2742 (0,79 %) |
| 23 | ett | 2553 (0,74 %) |
| 24 | men | 2107 (0,61 %) |
| 25 | så | 1968 (0,57 %) |
| 26 | skulle | 1797 (0,52 %) |
| 27 | har | 1692 (0,49 %) |
| 28 | när | 1664 (0,48 %) |
| 29 | du | 1550 (0,45 %) |
| 30 | mig | 1545 (0,45 %) |
| 31 | honom | 1404 (0,41 %) |
| 32 | sa | 1346 (0,39 %) |
| 33 | vi | 1324 (0,38 %) |
| 34 | där | 1258 (0,36 %) |
| 35 | från | 1169 (0,34 %) |
| 36 | hans | 1154 (0,33 %) |

| | | |
|----|--------|---------------|
| 37 | man | 1107 (0,32 %) |
| 38 | eller | 1090 (0,31 %) |
| 39 | ut | 1012 (0,29 %) |
| 40 | sin | 1009 (0,29 %) |
| 41 | vad | 989 (0,29 %) |
| 42 | upp | 963 (0,28 %) |
| 43 | kunde | 936 (0,27 %) |
| 44 | dem | 876 (0,25 %) |
| 45 | här | 855 (0,25 %) |
| 46 | ha | 853 (0,25 %) |
| 47 | henne | 830 (0,24 %) |
| 48 | in | 805 (0,23 %) |
| 49 | över | 790 (0,23 %) |
| 50 | kan | 778 (0,22 %) |
| 51 | bara | 777 (0,22 %) |
| 52 | hennes | 740 (0,21 %) |
| 53 | efter | 739 (0,21 %) |
| 54 | nu | 737 (0,21 %) |
| 55 | hur | 709 (0,20 %) |
| 56 | något | 703 (0,20 %) |
| 57 | min | 697 (0,20 %) |
| 58 | vid | 691 (0,20 %) |
| 59 | alla | 671 (0,19 %) |
| 60 | sedan | 665 (0,19 %) |
| 61 | ni | 639 (0,18 %) |
| 62 | mot | 635 (0,18 %) |
| 63 | kom | 624 (0,18 %) |
| 64 | än | 623 (0,18 %) |
| 65 | någon | 611 (0,18 %) |
| 66 | vara | 608 (0,18 %) |
| 67 | mycket | 600 (0,17 %) |
| 68 | såg | 597 (0,17 %) |
| 69 | andra | 568 (0,16 %) |
| 70 | då | 562 (0,16 %) |
| 71 | gick | 550 (0,16 %) |
| 72 | under | 530 (0,15 %) |
| 73 | få | 530 (0,15 %) |
| 74 | åt | 510 (0,15 %) |
| 75 | sina | 504 (0,15 %) |
| 76 | aldrig | 483 (0,14 %) |
| 77 | blev | 473 (0,14 %) |
| 78 | utan | 472 (0,14 %) |
| 79 | göra | 469 (0,14 %) |
| 80 | varit | 464 (0,13 %) |
| 81 | sade | 462 (0,13 %) |
| 82 | kommer | 462 (0,13 %) |
| 83 | fick | 442 (0,13 %) |
| 84 | måste | 441 (0,13 %) |
| 85 | sitt | 435 (0,13 %) |
| 86 | också | 435 (0,13 %) |
| 87 | ska | 435 (0,13 %) |
| 88 | fram | 428 (0,12 %) |
| 89 | tog | 427 (0,12 %) |

| | | |
|-----|--------|--------------|
| 90 | gjorde | 419 (0,12 %) |
| 91 | mer | 416 (0,12 %) |
| 92 | dig | 412 (0,12 %) |
| 93 | genom | 410 (0,12 %) |
| 94 | allt | 407 (0,12 %) |
| 95 | kanske | 406 (0,12 %) |
| 96 | år | 398 (0,11 %) |
| 97 | fanns | 396 (0,11 %) |
| 98 | se | 390 (0,11 %) |
| 99 | ingen | 388 (0,11 %) |
| 100 | bli | 388 (0,11 %) |
| 101 | ner | 387 (0,11 %) |
| 102 | två | 364 (0,11 %) |
| 103 | ta | 363 (0,10 %) |
| 104 | alltid | 359 (0,10 %) |
| 105 | oss | 351 (0,10 %) |
| 106 | själv | 348 (0,10 %) |
| 107 | deras | 347 (0,10 %) |
| 108 | vet | 347 (0,10 %) |

Frekvensordlista SOnfict

| | | |
|----|-------|----------------|
| 1 | och | 11919 (3,37 %) |
| 2 | i | 10507 (2,97 %) |
| 3 | att | 8490 (2,40 %) |
| 4 | av | 6806 (1,93 %) |
| 5 | som | 6571 (1,86 %) |
| 6 | en | 6392 (1,81 %) |
| 7 | det | 5535 (1,57 %) |
| 8 | för | 5451 (1,54 %) |
| 9 | är | 4466 (1,26 %) |
| 10 | på | 4372 (1,24 %) |
| 11 | till | 4243 (1,20 %) |
| 12 | den | 4080 (1,15 %) |
| 13 | med | 3996 (1,13 %) |
| 14 | de | 3490 (0,99 %) |
| 15 | har | 3241 (0,92 %) |
| 16 | ett | 2920 (0,83 %) |
| 17 | om | 2641 (0,75 %) |
| 18 | inte | 2304 (0,65 %) |
| 19 | var | 2017 (0,57 %) |
| 20 | kan | 1634 (0,46 %) |
| 21 | vi | 1573 (0,45 %) |
| 22 | eller | 1538 (0,44 %) |
| 23 | under | 1492 (0,42 %) |
| 24 | jag | 1463 (0,41 %) |
| 25 | han | 1387 (0,39 %) |
| 26 | från | 1387 (0,39 %) |
| 27 | man | 1383 (0,39 %) |
| 28 | sig | 1217 (0,34 %) |
| 29 | detta | 1207 (0,34 %) |
| 30 | också | 1147 (0,32 %) |
| 31 | men | 1103 (0,31 %) |

| | | |
|----|---------|---------------|
| 32 | så | 1098 (0,31 %) |
| 33 | vid | 1077 (0,30 %) |
| 34 | år | 946 (0,27 %) |
| 35 | skall | 826 (0,23 %) |
| 36 | andra | 824 (0,23 %) |
| 37 | hade | 782 (0,22 %) |
| 38 | denna | 773 (0,22 %) |
| 39 | skulle | 739 (0,21 %) |
| 40 | inom | 723 (0,20 %) |
| 41 | när | 711 (0,20 %) |
| 42 | än | 708 (0,20 %) |
| 43 | där | 696 (0,20 %) |
| 44 | vara | 694 (0,20 %) |
| 45 | finns | 664 (0,20 %) |
| 46 | mycket | 650 (0,18 %) |
| 47 | måste | 649 (0,18 %) |
| 48 | sin | 647 (0,18 %) |
| 49 | sverige | 645 (0,18 %) |
| 50 | alla | 642 (0,18 %) |
| 51 | genom | 634 (0,18 %) |
| 52 | även | 613 (0,17 %) |
| 53 | dessa | 601 (0,17 %) |
| 54 | över | 596 (0,17 %) |
| 55 | mot | 589 (0,17 %) |
| 56 | mellan | 557 (0,16 %) |
| 57 | efter | 556 (0,16 %) |
| 58 | kommer | 556 (0,16 %) |
| 59 | då | 551 (0,16 %) |
| 60 | första | 516 (0,15 %) |
| 61 | svenska | 504 (0,14 %) |
| 62 | får | 498 (0,14 %) |
| 63 | olika | 494 (0,14 %) |
| 64 | många | 493 (0,14 %) |
| 65 | utan | 492 (0,14 %) |
| 66 | nu | 486 (0,14 %) |
| 67 | nya | 456 (0,13 %) |
| 68 | stora | 447 (0,13 %) |
| 69 | mer | 443 (0,13 %) |
| 70 | samt | 434 (0,12 %) |
| 71 | allt | 421 (0,12 %) |
| 72 | del | 420 (0,12 %) |
| 73 | vad | 415 (0,12 %) |
| 74 | upp | 414 (0,12 %) |
| 75 | sina | 414 (0,12 %) |
| 76 | ha | 405 (0,11 %) |
| 77 | blev | 395 (0,11 %) |
| 78 | här | 390 (0,11 %) |
| 79 | något | 379 (0,11 %) |
| 80 | vill | 375 (0,11 %) |
| 81 | gäller | 374 (0,11 %) |
| 82 | hur | 369 (0,10 %) |
| 83 | hans | 368 (0,10 %) |
| 84 | in | 353 (0,10 %) |

| | | |
|-----|--------|--------------|
| 85 | få | 353 (0,10 %) |
| 86 | kunde | 352 (0,10 %) |
| 87 | därför | 351 (0,10 %) |
| 88 | enligt | 346 (0,10 %) |
| 89 | ut | 345 (0,10 %) |
| 90 | sedan | 345 (0,10 %) |
| 91 | bara | 340 (0,10 %) |
| 92 | dock | 329 (0,09 %) |
| 93 | två | 327 (0,09 %) |
| 94 | fick | 326 (0,09 %) |
| 95 | sätt | 323 (0,09 %) |
| 96 | någon | 320 (0,09 %) |
| 97 | stor | 317 (0,09 %) |
| 98 | några | 314 (0,09 %) |
| 99 | större | 309 (0,09 %) |
| 100 | hela | 309 (0,09 %) |
| 101 | tid | 309 (0,09 %) |
| 102 | ska | 308 (0,09 %) |
| 103 | kunna | 306 (0,09 %) |
| 104 | kom | 301 (0,09 %) |
| 105 | ju | 301 (0,09 %) |
| 106 | sitt | 299 (0,08 %) |
| 107 | bland | 298 (0,08 %) |
| 108 | vilket | 298 (0,08 %) |

Frekvensordlista STnfect

| | | |
|----|------|----------------|
| 1 | och | 11810 (4,10 %) |
| 2 | att | 9322 (3,20 %) |
| 3 | i | 8956 (3,10 %) |
| 4 | som | 7086 (2,43 %) |
| 5 | av | 7044 (2,42 %) |
| 6 | en | 6274 (2,15 %) |
| 7 | för | 5365 (1,84 %) |
| 8 | det | 5253 (1,80 %) |
| 9 | på | 4674 (1,61 %) |
| 10 | de | 4550 (1,56 %) |
| 11 | den | 4277 (1,47 %) |
| 12 | till | 4009 (1,38 %) |
| 13 | med | 3960 (1,36 %) |
| 14 | är | 3824 (1,31 %) |
| 15 | om | 2721 (0,93 %) |
| 16 | ett | 2679 (0,92 %) |
| 17 | inte | 2428 (0,83 %) |
| 18 | var | 2368 (0,81 %) |
| 19 | han | 2355 (0,81 %) |
| 20 | har | 2258 (0,78 %) |
| 21 | jag | 1925 (0,66 %) |
| 22 | vi | 1792 (0,62 %) |
| 23 | sig | 1661 (0,57 %) |
| 24 | från | 1437 (0,49 %) |
| 25 | men | 1428 (0,49 %) |
| 26 | hade | 1378 (0,47 %) |

| | | |
|----|---------|---------------|
| 27 | eller | 1338 (0,46 %) |
| 28 | detta | 1299 (0,45 %) |
| 29 | under | 1250 (0,43 %) |
| 30 | kan | 1149 (0,39 %) |
| 31 | så | 1123 (0,39 %) |
| 32 | när | 1008 (0,35 %) |
| 33 | skulle | 1004 (0,34 %) |
| 34 | skall | 968 (0,33 %) |
| 35 | sin | 872 (0,30 %) |
| 36 | andra | 840 (0,29 %) |
| 37 | man | 840 (0,29 %) |
| 38 | vid | 758 (0,26 %) |
| 39 | kommer | 728 (0,25 %) |
| 40 | denna | 715 (0,25 %) |
| 41 | dessa | 702 (0,24 %) |
| 42 | genom | 676 (0,23 %) |
| 43 | inom | 666 (0,23 %) |
| 44 | alla | 661 (0,23 %) |
| 45 | över | 657 (0,23 %) |
| 46 | där | 657 (0,23 %) |
| 47 | vara | 651 (0,22 %) |
| 48 | än | 648 (0,22 %) |
| 49 | år | 643 (0,22 %) |
| 50 | hans | 624 (0,21 %) |
| 51 | mycket | 620 (0,21 %) |
| 52 | sina | 611 (0,21 %) |
| 53 | måste | 596 (0,20 %) |
| 54 | efter | 577 (0,20 %) |
| 55 | vad | 575 (0,20 %) |
| 56 | dem | 554 (0,19 %) |
| 57 | mot | 539 (0,19 %) |
| 58 | finns | 535 (0,18 %) |
| 59 | också | 526 (0,18 %) |
| 60 | ha | 519 (0,18 %) |
| 61 | utan | 519 (0,18 %) |
| 62 | bara | 514 (0,18 %) |
| 63 | ut | 495 (0,17 %) |
| 64 | mer | 490 (0,17 %) |
| 65 | många | 470 (0,16 %) |
| 66 | upp | 468 (0,16 %) |
| 67 | första | 459 (0,16 %) |
| 68 | kunde | 459 (0,16 %) |
| 69 | deras | 439 (0,15 %) |
| 70 | nya | 428 (0,15 %) |
| 71 | artikel | 423 (0,15 %) |
| 72 | honom | 421 (0,14 %) |
| 73 | stora | 419 (0,14 %) |
| 74 | hur | 409 (0,14 %) |
| 75 | mellan | 387 (0,13 %) |
| 76 | oss | 382 (0,13 %) |
| 77 | då | 381 (0,13 %) |
| 78 | mig | 379 (0,13 %) |
| 79 | få | 378 (0,13 %) |

| | | |
|-----|--------------|--------------|
| 80 | något | 372 (0,13 %) |
| 81 | nu | 372 (0,13 %) |
| 82 | göra | 369 (0,13 %) |
| 83 | in | 369 (0,13 %) |
| 84 | sätt | 368 (0,13 %) |
| 85 | här | 362 (0,12 %) |
| 86 | del | 361 (0,12 %) |
| 87 | sitt | 352 (0,12 %) |
| 88 | hela | 345 (0,12 %) |
| 89 | allt | 339 (0,12 %) |
| 90 | dess | 339 (0,12 %) |
| 91 | någon | 338 (0,12 %) |
| 92 | blev | 337 (0,12 %) |
| 93 | varje | 333 (0,11 %) |
| 94 | fram | 325 (0,11 %) |
| 95 | kommissionen | 321 (0,11 %) |
| 96 | även | 312 (0,11 %) |
| 97 | sedan | 305 (0,10 %) |
| 98 | se | 300 (0,10 %) |
| 99 | hon | 298 (0,10 %) |
| 100 | vår | 298 (0,10 %) |
| 101 | får | 296 (0,10 %) |
| 102 | bli | 288 (0,10 %) |
| 103 | ta | 287 (0,10 %) |
| 104 | två | 284 (0,10 %) |
| 105 | några | 283 (0,10 %) |
| 106 | hos | 278 (0,10 %) |
| 107 | vilket | 276 (0,09 %) |
| 108 | miljoner | 271 (0,09 %) |

Bilaga 3. Lemmatiserade listor

| Lemmatiserad lista SOfict | |
|----------------------------------|--------------------------------|
| 1 | och |
| 2 | den det de dem dom |
| 3 | att |
| 4 | i |
| 5 | han honom hans |
| 6 | en ett |
| 7 | som |
| 8 | jag mig vi |
| 9 | på |
| 10 | inte |
| 11 | hon honne hennes |
| 12 | med |
| 13 | för |
| 14 | till |
| 15 | om |
| 16 | av |
| 17 | sig |
| 18 | men |
| 19 | så |
| 20 | när |
| 21 | sa |
| 22 | du |

23 där
24 man
25 ut
26 då
27 nu
28 vad
29 från
30 bara
31 över
32 upp
33 eller
34 sin
sitt
sina
35 in
36 här
37 efter
38 mot
39 se
såg
40 kom
41 kan
kunde
42 vid
43 gå
gick
44 alla
allt
45 min
mitt
46 hur
47 ha
har
hade
48 någon
något
några
49 mycket
50 vara
är
var
varit
51 få
fick
52 än
53 blev
54 ner
55 ju
56 utan
57 åt
58 måste
59 aldrig
60 sedan

| | |
|----|-----------|
| | sen |
| 61 | också |
| 62 | ingen |
| 63 | andra |
| 64 | fram |
| 65 | under |
| 66 | själv |
| 67 | tog |
| 68 | stod |
| 69 | ville |
| 70 | igen |
| 71 | ska |
| | skulle |
| 72 | vet |
| | visste |
| 73 | kanske |
| 74 | två |
| 75 | gång |
| 76 | hela |
| 77 | ur |
| 78 | wallander |
| 79 | satt |

Lemmatiserad lista STfict

| | |
|----|-------|
| 1 | och |
| 2 | att |
| 3 | den |
| | det |
| | de |
| | dem |
| | deras |
| 4 | i |
| 5 | en |
| | ett |
| 6 | som |
| 7 | på |
| 8 | han |
| | honom |
| | hans |
| 9 | jag |
| | mig |
| | vi |
| | oss |
| 10 | vara |
| | är |
| | var |
| | varit |
| 11 | inte |
| 12 | med |
| 13 | av |
| 14 | ha |
| | har |

15 hade
16 för
17 hon
18 henne
19 hennes
20 till
21 om
22 sig
23 men
24 så
25 ska
26 skulle
27 när
28 du
29 dig
30 ni
31 sade
32 sa
33 där
34 från
35 man
36 eller
37 ut
38 sin
39 sitt
40 sina
41 vad
42 upp
43 kan
44 kunde
45 här
46 in
47 över
48 bara
49 efter
50 nu
51 hur
52 någon
53 något
43 min
44 vid
45 allt
46 alla
47 sedan
48 mot
49 kommer
50 kom
51 än
52 mycket
53 se
54 såg
55 andra
56 då

| | |
|----|--------|
| 54 | gick |
| 55 | under |
| 56 | få |
| | fick |
| 57 | åt |
| 58 | aldrig |
| 59 | bli |
| | blev |
| 60 | utan |
| 61 | göra |
| | gjorde |
| 62 | måste |
| 63 | också |
| 64 | fram |
| 65 | ta |
| | tog |
| 66 | mer |
| 67 | genom |
| 68 | kanske |
| 69 | år |
| 70 | fanns |
| 71 | ingen |
| 72 | ner |
| 73 | två |
| 74 | alltid |
| 75 | själv |
| 76 | vet |

Lemmatiserad lista SOnfict

| | |
|----|-------|
| 1 | och |
| 2 | i |
| 3 | att |
| 4 | av |
| 5 | som |
| 6 | en |
| | ett |
| 7 | den |
| | det |
| | de |
| 8 | för |
| 9 | vara |
| | är |
| | var |
| 10 | på |
| 11 | till |
| 12 | med |
| 13 | ha |
| | har |
| | hade |
| 14 | om |
| 15 | inte |
| 16 | kunna |

| | |
|----|---------|
| | kan |
| | kunde |
| 17 | jag |
| | vi |
| 18 | eller |
| 19 | under |
| 20 | han |
| | hans |
| 21 | från |
| 22 | man |
| 23 | sig |
| 24 | också |
| 25 | men |
| 26 | så |
| 27 | vid |
| 28 | år |
| 29 | skall |
| | ska |
| | skulle |
| 30 | andra |
| 31 | inom |
| 32 | när |
| 33 | än |
| 34 | där |
| 35 | finns |
| 36 | mycket |
| 37 | måste |
| 38 | sin |
| | sitt |
| | sina |
| 39 | sverige |
| 40 | alla |
| | allt |
| 41 | genom |
| 42 | även |
| 43 | över |
| 44 | mot |
| 45 | mellan |
| 46 | efter |
| 47 | kommer |
| | kom |
| 48 | då |
| 49 | första |
| 50 | detta |
| | denna |
| | dessa |
| 51 | svenska |
| 52 | för |
| 53 | olika |
| 54 | många |
| 55 | utan |
| 56 | nu |
| 57 | nya |

| | |
|----|--------|
| 58 | stor |
| | stora |
| | större |
| 59 | samt |
| 60 | del |
| 61 | vad |
| 62 | upp |
| 63 | blev |
| 64 | här |
| 65 | någon |
| | något |
| | några |
| 66 | vill |
| 67 | gäller |
| 68 | hur |
| 69 | in |
| 70 | få |
| | får |
| | fick |
| 71 | därför |
| 72 | enligt |
| 73 | ut |
| 74 | sedan |
| 75 | bara |
| 76 | dock |
| 77 | två |
| 78 | sätt |
| 79 | hela |
| 80 | tid |
| 81 | ju |
| 82 | bland |
| 83 | vilket |
| 84 | mer |

Lemmatiserad lista STnfict

| | |
|----|-------|
| 1 | och |
| 2 | att |
| 3 | i |
| 4 | som |
| 5 | av |
| 6 | en |
| | ett |
| 7 | för |
| 8 | den |
| | det |
| | de |
| | dem |
| | dess |
| | deras |
| 9 | på |
| 10 | till |
| 11 | med |

12 vara
är
var
13 om
14 inte
15 han
honom
hans
16 ha
har
hade
17 jag
mig
vi
oss
18 sig
19 från
20 men
21 eller
22 denna
detta
dessa
23 under
24 kan
kunde
25 så
26 när
27 skall
skulle
28 sin
sitt
sina
29 andra
30 man
31 vid
32 kommer
33 genom
34 inom
35 alla
allt
36 över
37 där
38 än
39 år
40 mycket
41 efter
42 vad
43 mot
44 finns
45 också
46 utan
47 bara
48 ut

| | |
|----|--------------|
| 49 | måste |
| 50 | många |
| 51 | upp |
| 52 | första |
| 53 | nya |
| 54 | artikel |
| 55 | stora |
| 56 | hur |
| 57 | mellan |
| 58 | då |
| 59 | få |
| | får |
| 60 | någon |
| | något |
| | några |
| 61 | nu |
| 62 | göra |
| 63 | in |
| 64 | sätt |
| 65 | här |
| 66 | del |
| 67 | hela |
| 68 | bli |
| | blev |
| 69 | varje |
| 70 | fram |
| 71 | kommissionen |
| 72 | även |
| 73 | sedan |
| 74 | se |
| 75 | hon |
| 76 | vår |
| 77 | ta |
| 78 | två |
| 79 | hos |
| 80 | vilket |
| 81 | miljoner |
| 82 | mer |

Bilaga 4. Konfidensintervall

Mönster 1. Lexikal densitet

Grundvärden:

STfict: 6,21 %

SOfict: 6,92 %

STnfict: 5,89 %

SONfict: 7,47 %

Konfidensintervall:

STfict: $6,21 \pm 0,0003$

Värdet ligger mellan 6,2097 % och 6,2103 % med 99 % säkerhet

SOfict: $6,92 \pm 0,0004$

Värdet ligger mellan 6,9196 % och 6,9204 % med 99 % säkerhet
Ingen överlappning

STnfict: $5,89 \pm 0,0004$

Värdet ligger mellan 5,8896 % och 5,8904 % med 99 % säkerhet

SONfict: $7,47 \pm 0,0004$

Värdet ligger mellan 7,4696 % och 7,4704 % med 99 % säkerhet

Ingen överlappning

Mönster 2. Andel vanliga lexikonord

Grundvärden:

STfict: 94,82 %

SOfict: 93,24 %

STnfict: 99,34 %

SONfict: 99,52 %

Konfidensintervall:

STfict: $94,82 \pm 0,0003$

Värdet ligger mellan 94,8197 % och 94,8203 % med 99 % säkerhet

SOfict: $93,24 \pm 0,0004$

Värdet ligger mellan 93,2396 % och 93,2404 % med 99 % säkerhet

Ingen överlappning

STnfict: $99,34 \pm 0,0001$

Värdet ligger mellan 99,3399 % och 99,3401 % med 99 % säkerhet

SONfict: $99,52 \pm 0,0001$

Värdet ligger mellan 99,5199 % och 99,5201 % med 99 % säkerhet

Ingen överlappning

Mönster 3. Upprepning av vanligaste korpusord

Grundvärden:

STfict: 52,04 %

SOfict: 51,82 %

STnfict: 53,12 %

SONfict: 43,65 %

Konfidensintervall:

STfict: $52,04 \pm 0,0004$

Värdet ligger mellan 52,0496 % och 52,0404 % med 99 % säkerhet

SOfict: $51,82 \pm 0,0004$

Värdet ligger mellan 51,8196 % och 51,8204 % med 99 % säkerhet

Ingen överlappning

STnfict: $53,12 \pm 0,0004$

Värdet ligger mellan 53,1196 % och 53,1204 % med 99 % säkerhet

SONfict: $43,65 \pm 0,0004$

Värdet ligger mellan 43,6496 % och 43,6504 % med 99 % säkerhet

Ingen överlappning

Mönster 4. Antal lemman

För att kunna räkna ut konfidensintervall för mönster 4 har jag varit tvungen att räkna ut hur många lemman som har funnits i varje frekvenslista i procent. Jag har alltså dividerat antalet olika lemman med 108 för respektive delkorpus resultat.

Grundvärden:

STfict: 70,4 %

SOfict: 73,15 %

STnfict: 75,9 %

SONfict: 77,8 %

Konfidensintervall:

STfict: $70,4 \pm 1,09$

Värdet ligger mellan 69,31 % och 71,49 % med 99 % säkerhet

SOfict: $73,15 \pm 1,06$

Värdet ligger mellan 72,09 % och 74,21 % med 99 % säkerhet
Ingen överlappning

STnfict: $75,9 \pm 1,02$

Värdet ligger mellan 74,88 % och 76,92 % med 99 % säkerhet

SONfict: $77,8 \pm 0,99$

Värdet ligger mellan 76,81 % och 78,99 % med 99 % säkerhet

I detta fall finns det en överlappning mellan de två intervallen (76,92 % överlappar 76,81 %).

Därför provade jag att räkna ut konfidensintervallen med 95 % säkerhet och detta blev resultatet av det:

STnfict: $75,9 \pm 0,78$

Värdet ligger mellan 75,12 % och 76,68 % med 95 % säkerhet

SONfict: $77,8 \pm 0,75$

Värdet ligger mellan 77,05 % och 78,55 % med 95 % säkerhet

Ingen överlappning