

LUNDS UNIVERSITET

Språk- och litteraturcentrum

Nordiska språk

Max Sandberg

C-UPPSATS

Svenska: språklig inriktning

SVEK10

HT 2024

Spöklika fotspår i snön

Författarskapsanalys av skönlitterärt material genom
rättslingvistisk metod



LUNDS
UNIVERSITET

Handledare: Anna Smålander

Innehållsförteckning

| | | |
|-------|---|----|
| 1 | Inledning..... | 4 |
| 2 | Syfte och frågeställningar..... | 4 |
| 3 | Bakgrund..... | 5 |
| 3.1 | Rättslingvistik som forskningsområde..... | 5 |
| 3.2 | Variabler..... | 6 |
| 3.2.1 | Vanligaste ord..... | 6 |
| 3.2.2 | N-gram..... | 7 |
| 3.2.3 | Konjunktioner..... | 8 |
| 4 | Material..... | 9 |
| 4.1 | Korpusar..... | 9 |
| 4.2 | Städning av material..... | 9 |
| 5 | Metod..... | 10 |
| 5.1 | Variabler..... | 10 |
| 5.2 | Korpusverktyg..... | 10 |
| 5.3 | Metod för varje variabel..... | 11 |
| 5.3.1 | Vanligaste ord..... | 11 |
| 5.2.3 | N-gram..... | 12 |
| 5.3.3 | Konjunktioner..... | 12 |
| 6 | Resultat..... | 13 |
| 6.1 | Vanligaste ord..... | 13 |
| 6.1.1 | Jämförelse av målkorpusarnas listor..... | 14 |
| 6.1.2 | Ord som saknas i målkorpusarna..... | 15 |
| 6.1.3 | Ord som saknas i referenskorpusen..... | 16 |
| 6.1.4 | Innehållsord..... | 16 |
| 6.1.5 | Sammanfattning..... | 18 |
| 6.2 | N-gram..... | 19 |
| 6.2.1 | Likheter och skillnader mellan målkorpusarna..... | 19 |

| | |
|---|----|
| 6.2.2 Återkommande fyrgram..... | 22 |
| 6.2.3 Sporadiska fyrgram..... | 23 |
| 6.2.5 Sammanfattning..... | 24 |
| 6.3 Konjunktioner..... | 25 |
| 6.3.1 Konjunktioner per 1000 ord..... | 26 |
| 6.3.2 Konjunktioner per grafisk mening..... | 26 |
| 7 Diskussion..... | 27 |
| 7.1 Vanligaste ord..... | 28 |
| 7.2 N-gram..... | 30 |
| 7.3 Konjunktioner..... | 31 |
| 8 Avslutande ord..... | 32 |

1 Inledning

Det blåser på toppen. Camilla Läckberg är en av Sveriges största författare, med över 38 miljoner böcker sålda världen över (Shimoda 2024). År 2003 slog hon igenom med debutromanen *Isprinsessan*, som senare kom att bli det första verket i den välkända deckarserien *Fjällbacka*.

Under 2021, och nio tillskott i *Fjällbacka*-serien senare, ställs Läckberg inför kritik efter att författaren Håkan Lindberg publicerat boken *Spökskrivaren*. I verket gestaltas en man som arbetar som spökskrivare åt den framgångsrika kvinnliga deckarförfattaren Milli Månsson. Trots att boken är menad som satir, leder de många slående likheterna mellan Månsson och Läckberg till att Läckbergs författarskap ifrågasätts (Sveriges Radio 2023, 5:19). Läckberg förnekar dock att hon själv skulle ha använt en spökskrivare för någon av sina böcker (Dahlgren & Ek 2021).

Under 2023 tycks misstankarna bekräftas. I en undersökning publicerad i *Kvartal* utför kulturskribenten Lapo Lappin en stilistisk undersökning av Läckbergs böcker, och finner att flera av verken avviker från hennes etablerade språkliga stil. Efter vidare analys fastställs att stilen liknar den av Läckbergs redaktör: författaren Pascal Engman (Lapin 2023). När Läckberg konfronteras med fynden förnekar hon återigen att hon skulle ha använt en spökskrivare, men medger att Engman var mycket delaktig under böckernas skriv- och redigeringsprocess (TT 2023).

Händelseförloppet väcker intressanta frågor: har problemet djupare rötter? Kan liknande tecken på spökskrivning hittas i Läckbergs tidigare verk, mer specifikt i *Fjällbacka*-serien? Ifall Läckberg har skrivit dessa verk, borde de innehålla genomgående och mätbara drag som är typiska för Läckbergs egna språkbruk. I verken blottas, i sådana fall, Läckbergs idiolekt.

2 Syfte och frågeställningar

Syftet med föreliggande undersökning är att, med hjälp av forensisk-lingvistiska metoder, analysera hur tre skönlitterära böcker förhåller sig till varandra i fråga om tre variabler: vanligaste ord, n-gram och användning av konjunktioner. Genom uppsatsen kommer följande frågor att besvaras:

1. Hur skiljer sig de vanligaste orden åt i *Isprinsessan*, *Fyrvaktaren* och *Änglamakerskan*?

2. Hur skiljer sig n-gram åt i böckerna?
3. Hur skiljer sig antalet konjunktioner åt?
4. Vad indikerar resultaten kring huruvida någon av böckerna är spökskrivna?

Mot bakgrund av resultatet är målet att diskutera huruvida Camilla Läckberg skrev samtliga 3 böcker, eller om någon eller några kan ha spökskrivits.

3 Bakgrund

3.1 Rättslingvistik som forskningsområde

I artikeln ”Ordet fanger. Retslingvistik i en dansk kontekst” (2017) av den danska lingvisten Tanya Karoli Christensen kartläggs den *forensiska lingvistikens* begynnelse. Enligt Christensen myntades begreppet, även känt som *rättslingvistik*, av den svenska lingvisten Jan Svartvik i sin artikel ”The Evans Statements” från 1968. Där utför Svartvik en språklig analys på 2 brev, i vilka den påstådda författaren – Timothy John Evans – erkänner mordet på sin fru och dotter. I Evans rättegång spelade breven en avgörande roll, och resulterade i att han dömdes till döden och avrättades år 1950 (Christensen 2017, s. 172). Christensen beskriver hur Svartvik fann att språket i brevens mest komprometterande avsnitt avvek från resterande delar, exempelvis i form av placeringen av tidsadverbial: det vardagliga *Then I (...)* blev det mer formella *I then (...)* (s. 172). Utifrån sina resultat landade Svartvik i att Evans erkännande sannolikt var förfalskat, och att han avrättats på felaktiga grunder (Christensen 2017, s. 172).

Sedan Svartviks uppmärksammade artikel har rättslingvistik som forskningsområde utvecklats relativt långsamt, vilket Christensen (2017) tillskriver dess avsaknad av en allmänt erkänd metod (s. 206). Det finns idag ingen universell mall som beskriver hur rättslingvistiska analyser av text ska gå till, och därför kommer metoden för varje ny analys att variera beroende på kvalitén och omfattningen av materialet. Det är därmed upp till analytikern att bestämma dels vilka variabler som ska studeras, dels hur variablernas värden ska beräknas och skattas (Coulthard & Johnson, s. 14).

Ett vanligt tillvägagångssätt för en rättslingvistisk analys, som är relevant för föreliggande uppsats, är att utföra en *författarskapsanalys* (Christensen 2017, s. 187). I en sådan utgörs materialet av två eller flera olika texter: en text med ifrågasatt (*Q*) eller okänd (*U*) författare, och en med en känd (*K*) författare. Man studerar författarens språkliga spår i den

kända texten, med andra ord: man studerar författarens *idiolekt*. Därefter undersöker man ifall samma drag återfinns i de okända texterna. Ifall dragen är genomgående mellan texterna, indikerar detta att samma författare ligger bakom. Ifall dragen saknas i Q eller U, indikerar detta att texterna har olika författare.

3.2 Variabler

I detta avsnitt introduceras tre studier som ligger till grund för undersökningen. Varje studie presenterar en variabel som är bevisat betydelsefull för en språkbrukares idiolekt.

3.2.1 Vanligaste ord

I artikeln ”Towards a corpus-based methodology of forensic authorship attribution: a comparative study of two idiolects” (2003) undersöker lingvisten Krzysztof Kredens begreppet *idiolekt*. Enligt honom kan många acceptera idén av ett eget språkbruk, trots att konkreta bevis på ett sådant ännu inte kan framställas och presenteras på grund av avsaknaden av en etablerad metod (s. 405). Av denna anledning uppfattas rättslingvistisk bevisning ofta som spekulativ, vilket gör den mindre slagkraftig jämfört med andra forensiska bevis såsom DNA-spår (ibid.). Kredens menar att en mer systematisk och empiriskt inriktad metod hade lett till mindre spekulativa resultat. Frågan är då vilka variabler som är mest lämpliga för en sådan undersökning: vilka språkliga drag går att systematiskt framställa, och är samtidigt betydande för en författares idiolekt (s. 406)?

Studien består av två separata, cirka 15 minuter långa, sekvenser av talspråk från två olika musiker. Både musikerna och deras respektive tal har mycket gemensamt. Talarna – Robert James Smith respektive Steven Patrick Morrissey – är båda födda 1959, härstammar från arbetarklassen och har spelat i kända rockband som turnerade runt jorden under 70- och 80-talet. Musikernas talsekvenser, som utgör varsin delkorpus, är hämtade från diverse intervjuer. Materialet blir i hög grad homogent, vilket innebär att eventuella skillnader utmärker sig tydligare och blir enklare att uppmärksamma (Kredens 2003, s. 407). I slutändan landar Kredens i att talen skiljer sig åt i 5 avseenden: vanligast använda ord, frekvensen av gradadverbial och modifierare, samt frekvensen och användningen av diskurspartiklar och bisatser. Två av dessa gav särskilt starka utslag: vanligast använda ord och diskurspartiklar.

Dessa två uppnådde ett p-värde på 0,001, vilket klassificeras som kraftigt statistiskt signifikant. Skillnaderna mellan talarnas vanligast använda ord och diskurspartiklar är alltså extremt osannolika att bero på slumpen (s. 440).

För min undersökning är Kredens (2003) artikel viktig i två avseenden. Hans studie liknar min i att båda är en typ av *simulation*, där analytikern väljer materialet de ska studera (s. 407). På så vis skiljer sig denna undersökning från rättslingvistiska analyser i praktiken, då man ofta tilldelas materialet som ska granskas, såsom i fallet av en utredare som ska analysera ett hotbrev. I och med valfriheten ges möjlighet att homogenisera materialet i syfte att framställa delkorpusarnas eventuella skillnader på ett så tydligt sätt som möjligt. Denna idé låg till grund för mina val av material, till exempel avseende böckernas genre och omfång.

Studien är även värdefull genom dess resultat, och de fem variablerna som utkristalliseras: vanligast använda ord, frekvensen av gradadverbial och modifierare, samt frekvensen och användningen av diskurspartiklar och bisatser. Två av dessa – vanligast använda ord och diskurspartiklar – gav tydligast resultat, vilket innebär att de är mest lämpliga för denna undersökning. Viktigt att tänka på är att materialet i Kredens studie utgörs av naturligt tal, vilket är vitt skilt från mitt materials redigerade och korrekturlästa text. Vanliga exempel på diskurspartiklar, såsom *typ, ju, eller alltså*, bör därför vara extremt sällsynta i mitt material. Att utgå från dem i min analys hade alltså inte varit meningsfullt, och därför är de vanligast använda orden mer lämpliga.

3.2.2 N-gram

I artikeln ”Using n-grams to identify authors and idiolects: A corpus approach to a forensic linguistic problem” (2017) studerar den forensiska lingvisten David Wright *n-gram* och deras användbarhet i författarskapsanalyser. N-gram innebär *en sekvens av n enheter i följd* (s. 11). Man intresserar sig alltså inte för användningen av enskilda ord, som i 3.2.1, utan snarare för användningen av specifika ordföljder. I analysen byts *n* ut mot en siffra som motsvarar ordföljdens totala mängd ord. Till exempel kan ett tvågram innefatta fraser som *jag vill, en katt* och *ibland så*.

Wrights (2017) analys utgår från en korpus på totalt 2,5 miljoner ord skrivna av 176 författare (s. 13). Av dessa författare väljs tolv slumpvis ut. Utifrån de utvalda författarnas texter görs slumpvalda urval. Utifrån dessa urval bildas en uppfattning om författarens mest använda n-gram. Wright undersöker sedan ifall insikterna kring författarens vanliga n-gram kan

användas för att identifiera okända författare i korpusens övriga texter. Exempelvis: säg att författare A använder tvågrammet *jag tycker* ovanligt ofta, sett till övriga författare. Kan denna insikt användas för att identifiera författaren av texter där samma tvågram förekommer? Om ja, hur ofta är svaret korrekt? Det vill säga: vilken typ av n-gram har starkast implikationer för en persons idiolekt?

Av de sex typerna visade sig fyrgram, tätt följt av tregram, vara säkrast för att identifiera författare (Wright 2017, s. 20). Enligt Wright är fyrgrammens framgång resultatet av två faktorer: de är tillräckligt långa för att inte förekomma i de flesta texter – ett problem som tvågram ofta stöter på – och de är tillräckligt korta för att kunna skrivas i flera sammanhang. Fastän längre ordföljder, såsom sexgram, ofta har starkare betydelse för en idiolekt, är de ofta för specifika för att kunna anpassas till olika texter med olika sammanhang. Fyrgram utgör därför en bra balans mellan att vara tillräckligt vanliga för att kunna studeras, men samtidigt tillräckligt distinkta (s. 21).

Wrights artikel har varit viktig för mitt arbete i några avseenden. Först och främst visar artikeln ett samband mellan n-gram och idiolekt. Vidare är artikeln informativ kring vilket sorts n-gram som det är mest lämpligt att utgå ifrån, nämligen fyrgram.

3.2.3 Konjunktioner

I sin undersökning ”Assessing the Level of Stability of Idiolectal Features across Modes, Topics and Time of Text Production” (2018) granskar Tatiana Litvinova, Olga Litvinova och Pavel Seredin stabiliteten av idiolekter i både text och tal.

Studien sker genom fyra lingvistiska experiment som testar hur olika omständigheter, såsom tid eller ämne, påverkar subjektens idiolekter. I sin analys lägger forskarna stor vikt vid språkliga drag som visar *låg intravariabilitet* och *hög intervariabilitet* (Litvinova m.fl. 2018, s. 226). Detta kännetecknar drag som är konstanta för en enskild författare men som samtidigt varierar kraftigt från person till person. Av detta bedöms dragen utgöra sannolika aspekter av författarens idiolekt.

Resultatet visar att tre drag var mest stabila genom tids- och ämnesbyte: antal konjunktioner, diskurspartiklar och punkter (Litvinova m.fl. 2018, s. 228). Detta fastställs efter att antalet av, exempelvis, konjunktioner divideras med *Elementary Discourse Units* (EDU). En EDU kan beskrivas som någon typ av sats, alltså en meningsbärande enhet som antingen

fungerar självständigt (huvudsats) eller inte (bisats). På så vis mäter man hur ofta konjunktioner förekommer i förhållande till textens totala antal satser.

För undersökningen är de tre variablerna – konjunktioner, diskurspartiklar och punkter – värdefulla. I fråga om skönlitterärt material är diskurspartiklar svåra att studera eftersom de sällan förekommer i publicerad och korrekturläst litteratur, vilket redogjordes för i avsnitt 3.2.1. Av de två kvarstående variablerna – konjunktioner och punkter – kan författarens användning av punkter tänkas påverkas mest av det faktum att böckerna är korrekturläst och publicerad skönlitteratur. I och med att böckerna inriktar sig till den breda allmänheten, finns särskilda krav på meningslängd och lätläsighet. Antalet punkter i böckerna ger således inte en inblick i Läckbergs egna interpunktion. Att istället granska konjunktioner kan därför tänkas ge större insikter i Läckbergs egna språkbruk.

4 Material

4.1 Korpusar

Materialet utgörs i första hand av tre skönlitterära böcker: *Isprinsessan* (2003), *Fyrvaktaren* (2009) och *Änglamakerskan* (2011). Böckerna utgör varsin målkorpus: *Isprinsessan* är målkorpus A, *Fyrvaktaren* är målkorpus B och *Änglamakerskan* är målkorpus C. Valet av *Isprinsessan* motiveras av att detta var Läckbergs första bok. Därmed kan den tänkas innehålla många spår av Läckbergs naturliga språk. Hennes idiolekt blir på så vis tydligare, vilket gynnar undersökningen.

Fyrvaktaren och *Änglamakerskan* motiveras utifrån Läckbergs tilltagande popularitet. 2009 till 2011 var bland hennes mest framgångsrika år som författare. Vid detta laget hade hon slagit igenom nationellt, då hon 2009 blev den bäst säljande författaren i Sverige. Därtill hade hon etablerat sig inom Europa, då hon samma år var Europas sjätte bäst säljande författare (TT, 2010).

I materialet ingår även en skönlitterär referenskorpus som är hämtad från *Svensk diakronisk korpus* (2022). Denna korpus utgörs av 284 svenska romaner på totalt 19.545.718 ord, som publicerats mellan 1910 och 2002.

4.2 Städning av material

Referenskorpusen innehöll ursprungligen 384 verk på totalt 29 miljoner ord, men cirka 100 romaner valdes att tas bort eftersom de skrevs innan eller under stavningsreformen 1906. I och med elimineringen överensstämmer en högre grad av romanernas språk med moderna, svenska språknormer. De blir således mer jämförbara med Läckbergs böcker, vilket kan tänkas resultera i mer precisa resultat.

Efter att referenskorpusen kortats ner städades både målkorpusarna och referenskorpusen. I denna process raderas text som inte hör till romanerna i sig, såsom förord, efterord, baksidestext, tidigare verk och övrig metatext. Även romerska siffror raderas för att undvika att dessa räknas som prepositionen *i* när materialet analyseras. Dessa ändringar motiveras på samma vis som i föregående avsnitt: de är tänkta att leda till mer precisa resultat.

5 Metod

5.1 Variabler

Undersökningen är avgränsad till tre variabler: mest frekventa ord, n-gram i form av fyrgram och frekvens av konjunktioner. Varje variabel mäts och skattas på liknande vis som i studierna de hämtades ifrån, fast på grund av uppsatsens begränsade omfång kommer inga metoder helt kunna efterliknas. Exempelvis så skattar Kredens sitt resultat utifrån mer avancerade statistiska mått såsom p-värden och Likelihood, vilket i denna undersökning inte är möjligt.

Variablerna uppfyller två villkor: de är bevisat viktiga för en författares idiolekt och de är mätbara i skönlitterärt material. För att uppfylla det första kravet valdes variablerna utifrån tidigare studier som befäster deras betydelse för en idiolekt, vilket redovisades i avsnitt 3.2. Det andra kravet innebär att urvalet av valbara variabler minskar avsevärt. Som nämnts i avsnitt 3.2.1 och 3.2.3 är publicerad skönlitteratur sällan spontan och oredigerad. Detta innebär att många avvikande drag som traditionellt sett är intressanta i en forensisk-lingvistisk undersökning, såsom stavfel, städas bort under böckernas skriv- och redigeringsprocess. Jag var således tvungen att välja textuella drag som inte bryter mot språkregler, utan som istället kan betraktas som variation. Ett sådant exempel är synonymer: huruvida en författare väljer att

skriva *om* eller *ifall* beror på författarens egna språkbruk. Ordvalet är därmed mer osannolikt att redigeras bort under skrivprocessen, vilket gör det möjligt att undersöka trots att materialet är skönlitterärt.

5.2 Korpusverktyg

För att besvara frågeställningarna genomförs en korpusundersökning med hjälp av *AntConc*

4.3.1. *AntConc* är ett korpuslingvistiskt verktygsprogram som kan skapa textkorpora och analysera och framställa statistik över dem. För varje variabel spelar *AntConc* en central roll i att söka efter och presentera data i korpuserna. Datan exporteras sedan till ett Microsoft Excel-dokument och blir på så vis överskådlig.

Textredigerarprogrammet *Sublime Text* används för städning av materialet, som beskrevs i avsnitt 4.2. *Sublime Text* används också för att mäta korpusernas antal grafiska meningar. Detta utförs genom att söka efter stora skiljetecken följt av ett mellanslag, vilket i de allra flesta fall kännetecknar slutet av en grafisk mening. Antalet grafiska meningar kan alltså mätas i varje delkorpus, vilket är relevant för granskningen av konjunktioner.

Utöver *AntConc* används två ordforskningsplattformar från Språkbanken: *Mink* och *Korp*. *Mink* omvandlar en samling text till en korpus, vilket innebär att den kan laddas upp i *Korp*. I *Korp* analyseras korpuserna. Fördelen med *Korp*, jämfört med *AntConc*, är att man kan söka efter specifika ordklasser. Detta underlättar för granskningen av konjunktioner, vars ord ofta har flera lexikala funktioner. Ett exempel är *om*, som kan fungera som både subjunktion och preposition, såsom i följande satser: *jag vet inte om han kommer*, och *han pratade om vädret*.

5.3 Metod för varje variabel

5.3.1 Vanliga ord

Varje korpus 50 vanligaste ord framställs med hjälp av *AntConc*. När alla fyra listor har skapats så jämförs ordens förekomst mellan varje korpus. På grund av det stora antalet ord kommer fokus att ligga på övergripande trender samt under- och överrepresenterade ord i målkorpuserna jämfört med referenskorpuserna. Vilka ord finns med i referenskorpuserna, men saknas från

delkorpusarna? Eller tvärtom: vilka ord finns i delkorpusarna, men saknas från referenskorpusen? Ifall ett ord finns med i samtliga listor, bibehåller det sin under- eller överrepresentation i förhållande till referenskorpusen?

Därefter utförs en ytterligare sökning på varje delkorpus, fast denna gång med en stopplista införd. Med hjälp av stopplistan listas endast delkorpusarnas vanligaste innehållsord, som sedan framställs i ett urval av 15. Även detta resultat studeras i syfte att finna likheter och skillnader mellan målkorpusarna. Resultatet jämförs sedan med referenskorpusens förekomster, och då läggs särskilt fokus vid normavvikelser, det vill säga ord som är kraftigt över- eller underrepresenterade i målkorpusarna eller referenskorpusen. Tyder datan på att dessa ord används konsekvent mellan korpusarna? Vilka orsaker kan tänkas ligga bakom särskilda ords oregelbundna användning?

5.3.2 N-gram

AntConc används för att framställa korpusarnas 20 vanligaste fyrgram. Likt föregående variabel resulterar urvalsprocessen i att fyra listor skapas. Först studeras målkorpusarnas listor i syfte att hitta återkommande, såväl som sporadiska, fyrgram. Målkorpusarnas fyrgram jämförs sedan med referenskorpusens, i syfte att finna likheter och skillnader. Särskilt fokus läggs vid eventuella avsaknader mellan listorna: förekommer något fyrgram särskilt ofta i målkorpusarna, men inte i referenskorpusen? Detsamma gäller motsatsen: vilka fyrgram förekommer frekvent i referenskorpusen, men saknas i målkorpusarna?

5.3.3 Konjunktioner

Med *konjunktion* avses ord som binder samman satser. I undersökningen kommer Litvinova m.fl.:s (2018) begreppsdefinition att anammas. *Konjunktion* innefattar ord som binder samman både huvudsatser med andra huvudsatser, såväl som huvudsatser med bisatser. Med begreppet åsyftas alltså både konjunktioner och subjunktioner. Dessa beräknas med hjälp av Mink och Korp. Till skillnad från Litvinova m.fl.:s undersökning kommer värdet inte att skattas utifrån EDU. Anledningen till detta är att enheten sällan används i svensk språkforskning, och kräver mer avancerade program för att beräknas. EDU byts därför ut mot *grafiska meningar*.

AntConc används för att lista målkorpusens 100 vanligast förekommande ord och deras frekvens. Därefter eliminerar jag samtliga icke-konjunktioner från listan så att endast konjunktionerna, och deras frekvenser, kvarstår. Utifrån varje respektive ords frekvens

beräknas två normaliserade värden: *förekomst per 1000 ord* och *förekomst per grafisk mening*. *Förekomst per 1000 ord* beräknas genom att dividera frekvensen med det totala antalet ord, och sedan multiplicera kvoten med 1000. *Förekomst per grafisk mening* beräknas genom att dividera frekvensen med målkorpusens totala antal meningar. Genom normaliseringen kan värdena jämföras mellan olika målkorpusar, trots att målkorpusarna är olika stora. Exempelvis kan användningen av konjunktionen *och* i *Isprinsessan* jämföras med användningen i *Änglamakerskan*. Fokuset för studien är tvådelat då både den allmänna användningen av konjunktioner, och specifika varianter av konjunktioner, granskas.

6 Resultat

6.1 Vanligaste ord

I tabellen nedan presenteras de 50 vanligaste orden i var och en av de fyra korpusarna: Läckbergs tre böcker (A, B respektive C) och referenskorpusen. Orden anges under *Type* och är sorterade efter förekomst med högst förekomst först. Ordens bakgrunder är färglagda, och färgerna har olika innebörder. Grön kännetecknar ord som förekommer i samtliga listor. Gul kännetecknar ord som finns i målkorpus A och antingen B eller C. Blå kännetecknar ord som förekommer i B och C. Röd kännetecknar ord som förekommer i endast en korpus. Under *Frek* står ett ords normaliserade frekvens, alltså dess förekomst per 1000 ord avrundat till heltal.

| A | | | B | | | C | | | Ref | |
|------|------|--|------|------|--|------|------|--|------|------|
| Type | Frek | | Type | Frek | | Type | Frek | | Type | Frek |
| och | 32 | | och | 35 | | och | 35 | | och | 37 |
| att | 30 | | det | 28 | | det | 29 | | det | 24 |
| det | 27 | | att | 26 | | att | 27 | | att | 22 |
| hon | 21 | | hon | 24 | | hon | 22 | | i | 21 |
| i | 20 | | han | 19 | | på | 19 | | han | 19 |
| han | 20 | | på | 19 | | han | 19 | | en | 17 |
| på | 19 | | i | 17 | | i | 18 | | som | 16 |
| var | 18 | | som | 15 | | var | 16 | | jag | 16 |
| som | 17 | | var | 15 | | som | 15 | | på | 15 |

| | | | | | | | | | | |
|--------|----|--|--------|----|--|--------|----|--|--------|----|
| en | 16 | | hade | 15 | | inte | 14 | | hon | 12 |
| inte | 14 | | inte | 15 | | hade | 14 | | var | 12 |
| med | 13 | | en | 12 | | sig | 13 | | inte | 12 |
| jag | 12 | | sig | 12 | | en | 13 | | med | 11 |
| sig | 12 | | med | 12 | | jag | 12 | | den | 11 |
| för | 11 | | de | 11 | | med | 11 | | för | 10 |
| hade | 11 | | om | 10 | | om | 11 | | är | 9 |
| om | 9 | | jag | 10 | | de | 10 | | till | 9 |
| av | 9 | | men | 9 | | för | 9 | | sig | 9 |
| till | 9 | | till | 9 | | men | 9 | | av | 9 |
| men | 8 | | för | 8 | | till | 8 | | om | 8 |
| så | 8 | | vi | 8 | | är | 8 | | de | 8 |
| den | 7 | | är | 8 | | sa | 8 | | hade | 8 |
| ett | 7 | | sa | 7 | | så | 7 | | men | 8 |
| de | 6 | | så | 7 | | har | 7 | | så | 8 |
| skulle | 6 | | av | 7 | | skulle | 7 | | ett | 7 |
| du | 6 | | den | 7 | | av | 7 | | du | 6 |
| erica | 5 | | har | 6 | | den | 6 | | har | 5 |
| är | 5 | | skulle | 6 | | du | 6 | | skulle | 5 |
| henne | 5 | | henne | 5 | | henne | 5 | | honom | 4 |
| har | 5 | | ett | 5 | | vi | 5 | | mig | 4 |
| honom | 4 | | patrik | 5 | | ett | 5 | | där | 4 |
| när | 4 | | du | 5 | | här | 4 | | när | 4 |
| patrik | 4 | | honom | 5 | | när | 4 | | vi | 4 |
| kunde | 4 | | när | 4 | | patrik | 4 | | man | 4 |
| här | 4 | | här | 4 | | honom | 4 | | nu | 3 |
| vad | 4 | | ut | 4 | | vad | 4 | | då | 3 |
| ut | 3 | | något | 4 | | något | 4 | | henne | 3 |
| vi | 3 | | hur | 4 | | där | 4 | | hans | 3 |
| hennes | 3 | | vad | 4 | | upp | 4 | | sin | 3 |
| bara | 3 | | där | 3 | | ut | 3 | | vad | 2 |
| något | 3 | | upp | 3 | | nu | 3 | | ut | 3 |
| hur | 3 | | kan | 3 | | hur | 3 | | från | 3 |
| alex | 3 | | nu | 3 | | ebba | 3 | | upp | 3 |
| sin | 3 | | dem | 3 | | erica | 3 | | kunde | 3 |

| | | | | | | | | | | |
|------|---|--|--------|---|--|-------|---|--|-------|---|
| nu | 3 | | hennes | 3 | | kunde | 3 | | över | 3 |
| upp | 3 | | bara | 3 | | mot | 3 | | kan | 3 |
| från | 3 | | in | 3 | | kan | 3 | | eller | 2 |
| såg | 3 | | kunde | 3 | | ha | 3 | | här | 2 |
| hans | 3 | | ha | 3 | | dem | 3 | | sa | 2 |
| in | 3 | | mot | 3 | | hans | 3 | | bara | 2 |

Tabell 1. Vanligaste ord i de fyra korpusarna.

6.1.1 Jämförelse av målkorpusarnas listor

Målkorpusarnas vanligaste ord visar på många likheter. Till att börja med har listorna samma 9 vanligaste ord: *och, att, det, hon, i, han, på, var* och *som*. Ordens ordning varierar dock, då exempelvis *det* intar plats i 3 målkorpus A, men plats 2 i B och C. Även sett till helheten är listorna i hög grad lika. A delar totalt 45 ord med B, det vill säga 90 % av orden. Undantagen är orden *Erica, Alex, sin, från* respektive *hans* på platserna 27, 43, 44, 47 respektive 49 i A. Vidare delar A 44 ord med C, motsvarande 88 %.

De 6 resterande orden, som inte överensstämmer mellan A och C, är: *hennes, bara, Alex, sin, från* och *in*. B och C visar på flest gemensamma ord, nämligen 47. Detta motsvarar 94 %. Endast 3 ord skiljer dem åt: *hennes, bara* respektive *in* i målkorpus B på platser 45, 46 respektive 47. I C finns istället orden *Ebba, Erica* respektive *hans* på platser 43, 44 respektive 50.

Trots att jämförelsen visar på slående likheter mellan målkorpusarna så kommer resultatet, i följande avsnitt, att jämföras med referenskorpusens 50 vanligaste ord i syfte att upptäcka vilka ords förekomster som kan betraktas som normavvikande.

6.1.2 Ord som saknas i målkorpusarna

Referenskorpusens lista innehåller fem ord som inte dyker upp i någon annan lista: *mig, man, då, över* och *eller*. Varför just dessa ord saknas från målkorpusarnas listor är oklart. *Mig* och *man* är pronomen, och deras låga förekomst i målkorpusarna hade kunnat förklaras med att de tre böckerna är skrivna utifrån tredjeperson. Denna tes stärks av att målkorpusarna har högre frekvens av tredjepersonspronomen, såsom i fallet av *hon, han, hennes, sig* och *det*. Den låga andelen *mig* och *man* i målkorpusarna är alltså en trolig följd av böckernas berättarperspektiv. Detta innebär dock att orden är svaga bevis på ett unikt språkbruk hos författare, eftersom

många författare intar tredjeperson när de skriver sina berättelser: perspektivet är inte unikt för Läckberg.

Avsaknaden av *då*, *över* och *eller* är svårare att förklara. Orden hör till vitt skilda ordklasser, där *då* kan fungera som adverb, subjunktion eller interjektion. *Över* kan fungera som preposition, adverb eller adjektiv, och *eller* är en konjunktion. Att studera exakt hur orden används i varje delkorpus – exempelvis vilken andel *då* som är adverb respektive subjunktion – hade varit nästa steg i en mer omfattande analys. Ifall *då* är fem gånger mer sannolik att användas som subjunktion i målkorpusarna jämfört med referenskorpusen så hade detta kunna betraktas som betydande för en författares idiolekt, vilket underbygger uppfattningen att Läckberg skrivit samtliga tre böcker. I denna analys får absaknaden av *då*, *över* och *eller* istället direkt antas stärka samma uppfattning, trots att ingen djupare analys har utförts.

6.1.3 Ord som saknas i referenskorpusen

Målkorpusarnas listor innehåller två ord som saknas från referenskorpusens lista, nämligen *något* och *hur*. *Något* visar på en konstant användning genom samtliga tre verk: 3 i A, 3 i B och 3 i C. *Hur* uppvisar en relativ stabilitet: 3, 4 och 3. Att orden finns i varje delkorpus men saknas i referenskorpusen talar för att de är tämligen normavvikande. I kombination med att de bibehåller en stabil frekvens genom samtliga målkorpusar stärks uppfattningen att de ingår i en och samma författares idiolekt.

6.1.4 Innehållsord

Nedan presenteras Tabell 2 som utgörs av korpusarnas 15 vanligaste innehållsord. Färgerna har samma innebörd som innan, där grön kännetecknar ord som förekommer i samtliga listor. Gul kännetecknar ord som finns i målkorpus A och antingen B eller C. Blå kännetecknar ord som förekommer i B och C. Röd kännetecknar ord som förekommer i endast en korpus. Siffran bredvid ett ord indikerar dess förekomst per 1000 ord, men på grund av ordens låga förekomster avrundas de inte till heltal utan till en decimal.

| A | | B | | C | | Ref | |
|------|------|------|------|------|------|------|------|
| Type | Frek | Type | Frek | Type | Frek | Type | Frek |
| såg | 2,8 | sa | 7,4 | sa | 7,6 | sa | 2,3 |

| | | | | | | | |
|---------|-----|---------|-----|---------|-----|-------|-----|
| kände | 1,9 | såg | 2,5 | såg | 2,5 | sade | 2,1 |
| gick | 1,9 | gick | 1,9 | gick | 1,9 | såg | 1,9 |
| vet | 1,8 | vet | 1,8 | ta | 1,8 | min | 1,9 |
| tog | 1,6 | tittade | 1,8 | tog | 1,7 | än | 1,8 |
| visste | 1,6 | kände | 1,7 | tittade | 1,6 | gick | 1,7 |
| huset | 1,6 | visste | 1,7 | kände | 1,5 | också | 1,5 |
| ta | 1,5 | tog | 1,7 | göra | 1,4 | se | 1,3 |
| år | 1,4 | ta | 1,6 | vet | 1,4 | vet | 1,2 |
| fanns | 1,4 | göra | 1,4 | visste | 1,4 | vill | 1,2 |
| tittade | 1,4 | började | 1,3 | började | 1,3 | gå | 1,2 |
| gjorde | 1,4 | se | 1,2 | gjorde | 1,2 | göra | 1,2 |
| göra | 1,3 | gjorde | 1,2 | se | 1,1 | tog | 1,1 |
| se | 1,3 | gå | 1,1 | huvudet | 1,1 | stod | 1,1 |
| gå | 1,1 | huvudet | 1 | stod | 1,1 | gång | 1 |

Tabell 2. Vanligaste innehållsord i de fyra korpusarna.

Målkorpusarnas vanligaste innehållsord visar på många likheter. A delar 12 av 15 ord med B, det vill säga 80 % av orden. Undantagen är orden *huset*, *år* och *fanns*. Vidare delar A 11 ord med C, motsvarande 73 %. De fyra resterande orden, som inte överensstämmer mellan A och C, är: *huset*, *år*, *fanns*, *ur* och *gå*. B och C visar på flest gemensamma ord, nämligen 14. Detta motsvarar 93 %. Endast ett ord skiljer dem åt: *gå* i B och *stod* i C.

Utifrån referenskorpusen saknas *sade* från samtliga målkorpusar. *Sade* är en äldre formulering av ordet *sa*. Dess avsaknad i målkorpusarna bör alltså inte betraktas som särskilt betydelsefull för författarens idiolekt. Att det inte finns med i listan innebär endast att samtliga tre verk skrevs efter att *sa* hade populariserats i svensk skönlitteratur. Den senare ordformen kan alltså antas ingå i författarens idiolekt, men även i nästan varje modern författares idiolekt. Vidare är avsaknaden av förstapersonspronomen *min* det sannolika resultatet av att böckerna är skrivna utifrån tredjepersonsperspektiv, vilket redogjordes för i tidigare avsnitt. Precis som i fallet med *sade* är *min* föga intressant i fråga om Läckbergs idiolekt.

Sa är särskilt intressant med tanke på dess popularitet i B respektive C: 7,4 respektive 7,6. Referenslistan innehåller både *sade* och *sa*, med en gemensam frekvens på 4,4. Ordet saknas dock från A. Efter sökning i AntConc visar det sig inta plats 16 i A:s lista med en normaliserad frekvens på 1. Denna drastiska förändring mellan målkorpusarna kan förklaras med B och C:s författares sätt att skriva repliker. Ta delkorpus B som exempel. *Sa* förekommer

996 gånger totalt, och 642 av dessa gånger föregås ordet av ett citattecken följt av ett kommatecken, likt i följande utdrag:

(1) ”Jag åker förbi kommunen senare”, sa han och gick ut ur köket (s. 63)

I delkorpus C förekommer *sa* i 983 exempel, varav 619 har samma utformning som i B:

(2) ”Det här är min enda möjlighet”, sa han och nickade mot papperen (s. 86)

I delkorpus A, däremot, används *sa* 136 gånger, och i endast 21 av dessa fall i samma format som ovan. Detta beror på att deras repliker presenteras på andra sätt. Ett exempel är följande, där repliken presenteras utan något efterled:

(3) ”Jo, jag tänkte ta en tur till Evas och handla lite” (s. 5)

Att *sa* inte har en liknande frekvens i A som i B respektive C är alltså, åtminstone delvis, resultatet av ett skrivtekniskt grepp kopplat till karaktärernas repliker. Förändringen visar att författaren av A skriver repliker på ett annorlunda sätt än författaren av B och C. Fyndets betydelse för undersökningen kommer att diskuteras vidare i avsnitt 7.2.

Stod saknas helt från A och B:s listor, och är svagt överrepresenterad i C. Att ordet saknas från A och B är dock aningen missvisande eftersom dess normaliserade frekvens är 1 i både A och B. Sett till ordets frekvens i C är dess förekomst konstant genom samtliga tre målkorporusar. Därtill kretsar de tre värdena kring den normala frekvensen, vilket i slutändan innebär att ordets användning inte är särskilt uppseendeväckande.

Huvudet saknas från referenslistan och målkorpus A men återfinns i B och C. Ordets förekomst i B och C är ett resultat av det fasta uttrycket *skakade på huvudet*. I delkorpus A förekommer *huvudet* 70 gånger totalt, varav 10 ingår i *skakade på huvudet*. I B förekommer det 141 gånger, och då 62 gånger i det fasta uttrycket. I C förekommer ordet 139 gånger, och 68 gånger i den givna ordföljden. Ifall det fasta uttrycket exkluderas från målkorporusarna förekommer *huvudet* 60 gånger i A, 79 i B och 71 i C. Ordet får därmed normaliserade frekvenser på 0,42 i A, 0,49 i B och 0,47 i C. Sett till *huvudet* i referenskorpusen – 0,36 – är dessa nya frekvenser mer normala, fast konstant överrepresenterade. Resultatet diskuteras vidare i avsnitt 7.2.

6.1.5 Sammanfattning

I det första urvalet, bestående av korpusarnas 50 vanligaste ord, visar resultatet på stora likheter mellan målkorpusarna. Samtliga målkorpusar har samma 9 vanligaste ord: *och, att, det, hon, i, han, på, var* och *som*. Sett till helheten delar A 90 % av sina ord med B, och 88 % med C. B och C delar 94 %. Antalet gemensamma drag blir särskilt uppseendeväckande i jämförelse med referenskorpusen, som delar 82 %, 84 % respektive 84 % av sina ord med målkorpus A, B respektive C. Vissa ord förekommer oftare i målkorpusarna än i referenskorpusen, fast i fråga om författarens idiolekt är alla ord inte lika intressanta eftersom somliga – *mitt, min, icke* och *sade* – kan förklaras utifrån böckernas berättarperspektiv eller upphovsår. Orden som uppvisade särskilt starka tecken på författarens idiolekt är följande: *då, över, eller, något* och *hur*.

I den andra sökningen begränsades urvalet till korpusarnas 15 vanligaste innehållsord. Utifrån böckernas nya listor visade sig A och B dela 12 ord, A och C 11, och B och C 14. Dessa värden är än en gång högre än referenskorpusens gemensamma drag med målkorpusarna: 10 med A, 10 med B och 10 med C.

6.2 N-gram

6.2.1 Likheter och skillnader mellan målkorpusarna

I Tabell 3 nedan presenteras de 20 vanligaste fyrgrammen i Läckbergs böcker (A, B respektive C). Under *Type* anges fyrgrammen med färglagd bakgrund. Färgerna har samma betydelse som i Tabell 1 och 2. Grön kännetecknar fyrgram som förekommer i samtliga listor. Gul kännetecknar fyrgram som finns i målkorpus A och antingen B eller C. Blå kännetecknar fyrgram som förekommer i B och C. Röd kännetecknar fyrgram som förekommer i endast en korpus. Siffran bredvid ett ord, under *Frek*, anger ordets förekomst per 1 miljon ord avrundat till heltal.

| A | | B | | C | |
|----------------------|------|-------------------|------|-------------------|------|
| Type | Frek | Type | Frek | Type | Frek |
| bestämde sig för att | 215 | det var som om | 260 | inte låta bli att | 124 |
| inte låta bli att | 151 | det kändes som om | 186 | det var som om | 116 |

| | | | | | | | |
|----------------------|-----|--|-----------------------|-----|--|------------------------|----|
| vid det här laget | 119 | | hon visste att hon | 141 | | kunde inte låta bli | 93 |
| det kändes som om | 111 | | inte låta bli att | 126 | | och slog sig ner | 93 |
| det var som om | 111 | | bestämde sig för att | 111 | | tog ett djupt andetag | 93 |
| såg ut som om | 103 | | såg ut som om | 111 | | var på väg att | 93 |
| det var något som | 96 | | tog ett djupt andetag | 111 | | att det inte var | 85 |
| hon var tvungen att | 96 | | hon visste att han | 97 | | i och för sig | 77 |
| det måste ha varit | 80 | | kunde inte låta bli | 97 | | sig om för att | 77 |
| hon bestämde sig för | 80 | | hur är det med | 89 | | vad är det som | 77 |
| kunde inte låta bli | 80 | | jag tror att det | 89 | | vände sig om och | 77 |
| att det var något | 72 | | att det inte var | 82 | | det är klart att | 70 |
| han var tvungen att | 72 | | i och för sig | 82 | | jag tror att jag | 70 |
| jag vet inte om | 72 | | och slog sig ner | 82 | | reda på vad som | 70 |
| kanske till och med | 72 | | sträckte sig efter en | 82 | | skakade på huvudet det | 70 |
| om att det var | 72 | | på ett sätt som | 74 | | slog ut med händerna | 70 |
| att det inte var | 72 | | på sig själv med | 74 | | vände sig om för | 70 |
| att hon inte skulle | 64 | | reste sig och gick | 74 | | bestämde sig för att | 62 |
| det var dags att | 64 | | sig för att inte | 74 | | brydde sig inte om | 62 |
| det var inte förrän | 64 | | vände sig om och | 74 | | han var tvungen att | 62 |

Tabell 3. Vanligaste fyrgram i varje målkörpus.

Fem fyrgram förekommer i samtliga målkörpusar: *bestämde sig för att*, *inte låta bli att*, *det var som om*, *kunde inte låta bli* och *att det inte var*. *Bestämde sig för att* är delkörpus A:s vanligaste fyrgram med en normaliserad frekvens på 215. I B har frekvensen minskat till 111, och i C till 62. Förekomsten i A är alltså mer än tredubbelt så hög som i C. *Inte låta bli att* har en normaliserad frekvens på 151 i A, 126 i B och 124 i C. Trots att förekomsten minskar mer från A till B jämfört med B till C, håller sig värdena relativt konstanta. *Att det inte var* ökar successivt från A till C: 72 till 82 och slutligen till 85. Dess förekomst är alltså relativt konstant mellan målkörpusarna. Även *kunde inte låta bli* visar på en relativ stabilitet. Det börjar med 71 i A, ökar till 96 i B och minskar till 92 i C. *Det var som om* är återkommande, men visar inte på samma stabilitet. Ordföljden startar med en normaliserad frekvens på 111 i delkörpus A, och är därmed delkörpusens femte vanligaste fyrgram. I B ökar användningen kraftigt och fyrgrammet intar plats 1 med en frekvens på 260. Till sist, i C, sjunker frekvensen till 116. Vid detta laget har värdet nästan återställts till sitt ursprungliga värde i A.

Två fyrgram är gemensamma för endast A och B: *det kändes som om* och *såg ut som om*. *Det kändes som om* har i A en normaliserad frekvens på 111 och 185 i B. *Såg ut som om* är

snäppet ovanligare: 103 respektive 111. A och C har endast ett eget fyrgram: *han var tvungen att*, vars frekvens är 71 i A och 62 i C. B och C har totalt fyra egna fyrgram: *tog ett djupt andetag*, *i och för sig*, *och slog sig ner* och *vände sig om och*. *Tog ett djupt andetag* har frekvenserna 111 respektive 92 i B respektive C. *I och för sig* ligger på 81 respektive 77, medan *och slog sig ner* ligger på 81 respektive 92. *Vände sig om och* har värdena 74 respektive 77. Totalt sett delar A sju fyrgram med B och sex med C. B och C delar totalt nio fyrgram.

Likt den förra variabeln, ska fyrgrammens normaliserade frekvenser analyseras mot bakgrund av referenskorpusen. I och med detta framgår hur normenliga eller normavvikande målkorpusarnas fyrgram är. Resultaten analyseras i samma ordning som i detta avsnitt: först återkommande fyrgram, sedan sporadiska fyrgram i A och B respektive A och C, följt av B och C. Till sist, med resultaten framställda, ges ett överblickande perspektiv på delkorpusarnas totala antal gemensamma fyrgram.

6.2.2 Återkommande fyrgram

Nedan presenteras Tabell 4. Jämfört med Tabell 3 innehåller Tabell 4 även statistik över referenskorpusens 20 vanligaste fyrgram. Därtill introduceras även två nya färger: mörkgrön och mörkgul. Mörkgrön kännetecknar fyrgram som förekommer i samtliga korpusar. Fyrgram som förekommer i målkorpusarna och saknas från referenskorpusen kännetecknas av ljusgrön färg. Mörkgul kännetecknar fyrgram som finns i referenskorpusen och i en eller två målkorpusar. Ljusgul kännetecknar fyrgram som förekommer i två målkorpusar, men saknas från referenskorpusen. Kort sagt: ifall ett fyrgram är mörkgrönt eller mörkgult så finns det i referenskorpusen. Syftet med de mörkare färgerna är att överskådliggöra vilka fyrgram som är unika för målkorpusarna.

| A | | B | | C | | Referens | |
|----------------------|------|-----------------------|------|-----------------------|------|-------------------|------|
| Type | Frek | Type | Frek | Type | Frek | Type | Frek |
| bestämde sig för att | 215 | det var som om | 260 | inte låta bli att | 124 | det var som om | 85 |
| inte låta bli att | 151 | det kändes som om | 186 | det var som om | 116 | i stället för att | 33 |
| vid det här laget | 119 | hon visste att hon | 141 | kunde inte låta bli | 93 | att det var en | 32 |
| det kändes som om | 111 | inte låta bli att | 126 | och slog sig ner | 93 | en känsla av att | 32 |
| det var som om | 111 | bestämde sig för att | 111 | tog ett djupt andetag | 93 | såg ut som om | 32 |
| såg ut som om | 103 | såg ut som om | 111 | var på väg att | 93 | vad är det för | 30 |
| det var något som | 96 | tog ett djupt andetag | 111 | att det inte var | 85 | att det inte var | 29 |
| hon var tvungen att | 96 | hon visste att han | 97 | i och för sig | 77 | men det var inte | 27 |

| | | | | | | | |
|----------------------|----|-----------------------|----|------------------------|----|-------------------|----|
| det måste ha varit | 80 | kunde inte låta bli | 97 | sig om för att | 77 | på samma sätt som | 26 |
| hon bestämde sig för | 80 | hur är det med | 89 | vad är det som | 77 | inte låta bli att | 25 |
| kunde inte låta bli | 80 | jag tror att det | 89 | vände sig om och | 77 | jag vet inte vad | 24 |
| att det var något | 72 | att det inte var | 82 | det är klart att | 70 | vände sig om och | 24 |
| han var tvungen att | 72 | i och för sig | 82 | jag tror att jag | 70 | det är som om | 23 |
| jag vet inte om | 72 | och slog sig ner | 82 | reda på vad som | 70 | såg ut som en | 23 |
| kanske till och med | 72 | sträckte sig efter en | 82 | skakade på huvudet det | 70 | på sätt och vis | 22 |
| om att det var | 72 | på ett sätt som | 74 | slog ut med händerna | 70 | ut som om han | 22 |
| att det inte var | 72 | på sig själv med | 74 | vände sig om för | 70 | jag tror inte att | 21 |
| att hon inte skulle | 64 | reste sig och gick | 74 | bestämde sig för att | 62 | var det som om | 21 |
| det var dags att | 64 | sig för att inte | 74 | brydde sig inte om | 62 | det såg ut som | 20 |
| det var inte förrän | 64 | vände sig om och | 74 | han var tvungen att | 62 | det är klart att | 20 |

Tabell 4. De fyra korpusarnas 20 vanligaste fyrgram.

Bestämde sig för att finns inte med i referenslistan, vilket innebär att dess frekvens inte överstiger 20, för annars hade det syntts i Tabell 4. Fyrgrammets relativa förekomst i C är alltså minst tredubbelt så hög som dess relativa förekomst i referenslistan. Fyrgrammet är därmed fortsatt kraftigt överrepresenterat i C, fast inte lika överrepresenterat som i A och B.

Sett till *inte låta bli att* och dess normaliserade förekomst i referenskorpusen – 25 – är det kraftigt överrepresenterat i samtliga målkorpor: 151, 126 och 124. Detsamma gäller för *kunde inte låta bli*. Med värden på 80, 97 respektive 93 i A, B respektive C är fyrgrammet åtminstone tre gånger vanligare i målkorpusarna än i referenskorpusen.

Ordföljden finns även i referenslistan på plats 7 med ett värde på 29, vilket innebär att fyrgrammet är relativt vanligt i svensk skönlitteratur.

Det var som om är ett svårtolkat fall. Ordföljden startar med en normaliserad frekvens på 111 i delkorpus A, och är därmed delkorpusens femte vanligaste fyrgram. I B ökar användningen kraftigt och fyrgrammet intar plats 1 med en frekvens på 260. Till sist, i C, sjunker frekvensen till 116. Vid detta laget har värdet nästan återställts till sitt ursprungliga värde i A. I referenskorpusens lista är fyrgrammet överlägset först, för med en frekvens på 85 är det mer än dubbelt så vanligt som nästföljande fyrgram. Målkorpusarnas frekvenser tyder alltså på att fyrgrammet är konstant överrepresenterat, samtidigt som den kraftiga ökningen i B är uppseendeväckande.

6.2.3 Sporadiska fyrgram

Sju fyrgram finns med i endast två målkorpor: *det kändes som om*, *såg ut som om*, *han var tvungen att*, *tog ett djupt andetag*, *i och för sig*, *och slog sig ner* och *vände sig om och*. *Det*

kändes som om uppvisar höga förekomster i A och B: 111 och 185. I C är fyrgrammet betydligt mer sällsynt med en avrundad frekvens på 31. Sett till dess stora variationer är förekomsterna uppseendeväckande, fast dess låga förekomst i referenskorpusen – 14 – förmildrar fyrgrammets allvarsgrad. Trots dess låga värde i C, är ordföljden där mer än dubbelt så vanlig som i referenskorpusen. Det är alltså fortsatt överrepresenterat – som i A och B – fast inte lika kraftigt.

Såg ut som om har en normaliserad frekvens på 103 respektive 111 i målkorpus A respektive B. Jämfört med dess normaliserade frekvens i referenskorpusen – 32 – är den alltså kraftigt överrepresenterad. I C, däremot, saknas den från listan. Efter sökning i AntConc visade den sig ligga på 31, alltså betydligt lägre än i övriga målkorpor. Att fyrgrammet är det sjätte vanligaste i A och B, och femte vanligaste i referenslistan, gör dess avsaknad i C mer uppseendeväckande.

Han var tvungen att finns i A och C, men saknas från B. Det saknas även från referenskorpusens lista. I A och C har ordföljden normaliserade frekvenser på 72 respektive 62, och i B 30. I referenskorpusen är förekomsten betydligt lägre: 7. Trots dess minskade förekomst i B, är fyrgrammet mer än 4 gånger vanligare där än i referenskorpusen.

Med normaliserade frekvenser på 111 respektive 93 är *tog ett djupt andetag* ett tämligen vanligt fyrgram i B och C. I A ligger dess värde på 48, och i referenskorpusen 3. Sett till fyrgrammets extremt låga värde i referenskorpusen är det kraftigt överrepresenterat i samtliga målkorpor.

I och för sig har värdena 82 i B och 77 i C. I A är det mer ovanligt, med ett värde på 32. Sett till dess värde i referenskorpusen – 15 – är det alltså fortfarande överrepresenterat i samtliga målkorpor.

Och slog sig ner har värdena 82 respektive 93 i B respektive C, men saknas från A och referenskorpusens listor på grund av dess låga förekomst: 8 respektive 3. Användningen i B och C är alltså kraftigt normavvikande, fast samma kraftiga normavvikelse saknas i A.

Vände sig om och finns i samtliga delkorpor bortsett från A. Med en normaliserad frekvens på 74 respektive 77 i B respektive C, och 24 i referenslistan, är den överrepresenterad. Dess normaliserade frekvens i A ligger dock på 32, vilket innebär att den är överrepresenterad i samtliga målkorpor.

Sammantaget visar sju fyrgram på sporadisk användning. Fastän de avsaknas från en målkorpus, visar närmare granskning att de flesta fyrgrammen följer samma trend som i resterande målkorpor. Exempelvis: ifall ett fyrgram var överrepresenterat i A och B, men

saknas från C, är det sannolikt fortsatt överrepresenterat i C fast inte lika kraftigt. Två fyrgram följer inte en genomgående trend, nämligen *såg ut som om* och *och slog sig ner*. Resultatet diskuteras vidare i avsnitt 7.2.

6.2.5 Sammanfattning

Tre fyrgram som är vanliga i svensk skönlitteratur är även vanliga i Läckbergs böcker: *det var som om*, *att det inte var* och *inte låta bli att*. Av dessa var *det var som om* mest uppseendeväckande, sett till dess betydligt högre frekvens i delkorpus B jämfört med A och C. Just detta fyrgrammet kan alltså beskrivas som både normenligt och normavvikande: att det är vanligt i samtliga målkorporusar gör det normenligt, men dess kraftiga ökning i B gör det normavvikande.

Målkorpusarna visar också på andra normavvikande trender. Referenslistans andra, tredje, fjärde och sjätte vanligaste fyrgram återfinns inte i någon målkorpus. Därtill har målkorpusarnas listor två unika fyrgram som inte finns med i referenslistan: *kunde inte låta bli* och *bestämde sig för att*. Dessa två exempel är särskilt uppseendeväckande med tanke på deras extremt låga frekvenser i referenskorpusen. Till sist visade två fyrgram normavvikande trender som inte upprätthölls genom samtliga målkorporusar: *och slog sig ner* och *såg ut som om*.

Sammantaget visar resultatet att referenskorpusen delar 4 av sina 20 vanligaste fyrgram med A, 5 med B och 5 med C. Därmed uppvisar målkorpusarna fler gemensamma fyrgram än de visar i förhållande till referenskorpusen: A delar 7 fyrgram med B och 6 med C. B och C delar 9 fyrgram.

6.3 Konjunktioner

6.3.1 Konjunktioner per 1000 ord

I Tabell 5 nedan presenteras de 10 vanliga konjunktionerna. *A*, *B* respektive *C* representerar Läckbergs tre böcker, och *Ref* representerar referenskorpusen. I kolumn 1 listas konjunktionerna, och sorteras efter deras förekomst i referenskorpusen. Till höger om kolumn 1 och nedanför rad 2 finns siffror. Dessa siffror anger en specifik konjunktions normaliserade värde – i detta fall dess förekomst per 1000 ord – i en given korpus. Värdena har avrundats till

en decimal. Vissa av konjunktionerna har understrukits. Understrykning indikerar att konjunktionen är både över- och underrepresenterad i målkorpusarna. Exempelvis är *och* understruken eftersom den är underrepresenterad i A och B, och överrepresenterad i C. Dess avvikande värde – värdet i C – är understruket för att markera var diskrepansen finns. En konjunktion som saknar understrykning är genomgående över- eller underrepresenterad i målkorpusarna.

| Per 1000 ord | | | | |
|--------------|------------|----------|-------------|------------|
| | A | B | C | Ref |
| <u>och</u> | 28,3 | 29,4 | <u>30,1</u> | 29,5 |
| att | 13,4 | 11,1 | 12,0 | 9,9 |
| men | 7,1 | 8,0 | 7,3 | 6,5 |
| <u>som</u> | <u>4,8</u> | 3,9 | 3,1 | 4,5 |
| när | 0,4 | 0,4 | 0,4 | 3,4 |
| om | 3,4 | 3,9 | 4,4 | 2,9 |
| än | 1,6 | 1,3 | 1,0 | 2,0 |
| <u>eller</u> | <u>1,3</u> | 1,2 | 1,0 | 1,2 |
| <u>utan</u> | <u>0,5</u> | 0,4 | 0,3 | 0,4 |
| så | 0,6 | 0,6 | 0,7 | 0,3 |

Tabell 5. Vanligaste konjunktioner och deras förekomst per 1000 ord i de fyra korpusarna.

Utifrån Tabell 5 används fyra ord mer i samtliga målkorpusar jämfört med referenskorpusen: *att*, *men*, *om* och *så*. Två ord visar på en motsatt trend då de förekommer mer sällan: *när* och *än*. *När* är särskilt intressant eftersom dess förekomster i målkorpusarna (0,4, 0,4 respektive 0,4) är betydligt lägre än i referenskorpusen (3,4). Konjunktionen är alltså mer än åtta gånger vanligare i referenskorpusen än i någon av målkorpusarna.

Fyra av konjunktionerna bibehåller inte sin representation i förhållande till referenskorpusen: *och*, *som*, *eller* respektive *utan*. *Och* är underrepresenterad i både A och B med frekvenser på 28,4 och 29,4 i förhållande till referenskorpusens 29,5, för att sedan överrepresenteras i C med ett värde på 30,1. *Som*, *eller* och *utan* visar på en annan trend: de är först överrepresenterade i A, för att sedan underrepresenteras i B och C. Exempelvis har *som* ett värde på 4,5 i referenskorpusen, och startar med ett värde på 4,8 i A som sedan sjunker till 3,9 respektive 3,1 i B och C.

6.3.2 Konjunktioner per grafisk mening

I Tabell 6 nedan presenteras de tio vanliga konjunktionerna i samma format som i Tabell 5. Siffrorna indikerar konjunktionernas normaliserade frekvenser – dess förekomst per grafisk mening – i en given korpus. På grund av att värdena är lägre än i tidigare tabeller, avrundas de till tre decimaler.

| Per graf. men. | A | B | C | Ref |
|----------------|-------|-------|--------------|-------|
| och | 0,600 | 0,704 | 0,750 | 0,939 |
| att | 0,300 | 0,265 | 0,299 | 0,314 |
| men | 0,200 | 0,190 | 0,182 | 0,207 |
| som | 0,100 | 0,093 | 0,078 | 0,144 |
| när | 0,009 | 0,009 | 0,009 | 0,100 |
| om | 0,075 | 0,092 | <u>0,109</u> | 0,092 |
| än | 0,035 | 0,031 | 0,025 | 0,064 |
| eller | 0,030 | 0,027 | 0,025 | 0,040 |
| utan | 0,009 | 0,009 | 0,007 | 0,013 |
| så | 0,014 | 0,014 | 0,018 | 0,009 |

Tabell 6. Vanligaste konjunktioner och deras förekomster per grafisk mening i korpusarna.

Tabell 6 visar på en mer enhetlig representation jämfört med föregående tabell. Samtliga konjunktioner, bortsett från *om*, är konstant under- eller överrepresenterade i målkorpusarna i förhållande till referenskorpusen. *Så* är här unik, i att det är den enda konjunktionen som är överrepresenterad i delkorpusarna: *0,014*, *0,014* respektive *0,018* jämfört med *0,009*. De övriga åtta är konstant underrepresenterade, och *när* är återigen det mest tydliga exemplet: *0,009*, *0,009* respektive *0,009* jämfört med *0,1*. Precis som innan bibehåller variabeln sin underrepresentation och är numeriskt konstant mellan målkorpusarna.

6.3.3 Sammanfattning

Sammantaget visar granskningen av konjunktioner på en övergripande stabilitet mellan målkorpusarna. Sett till deras förekomst per 1000 ord är sex av tio konstant under- eller överrepresenterade i förhållande till referenskorpusen. Fyra ord används mer i samtliga målkorpusar än i referenskorpusen: *att*, *men*, *om* och *så*. Dessa ord är konstant

överrepresenterade. Två ord används mindre i målkorpusarna, och är således konstant underrepresenterade: *när* och *än*. Sett till konjunktionernas förekomst per grafisk mening uppvisas en större stabilitet än tidigare, då nio av tio är konstant under- eller överrepresenterade. *Om* utgör det enda undantaget: konjunktionen är underrepresenterad i A och B, och överrepresenterad i C. I slutändan är *när* mest betydelsefull för studien, eftersom denna konjunktion är konstant ovanlig i jämförelse med referenskorpusen sett till både förekomst per 1000 ord, såväl som förekomst per grafiska meningar i målkorpusarna.

7 Diskussion

7.1 Vanligaste ord

Samtliga målkorpusars ord uppvisar en genomgående stabilitet, både med och utan stopplistan. I det breda urvalet, bestående av varje respektive delkorpus 50 vanligaste ord, representeras 36 av referenslistans ord konstant i målkorpusarna. Detta motsvarar 76 % av dess ord. Därtill innehåller referenslistan tre ord som helt saknas från samtliga målkorpusar: *då*, *över* och *eller*. Målkorpusarna innehöll två ord som saknades från referenslistan: *något* och *hur*. Frekvensen av dessa fem ord tycks indikera att de ingår, eller inte ingår som i fallet med *något* och *hur*, i en och samma författares idiolekt.

Det begränsade urvalet bestod av varje respektive korpus 15 vanligaste innehållsord. *Sa* och *huvudet* är här mer intressanta. Båda har oregelbundna frekvenser till följd av nya författartekniska grepp i B och C jämfört med A. *Sa* har höga förekomster i B och C eftersom karaktärernas repliker ofta följs av *sa X*, där *X* motsvarar karaktären som yttrade sig. I samma målkorpusar förekommer *huvudet* mer på grund av en ökad användning av det fasta uttrycket *skakade på huvudet*.

Ifall samtliga tre böcker var skrivna under samma tidpunkt hade diskrepanserna antytt att olika författare ligger bakom. Men eftersom A och B skrevs med cirka sex års mellanrum så blir diskrepanserna mer svårtolkade. Att lära sig nya författartekniska grepp eller nya fraser som man sedan utnyttjar i sitt skrivande är en naturlig del i att utvecklas som författare. Att betrakta tillskotten av *sa* och *huvudet* som direkta bevis på spökskrivning hade därför inte varit rättvist. Samtidigt bör de inte bortses ifrån, eftersom de i slutändan utgör mätbara skillnader som är relevanta för undersökningens syfte. Frågan är hur betydelsefulla sådana tillskott är: är

generella ord, som inte har ökat till följd av nya författartekniska grepp eller fasta uttryck, mer värdefulla?

En av nyckelinsikterna från Kredens undersökning var att det är viktigt att homogenisera sitt material för att enklare kunna uppmärksamma skillnader. Kanske gäller detta inte bara textuella aspekter såsom genre, omfattning och målgruppsanpassning, utan även aspekter utanför texten såsom tidsskillnader. Trots att samtliga böcker överensstämmer med varandra i många avseenden, verkar tidsskillnaderna mellan A och B respektive C utgöra ett stort problem för undersökningen. I nuläget leder variabeln *vanligast använda ord* till tvetydiga resultat, där man omöjligt kan urskilja naturlig förändring från onaturlig förändring. Hur kan man veta om förändringarna beror på att samma författare har skrivit samtliga tre böcker och lärt sig nya fraser under sex års tid, eller att bok A har en annan författare än B och C?

Kredens material utgörs av icke-fiktiv text och naturligt språk. Min undersökning, däremot, utgörs av fiktiv text och icke-naturligt språk. Det är möjligt att textens fiktiva beskaffenhet utgör komplikationer, trots att materialet har homogeniserats sett till både dess litterära kategori – skönlitteratur – såväl som dess genre – deckarroman. I sådana fall ligger problemet inte i den tidsmässiga skillnaden mellan A och B respektive C, utan i böckernas skönlitterära slag. Jag är nyfiken kring hur resultaten sett ut ifall fler statistiska mått tillämpats, såsom *Likelihood* och *p-värde*. Därutöver väcks frågor kring rättslingvistiska metoders applicerbarhet, och hur väl de fungerar i fall av material med redigerat och korrekturläst språk.

Sammantaget visar variabeln *vanligast använda ord* på en genomgående relativ stabilitet i samtliga tre målkorpor. Det breda urvalet visade att A och B delar 45 ord (90 %). A och C delar 44 ord (88 %). B och C visar flest likheter: 47 ord (96 %). Ord som visade på sporadisk användning analyserades och bedömdes vara relativt ovidkommande för författarens idiolekt, med undantag för *sa* och *huvudet*. Vidare analys visade att förekomsterna av *sa* och *huvudet* resulterade från författartekniska tillskott respektive det fasta uttrycket *skakade på huvudet*. Sett till författarens idiolekt och eventuell spökskrivning är resultatet tvetydigt, fast med tanke på den övergripande stabiliteten i de resterande värdena är jag mer benägen att tolka ordens ökade förekomst en följd av Läckbergs utvecklande författarskap. I sin helhet talar resultatet alltså för att Läckberg skrivit böckerna.

7.2 N-gram

För fyrgrammen finns tre drag i målkorpusarna som, i min mening, talar för genomgående bevis för en och samma författares idiolekt. Det första är fyrgrammen som är unika för målkorpusarna: *kunde inte låta bli* och *bestämde sig för att*. Sett till referenskorpusen är dessa mycket mer sällsynta i svensk skönlitteratur i stort, vilket är varför deras återkommande höga frekvens i målkorpusarna blir betydelsefulla: de utgör normbrytande drag som förekommer konstant. Chansen att deras placeringar i mållistorna beror på slumpen kan således tänkas vara väldigt liten.

Det andra draget är målkorpusarnas avsaknad av det andra, tredje, fjärde och sjätte vanligaste fyrgrammen: *istället för att*, *att det var en*, *en känsla av att* och *vad är det för*. Detta drag kan argumenteras vara betydelsefullt genom samma resonemang som ovan. Ifall olika författare skrev böckerna A, B och C, kan chansen tänkas vara stor att fler av dessa fyrgram syns till i målkorpusarnas listor. Med andra ord: referenslistan hade varit mer gul.

Det tredje draget är antalet gemensamma fyrgram mellan varje delkorpus. I denna undersökning kan referenslistans gemensamma fyrgram med en målkorpus – fyra till fem – antas vara normalt. Ifall två korpusar delar fler än fem fyrgram bland sina 20 vanligaste, bör det alltså tolkas som betydelsefullt. Ju högre antal gemensamma drag, desto större blir sannolikheten att böckerna som utgör korpusarna har samma författare. Att målkorpus A delar sju fyrgram med delkorpus B indikerar alltså att samma författare ligger bakom. Korrelationen blir tydligare i fallet av B jämfört med C, som delade nio fyrgram. Sett till antalet gemensamma ord mellan referenskorpusen och en av målkorpusarna – fyra till fem – är det alltså sannolikt att bok A och B har samma författare, och mycket sannolikt att så är fallet även med B och C.

Resultatet är dock inte entydigt. I två fall hittades diskrepanser som, trots närmare granskning, förblir oförklarliga. Detta berör de oregelbundna förekomsterna av *och slog sig ner* och *såg ut som om*. Varför just dessa fyrgram användes sporadiskt är okänt, och vidare efterforskning av fyrgrammen ligger dessvärre utanför denna uppsats ramar. Viktigt att påpeka är att en begränsad mängd diskrepanser är väntade. Även Wrights (2017) undersökning, som hade en betydligt större omfattning samt en komplex och robust urvalsprocess, lyckades inte att felfritt identifiera författares idiolekter utifrån n-gram (s. 20). Att ett fåtal diskrepanser uppstår även i denna undersökning är därmed inte särskilt uppseendeväckande.

I fråga om framtida forskning anser jag att det finns många intressanta infallsvinklar. En vinkel, för en undersökning av samma omfattning som denna, hade varit att undersöka resterande n-gram. Trots att fyrgram var de mest träffsäkra i Wrights (2017) studie, var både

tregram och femgram nästan lika effektiva. Eftersom Wrights material utgjordes av icke-fiktiv text, väcks även frågan kring vilket sorts n-gram som är mest effektivt för skönlitterär text. En annan vinkel, för en mer omfattande undersökning, hade varit att analysera böckernas n-gram i närmare detalj, samt med en metod och urvalsprocess som närmare liknar Wrights.

7.3 Konjunktioner

I fråga om konjunktioner per 1000 ord visar många variabler en relativ stabilitet mellan de tre målkorpusarna. Majoriteten av variablerna – sex av tio – är dessutom konstant representerade i förhållande till referenskorpusen. Om en konjunktion är överrepresenterad i A, är sannolikheten alltså sex av tio att den är överrepresenterad i B och C. Mest slående är användningen av *när* som var dels sällsynt i förhållande till referenskorpusen, dels stabil mellan samtliga delkorpusar. Variabeln var konstant normavvikande, vilket talar starkt för att användningen är en del av endast en idiolekt, och därmed att verken skrivits av en och samma person.

Mätvärdena för konjunktioner per grafisk mening är mer entydiga. nio av tio variabler uppvisar en konstant representation mellan målkorpusarna. Undantaget är konjunktionen *om*, som i målkorpus C överstiger den genomsnittliga förekomsten från att ha underrepresenterats i A och normalrepresenterats i B. *När* är återigen den mest slående variabeln, med värden som är både extremt konstanta och kraftigt normavvikande.

Beskrivningen av den ideala variabeln i Litvinova m.fl.:s (2018) undersökning – den med *låg intravariabilitet* och *hög intervariabilitet* (s. 226) – är relevant för resultaten. Detta kännetecknar en variabel som kan förväntas variera mycket från person till person, men som samtidigt är konstant hos varje individ. Ifall en variabel, såsom förekomsten av en särskild konjunktion, varierar mycket mellan olika verk så indikerar detta att verken inte skrivits av samma person.

En viktig fråga kring Litvinova m.fl.:s (2018) definition är vad som klassas som *låg intravariabilitet*. Hur konstant måste en variabel vara för att uppfylla detta kriterium? Ifall man stipulerar en hård gräns, exempelvis att variabeln inte får öka eller minska med mer än 0.01 procentenheter mellan varje målkorpus, hur bör man betrakta en variabel som ökar eller minskar med 0.011 procentenheter? Utifrån stipulationen är sådana variabler fullständigt ovidkommande: 0.011 skulle lika gärna kunna betyda 0.9. Ifall man istället väljer att betrakta variabeln som relevant, trots att den överskrider gränsen: vad är syftet med att sätta en hård

gräns till att börja med? Detsamma gäller för *hög intervvariabilitet*: hur stor bör variationen vara för att kriteriet ska uppfyllas?

För min undersökning är alltså både beskrivningen av den ideala variabeln, såväl som dess ovan nämnda komplikationer, relevanta. Såvida en variabel inte uppvisar samma exakta värde genom samtliga målkorporusar, saknar jag grund för att betrakta variabeln som konstant. Till följd valde jag att fokusera på variabelernas representation, alltså hur de förhåller sig till ett genomsnittligt värde hämtat från en bred referenskorpus. Bedömningsprocessen blir därmed enklare, på bekostnad av att slutsatserna blir mindre säkra. Att sex av tio konjunktioner representerades konstant sett till deras förekomst per 1000 ord, och att nio av tio representerades konstant sett till förekomst per grafiska meningar, är onekligen talande för att en och samma idiolekt ligger till grund för samtliga tre böcker. Resultatet indikerar alltså att Läckberg, eller åtminstone en och samma författare, har skrivit böckerna.

I framtida undersökningar hade en djupare lingvistisk analys kunnat utföras i kombination med statistiska analyser. Variablernas inter- och intravariabilitet hade kunnat sammanställas även från referenskorpusen. Värdena från målkorpusen hade därmed kunnat ses i ljuset av referenskorpusens, vilket hade resulterat i mer nyanserade och välgrundade slutsatser.

8 Avslutande ord

I denna undersökning har jag studerat tre av Camilla Läckbergs böcker i syfte att hitta likheter och skillnader i språkbruket. Målet var att använda rättslingvistisk metodik och utifrån resultatet diskutera huruvida någon, eller några, av böckerna kan ha spökskrivits. Tre variabler granskades: *vanligaste ord*, *n-gram* i form av *fyrgram* och *konjunktioner*. Sammantaget visar resultatet på genomgående trender i böckernas språk. Efter att ha jämfört resultatet med en bred, skönlitterär referenskorpus så upptäcktes att en del trender dessutom är normavvikande. Dessa trender betraktas som starkast bevis på författarens idiolekt. I fråga om spökskrivning presenterar undersökningen inga absoluta svar: det är inte omöjligt att böckerna kan ha spökskrivits, fast det slutliga resultatet indikerar att så inte är fallet. Utifrån denna undersökning har Läckberg sannolikt skrivit samtliga tre böcker.

Undersökningen är långt ifrån allomfattande. För att få en mer nyanserad bild av Läckbergs, eller någon annan författares, författarskap så kan fler variabler studeras och

dessutom i närmare detalj. För en undersökning av samma omfattning som denna så kan andra ordlistor framställas och analyseras, andra typer av n-gram granskas och andra ordklasser studeras. För ett mer omfattande arbete kan fler statistiska mått tillämpas, vilket hade möjliggjort för den nyfikne att närmare följa en rättslingvistisk metod.

Referenslista

Christensen, T.K. (2017). Ordet fanger. Retslingvistik i en dansk kontekst. *NyS, Nydanske Sprogstudier* 1(52-53): s. 169-206. <https://doi.org/10.7146/nys.v1i52-53.26334>

Coulthard, M. & Johnson, A. (2007). *An Introduction to Forensic Linguistics: Language in Evidence*. 1. uppl. Routledge.

Dagens eko (2023). Jakten på Camilla Läckbergs spökskrivare [Podcast]. Sveriges Radio. 17 oktober. <https://sverigesradio.se/avsnitt/jakten-pa-camilla-lackbergs-spokskrivare> (Hämtad 2024-01-24).

Dahlgren, S. & Ek, T. (2021). Läckberg skrattar åt spökskrivarryktena. *Aftonbladet*. 21 augusti. <https://www.aftonbladet.se/nojesbladet/a/66PG00/lackberg-skrattar-at-spokskrivarryktena> (Hämtad 2025-01-24).

Kredens, K. (2003). Towards a corpus-based methodology of forensic authorship attribution: a comparative study of two idiolects. I Lewandowska-Tomaszczyk, B. & Melia, P.J. (red.). *PALC 2001: Practical Applications in Language Coropra*. Peter Lang, s. 405-447.

Lapin, L. (2023). AI visar: De här romanerna av Läckberg är spökskrivna. *Kvartal*. 23 september. <https://kvartal.se/artiklar/ai-visar-de-har-romanerna-av-lackberg-ar-spokskrivna/> (Hämtad 2025-01-24).

Litvinova T., Litvinova, O. & Seredin, P. (2018). Assessing the Level of Stability of Idiolectal Features across Modes, Topics and Time of Text Production. I: *Proceedings of the 23rd*

Conference of Open Innovations Association FRUCT (s. 223-230).

<https://doi.org/10.23919/fruct.2018.8588092>

Shimoda, A. & Shimoda, H. (2024). Camilla Läckberg vill ha ursäkt av Jan Guillou – efter påhoppet: ”Djupt besviken”. *Aftonbladet*. 17 januari.

<https://www.aftonbladet.se/nojesbladet/a/rlwgKA/camilla-lackberg-om-jan-guillous-pahopp> (Hämtad 2024-01-24).

Spektra, TT. (2010). Läckberg tjänar mest av författarna. *Svenska Dagbladet*. 2 december.

<https://www.svd.se/a/0cbbc4a7-1cd3-3382-9558-8b5bd2ab6538/lackberg-tjanar-mest-av-forfattarna> (Hämtad 2025-01-24).

TT. (2023). Camilla Läckberg om ryktena: ”Han petade i meningarna”. *Aftonbladet*. 18 december.

<https://www.aftonbladet.se/nojesbladet/a/LIV689/lackberg-om-ryktena-han-petade-i-meningarna> (Hämtad 2025-01-24).

TT. (2010). Läckberg tjänar mest av författarna. *Svenska Dagbladet*. 2 december.

<https://www.svd.se/a/0cbbc4a7-1cd3-3382-9558-8b5bd2ab6538/lackberg-tjanar-mest-av-forfattarna> (Hämtad 2025-01-24)

Wright, D. (2017). Using word n-grams to identify authors and idiolects: A corpus approach to a forensic linguistic problem. *International Journal of Corpus Linguistics*, 22(2), s. 212-241.

<https://doi.org/10.1075/ijcl.22.2.03wri>

Materialkällor

Borin, L & Pettersson, E. (2022). *Svensk diakronisk korpus* [Dataset].

<https://www2.lingfil.uu.se/person/pettersson/svediakorp/index.html> (Hämtad 2025-01-24)

Läckberg, C. (2003). *Isprinsessan*. Bokförlaget Forum.

Läckberg, C. (2009). *Fyrvaktaren*. Bokförlaget Forum.

Läckberg, C. (2011). *Änglamakerskan*. Bokförlaget Forum.