

Justus Vuorio

TORDE VARA I ÖVRE TÖR

Användningen av hjälpverbet *torde* i olika sammanhang

Fakulteten för informationsteknologi och kommunikation
Studieinriktningen i nordiska språk
Avhandling pro gradu
Maj 2021

SAMMANDRAG

Justus Vuorio: Torde vara i Övre Tör. Användningen av hjälpverbet *torde* i olika sammanhang.
Avhandling pro gradu
Tampereen yliopisto
Studieinriktningen i nordiska språk
Maj 2021

I denna avhandling pro gradu undersöks användningen av det epistemiska hjälpverbet *torde* i olika genrer. Den första forskningsfrågan är att utreda hur ofta verbet och dess olika uttrycksformer förekommer i olika genrer och olika slags kontexter. Av de mindre vanliga uttrycksformerna är främst den morfologiska presensformen *tör* att notera, vilken enligt litteraturen har gett vika för formen *torde*, som ursprungligen är ett preteritum men som fått en presensbetydelse, samt som ett annat särfall den normala formen *torde* i betydelsen 'bör lämpligen, begärs'. Den andra forskningsfrågan är att identifiera de typiska situationer där det perfekta hjälpverbet *ha* utelämnas i dåtidsformer mellan *torde* och supinum.

Studien är av synkronisk natur och fokuserar på språkbruket på 1990-talet och framåt. Metoden är korpuslingvistik, särskilt frekvensanalys. Förutom den kvantitativa utgångspunkten behandlas materialet även kvalitativt. För att kunna jämföra undersöks frekvensen av *torde* i förhållande till adverb med liknande betydelse, exempelvis *förmodligen*, *sannolikt* osv.

Materialet består av Språkbankens korpusar, som användes genom tjänsten Korp. Språkbanken upprätthålls av Göteborgs universitet. Å ena sidan söktes i alla Korps fritt tillgängliga korpusar i det s.k. Moderna-läget. Å andra sidan utvaldes sju korpusar som representerar var sin genre och vilkas träffar jag gått noggrannare manuellt igenom och räknat felaktiga träffar, träffar som är för gamla med tanke på studien samt dubletter. Som felaktiga klassificerades sådana fall där skribenten tydligt verkade ha avsett ett annat lexem samt fall där XML-texten bakom korpusen hade annoterats fel.

Studien visar att *torde* används på högre stilnivåer mera än på andra stilnivåer, vilket var förväntat. Verbet förekommer oftare i akademiska än i skönlitterära texter fast bägge kan uppvisa högre stil. Presensformen *tör* förekommer sällan i materialet, likaså formen *torde* i betydelsen 'bör lämpligen, begärs'. Allt som allt kan *torde* anses vara tämligen vanlig i sakprosa; i somliga korpusar var det näst vanligast av de alternativa uttryckssätten. Vad gäller bortfallet av det perfekta hjälpverbet *ha*, korrelerar antalet bestämningar mellan *torde* och supinum med sannolikheten av bortfallet i materialet: *ha* utelämnas sannolikare om det finns tunga bestämningar. Större genreskillnader hittades inte för denna forskningsfråga; vanligast förekom bortfallet, kanske lite förvånande, i korpusen *Poeter.se*, som innehåller amatörlyrik.

Nyckelord: korpuslingvistik, frekvens, epistemisk, genrer, hjälpverben, supinum

Originaliteten av denna avhandling har granskats med Turnitin OriginalityCheck-programmet.

TIIVISTELMÄ

Justus Vuorio: Torde vara i Övre Tör. Användningen av hjälpverbet *torde* i olika sammanhang.
Pro gradu -tutkielma
Tampereen yliopisto
Pohjoismaisten kielten tutkinto-ohjelma
Toukokuu 2021

Tässä pro gradu -tutkielmassa selvitetään ruotsin kielen episteemisen modaaliapuverbin *torde* käyttöä eri genreissa. Ensimmäisenä tutkimuskysymyksenä on selvittää, missä genreissa ja millaisissa käyttöyhteyksissä verbi ja sen eri ilmenemismuodot esiintyvät enemmän ja missä vähemmän. Erikoisemmista ilmenemismuodoista kyseeseen tulevat ennen kaikkea morfologinen preesensmuoto *tör*, joka jo kirjallisuuden mukaan on joutunut antamaan tilaa muodolle *torde*, joka on alun perin preteriti, mutta saanut preesensin merkityksen, sekä toisena erikoistapauksena normaalin muodon *torde* käyttö merkityksessä 'pyydetään'. Toisena tutkimuskysymyksenä on selvittää, missä tilanteissa perfektuaalinen apuverbi *ha* tyypillisimmin jää pois menneestä aikaa ilmaisevissa rakenteissa *torde*-apuverbin ja supiin välillä.

Tutkimus on luonteeltaan pääosin synkroninen ja keskittyy 1990-luvun ja sitä nuorempaan kielenkäyttöön. Menetelmänä on korpuslingvistiikka, erityisesti frekvenssianalyysi. Määrällisen lähestymistavan lisäksi aineistoa käsitellään myös laadullisesti. Vertailukohtaan saamiseksi *torde*-verbin frekvenssiä suhteutetaan suunnilleen samanmerkityksisiin adverbisiin, kuten *förmodligen*, *sannolikt* jne.

Aineistona on joukko Språkbankenin korpuksia, joita käytettiin Korp-nimisen palvelun kautta. Språkbanken ylläpitää Göteborgin yliopisto. Yhtäältä haettiin kaikista Korpin vapaasti haettavissa olevista korpuksista ns. Moderna-tilassa. Toisaalta valittiin tarkemmin seitsemän eri genreja edustavaa korpusta, joiden osalta olen käynyt osumat tarkemmin manuaalisesti läpi ja laskenut osumista virheelliset osumat, tutkimuksen kannalta liian vanhat osumat ja kaksoiskappaleet. Virheellisiksi määriteltiin tapaukset, joissa kirjoittaja selvästi näytti tarkoittaneen jotakin muuta lekseemiä, sekä tapaukset, joissa korpuksen taustalla oleva XML-teksti oli annotoitu väärin.

Tutkimus osoittaa, että *torde*-verbiä käytetään korkeammassa tyyllilajeissa enemmän kuin muissa, kuten oli oletettakin. Tieteellisissä teksteissä verbiä käytetään enemmän kuin kaunokirjallisissa, vaikka molemmissa voi esiintyä korkeampaa tyyliä. Preesensmuotoa *tör* ei aineistossa juuri esiinny, ei liioin muotoa *torde* merkityksessä 'pyydetään'. Kaiken kaikkiaan *torde*-verbiä voidaan pitää asiateksteissä melko yleisenä; muutamissa korpuksissa se oli toiseksi yleisin vastaavista ilmaisutavoista. Perfektuaalisen apuverbin *ha* poisjätöstä havaittiin, että *torde*-verbin ja supiin välissä olevien määritteiden määrä korreloi aineistossa poisjätön todennäköisyyden kanssa: *ha* jää pois todennäköisemmin, jos mukana on raskaita määritteitä. Genre-eroja ei tämän tutkimuskysymyksen osalta juuri löytynyt; eniten poisjätettä esiintyi, ehkä hieman yllättäen, amatööriyleisyyttä sisältävässä *Poeter.se*-korpuksessa.

Avainsanat: korpuslingvistiikka, frekvenssi, episteemisyys, genret, apuverbit, supiini

Tämän julkaisun alkuperäisyys on tarkastettu Turnitin OriginalityCheck -ohjelmalla.

Innehållsförteckning

1 Inledning.....	1
1.1 Syfte.....	2
1.2 Disposition.....	3
1.3 Tidigare forskning.....	3
2 Teori.....	7
2.1 Modalitet och hjälpverb.....	8
2.2 Etymologin och morfologin av <i>torde</i>	15
2.3 Det perfektuella hjälp verbet <i>ha</i>	16
2.4 Reguljära uttryck.....	18
2.5 CQP-språket och andra formella språk.....	20
3 Korpuslingvistik.....	27
3.1 Närmare metod.....	30
3.2 Allmänt om materialet.....	34
3.3 De sju utvalda korpusarna.....	35
4 <i>Torde</i> i korpusarna.....	42
4.1 <i>Torde</i> i samtliga korpusarna.....	42
4.2 <i>Torde</i> i olika genrer i de utvalda korpusarna.....	47
4.3 <i>Ha</i> med <i>torde</i> + supinum.....	60
5 Diskussion.....	66
6 Avslutning.....	71
Litteratur.....	73

Diagram

Diagram 1. Andelarna av <i>torde</i> och alternativa uttryckssätt i alla fritt tillgängliga korpusar (<i>n</i> %) i förhållande till varandra	45
Diagram 2. Andelarna av <i>torde</i> samt motsvarande uttryckssätt (<i>n</i> ppm) per år	46
Diagram 3. Andelen av <i>torde</i> av motsvarande uttryckssätt (<i>n</i> %) per år	47
Diagram 4. Andelarna av <i>torde</i> och alternativa uttryckssätt (<i>n</i> %) i <i>Poeter.se</i>	50
Diagram 5. Andelarna av <i>torde</i> och alternativa uttryckssätt (<i>n</i> %) i <i>Skönlitteratur 2000–2013</i>	51
Diagram 6. Andelarna av <i>torde</i> och alternativa uttryckssätt (<i>n</i> %) i <i>Samhällsvetenskap</i>	53
Diagram 7. Andelarna av <i>torde</i> och alternativa uttryckssätt (<i>n</i> %) i <i>Läkartidningen 2003</i>	54
Diagram 8. Andelarna av <i>torde</i> och alternativa uttryckssätt (<i>n</i> %) i samtliga <i>Läkartidningen</i> -korpusar	55
Diagram 9. Andelarna av <i>torde</i> och alternativa uttryckssätt (<i>n</i> %) i <i>GP 2012</i>	57
Diagram 10. Andelarna av <i>torde</i> och alternativa uttryckssätt (<i>n</i> %) i samtliga <i>GP</i> -korpusar.	58
Diagram 11. Andelarna av de alternativa uttryckssätten (<i>n</i> %) i <i>8 sidor</i>	59
Diagram 12. Andelarna av de alternativa uttryckssätten (<i>n</i> %) i <i>Barnlitteratur 1988–2013</i>	60
Diagram 13. Andelarna av rensade förekomster av lemgrammet <i>torde</i> (<i>verb</i>) och dubletter, oäkta, tvivelaktiga och för gamla förekomster i de sju utvalda korpusarna	67

Tabeller

Tabell 1. Kategorierna av epistemiska hjälpverb enligt SAG (<i>p</i> = proposition)	12
Tabell 2. Allmänna upprepningsoperatorer i reguljära uttryck	20
Tabell 3. De sju närmare genomgångna korpusarnas indelning och egenskaper	40
Tabell 4. Antalet bestämningar mellan <i>torde</i> och supinum i fall där perfektuellt hjälpverb utelämnats respektive i fall där det utsatts	64
Tabell 5. De vanligaste alternativa uttryckssätten i respektive genre	68

1 Inledning

Det modala hjälp verbet *torde*, som motsvarar finskans modus potentialis är en del av det svenska språket som inte rönt mycket uppmärksamhet i den vetenskapliga litteraturen. I mina egna iakttagelser verkar verbet ha blivit mera sällsynt under de senaste årtiondena. Svenska Akademiens ordbok, (s.v. *tör_{verb}*), Svenska Akademiens ordlista (s.v. *torde*) och Svensk Ordbok (s.v. *torde*) nämner visserligen ingenting om att verbet i sig vore sällsynt eller skulle ha något slags särskilt stilvärde. Det nämns bara att presensformen *tör* har blivit mindre bruklig och att ”**torde** är nu helt övervägande” (SAOB) och att formen blir ”i pres. ibl. *tör*” (SO och SAOL). Dessutom befinnes användningen av *torde* i betydelsen ’bör lämpligen, begärs’ ’hövlig’ i SAOB och SO, och nämns inte alls i SAOL.

Om man ser på *torde* ur synvinkeln av skolböcker, finns det indicier på att verbet skulle ha blivit mera sällsynt eller att skribenterna av böckerna uppfattar att det förhåller sig så. *Torde* har försvunnit ur en del gymnasieböcker och det anses vara rätt så sällsynt i dem, fast det förekommer. Exempelvis förlaget Otavas *Lukion ruotsin kielioppi*, det vill säga ’Gymnasiesvenskans grammatik’ från 1980-talet nämner *torde* men konstaterar att det ”används främst i skriftspråket” samt att ”i naturligt språkbruk uttrycks förmodande och antagande med hjälp av t.ex. adverb [...]” (min översättning). Ur retorisk synvinkel är *naturligt* ett värdeladdat uttryck. Det kan ge intrycket av att det vore något fel, konstlat eller eventuellt rentav perverst med att använda *torde*. Dessutom är den paragraf som behandlar *torde* skriven med petit. Samma förlags nyare läroboksserie för svenska som andra inhemska språket i gymnasier, *Fokus*, behandlar modala hjälpverb i läroboken för kurs 4 under rubriken ”Apuverbit” med en koncis överblick på två sidor och nämner ingenting om *torde* (Blom & al. 2017: 124f).

Jag undersöker i denna studie användningen av *torde* i olika genrer och kontexter samt det optionella utelämnandet av det perfektuella hjälp verbet *ha* i samband med *torde* och supinum. Min metod är korpuslingvistik, och en särskilt viktig del av den metoden är frekvensanalys. Jag söker i Språkbankens korpusar och analyserar träffarna både kvantitativt och kvalitativt.

Bruket av *torde* är ett intressant ämne att forska på i och med att det är ett alternativt uttryckssätt för sådant som eljest brukar uttryckas i svenskan genom adverb, alltså en helt annan ordklass. Det är intressant att se i hurdana situationer skribenter väljer detta verbala alternativ. Det verkar inte

forskats mycket på temat, så det är dags att kartlägga situationen. Detta är viktigt ur synvinkeln av vetenskaplig basforskning. Min forskning kan komplettera bilden av ett delvis förglömt och försummat modalverb bland lingvister. Svenska Akademiens grammatik (Teleman & al. 1999, band 4: 282ff) behandlar modala hjälpverb i överlag men *torde* i sig endast kursoriskt. Lagervall (2014) har i sin doktorsavhandling behandlat svenskans modala hjälpverb och ägnat utrymme även åt just *torde*, men inte så noggrant och hennes tyngdpunkt har legat snarare i den historiska utvecklingen än i t.ex. semantik och stilistik, vilka jag ser närmare på.

1.1 Syfte

Syftet i min forskning är att studera användningen av det modala hjälp verbet *torde* i olika kontexter. Dess frekvens och hur och i vilka genrer det används intresserar mig. Min undersökning gäller även felaktig användning av *torde* i andra betydelser än de i ordböckerna föreskrivna¹, såtillvida sådana fall kommer fram i korpusarna. Ett särskilt underfenomen med *torde* är användningen av *ha* i samband med supinum eller snarare utelämningen av det. Enligt Svenska akademins grammatik (Teleman & al. 1999, band 4: 286) går det nämligen att lämna bort det perfektuella hjälp verbet *ha* efter *torde* och *måtte*. Jag har drivit korpuslingvistisk forskning med hjälp av korpusverktyget Korp. Jag har två huvudsakliga forskningsfrågor:

1. Hur används *torde* i olika genrer och korpusar?
2. I vad mån och i vilka kontexter brukar det perfektuella hjälp verbet *ha* utelämnas i samband med *torde* och supinum av huvud verbet?

Jag har tre hypoteser. Min första hypotes är att *torde* finner användning främst på högre stilnivåer, eftersom dess användning i betydelsen 'bör lämpligen, begärs' kallas hövlig och formell i Svenska Akademiens ordbok och redan en snabb bakgrundsforskning i Korp ger hänvisning till att verbet har varit vanligare tidigare.

Min andra hypotes, som gäller den andra forskningsfrågan, är att *ha* utelämnas mest typiskt i koncisa tidningstexter, t.ex. i rubriker eller i tabellartade framställningar eftersom korthet är ofta eftersträvad i sådana texter. Dessutom har jag en annan hypotes om *ha*-bortfallet, nämligen att *ha* utelämnas hellre i situationer där bestämmningar har placerats mellan *torde* och supinumet. Detta

¹ Med felaktig användning avses här sådana fall, där skribenten utan större tvivel avsett något annat ord, t.ex. *trodde* eller *tordes*.

argumenterar jag med att en av reglerna till bortfallet är att det kan ske i samband med en infinitivfras inledd av *att* om en bestämning föregår supinumet (se närmare på s. 17), och därför kunde man inducera att en (eventuellt tung) bestämning före supinumet kunde locka till utelämningen av *ha* även i samband med *torde*.

1.2 Disposition

Avhandlingen är disponerad så att en bred teoridel (kapitel 2) följer Inledning, Tidigare forskning och Syfte i kapitel 1. I Teori i kapitel 2 börjar jag med en liten filosofisk inblick och presenterar några centrala grammatiska begrepp och distinktioner. Jag belyser också *torde* som verb, supinum, det perfektuella hjälpverbet *ha* och dess utelämnande. Därefter belyser jag några teknologiska aspekter som är viktiga i korpuslingvistik, framför allt reguljära uttryck, CQP-språket och XML-språket. I kapitel 3, Korpuslingvistik, skriver jag allmänt om korpuslingvistik samt beskriver min egen metod och mitt material. I kapitel 4, *Torde* i korpusarna, beskriver jag resultaten och i kapitel 5, Diskussion, sammanfattar jag dem och ger tolkningar. I kapitel 6, Avslutning, diskuterar jag ännu det hela.

1.3 Tidigare forskning

Maria Bylin har i sin doktorsavhandling *Aspektuella hjälpverb i svenskan* (Bylin 2013) studerat grammatikaliseringsprocesser i den historiska utvecklingen av svenskans aspektuella hjälpverb. Med aspektuella hjälpverb avses hjälpverb som uttrycker aspektkategorin, medan modala hjälpverb uttrycker moduskategorin och temporala hjälpverb uttrycker tempuskategorin. Bylin (2013: 66, passim) behandlar hjälpverb som liknande grammatiska former som t.ex. modus- och tempusändelser är. Skillnaden mellan aspekt- och tempuskategorin lär vara svår att dra, men ett distinktionsförsök som särskilt lyfts fram är att tempus har deiktisk tidsbetydelse och aspekt en icke-deiktisk tidsbetydelse (op. cit.: 65ff). De forskade ”hjälpverbskandidat” som ingått i studien har varit *bruka*, *tendera*, *börja*, *fortsätta*, *sluta* och *hota* (op. cit.: 19). Metoden är olika hjälpverbstest med vilka hon analyserat dessa verb (ibidem). Teorin och empirin överlappar och har behandlats delvis ihopblandade på grund av de använda testens karaktär; teoretisk referensram varvas med analysen av verbet *bruka*. (Op. cit.: 22).

Bylins material består av texter från den fornsvenska tiden ända till yngre texter. Närmare bestämt är det fråga om hela den *Fornsvenska textbanken*, exempelmeningar i *Svenska Akademiens ordbok*, *Litteraturbanken*, *Svensk Dramadialog 1725–2000* samt ett urval korpusar i Språkbanken: några presstextskorpusar och *Bonniersromaner I och II*. Dessa tillsammans kallar hon sin korpus. (Bylin 2013: 75ff.) Ibland har hon dock använt sig av andra korpusar också (jfr op. cit.: 98). *Korpuslingvistik* nämns inte fast studien tydligen har drag av korpuslingvistik. Studien har kvantitativa dimensioner; t.ex. innehåller den 15 tabeller, men huvudsakligen är den en kvalitativ studie. Hon konstaterar att hennes material inte är och med tanke på temat inte heller kunde vara statistiskt representativ och balanserad (op. cit.: 75, 79f). Studien har likheter med min i och med att hon behandlat hjälpverb, behandlat dem ur flera synvinklar, räknat frekvenser, analyserat exempelmeningar och iakttagit genreskillnader.

Marika Lagervall har i sin doktorsavhandling *Modala hjälpverb i språkhistorisk belysning* (Lagervall 2014) forskat på den historiska utvecklingen av svenskans modala hjälpverb samt utfört en jämförelse till andra germanska språk. För att kunna göra detta har hon använt och bedömt olika kriterier, alltså ungefär detsamma som Bylin (2013) gjorde i sin studie med aspektuella hjälpverb. Lagervalls resultat visar tecken på en grammatikaliseringsprocess i germanska språk där de modala hjälpverben orienterar sig allt mera från att vara lexikala ord till att vara grammatiska ord, och dessa diakrona processer har betydelse för t.ex. huruvida ett infinitivmärke (*att* i svenskan) används i samband med ett visst hjälpverb eller ej.

En väsentlig faktor i utvecklingen och samtidigt en förenande egenskap för många av de modala hjälpverben är att många är s.k. *preterito-presentiska verb*, som ursprungligen har (singularis) presensformer som morfologiskt påminner om starka preteritumformer och preteritumformer som morfologiskt är svaga preteritumformer. Ett exempel är *kunna*, som har starkt preteritum *kan* i presensbetydelse (tidigare dock *kunna* i presens pluralis) och svagt preteritum *kunde* i preteritumbetydelse. (Lagervall 2014: 2, 50ff.) Dessutom har många modala hjälpverb mycket defekt böjning, som *måste*, *lär*. Detta gäller även mitt forskningsobjekt *torde*.

Utgångspunkten för identifierandet av de svenska verb som skulle forskas har varit korpusen *PAROLE*, som innehåller nutidssvenska och ingår numera i Korp. Den egentliga analysen baserar sig på ett stort elektroniskt sökbart material som består av åtskilliga korpustexter, främst moderna och medeltida sådana. En särskild tyngdpunkt ligger i moderna och medeltida lagtexter samt moderna presstexter och medeltida krönikor; alltså en jämförelse mellan konservativa texter och

friare skrivna texter. Lagervalls studie har likheter med min i och med att hon behandlat modala hjälpverb i ett stort korpusmaterial, iakttagit genreskillnader och stilistik och använt tid för att gå igenom träffar och gallra bort oäkta träffar. (Lagervall 2014.) Däremot har hon inte fäst stor uppmärksamhet vid semantiska aspekter. En skillnad är naturligtvis också att jag fokuserar på endast ett verb.

Linnéa Bäckström har i sin doktorsavhandling *Etableringen av ha-bortfall i svenskan. Från kontaktfenomen till inhemsk konstruktion* forskat på upprinnelsen av utelämnandet av det perfekta hjälp verbet *ha* i svenskan. Hennes huvudsakliga resultat är att bortfallet av *ha* blev först vanligt i finit form (*har, hade*) och först därefter i infinitivform. Se också kapitel 2.3. Enligt Bäckström bekräftar hennes avhandling den redan tidigare föreslagna teorin om att fenomenet med *ha-bortfall* uppstått som ett lån från tyskan. (Bäckström 2019.) Bäckströms avhandling har relevans med tanke på min avhandling, särskilt gällande den andra forskningsfrågan. Även hon har använt sig av korpusmaterial, och även Korp.

Nuutti Harmo har i sin pro gradu-avhandling *Vore i finlandssvenskt och sverigesvenskt tidningsspråk* (Harmo 2016) forskat på verbformen *vore*. Grundtanken har varit att kartlägga användningen av *vore* i finlandssvenska och sverigesvenska tidningstexter. Enligt resultaten är *vore* vanligare i sverigesvenskan än i finlandssvenskan. Analogt befinner *skulle vara* vanligare i finlandssvenskan. Harmo har sökt i alla fritt tillgängliga korpusar² som Språkbanken erbjuder, så att han har ungefär samma material som jag. Noggrannare har han forskat bl.a. på tidningskorpusar och jämfört frekvenserna av strukturerna *[Oo]m X vore*³ respektive *[Oo]m X skulle vara* eller *[Oo]m X var*. Harmo ställer fast (2016: 36) att det sverigesvenska tidningsmaterialet vid sidan om alldeles moderna texter även omfattar äldre texter än det finlandssvenska tidningsmaterialet. Det kommer inte explicit fram hur han behandlat denna skevhet i sin analys.

Vid jämförandet av förekomsttäteten i olika korpusar/genrer kan det spela en stor roll vad innehållet i texterna är; *vore* och dess alternativa uttrycksätt förekommer nämligen säkert mindre i vissa texttyper än i andra. En korpus kan ha ringa förekomster av t.ex. *vore*, men av innehållsliga, och inte stilistiska orsaker, nämligen på grund av att alla propositioner uttrycks som reella i stället för konditionala eller villkorliga i kontexten. Så kan vara fallet t.ex. i texter där det helt enkelt

2 Moderna svenska korpusar i läget *Moderna*.

3 Ett s.k. reguljärt uttryck *[Oo]* anger två alternativ: *<O>* och *<o>*. Harmo (2016: 25) konstaterar att han ej använt denna formulering utan två med den logiska operatoren *|* förbundna element i en CQP-fras, men jag gjorde det dock här för korthetens skull. Resultaten borde i princip vara desamma.

beskrivs olika slags reella system eller ges befäl utan särskild artighet, eller redogörs om reella händelser, och där det inte bedrivs jämförelser eller spekulationer. Budskapet avgör. Detta har Harmo också delvis iakttagit (2016: 32f), men det är ju svårt att estimeras sådana nyanser. När det gäller *torde* i sin tur, torde det vara betydligt lättare att undvika denna typ av skevhet i resultaten. Den typen av epistemisk modalitet som *torde* uttrycker, nämligen att propositionens sanningsvärde inte tas för givet utan framställs som sannolikt men möjligen osant (jfr Teleman & al. 1999, band 4: 284ff), kan förknippas med nästan vilket slags innehåll som helst. Däremot skulle jag intuitivt antaga att konditional modalitet kan vara häpnadsväckande sällsynt utom i begränsade situationer, framför allt sådana där det spekuleras med olika scenarier eller mottagare ombes göra någonting.

Sofia Alander har i sin pro gradu-avhandling *MÅNNE. En korpusbaserad studie av satsadverbialet männe i finlandssvenska och sverigesvenska tidningstexter* (Alander 2019) forskat på det epistemiska satsadverbialet *männe* i Språkbankens korpusar. Särskild fokus har också denna gång legat på tidningstexter. Alander har utfört en omfattande syntaktisk analys av hurdan ordföljd som råder i satser med *männe*. Sedan har hon jämfört frekvensen av olika ordföljdmöjligheter i finlandssvenska tidningstexter å ena sidan och i sverigesvenska dito å andra sidan.

I Alanders studie kom det fram att finlandssvenska texter i undersökningen har *männe* vanligast i initialfältet och mest sällan i slutet av satsen. I sverigesvenska texter förekommer *männe* mest frekvent i mittfältet medan ungefär samma frekvens gäller för initialfältet och efterfältet. I finlandssvenska texter placeras *männe* typiskt i samband med ett frågeord i initialfältet. Oftast har satsen då rak ordföljd. I sverigesvenska satser förekommer att *männe* används i mittfältet som inskjutning, separerat av kommatecken. Min studie liknar Alanders såtillvida att jag har delvis samma material och jag forskar på ett visst ords användning i korpusar. Syntaxen analyserar jag också när det gäller utelämnningen av *ha* i samband med *torde* och supinum. Däremot forskar jag inte på skillnaden mellan sverigesvenska och finlandssvenska som hon.

2 Teori

Torde är ett modalt hjälpverb som jag forskar på i ett datatekniskt sammanhang, och som uttrycker ett moderat ställningstagande till en propositions sanningsvärde. Man vill inte garantera att något – propositionen – är sann, men vill säga att den är sannolik. Vi har alltså att göra med sanningen. Därför börjar jag avdelningen om teori med en liten filosofisk inblick i sanningens koncept.

Torde räknas som ett *epistemiskt* modalt hjälpverb. Svenska Akademiens grammatik (Teleman & al. 1999, band 4: 282) ger följande definition: ”De epistemiska hjälpverben anger i vilken mån satsen enligt talaren är en sann beskrivning av den faktiska verkligheten [...]” Denna definition förutsätter *korresponsten teorin* för sanningen – fast detta ej nämns rakt ut –, det vill säga att propositionen är sann om och endast om den motsvarar den faktiska verkligheten. Detta i sin tur förutsätter 1) att verkligheten existerar och 2) att den är ontologiskt oberoende av språket.

Aristoteles har formulerat en urform av korresponsten teorin: ”Att säga om vad är, att det inte är, eller om vad inte är, att det är, är falskhet. Men att säga om vad är, att det är, eller om vad inte är, att det inte är, är sanning.” (Aristoteles 320-tal f.Kr. (1990): 1011b26.)

Det har påpekats bland filosofer att det går att utreda sanningsvärden hos satser i naturliga språk enbart efter att satsen har tolkats på något sätt. Därför kunde satsen aldrig nå verkligheten för att kunna vara vare sig sann eller falsk. I Alfred Tarskis version av korresponsten teorin har detta emellertid explicit iakttagits. Tolkningen kan nämligen ha med verkligheten ett samband som kan uttryckas som sant eller falskt. (Niiniluoto 1980: 108–109.)

Ett annat motargument till korresponsten teorin är ungefär följande. För att kunna jämföra propositionen med verkligheten måste man känna både propositionen och verkligheten. En sådan jämförelse ter sig onyttig om man redan känner till verkligheten, och om man inte känner till verkligheten, hur kan man jämföra? Häremot kan ändå påpekas att det har skett en förväxling mellan sanningens definition och sanningens kriterier. Att vi kan definiera sanningen betyder nämligen inte att vi kan urskilja sanna propositioner från falska propositioner. (Niiniluoto 1980: 110.)

En alternativ sanningsteori är *koherensteorin*. Anhängare var bl.a. Baruch Spinoza och Otto Neurath. Enligt koherensteorin är propositionen sann om den passar ihop med ett system av propositioner. Ett problem med teorin är att det kan finnas i alla situationer talrika möjliga propositionssystem eller satsgrupper som inte är motstridiga och vars sanningsvärde koherensteorin således inte kan ta ställning till. Inre koherens kan visserligen vara ett bra, t.o.m. nödvändigt – men otillräckligt – kriterium för att en proposition kan vara sann. (Niiniluoto 1980: 110.)

Därtill har William James, John Dewey och F. C. S. Schiller likställt sanningen med funktionalitet eller framgång. Sant är det som befinner sig instrumentellt värdefullt. Detta kallas *pragmatistisk sanningsteori*. Enligt Niiniluoto är ett problem med denna pragmatism att en indikator av sanning görs till definition av sanning. Även falska uppfattningar kan leda till nyttiga resultat. Ett annat problem med teorin vore att det förväxlas semantik och metodologi. Enligt Dewey kunde en sats först bli sann eller falsk när den verifieras respektive falsifieras. Enligt Niiniluoto är verifiering endast konstaterande av sanningen, som i sin tur är oberoende av huruvida det konstateras eller ej. Verifiering och falsifiering hör således under metodologi snarare än under semantik eller definitioner. (Niiniluoto 1980: 111.)

2.1 Modalitet och hjälpverb

Med modalitet avses i grammatisk teori en semantisk dimension som har att göra med en propositions sannolikhet eller förutsättningar att förverkligas eller talarens inställning till propositionen (Tieteen termipankki). Enligt Svenska Akademiens grammatik eller SAG (Teleman & al. 1999, band 1: 198) har en sats modal betydelse när den inbegriper att ”talaren ifrågasätter eller bekräftar satsinnehållets sanning, anger dess lämplighet eller önskvärdhet eller uttalar sig om subjektreferentens möjlighet eller vilja att förverkliga satsens aktion”.

Modalitet kan uttryckas med hjälp av t.ex. modala satsadverbial som *troligtvis*, modala konstruktioner som *något är att göra* (= ’något måste/ska göras’), modala adjektiv som *möjlig*, *verklig*, *förment* (dessa tre exempel ur SAG) och olika modus, t.ex. moduset *imperativ* uttrycker att satsens aktion befallas. Dessutom uttrycks modalitet även med så kallade modala hjälpverb eller modalverb. (Tieteen termipankki; Teleman & al. 1999, band 1: 198.)

För att ge en bild av modala hjälpverb som fenomen, torde det vara befogat att skissera vad ett hjälpverb är och vad som inte brukar anses som hjälpverb. Gränsen mellan hjälpverb och huvudverb är tyvärr inte alltid lätt att dra. SAG definierar att hjälpverb är verb ”som vanligtvis har en annan verbfras eller participfras som obligatorisk och som enda bundna bestämning”. SAG indelar hjälpverben i fem grupper: modala, *temporala* (t.ex. *det har regnat, det kommer att regna*), *passivbildande* (*lusthuset blev vandaliserat, hon är respekterad av alla*), *aktionella* (*det började regna, det råkade regna*) och *kausativa* (*hon lät bygga ett lusthus*). Av grupperna har de temporala och de passivbildande hjälpverben den särarten att de inte kompletteras med en infinitivfras utan med supinum respektive perfektparticip. (Teleman & al. 1999, band 2: 536. Exemplen SAG:s.) Bolander (2001: 124f) nämner tre grupper – modala, temporala och passivbildande – och talar dessutom om *verbkedjor* som ett överbegrepp. Med aktionella förstår SAG ungefär detsamma som Bylin (2013: passim) med *aspektuella hjälpverb*.

Det brukar klassificeras inom litteraturen somliga kriterier för hur man kan göra skillnad mellan huvudverb och hjälpverb. Lagervall (2014) och Bylin (2013) behandlar också frågan om sådana hjälpverbskriterier. I det följande ska jag emellertid belysa hjälpverbskriterierna på grundval av SAG:s beskrivning. SAG (Teleman & al. 1999, band 2: 536f) ger fem huvudsakliga för hjälpverb typiska egenskaper vilka ”uppvisas alla av de allra mest typiska hjälpverben men i varierande grad av andra verb”.

Det första av SAG:s kriterier definierar att en underordnad verbfras är det enda som kan fungera som bunden bestämning till ett typiskt hjälpverb. Därför säger man t.ex. inte ”*Jag vill gröt” utan ”Jag vill ha gröt” (mitt exempel). Verbet *kunna* kan emellertid uppvisa ett vanligt nominalt objekt som i satsen ”Jag kan forngrekiska” (mitt exempel), vilket SAG inte nämner i detta sammanhang. Däremot nämns det att sätts- och gradadverbial kan fungera som fria (alltså icke-bundna) bestämningar till hjälpverbet som i exemplen (1), (2) och (3). I ex. (1) är det inte regnet som var plötsligt (syftning på huvudverbet) utan det att det slutade (syftning på hjälpverbet). På samma sätt syftar i ex. (2) adverbialet *med knapp marginal* på hjälpverbet *hinna* och adverbialet *i hög grad* i ex. (3) på hjälpverbet *behöva*. Hos det perfektuella hjälpverbet *ha* kan det uppstå en intressant betydelskillnad och tolkningsproblem som åskådliggörs av ex. (4): om tidsadverbialet *klockan 12* syftar på hjälpverbet – som kursiveringen i exemplet antyder –, har handlingen ägt rum tidigare än klockan 12 och tillståndet bara konstaterades klockan 12, men om tidsadverbialet syftar på huvudverbet, ägde handlingen rum just klockan 12 och inte senare. (Teleman & al. 1999, band 2: 537f.)

- (1) Det *slutade plötsligt* regna. [SAG:s exempel.]
- (2) Det *hann med knapp marginal* bli sommar innan torget stod färdigt. [SAG:s exempel.]
- (3) Torget *behövde i hög grad* få en uppfräschning. [SAG:s exempel.]
- (4) Han *hade gått klockan 12*. [SAG:s exempel.]

Det andra kriteriet bestämmer att den underordnade verbfrasen ska ha föga nominala karakteristika, det vill säga svag objektskaraktär. Detta syns bland annat i att den underordnade verbfrasen inte kan ersättas med en icke-pronominell nominalfras hos hjälpverb – se ex. (5) –, medan det hos huvudverb ofta går an – se ex. (6). Typiska hjälpverb uppvisar vanligtvis inget *att* (infinitivmärke) framför infinitiven. Den underordnade verbfrasen kan inte heller ställas som subjekt till hjälpverbet på så sätt att hjälpverbet vore i passiv, vilket åskådliggörs i ex. (7). Däremot går det an hos många huvudverb, som i ex. (8). (Teleman & al. 1999, band 2: 537–539.)

- (5) Han *måste* måla verandan. *Han *måste* {målningen / denna verksamhet / boken} [SAG:s exempel.]
- (6) Han *glömde* att resa till Berlin. Han *glömde* {resan / sitt löfte / boken} [SAG:s exempel.]
- (7) *Det {*kundes/villes/behövdes*} inte acceptera förslaget av styrelsens majoritet. [SAG:s exempel.]⁴
- (8) Att acceptera förslaget försöktes inte ens av styrelsens majoritet. [SAG:s exempel.]

Det tredje kriteriet heter att det typiska hjälpverbet inte kan fungera som korrelerat till *göra (det)*, vilket åskådliggörs i ex. (9). Hos de flesta huvudverb brukar detta däremot göras, som i ex. (10). (Teleman & al. 1999, band 2: 539.)

- (9) – *Måste* Stina bli sekreterare? – Ja, det {*måste* /*gör} hon faktiskt. [SAG:s exempel.]
- (10) – Försökte hon komma in vid grafikskolan? – Nej, det {försökte / gjorde} hon inte. [SAG:s exempel.]

Det fjärde kriteriet heter att typiska hjälpverb ofta är *subjektsautonoma*, det vill säga att subjektet inte har semantisk roll i verbets aktion utan snarare i den till verbet underordnade andra verbets aktion. Mera om detta nedan. (Teleman & al. 1999, band 2: 508f, 537.) Det femte kriteriet heter att det typiska hjälpverbet har defekt eller oregelbunden böjning. (Teleman & al. 1999, band 2: 537.) *Torde* uppfyller alla kriterierna i SAG och är således ett typiskt hjälpverb. (Teleman & al. 1999, band 2: 541.)

4 Däremot kunde man alldeles väl säga: *Förslaget {kunde/ville/behövde} inte accepteras av styrelsens majoritet.* Egen anmärkning.

I alla fall indelar SAG de modala hjälpverben i fyra kategorier på semantiska grunder. Kategorierna är *epistemiska*, *deontiska*, *potentiella* och *intentionella* hjälpverb. (Teleman & al. 1999, band 4: 282.) *Torde* grupperas under epistemiska hjälpverb (Teleman & al. 1999, band 4: 284). För att ge en helhetsbild gör jag en liten överblick över de andra kategorierna innan jag går närmare in på de epistemiska hjälpverben.

De deontiska hjälpverben har att göra med normer, skyldigheter, funktionella eller praktiska nödvändigheter, befallningar, beslut, friheter och dylika. Till deontiska hjälpverb kan räknas, beroende på situationen, åtminstone *måste*, *bör*, *behöver (inte)*, och *ska(ll)*. (Teleman & al. 1999, band 4: 282.) Ordet *deontisk* kommer från forngrekiskans *δεῖν δεῖν* 'behöva, vara tvungen att' och särskilt dess particip presens neutrum *τὸ δέον to déon* (med artikel, pluralis *τὰ δέοντα ta déonta*) 'det nödvändiga, det som måste göras' (jfr Menge & al. 1986, 1993: s.v. δεῖ, δέω).

Potentiella hjälpverb ska inte förväxlas med det finska moduset *potentialis*, som faller under epistemisk modalitet (och som kan översättas med *torde*). Däremot har potentiell modalitet enligt SAG att göra med någons förmåga eller resurser att göra något. Potentiella hjälpverb är t.ex. *kan* och *orkar*. Ordet *potentiell* kommer från latinets *potentia* 'förmåga, makt', till *posse* 'kunna' (jfr Pitkäranta 2001 (2018): s.v. *potentia*, *posse*).

Intentionell modalitet syftar på någons vilja eller avsikt (*intention*) att göra något. Intentionella hjälpverb är t.ex. *vill*, *ämnar*, *strävar efter att*, ibland *ska(ll)* och varför inte även *försöker*. Ordet *intentionell* kommer från latinets *intentio* 'spänning; uppmärksamhet', till *intendere* 'spänna; sträva efter' (jfr Pitkäranta 2001 (2018): s.v. *intentio*, *intendere*).

Epistemisk modalitet syftar på talarens bedömning av i vilken mån propositionen är sann eller pålitlig i förhållande till verkligheten (Teleman & al. 1999, band 4: 282). Ordet *epistemisk* kommer från forngrekiskans *ἐπιστήμη epistēmē* 'vetenskap, kunskap; vetenskap' (jfr Menge & al. 1986, 1993: s.v. *ἐπιστήμη*). Epistemiska hjälpverb i svenskan är ofta såtillvida flexibla att de har en viss specifik grundbetydelse som anger på hurdana grunder talaren tror på propositionen – t.ex. om det är hörsägen eller en slutsats –, men att dessa verb också används vagare för att helt enkelt uttrycka ett förbehåll till propositionens sanningsvärde. SAG indelar de epistemiska hjälpverben i sju kategorier, som presenteras i tabell 1. (Teleman & al. 1999, band 4: 284.)

Tabell 1. Kategorierna av epistemiska hjälpverb enligt SAG (p = proposition)

a	Hänvisning till källa (evidentiell modalitet)	<i>lär, skall</i>
b	Förmodan (man tror att p är sann)	<i>torde</i>
c	Slutsats (det kan räknas ut att p är sann)	<i>måste, bör, kan, behöver inte</i> (i fallande sannolikhetsordning)
d	"Hänvisning till indicier" (några tecken tyder på att p är sann)	<i>Måtte, verkar, förefaller, tycks, ser ut</i>
e	Konstaterande (någon ställer fast att p är sann)	<i>befinnes</i>
f	Hypotes (p framställs som hypotetisk)	<i>skulle</i>
g	Medgivande ("p antas motvilligt av talaren vara sann för resonemangets skull")	<i>må</i>

Enligt SAG är de epistemiska hjälpverben *subjektsautonoma*. Härmed avses att satsens subjekt är semantiskt kopplat enbart till huvud verbet – så länge detta inte även är subjektsautonomt – och har ingen roll vid hjälp verbet. T.ex. satsen "Hon bör vara framme nu." kan omformuleras till: "Det är rimligt att anta att hon är framme nu." (SAG:s exempel.) Det är inte egentligen "hon" som bör utan talaren modifierar med *bör* propositionen att "hon" är framme "nu". Hjälp verbet kommenterar alltså hela propositionen. Däremot syftar *kan* i satsen "Hjördis kan köra bil." (mitt exempel) till Hjördis förmåga att köra ett fordon av det nämnda slaget. I detta exempel är *kan* potentiellt och inte epistemiskt; potentiella hjälpverb är vanligtvis *subjektsorienterade*, vilket är motsatsen till subjektsautonom. (Teleman & al. 1999, band 4: 283.) Undantagsvis är sådana epistemiska verb subjektsorienterade som syftar på att subjektsreferentens utseende är anledningen till att anta att propositionen är sann, t.ex. *verkar, förefaller, och ser ut att* (Teleman & al. 1999, band 4: 285).

Att epistemiska verb brukar vara subjektsautonoma och kommentera hela propositionen kan ha samband med att de framför allt används i deklarativsats och inte så mycket i andra satsmodi som interrogativsats, imperativsats eller exklamativsats. Man kunde tänka att dessa fem är alla olika sätt att använda en proposition p (t.ex. *a är b*): en vanlig deklarativsats konstaterar att p är sann ("A är b."), en interrogativsats frågar om p är sann ("Är a b?"), en imperativsats befäller att p ska förverkligas, m.a.o. bli sann ("A vare b."), en exklamativsats uttrycker talarens emotionella inställning till p ("A är b!" eller "Måtte a vara b!"), då sanninghalten av p blir ofta mindre viktigt, och en epistemiskt modifierad deklarativsats vill inte garantera att p är sann ("A torde vara b."). Om man lägger huvudtrycket på just hjälp verbet, kan det – beroende på kontexten – innebära att man

uttryckligen vill ta avstånd från att propositionen vore sann, som i ex. (11). Samma gäller vissa koncessiva satser som i SAG:s indelning kan falla under epistemisk modalitet och närmare gruppen g (se tabell 1 ovan), som i ex. (12). Ännu tydligare är detta hos det hypotetiska *skulle* (SAG:s grupp f). P skulle vara sann i andra omständigheter, men är det inte i verkligheten. Det finns alltså oftast en implicit negation med *skulle*. Således kunde en *epistemiskt modal sats* anses ha ett säreget förhållande till sanningen jämfört med de traditionella satsmodusen, lite som om den vore ett eget satsmodus i sin egen rätt. Å ena sidan finns det, som sagt, vissa likheter med satsmoduskategorin, fast epistemiskt modal sats inte brukar klassificeras som ett eget satsmodus. Å andra sidan går det bra att kombinera det hypotetiska epistemiska med interrogativsatser som i ex. (13), emellertid inte med imperativsatser. Såvitt andra epistemiska hjälpverb än *skulle* förekommer i interrogativsatser, är det enligt introspektion mest fråga om pleonasm⁵, se ex. (14).

(11) ”Hej, den där flaskan kan innehålla metanol!” – ”Jaja, flaskan **kan** innehålla metanol, men det tror jag inte, så jag ska dricka ur den.” [Eget exempel.]

(12) ”Hon må säga vad som helst, men jag förlåter henne inte.” [Eget exempel.]

(13) ”Skulle p vara sann om omständigheterna vore annorlunda?” [Eget exempel.]

(14) ”Kan det vara sant?” ~ ”Är det sant?” [Eget exempel.]

En ytterligare egenskap hos satser som har betydelse för modala hjälpverb är *satsnegationens räckvidd*. Med satsnegation avses att den av satsen innehållna propositionen negeras, vilket i formell logik kan uttryckas ”~ p”. När propositionen modifieras med ett modalt hjälpverb, lyder frågan, huruvida satsnegationen (som oftast förverkligas genom ordet *inte*) även negerar det modala hjälpverbets betydelse, ~ (*m(p)*) [*m* må vara det modala hjälpverbet, som modifierar *p*.], eller om den endast negerar själva propositionen medan det modala hjälpverbet modifierar den negerade propositionen, *m(~ p)*. I det förra fallet står hjälpverbet inom satsnegationens räckvidd. I det senare fallet står hjälpverbet utom satsnegationens räckvidd. (jfr Teleman & al. 1999, band 4: 285.)

De epistemiska hjälpverben står oftast utom satsnegationens räckvidd, på vilket SAG ger lämpligt ett exempel med *torde*: i exempel (15) modifierar modalverbet *torde* den negerade propositionen ”förhandlingarna kan **inte** slutföras i dag”, det vill säga *m(~ p)* – det är alltså inte törandet som negeras (~~Det är inte troligt att det är så~~) utan törandet modifierar den negerade propositionen.

5 SAOB:s definition om pleonasm: ”språkligt uttryck i vilket ett l. flera led äro (mer l. mindre) överflödiga, i det de kunna utgå, utan att den talandes l. skrivandes mening (väsentligt) förändras (t. ex.: ”ett gammalt fornminne”); äv. inskränkare, om dylik språkföreteelse som har ett visst stilistiskt l. praktiskt ändamål (i det den skänker större eftertryck l. dyl.; motsatt: tautologi); äv. abstraktare, om förhållandet att två l. flera led i ett uttr. göra varandra på detta sätt mer l. mindre överflödiga l. om förekomsten l. bruket av ett dylikt uttr.; stundom äv. (konkretare) om överflödigt ord l. led i ett dylikt uttr.”

Undantag gör *behöva*, *kunna* och *måste*, som står inom satsnegationens räckvidd: i exempel (16) förhåller det sig så att *inte*, negationen, negerar den av modal verbet *behöver* modifierade propositionen ”förhandlingarna **behöver** slutföras i dag”, det vill säga $\sim(m(p))$ – det är alltså inte negationen som modifieras (*Det är nödvändigt att det inte är så att p*) utan negationen negerar den av modal verbet modifierade propositionen. (Teleman & al. 1999, band 4: 285.)

(15) ”Förhandlingarna *torde* inte kunna slutföras i dag. [Det är troligt att det inte är så att p.]” SAG:s exempel.
Hakparenteserna och kursiv SAG:s.

(16) ”Förhandlingarna *behöver* inte slutföras i dag. [Det är inte så att det är nödvändigt att p.]” SAG:s exempel.
Hakparenteserna och kursiv SAG:s.

Epistemiska hjälpverb visar ofta bristfällig tempusböjning. Enligt SAG beror detta på att de – alltså själva hjälpverben – anger talarens bedömning i talögonblicket även om huvud verbet skulle ligga i förfluten tid, som i exempel (17). Själva hjälp verbet kan ändå ligga i förfluten tid hos verb som *kan* och *verkar*, i vilket fall bedömningen av propositionens sanningsvärde är gjord i förfluten tid som oftast är detsamma som huvudverbets tid. Se därpå exempel (18). (jfr Teleman & al. 1999, band 4: 284, 285.)

(17) ”Hans bror *lär* ha arbetat på Åland.” SAG:s exempel. Kursiv SAG:s.

(18) ”Det *verkade* kunna regna hur länge som helst. [Indicierna låg vid en dåtida tidpunkt.]” SAG:s exempel.
Hakparenteserna och kursiv SAG:s.

Enligt min introspektion vore det dessutom semantiskt alldeles logiskt att det skulle finnas dåtidsformer av hjälp verbet *lär* för att uttrycka att bedömningen av källan gjordes i förfluten tid och inte nödvändigtvis den modifierade propositionen. Detta åskådliggörs i exemplen (19) och (20): i bägge exemplen ligger huvudverbets aktion i dåtid; i ex. (19) ligger bedömningen av källan i nutid medan den i ex. (20) ligger i dåtid. I ex. (19) fungerar det att omskriva satsen med *lär*. Semantiskt vore det logiskt att kunna omskriva på motsvarande sätt i ex. (20) också, men detta är ogrammatiskt i svenskan. Att huvud verbet ligger i förfluten tid torde i sin tur vara helt oproblematiskt hos alla svenskans modala hjälpverb överhuvudtaget, också andra än epistemiska. När hjälpverbets betydelse är imperativisk, kan användningen av förfluten tid (perfekt) dock vara sällsynt – exempel (21), som är ett tämligen allmänt yttrande⁶, torde oftast syfta på framtiden snarare än dåtiden trots att huvud verbet formellt sett står i perfekt. I forngrekiskan är det relativt vanligt att ha förfluten tid i

6 Google ger ca 16 700 resultat för uttrycket inom anföringstecken. Sökt 15.2.2020. Korp (Moderna-läget) ger visserligen inga resultat.

imperativ; då är det fråga om imperativ perfekt, som i exempel (22). Visst används det främst i passiv eller medium⁷.

(19) ”Karlsson sägs ha sålt sin röda sportbil.” ~ ”Karlsson lär ha sålt sin röda sportbil.”

(20) ”Karlsson sades ha sålt sin röda sportbil.” ~ *”Karlsson lärde ha sålt sin röda sportbil.” Semantiskt logiskt men ogrammatiskt.

(21) ”Det må vara hänt!”

(22) ”Περίμωσο!” *peřimōso* ’Håll käften!’ Bokstavligt: ’Var stängd med en munkorg!’⁸

Enligt Lagervall (2014: 2, 50ff) spelar s.k. preterito-presentiska verb en väsentlig roll i utvecklingen av de svenska modala hjälpverben. När tempussystemet och verbböjningen utvecklades i urgermanskan, uppstod preteritum av starka verb på grundval av tidigare aorist- och perfektformer, och preteritum av svaga verb som en nybildning på grundval av sannolikt ett hjälpverb (eventuellt ett motsvarande engelskans *do* och tyskans *tun* ’göra’), som förvandlades till suffix. En grupp verb fick emellertid en presensbetydelse för sina tidigare perfektformer. För att även kunna använda dessa verb i dåtid bildades nya preteritumformer på samma sätt som hos svaga verb. (Op. cit.: 52f.) Detta förklarar t.ex. att verbet *skola* har en presensform som liknar preteritum av starka verb, nämligen *ska(ll)*, och på samma sätt i engelskan *shall*, även svenskt *kunna* : *kan*, engelskt *can*, tyskt *können* : *kann* (Jfr op. cit.: 61, 66f). Etymologin av *torde* har samband med preterito-presentiska verb som det konstateras i följande kapitel.

2.2 Etymologin och morfologin av *torde*

Det modala hjälp verbet *torde* har enligt SAOL betydelsen ’lär sannolikt’ (s.v. *torde*). SO ger beskrivningen: ”bör sannolikt i uttryck för förmodan, möjlighet o.d.”. Utöver denna epistemiska betydelse ger SO även en annan, deontisk betydelse: ”bör lämpligen i uttryck för hövlig uppmaning”. (S.v. *torde*.) Morfologiskt sett är <torde> preteritumformen och <tör> presensformen, men i praktiken fungerar <torde> semantiskt som presensform medan böjningsformen <tör> är nästan utdöd. I början blev formen <torde> presens i en försvagad betydelse med eventuellt mindre sannolikhetsgrad, men denna skillnad har försvunnit, enligt SAOB (s.v. *tör_{verb}*).

7 *Medium* kallas en tredje diates som grekiskan har mellan aktiv och passiv. Betydelsen är oftast reflexiv, reciprok eller intransitiv. Passiv- och mediumformerna är identiska i perfekt så att de ej går att åtskilja i detta fall.

8 Den som befäller på detta sätt vill alltså åstadkomma i åhöraren ett tillstånd som resulterat från att åhörarens mun har stängts i dåtiden med en munkorg. Det är dock ett bildligt uttryck och man får utgå från att det inte föreligger någon verklig munkorg.

Marika Lagervall (2014) har behandlat även *torde* och dess historiska utveckling vid sidan om andra modala hjälpverb i sin doktorsavhandling. Hjälpverbet är etymologiskt samma verb som deponensverbet *töras* (*tordes*) 'våga', och den senare betydelsen anses ursprungligare. Verbet har blivit deponensverb i den ursprungligare betydelsen genom *lexemspaltning* i svenskan. I danskan uppvisar det motsvarande *turde* däremot de bägge betydelseerna. Den fornsvenska och fornnorska formen var *þora*. Detta verbs ursprung råder det i sin tur ingen konsensus om. (Lagervall 2014: 59–61; Hellquist 1922: 1052f; Ernby 2008: 707, 729.)

Enligt några är *þora* en avvikande utveckling från det urgermanska **durzan* 'våga'. I alla fall verkar *þora* ha förväxlat med arvingarna till det urgermanska **þurbān* 'kräva, behöva', närbesläktade till nysvenska *tarva*. De två preterito-presentiska verben, **durzan* och **þurbān*, har förväxlat med varandra även i andra germanska språk och rönt många komplexa semantiska, fonologiska och morfologiska förändringar. Besläktade är bland andra nyengelskans *dare* 'våga' och tyskans *dürfen* 'få göra ngt' och *bedürfen* 'behöva'. (Lagervall 2014: 59–61; se även Hellquist 1922: 956.) Det nysvenska omljudsförhållandet *tör—torde* kan förklaras med förväxling med formerna av de gamla verben *þorva* och *þörva* (arvingar av **þurbān*) och med analogiskt inflytande från verb som *bör—borde* och *spör—sporde*. (Hellquist 1922: 1053.)

2.3 Det perfektuella hjälpverbet *ha*

Utom modala hjälpverb finns det i svenskan, som beskrivits ovan, även andra hjälpverb. Temporala hjälpverb har som uppgift att bestämma vad som är förhållandet av tiden för huvudverbets aktion till tiden för hjälpverbets aktion. Även temporala hjälpverb kan dock uppvisa modala bibetydelse, t.ex. *skola* förknippas ofta med tanken att någon ämnar eller vill göra någonting. Detta gäller emellertid inte *ha*, som är relevant på grund av min andra forskningsfråga. (Teleman & al. 1999, band 4: 229.)

Det perfektuella hjälpverbet *ha* bestäms av en verbfras med huvudverbet i supinum (Teleman & al. 1999, band 4: 230) eller ibland av ett anaforiskt *det* som syftar på en verbfras med supinum (op. cit., band 4: 231). Denna verbfras kan betraktas som objekt till *ha* (op. cit., band 3: 272). I finit form kan det perfektuella hjälpverbet förekomma antingen i presens (*har*), då det bildar tillsammans med supinum det sammansatta tempuset perfekt, eller i preteritum (*hade*), då det bildar tillsammans

med supinumet det sammansatta tempuset pluskvamperfekt (op. cit., band 4: 229). Konstruktionen kan dock förekomma med *ha* i infinitiv som i ex. (23) (op. cit., band 4: 231).

(23) Hon påstår sig *ha* läst Bibeln rätt igenom trettio gånger. [Exempel och kursivering SAG:s.]

Supinum är en verbform som klassificeras som icke-finit, men det har även finita egenskaper, särskilt i och med att supinum kan fungera som predikat i vissa satser med *ha*-bortfall. Supinum är unikt för svenskan bland de germanska språken. Formen uppstod på grundval av perfekt particip i början på 1700-talet; tidigare hade man använt perfekt particip i bägge funktioner. Uppkomsten av supinumet antas ha samband med att *ha*-bortfallet hade blivit vanligt. (Bäckström 2019: 21f.)

I en så kallad *supinumfras* är supinumet oundgängligt, men det perfektuella hjälpverbet kan däremot under vissa omständigheter utelämnas. Då talar man om *ha-bortfall*. En klassisk situation som detta sker i är satser med bisatsordföljd⁹. Det frånvarande *ha* kan beroende på kontexten lika bra vara presens <har> eller preteritum <hade>, se ex. (24). (Teleman & al. 1999, band 3: 272; band 4: 230.) Bortfallet kan också ske i infinitivformen <ha> både i bisats och i huvudsats om det fungerar som objekt till ett hjälpverb. SAG förutsätter i band 4 att detta hjälpverb måste stå i preteritum medan den i band 3 – som behandlar ämnet noggrannare – formulerar detta vagare och har en exempelmening med presens (*lär*). I ex. (25) kan *ha* utelämnas efter hjälpverb; supinumfrasen är betecknad med kursiv. (Op. cit., band 4: 230; band 3: 272f.) Bortfallet kan också ske vid så kallade nexusinfinitiv¹⁰, som i ex. (26). I samband med en infinitivfras inledd av *att* kan bortfallet ske om en bestämning föregår supinumet som i ex. (27). Enligt SAG rekommenderas bortfallet av *ha* inte av språkvårdare utom frivilligt i bisatser. (Op. cit., band 3: 273.) I samband med epistemiska hjälpverb nämns det specifikt om *torde* och *måtte* att det perfektuella hjälpverbet kan utelämnas efter dem (op. cit., band 4: 286). Bäckström (2019: 13–15) ”utmanar” SAG:s och vissa andras sätt att förknippa af-satsen och *ha*-bortfallet tätt med varandra, men konstaterar också att SAG:s definition är ”tydlig och koncentrerad”.

(24) Eftersom Otto {har/hade/Ø} besökt Paris ... [Sammanslaget från två exempel i SAG, band 4.]

(25) en bok jag måtte (*ha*) läst redan i ungdomen [Exempel och kursivering SAG:s, band 3.]

(26) Hundra personer uppges redan (*ha*) sänt in sina protester. [Exempel och kursivering SAG:s.]

9 Egentligen af-satser (adverb före fundamentet). Detta är ej helt samma sak som bisatser.

10 Nexusinfinitiv består av en infinitivfras utan infinitivmärke vilken utgör bunden bestämning till ett transitivt verb vars objekt blir infinitivfrasens predikationsbas eller till ett transitivt verb i passiv så att subjektet blir infinitivfrasens predikationsbas (Teleman & al. 1999, band 1: 201).

(27) [Hon] har fått många svarta rubriker efter att offentligt *kritiserat en av sina ämbetsmän*. (S) [Exempel och kursivering SAG:s.]

2.4 Reguljära uttryck

Ett *reguljärt uttryck* kan i databehandling definieras som en ändlig teckensekvens som har som uppgift att bestämma ett visst sökmönster (eng. search pattern). Begreppet härstammar från matematisk teori om formella språk men har fått stor spridning på datatekniska områden. Nu för tiden är inte längre tillgången till data det vanskliga utan filtreringen av data. Det hjälper de reguljära uttrycken med. (Nagy 2018: kap. 1.) De är ett väsentligt verktyg även då man gör korpuslingvistiska sökningar.

Reguljära uttryck heter på engelska *regular expressions* och förkortas *regex* eller *regexp*. Dessa förkortningar används ofta i svenskspråkiga texter också, men jag har valt att huvudsakligen använda i denna avhandling den svenska termen, som även är vedertagen.

I mångsidig databehandling, inklusive korpuslingvistik, räcker det inte till att bara söka ett visst ord när man ska söka något i en databas eller t.ex. i en korpus. Det som man söker kan ofta inte uttryckas med hjälp av bara ett ord i en form. Ord böjs eller uppvisar vacklande ortografi, särskilt när det är fråga om historiska texter, men även hos moderna texter. Det kan hända att man vill söka ett prefix eller ett suffix och veta just med vilka efterled respektive förled den förekommer. Ett behändigt redskap för att möjliggöra detta effektivt och snabbt är just reguljära uttryck.

Innan jag beskriver närmare vad reguljära uttryck i praktiken betyder, ger jag några exempel på vad de inte är. Ofta när man använder sökfunktionen (Ctrl + F) i bl.a. textbehandlingsprogram, textredigerare eller webbläsare, får man välja, huruvida man vill göra skillnad mellan gemener och versaler eller ej. Standardinställningen (default) är vanligast att skillnad inte görs, det vill säga sökningen är skiftlägesoberoende. Detta kan göra sökningen långsammare, och det kan finnas andra orsaker att söka skiftlägesberoende i stället för skiftlägesoberoende. Att förändra inställningen brukar man göra *grafiskt* (graphic user interface, GUI) genom att klicka på en grafisk ruta på skärmen med musen. Då är det inte fråga om reguljära uttryck. Reguljära uttryck är någonting man gör främst med tangentbordet.

Lite närmare koncepten reguljära uttryck kommer förfarandet med jokertecken i många enkla databaser och dylika. I Stora finsk–svenska ordboken (Martola (red.) 1997) fungerar procenttecknet `<%>` och understrecket `<_>` som jokertecken. Procenttecknet betyder vilket tecken som helst hur många gånger som helst i just det angivna stället, alltså noll eller flera gånger. Däremot betyder understrecket att vilket tecken som helst uppträder exakt en gång i just det angivna stället. T.ex. `ka_kki` ger träffar på både *kaakki* 'hästkrake', *kaikki* 'allt, alla', *kalkki* 'kalk' och *karkki* 'godis', men inte på *kajakki* 'kajak'. Sökningen `närä%` ger träffar på *närä* 'groll', *närästys* 'halsbränna' och *närästää* (opersonligt) 'ha halsbränna'. I många andra sammanhang fungerar asterisken `<*>` på samma sätt som procenttecknet ovan. Uttrycket `*.pdf` kan syfta på alla filer som har Portable Document Format.

Reguljära uttryck är en väsentlig del av många programmeringsspråk, men används också utanför dem, som det förhåller sig bl.a. i denna studie. Gemensamt med programmeringsspråken är att även reguljära uttryck följer reglerna (!) av en viss *syntax*. Med *syntax* avses inte här samma sak som i grammatiken, utan *syntaxen* i databehandling har kanske större sammanhörighet med vad man i lingvistik kallar för ortografi. Vissa tecken – s.k. *metatecken* (eng. *meta character, special character*) – har en viss betydelse, och vissa tecken och speciella ord¹¹ måste följa varandra i en viss ordning för att en maskin ska kunna tolka texten rätt. (Jfr Nagy 2018: kap. 1, 2; Nield 2019: kap. 1.)

Idén med reguljära uttryck är att med en viss *syntax* ange en mängd strängar för att se om de matchar någon delsträng i en större sträng som man är intresserad av. Samtidigt som maskinen gör detta, kan maskinen även räkna frekvensförekomster, det vill säga hur många gånger en sträng som det reguljära uttrycket angivit, eller alla av dem, förekommer i den större sträng som man söker i. Maskinen kan även göra övrig statistik. (Jfr Nagy 2018: kap. 1; Nield 2019: kap. 1.)

Syntaxen kan variera i och med att det finns lite olika versioner av reguljära uttryck. Vanligen används emellertid en punkt `<.>` som jokertecken. Olika vanliga upprepningsoperatorer, med vilka man kan låta det föregående tecknet eller flera tecken upprepas ett visst antal gånger, presenteras i tabell 2. Om man vill låta en grupp av flera tecken upprepas noll till flera gånger, kan man omge de önskade tecknen med runda parenteser. (Nield 2019: kap. 1.)

11 Särskilt s.k. *reserverade ord*, eng. *reserved words*. Sådana har dock knappast betydelse i reguljära uttryck, åtminstone på denna nivå, men exemplifierar vad *syntax* kan betyda i databehandling.

Tabell 2. Allmänna upprepningsoperatorer i reguljära uttryck

$xyz?$	z förekommer 0 eller 1 gång, alltså xyz eller xy
xyz^*	z förekommer 0 eller flera gånger, alltså hur ofta som helst
xyz^+	z förekommer 1 eller flera gånger
$xyz\{2, 7\}$	z förekommer 2–7 gånger
$x(yz)\{0, 2\}$	yz förekommer 0–2 gånger, alltså xyz , $xyzyz$ eller $xyzyzyz$

Inom hakparentes kan man ange alternativ som man ska välja emellan, t.ex. $[Tt]orde$ kan ge träffar på både $\langle Torde \rangle$ och $\langle torde \rangle$. Detsamma kan åstadkommas med hjälp av ett eller vid behov flera lodstreck $\langle | \rangle$, $(T|t)orde$. Om ett metatecken, som punkt, frågetecken, parenteser eller dylikt används verkligen som en del av det egentliga uttrycket och ska förstås bokstavligt och synas i text som är skriven på ett naturligt, mänskligt språk, måste det föregås av ett s.k. *avbrottstecken* (eng. escape character), vilket oftast är ett omvänt snedstreck $\langle \backslash \rangle$. Om man t.ex. vill leta fram alla punkter i en korpus, kan man ange detta med det reguljära uttrycket $\backslash \cdot$ behändigt. (Nield 2019: kap. 1.)

2.5 CQP-språket och andra formella språk

Sökfrågorna i Språkbankens sökgränssnitt Korp formuleras i bakgrunden på CQP-språket även om man vill använda det grafiska användargränssnittet (GUI), alltså Enkel och Utökad sökning. Därför är det viktigt att ge en inblick på temat i denna avhandling. CQP eller *Corpus Query Processor*¹² (Evert & CWB-utvecklingsgruppen 2005–2020) är ett formellt sökspråk som har utvecklats för att söka innehåll i korpusar och ge statistik om resultaten. En digitaliserad korpussamling är ett slags databas, och databaserna behöver ett *frågespråk* (eng. query language). CQP är ett frågespråk som har utvecklats i ett projekt som heter The IMS Open Corpus Workbench (CWB). Det har öppen källkod.¹³ Vad gäller Korp, ligger CQP dolt i Enkel sökning, och i Utökad sökning är användningen av reguljära uttryck en valbar möjlighet, men annat sköts genom GUI, medan i Avancerad sökning skriver man frågorna direkt som CQP. (Hjortstam 2018: 6.) I Avancerad sökning-fliken ser man ändå den aktiva sökfrågan i Enkel sökning och i Utökad sökning i CQP-form, såvitt man har ställt en sökfråga i Enkel sökning respektive Utökad sökning-fliken Jag återkommer till Enkel sökning

12 På några ställen står det att förkortningen är av *Corpus Query Protocol*, t.ex. i Kielipankki & al. 2015–2020.

13 Källkoden finns i: $\langle \text{https://github.com/spraakbanken/korp-backend} \rangle$.

och Utökad sökning i kapitlet 3.2. I reguljära uttryck används en version eller dialekt av reguljära uttryck som heter *PCRE (Perl Compatible Regular Expressions)*. (Hjortstam 2018: 8.)

CQP är avsett att användas för korpusar som har CWB-form, som är en version¹⁴ av XML (Extensible Markup Language). Det är fråga om ett s.k. *märkspråk* (eng. markup language). Med hjälp av ett märkspråk det går att (1) ge information om textens formatering för att kunna tolkas på rätt sätt oberoende av på vilken enhet eller plattform man använder texten, exempelvis `<i>viktigt</i>` betyder att ordet ”viktigt” ska återges med kursiv stil i t.ex. webbläsaren, samt (2) ge om texten och dess ord sådan information som en maskin förmår att läsa, i syfte att möjliggöra mångsidigare sökningar än bara ordformer och dylika. Man kan söka rena ord i en vanlig textfil också, och då behöver man varken CQP eller något annat korpusredskap; en textredigerare räcker till. Om man däremot vill söka på t.ex. substantiv eller supinum, måste man ha en text där information om ordklasser och böjning har tillagts för maskiner att läsas. I korpussammanhang är främst den senare användningen relevant. (Jfr Evert & CWB-utvecklingsgruppen 2005–2020: 6; Bray & al. 2008.)

Båda användningarna av märkspråk, som presenterats ovan, förverkligas genom uttryck som följer ifrågavarande märkspråks syntax och som inte syns till den normala mänskliga läsaren. Inom XML och dess arvingar förekommer dessa i form av bl.a. *taggar* och *entitetreferenser*. (Bray & al. 2008: kap. 3.1, 4.1.)

Entitetreferenser börjar oftast med et-tecknet `&` och slutar på semikolon `>`. De används bl.a. för att referera till ett visst tecken. Detta gör man av åtminstone tre orsaker: man vill säkerställa att ett mera sällsynt tecken syns korrekt till läsaren trots att olika enheter eller plattformar kan ha olika slags kodering (Unicode, ANSI, ASCII osv.); det önskade tecknet ser likartat ut med något annat tecken och man vill undvika misstolkningar hos den mänskliga läsaren, t.ex. bindestreck och tankstreck; eller tecknet har en annan speciell funktion i märkspråkets syntax som man vill undvika, det vill säga är ett metatecken, och man vill att tecknet ska förstås bokstavligt i stället för som metatecken. Et-tecknet kan återges med entitetreferensen `&`; (< eng. *ampersand* ’et-tecken’) osv. På samma sätt kan nästan vilket som helst Unicode-tecken återges med en s.k. *teckenreferens* baserad på tecknets Unicode-ordningsnummer: t.ex. det isländska och färöiska edh-tecknet `<ö>` kan

14 *Version* är ett avsiktligt vagt uttryck; källorna lämnar ganska oklart, vad CWB:s förhållande till XML är. I alla fall är Korps korpustexter, som man kan ladda ner på <https://spraakbanken.gu.se/resurser>, i grund och botten XML-filer, såvitt jag har laddat ner några av dem.

återges med $\&\#240$; (decimalsystem) eller $\&\#xF0$; (hexadecimalsystem¹⁵). (Jfr Bray & al. 2008: kap. 4.0, 4.1.)

Taggar placeras vanligtvis mellan vinkelparenteser $\langle \rangle$ och $\langle \rangle$, och förekommer parvis omkring ett eller flera egentliga ord, det vill säga ett *elementinnehåll*, eller som en ensam, innehållslös tagg¹⁶, t.ex. $\langle xxx \ / \rangle$, där xxx representerar själva taggtexten¹⁷, vad den än må vara. Om det finns innehåll, börjar man med en starttagg $\langle xxx \rangle$ och slutar med en sluttagg $\langle /xxx \rangle$. Ordet eller orden mellan dem kallas innehåll. Tillsammans bildar dessa tre ett *element*. (Bray & al. 2008: kap. 3.0, 3.1.) Element kan inkapslas inuti varandra så att ett element innehåller andra element, t.ex. $\langle mening \rangle$ Det är fråga om en $\langle specialterm \rangle \langle i \rangle$ dikotomi $\langle /i \rangle \langle /specialterm \rangle$. $\langle /mening \rangle$ (Evert & CWB-utvecklingsgruppen 2005–2020: 6f). De inkapslade taggarna får ändå inte vara identiska, m.a.o. får man inte använda dem *rekursivt*, t.ex. $\langle np \rangle$ the man $\langle pp \rangle$ with $\langle np \rangle$ the telescope $\langle /np \rangle \langle /pp \rangle \langle /np \rangle$ (exempel ur källan) är inte tillåtet. (Evert & CWB-utvecklingsgruppen 2005–2020: 6f). Taggarna får inte heller överlappa varandra så att ett element skulle börja inuti ett element och fortsätta vidare efter att detta har slutat.

Inom en starttagg eller en ensam innehållslös tagg kan man placera ett s.k. *attribut*, som har uppgiften att beskriva egenskaper av ifrågavarande elements innehåll eller dylikt. Attributet består av ett *namn* och ett *värde*, vanligtvis bundna med $=$ -tecknet, t.ex. $\langle citat \ sprak = latin \rangle$ där $sprak = latin$ är attributet, $sprak$ är namnet och $latin$ är värdet. Exemplet kan innebära att taggens innehåll (ett citat) ska tolkas som att vara latinspråkigt, men den som gör en XML-fil kan i princip oftast definiera detta själv. Det kan finnas många attribut inom samma tagg, och de brukar skiljas från varandra med mellanslag. Samma attribut kan förekomma och oftast förekommer i flera element inom samma text. (Bray & al. 2008: kap. 3.1 & passim.)

Evert och CWB-utvecklingsgruppen (2005–2020: 6) konstaterar att det finns i det av dem avsedda korpussammanhanget två typer av attribut: *positional attributes*, som torde motsvara det svenskspråkiga begreppet *ordattribut*, som används i Språkbanken; och *structural attributes*, som förvirrande nog torde avse taggar och inte det man vanligtvis kallar attribut. Evert & CWB-

15 Decimalsystemet är vårt vanliga talsystem, där talbasen är 10. Hexadecimalsystemet är behändigt i databehandling och har 16 som talbas. Bokstäverna A–F (skiftläget är betydelselöst) står för decimalsystemets 10–15. I Unicode anger man ordningsnumren av tecknen hellre i hexadecimalsystemet eftersom det sparar utrymme och kapacitet.

16 Då talar man om ett *tomt element*.

17 Denna kallas *typ*, eller dess *namn*, eng. *type*, *name*. Typens namn är inte samma sak som attributets namn.

utvecklingsgruppen (ibidem) låter en förstå att structural attributes markeras som vanliga XML-taggar men att positional attributes markeras i CWB-texter efter elementinnehållet, eller snarare tokenet¹⁸ med snedstreck, t.ex. An/DET/a där DET betyder att ordet i fråga är en artikel, och a är ordets lemma, det vill säga grundform. I Språkbanken förhåller det sig dock inte så, utan ordattributen verkar markeras inom vanliga XML-taggar, såsom också de s.k. *textattributen*, vilka gäller större områden än bara ett enskilda token.¹⁹ Ett exempel på XML-element i en av Språkbankens korpusar är `<w pos="VB" msd="VB.PRT.AKT" lemma="|finna|" lex="|finna..vb.1|" sense="|finna..2:0.515|finna..1:0.485|" prefix="|" suffix="|" compwf="|" complemgram="|" ref="11" dephead="08" deprel="MS">fann</w>` (Korpusen *Psalmboken 1937*: rad 87²⁰) där w är den egentliga taggtexten, fann är innehållet och största delen av resten är attribut. I Språkbankens korpusfiler ligger varje token på sin egen rad, taggad med w. Andra taggar är t.ex. sentence som annoterar, var meningsgränser antas ligga.

Sökningar i CQP består huvudsakligen av uttryck inom hakparenteser, som indikerar token. Enklast är att söka bara elementinnehåll och inte bry sig om taggar. Då kan man söka `[word = "tokenet"]`, som ger träffar på just denna grafiska form. I praktiken kan man bara skriva tokenet inom anföringstecken utan hakparenteserna osv. för att åstadkomma detsamma, men det fungerar enbart i denna enklaste form av sökningar. Om man vill söka flera ord som ska följa varandra, kan man helt enkelt lägga hakparentesuttrycken efter varandra: `[word = "token"][word = "efterföljande token"][word = "tredje tokenet"] osv.` (Hjortstam 2018: 9, 27.)

Om man vill ange alternativa token, kan man använda lodstreckets `<|>`: `[word = "hundar|katter"]` ger samma träffar som `[word = "hundar"]|[word = "katter"]` – alltså träffar på `<hundar>` och `<katter>`. Man kan också använda parenteser för att ange lodstreckets räckvidd, t.ex. `[word = "ble(v|f)"]` ger träffar på `<blev>` och den äldre stavningen `<blef>`. Tack vare detta är det lätt att iaktta alternativa skrivsätt, som t.ex. `[word = "förresten"]|([word = "för"][word = "resten"])` ger träffar på `<förresten>` och `<för resten>`. Alternativ kan även anges med hjälp av hakparentes som det konstaterats ovan i kapitel om reguljär

18 Token betyder ungefär 'ord', men även skiljetecken räknas som token.

19 Jag har visserligen kunnat öppna för läsning endast en korpusfil, eftersom de flesta är för stora filer för att öppnas med tillgängliga resurser. Den som jag har kunnat läsa, är korpusen *Psalmboken 1937*. Det torde ändå utgöra ett tillräckligt och dugligt exempel.

20 Filen *psalmboken.xml* kan laddas ner i bz2-komprimerad form i: `<https://spraakbanken.gu.se/resurser>`.

uttryck, t.ex. `[word = "ble[vf]"]`. Inom hakparenteser kan även ett spann anges²¹. `[0-9]` innebär vilket som helst nummer och (kombinerat med upprepningsoperatoren, se tabell 2 på s. 20) `[0-9]+` innebär vilken som helst kedja av nummer. `[a-zåäöA-ZÅÄÖ]` innebär vilken som helst av de vanliga svenska bokstäverna men utesluter t.ex. `<é>` och `<æ>`. På motsvarande sätt ger `[word = "[j-n]app"]` träffar på `<japp>`, `<kapp>`, `<lapp>`, `<mapp>` och `<napp>`. (Hjortstam 2018: 10–12, 29.)

Uppreppningsoperatorerna fungerar i CQP på det sätt som är vanligt i reguljära uttryck och som presenterats i tabell 2 på s. 20. Som jokertecken används punkt `<.>` när det är fråga om jokertecken inom ett token eller attributvärde. Om man däremot vill tillåta vilket som helst token mellan vissa andra token, använder man i stället ett tomt par av hakparenteser, t.ex. `[word = "den"] [] [word = "som"]` ger träffar på tokenen `<den>` och `<som>` med ett godtyckligt token mellan dem, och `[word = "den"] [] * [word = "som"]` tillåter hur många som helst godtyckliga token mellan `<den>` och `<som>`. Det gäller att beakta när man kombinerar upprepningsoperatorer med sådana jokertoken och jokertecken att det upprepade tokenet respektive tecknet inte behöver vara detsamma när det upprepas. Föregående exempel kunde alltså ge träffar på en sträng som `<den kalla kalla som>`²², men också sådan som `<den vackra, vita, snabbt växande kalla som>`. (Hjortstam 2018: 13–15.)

Det går också att negera värden med hjälp av operatoren `<!=>`, t.ex. `[word != "äro"]` ger träffar på alla token som inte är `<äro>` – och fyller eventuella andra krav som man kanske ställer. Detta kommer främst till nytta som en del av komplexare sökuttryck, jfr nedan. Inom hakparentesuttryck, med vilka man anger alternativ, sker negerandet med `<^>`, t.ex. `[^åäöÅÄÖ]+` ger inte träffar på ord som innehåller s.k. skandinaviska bokstäver *å*, *ä*, *ö*. (Hjortstam 2018: 9, 12f.) CQP är i princip skiftlägesberoende. Om man vill inbegripa versaler eller gemener utan att göra skillnad mellan dem, kan man ignorera skiftläget genom att tillägga `<%c>` i sökuttrycket efter värdet och de anföringstecken som omger värdet. På motsvarande vis kan man ignorera diakritiska tecken med hjälp av `<%d>`. T.ex. `[word = ".{0,2}frangen" %c %d]` kan ge träffar på `<refrängen>`, `<Refrängen>`, `<REFRÄNGEN>`, `<Frangen>`, `<Frangén>` osv. (Hjortstam 2018: 15f.) Om man vill söka på metatecken eller skiljetecken – mest bildar de redan ett token i sig – används det omvända snedstreckat `<\>` som avbrottstecken som det är vanligt i reguljära uttryck. Anföringstecken ska dock åstadkommas på ett annat sätt: för att söka på tokenet `<>` (ett enkelt citattecken) skriver man

21 Spannets räckvidd är alltid endast ett tecken på varje sida av bindestrecket. (Hjortstam 2018: 12.)

22 Ett icke-varmt exemplar av familjen kallaväxter, *Araceae*.

[word = '''] och för att söka på <"> (ett dubbelt citattecken) skriver man [word = """]. (Hjortstam 2018: 16f.)

Utöver själva tokenen kan man även söka i XML-taggar, närmare bestämt i ordattributen och textattributen. På samma sätt som `word` fungerar som s.k. *parameter* för token, använder man andra parameter för ord- och textattribut. Parameterns *värde* anger man sedan med <=> eller hos vissa attribut med `contains` samt anföringstecken. Urvalet av olika attribut varierar beroende på korpus. Det gäller att vara noggrann att man söker på attribut som existerar i de valda korpusarna. (Hjortstam 2018: 20ff.)

Ett av de väsentligaste ordattributen är *lemgram* (`lex`). Då är det fråga om alla böjningsformer av ett ord och man söker vilken som helst av dem. T.ex. lemgrammet *röra* (*verb*) kan sökas med [lex contains "röra\\.\\.vb\\.1"] (Exempel ur källan). I praktiken är det i Korp lättare att söka på lemgram med hjälp av det grafiska användargränssnitt som finns i Enkel eller Utökad sökning; då behöver man inte veta namnen på lemgrammen i CQP-form utan programmet föreslår lämpliga lemgram. Ett annat viktigt ordattribut är *msd* (`msd`, förkortning av eng. *morphosyntactic description*), som ger kunskap av ordets ordklass och böjningsform. T.ex. alla verbformer i supinum s-form kan sökas med [msd = "VB\\.SUP\\.SFO"]. Om man vill söka på alla supinum oberoende av diates, kan man skriva [msd = ".*SUP.*"]. Textattribut varierar korpus för korpus i större mån än ordattribut. Textattribut indikerar titel, utgivningsår eller dylikt för den text (hela korpusen eller oftare en del av korpusen) som det sökta tokenet ingår i. T.ex. token som har titeln *Sancti Joannisoppenbarilse*, som finns i korpussamlingen *Yngre religiös prosa*,²³ kan sökas med [_.text_title = "Sancti Joannisoppenbarilse"] (Exempel ur källan). I praktiken är det lättare att söka på textattribut med hjälp av det grafiska användargränssnitt som finns i Enkel eller Utökad sökning; då föreslår programmet alla de textattribut som förekommer i de valda korpusarna. (Hjortstam 2018: 22–26.)

Det går lätt att kombinera flera parametrar för samma token med hjälp av et-tecknet <&> (Hjortstam 2018: 17). T.ex. [word = ".*haf{0,2}[wvu]a" %c & word != ".*hava" %c & msd = ".*VB.*" & _.text_year != "18.."] söker på verbformen *hava* i andra ortografiska varianter än just <hava>, med eller utan förled, skiftläget ignorerat, i texter som inte

23 Ingår i läget *Fornsvenska*, och inte *Moderna*, som jag använder. Skrivformen "Sancti Joannisoppenbare[sic!]lse", som förekommer i källan (Hjortstam 2018: 26) torde vara en lapsus calami.

härstammar från 1800-talet. Det kan ge träffar på <hafva>, <haffwa>, <haffua> osv.²⁴ Vad man än må göra med Korp, gäller det att komma ihåg att om man använder CQP-uttryck med felaktig syntax eller söker på attribut som inte finns i just den valda korpusen eller som finns i bara en del av de valda korpusarna, eller om man gör andra dylika misstag, får man inte alltid felmeddelande. Det händer nämligen ofta att man får noll träffar eller för lite träffar fast det man har avsett verkligen skulle finnas i stor mån i materialet.

24 En exempelmening, som uppvisar humoristisk, pastischerande användning av äldre stavning i en modern text: @Muralgranskaren " *Then som haffwa kluven inställning till sitt gender och gå omkring i rosa särk skola själv klyfwas på den yttersta dagen* " (Korpus: *Twitter – 2016*). En autentisk gammal text skulle i princip ha singularis <haffwer> osv. i stället för pluralis <haffwa> osv.

3 Korpuslingvistik

Min metod är *korpuslingvistik*. En *korpus* är en textmassa som är utvald eller begränsad på något sätt och där man söker något, t.ex. ord och meningar. Stora och i bästa fall systematiskt uppbyggda korpusar kan möjliggöra statistiskt pålitligare resultat än introspektion eller arbiträr samlingsprocess av exempel. Latinets *corpus* (gen. corporis) har som huvudbetydelse 'kropp'. Kroppsmetaforen kan förstås som en hänvisning till någon sammanhängande helhet; jämför den senantika samlingen av romersk rätt, *Corpus iuris civilis*²⁵. Enligt McEnery och Wilson (1996, 2001: 103) ska en textsamling vara statistiskt representativ för att kunna kallas en korpus i egentlig bemärkelse. Många forskare lär ändå använda begreppet bredare för att syfta på någonting som McEnery och Wilson (ibidem) skulle vilja kalla helt enkelt för textsamlingar. Korpuslingvistik är alltid empirisk forskning, men inte all empirisk forskning har med korpusar att göra (ibidem). I sitt glossarium ger McEnery och Wilson (op. cit.: 197) följande tre definitioner till uppslagsordet *corpus*:

- (i) (loosely) any body of text; (ii) (most commonly) a body of machine-readable text; (iii) (more strictly) a finite collection of machine-readable text, sampled to be maximally representative of a language or variety

Korpusar är ett slags material, och korpuslingvistik är en metod som sysslar med stora textmassor, främst just korpusar. Centralt är t.ex. att räkna frekvenser av något ord eller annat fenomen, det vill säga hur ofta det förekommer i autentiska texter, som representeras av korpusmaterialet. Nuförtiden hjälper datorer stort i att söka i korpusarna. Tack vare korpusarna behöver lingvisterna inte endast ty sig till sin introspektion. McEnery och Wilson (1996, 2001: 1) betonar att det i korpuslingvistik är fråga om språkforskning som baserar sig på autentiska exempel av språkbruket i det verkliga livet.

Det har pågått en debatt inom lingvistikens om korpusarnas ställning och värde. Noam Chomsky har varit en prominent kritiker, som efterlyst en mera rationalistisk än empiristisk utgångspunkt till språket. En språkvetare borde enligt honom lita på sin egen introspektion – såvitt som möjligt – hellre än på korpusar, som skulle vara för arbiträrt insamlade och innehålla många skevheter. Ett bekant exempel av honom var att *I live in New York* skulle te sig som mera grammatiskt än *I live in Dayton Ohio*, endast eftersom New York har betydligt flera invånare än Dayton Ohio. Dessutom kan man utan introspektion inte urskilja mellan uttryck som är ogrammatiska och uttryck som vore grammatiska men som korpusen bara inte råkar innehålla. Dessutom innehåller korpusar även

25 'Den medborgerliga lagens kropp'.

ogrammatiska satser, bl.a. genom skrivfel, och enligt Chomsky i stor mån. En korpus kunde aldrig omfatta det hela oändliga språket. För Chomsky var åtminstone på den tiden sin av Ferdinand de Saussure inspirerade dikotomi mellan *competence* och *performance* av stor vikt, och enligt honom kan empirisk forskning som korpuslingvistik endast nå performance-nivån, vilket vore otillräckligt. (McEnery & Wilson 1996, 2001: 5–12.)

Enligt McEnery och Wilson är Chomskys teser överdrivna men innehåller värdefull kritik som har varit en viktig orsak till att korpuslingvistik har förändrats från ungefär 1950-talet och att det nuförtiden fästs så mycket uppmärksamhet vid representativiteten hos korpusarna. Att ty sig enbart till sin introspektion är mindre lämpligt när man forskar på en språkform som man inte är en infödd talare av, särskilt när det är fråga om historiska språkformer eller barnens språkinläring. Småbarn och döda människor saknar sådan metalingvistisk medvetenhet som språkvetenskaplig introspektion förutsätter. Korpusarna behövs alltså. Även frekvenser måste i praktiken räknas med hjälp av korpuslingvistik. Det är mycket lättare att verifiera och återtesta korpuslingvistiska resultat än resultat av rationalistisk introspektion. (McEnery & Wilson 1996, 2001: 13f.)

Tognini-Bonelli (2001: 65ff, 84ff, passim) gör en klar distinktion mellan två olika utgångspunkter för korpuslingvistik: *korpusbaserad* (*corpus-based*) och *korpusdriven* (*corpus-driven*). Korpusbaserad lingvistik testar sina teorier och hypoteser med hjälp av korpusar; teori är alltså primär i förhållande till empiri. Enligt Tognini-Bonelli kan detta leda till problemet att någonting viktigt förbigås obemärkt, då man koncentrerar sig på att söka det man vill och antar sig hitta. Överraskande fynd riskerar att hamna i skymundan. Termen *based* lär också vara så pass vag att den tillåter en forskare att låta bli att respektera den framkomna korpusevidensen. (Tognini-Bonelli 2001: 65f.)

I korpusdriven forskning är det i stället empiriska observationer som föregår och leder till hypoteser, vilka generaliseras, görs enhetliga och slutligen blir till teorier; teori är alltså sekundär till empiri. Utgångspunkten påminner om naturvetenskapernas och torde kunna klassificeras som empiristisk och eventuellt positivistisk. Forskaren ska befria sig från sina föreställningar före korpusen ("pre-corpus beliefs"). (Tognini-Bonelli 2001: 84–86.) Det kan dock vara svårt att göra objektiva tolkningar från korpusmaterialet helt utan att ens förförståelse inverkar på vad man hittar och hur man tolkar.

McEnery och Hardie kritiserar Tognini-Bonelli och andra anhängare av korpusdriven forskning, vilka de kallar ny-Firthska (*Neo-Firthian*), för att de fördunklar gränsen mellan teori och material ("data") då de kräver att korpusar ska ses som teori i sig och den primära källan av all språkvetenskaplig kunskap. Enligt McEnery och Hardie har det förblivit oklart vad de ny-Firthska korpusforskare avser med att korpus är en teori och annan teori tillbakavisas. Om det verkligen avses att korpusen (och inte korpuslingvistik) är en teori som är i sig tillräcklig och ingen (språkvetenskaplig) teori kan existera utan eller utanför korpusevidensen, har vi att göra med begreppsligt nonsens: samma data vore samtidigt det som behöver förklaras och det som har uppgiften att förklara det som ska förklaras. McEnery och Hardie konkluderar att det egentligen torde vara fråga om ett försök till att maximalt understryka korpusevidensens betydelse. I varje fall ter sig den ny-Firthska skolan som en motsatt extremitet till Chomskys position som behandlades ovan. Chomsky kritiserade ju dugligheten av korpusar som källor till språkvetenskaplig kunskap medan de ny-Firthska poängterar användningen av korpusar som källor till språkvetenskaplig kunskap och vill ge dem en ontologiskt överlägsen ställning. (McEnery & Hardie 2012: 6, 147f.)

McEnery och Hardie ser användningen av korpusar som en metod, och inte som en teori. Det utgår jag också ifrån i denna studie. För McEnery och Hardie är uppfattningen om att korpusar har teoretisk status ohållbar, och därför tar de också avstånd från hela dikotomin korpusbaserad–korpusdriven och konstaterar att all korpuslingvistisk forskning kan ses som korpusbaserad. Med tanke på detta är det häpnadsväckande att denna samma dikotomi sägs "löpa genom denna [McEnery & Hardies] bok"²⁶. Det konstateras också att Tognini-Bonelli lanserade termerna *corpus-based* och *corpus-driven* i sin bok (2001), men det verkar finnas en motsägelse: Tognini-Bonelli (2001: 65) skriver att dessa två termer är vedertagna²⁷. (McEnery & Hardie 2012: 5f.)

Ett *token* är ungefär samma som ett ord – ett grafiskt ord –, men det ska iakttas att även punkt, kolon och andra skiljetecken räknas som token. Mellanslaget är däremot en separator och inget token, men <punkt> och <. > i <punkt.> räknas som separata token fast ingen separator föreligger. (Jfr Hammarstedt & al. 2017: 3.)

Termen *kollokation* syftar på vilka ord (eng. *collocates*) som förekommer i samband med ett visst ord (eng. *node*). Termen är central i den ny-Firthska (*Neo-Firthian*) inriktningen av korpuslingvistik, men har också kritiserats kraftigt för att vara omöjlig att definieras exakt. Hur

26 "The difference [...] is a topic that runs through this book."

27 "The terms that are frequently used are corpus-based, as against corpus-driven".

många ord räknas som en kollokation? Rent mekaniskt och linjärt kan man ta upp n närmaste ord på bägge sidorna och på så sätt bilda ett s.k. n -gram. Det är emellertid inte alltid de relevantaste orden som förekommer grafiskt intill noden. Mitt eget exempel på detta är tyskans partikelverb av typen *einschlafen* 'somna in, falla i sömn', som splittras i två delar i många syntaktiska kontexter och det kan finnas i princip hur många ord som helst mellan dessa delar, t.ex. *Der Lehrer schlief heute während der Vorlesung unbemerkt ein.*²⁸ En lösning på att de relevanta kollokationerna är svåra att urskilja kan vara att läsaren utnyttjar sin intuition för att avgöra vilka ord som hör ihop och bildar kollokationer. *KWIC* betyder i sin tur nyckelord i kontext (eng. *keyword in context*). *Konkordans* används ibland som en synonym till detta. Ett nära begrepp till kollokationen är *kolligation* (McEnery & Wilson 1996, 2001: 18), som betyder i vilka grammatiska former och strukturer ordet förekommer mest. (McEnery & Hardie 2012: 122ff, 130f.)

En forskarvänlig korpus ska gärna vara annoterad. Det betyder i sin enklaste form att tokenen i korpusen har försetts med taggar som inte syns i den normala vyn i gränssnittet utan bara på de nedre nivåerna, det vill säga i den taggade versionen av korpustexten som skrivits på ett formellt språk, s.k. märkspråk. Klassiskt används det i korpusannotationer ett märkspråk som heter *XML* (Extended Markup Language),²⁹ eller dess arvingar. I XML kan man t.ex. annotera på följande sätt att två token bildar en nominalfras: `<np>The cat</np>`. (McEnery & Hardie 2012: 29f.) Se mera om detta i kapitlet 2.5.

3.1 Närmare metod

Min forskning baserar sig på maskinella frekvensanalyser i alla fritt tillgängliga korpusar i Korps Moderna-läge och jämförelse av olika korpusar med manuell inblick i alla meningar där *torde* förekommer i några utvalda korpusar. Jag har manuellt analyserat alla meningar i en del utvalda korpusar för att kunna se i vilka slags syntaktiska och semantiska kontexter och texttyper *torde* förekommer. Jag hänvisar också till mera komplexa sökuttryck på CQP-språk.

Då jag gör olika iakttagelser och analyserar dem, blir det en kvalitativ aspekt i min forskning, men till största delen är det fråga om kvantitativ forskning. Å ena sidan kan kvalitativ forskning tänkas vara av nöden för att kunna identifiera kategorierna som ska klassificeras och räknas i kvantitativ

28 "Läraren somnade in under föreläsningen idag och ingen märkte."

29 XML är nära besläktat med HTML (Hypertext Markup Language), som används för att göra och redigera webbsidor och som webbrowsersarna tolkar till normal text.

forskning, men å andra sidan kan det löna sig att betrakta det kvalitativa och det kvantitativa som två skilda synvinklar på språket, vilka ändå kan kombineras och komplettera varandra. Kvalitativ forskning behövs inte endast aprioriskt för att identifiera kategorierna utan även aposterioriskt för att kunna korrigera de fel – eller missledande resultat – som den mekaniska och fasta regler strängt följande kvantitativa arbetsprocessen har åstadkommit. (McEnery & Wilson 1996, 2001: 76.)

Det finns tämligen mycket statistik i min studie. En viktig skillnad är mellan *relativa* och *absoluta frekvenser*. Med absolut frekvens avses i praktiken hur många förekomster som hittats medan relativ frekvens återger den relativa andelen förekomster i förhållande till den helhet vars del förekomsterna är och till vilken förekomsterna jämförs. Relativ frekvens brukar uttryckas i form av träffar per miljon token (parts per million), förkortat *ppm*. Några diagram om resultaten tydliggör framställningen.

Jag anser det rimligt att forska på *torde* ur en korpuslingvistisk infallsvinkel. Det är tvivelaktigt hur pålitliga resultat man finge med hjälp av bara introspektion, särskilt med tanke på att jag inte är infödd talare av svenska. Stora textmassor och opartisk mjukvara möjliggör en klart större reliabilitet. McEnery och Hardie (2012: 28) berättar att korpuslingvistik hade bidragit med så mycken ny kunskap mellan den första utgivningen av Leechs *Meaning and the English Verb* år 1971³⁰ och dess tredje upplaga år 2004 att exempelvis kapitlet om modalverb har måst förnyas grundligt. Det är särskilt intressant med tanke på min studie att McEnery och Hardie nämner just modalverb som exempel.

Det har varit svårt att välja vilka korpusar att använda. Det finns många olika lägen: *Moderna*, *Parallella*, *Fornsvenska* osv. En del av dessa är ägnade åt andra språk som färöiska eller Sibirientyska. Jag har valt att använda det huvudsakliga läget, *Moderna*. Man kan kategorisera de olika korpusarna i större grupper på grundval av genretillhörigheten för att således kunna utreda i vilka genrer *torde* förekommer mest. För att nå en balanserad samling korpusar av olika genrer och därmed valida resultat har jag utvalt sådana korpusar som

1. härstammar från ungefär samma tid och närmare bestämt nära förflutna,³¹
2. är av ungefär samma storlek och
3. inte uppdateras regelbundet
4. innehåller bara en huvudsaklig genre var.

30 Leech, G 1971. *Meaning and the English Verb*. London: Longman. Hänvisning hos McEnery & Hardie.

31 Det som har dokumenterats ligger ju i praktiken i dåtiden medan samtidens räckvidd är vag.

Jag fokuserar på den synkrona situationen. Därför har jag valt att begränsa mig till korpusar vilkas texter härstammar från 1990-talet och framåt. Det är alltså fråga om tämligen moderna texter. En modern situation förmår bättre att svara på mina forskningsfrågor än en föråldrad situation. För att få pålitliga resultat har jag försökt att gallra bort sådana träffar i de utvalda korpusarna som ger anledning till att tro att de faktiskt är mycket äldre än vad texterna i korpusen i princip är. Naturligtvis kan också en del av de meningar i korpusen som inte ingår i träffarna vara av äldre upprinnelse, och därför bör man vara försiktig i att dra slutsatser. Man kan emellertid resonera att det finns en icke alldeles oväsentlig skevhet i resultaten om meningar med *torde* uppvisar äldre ursprung betydligt oftare än vad samma korpus över huvud verkar göra.

Det är viktigt att korpusarna inte uppvisar alldeles för stor variation beträffande storlek. I små korpusar ökar sannolikheterna att någonting inte finns enbart eftersom korpusen är så liten eller att någonting förekommer av slump mycket tätare än vad det skulle göra i en större och balanserad korpus. Det finns alltid en viss grad av skevhetspotential i sådana statistiska beräkningar som även i denna studie utförs. Skevhetspotentialens storlek påverkas av många faktorer, men en faktor är materialets storlek. Att korpusarna är lika stora innebär lättare jämförelser och skevheten kan antas vara på ungefär samma nivå. Att materialet är statistiskt representativt och har valts med eftertanke är centralt i korpuslingvistik (McEnery & Wilson 1996, 2001: 103).

Det tredje kriteriet bestämmer att korpusarna ska vara sådana som har stabilt innehåll i och med att de inte uppdateras eller utvidgas regelbundet. Annars skulle forskningsprocessen kunna bli krångligare när materialet och därifrån stammande resultat skulle förändras mitt i processen. Det vore då även svårare om någon skulle vilja utföra en likadan studie genom att upprepa de sökningar som jag gjort.

Det fjärde kriteriet kan naturligtvis inte följas strängt eftersom t.ex. tidningstexter kan innehålla kolumner, satir, på allvar menade insändare, nyheter, en egen avdelning för barn osv. T.o.m. romaner kan innehålla dikter som epigrafer och även mitt emellan, varpå Miguel Cervantes klassiker *Den snillrike riddaren Don Quijote av La Mancha* bjuder rikligt med exempel. Därför syftar *genre* här på en sådan genre eller texttyp som kan härledas av namnet eller beskrivningen på korpusen. I de flesta fall vore det i praktiken omöjligt att med säkerhet veta till vilken genre en enstaka mening i Korps KWIC-resultat hör. Därför kan den antagna genretillhörigheten i denna studie ibland vara annan än den verkliga genretillhörigheten. Detta är också en orsak till att jag inte

gått in på noggranna genreindelningar utan har nöjt mig med att operera med en tämligen enkel genreindelning.

Efter att ha definierat genrekategorierna har jag utfört en sökning med lemgrammet *torde* i de sagda korpusarna. Sedan har jag gått igenom alla KWIC-resultatmeningar manuellt och gallrat bort de förekomster som jag konstaterat vara oäkta. Det finns nämligen en hel del förekomster där skribenten uppenbarligen menat något annat än modal verbet *torde* eller *tör* (se exempel (28), (29) och (30) vidare nedan). Jag har också gallrat bort förekomster som är på andra språk än svenska även om de etymologiskt skulle motsvara svenskans *torde*, eller som är för gamla, såvitt jag identifierat dem.

Jag har också räknat dubbletterna. Med dubblett anses i princip ett fall där samma mening förekommer flera gånger i träffarna eller materialet. Även denna räkningsprocess har skett manuellt så det är möjligt att det fanns flera dubbletter än jag kunnat hitta. Det är emellertid vanskligt att dra slutsatser av dubbletterna, då det även kan och *torde* finnas dubbletter också i de meningar av samma korpus som inte ingår i träffarna. Det står inte heller klart hur en dubblett ska definieras exakt. Det kan nämligen hända att två satser, som uppvisar alldeles identisk form, härstammar från olika skribenter och har helt annat ursprung eller härstammar från samma skribent men hör till separata texter. Är det då fråga om samma meningar? Alander har övervägt detta i sin pro gradu-avhandling (2019: 37). Över huvud taget finns det nödvändigtvis en mänsklig misstagspotential i manuellt arbete. Detta gäller att noteras.

När jag analyserar resultaten från samtliga fritt tillgängliga korpusar som helhet, finns det också många äldre texter med i en del av korpusarna. Denna studie fokuserar huvudsakligen inte på diakrona synpunkter utan försöker att ge en noggrann blick på den synkrona situationen, men eftersom en stor del av korpusarna innehåller material som är för gammalt för mitt syfte har jag måst på något sätt iaktta detta och kartlägga även den diakrona situationen i någon mån. Jag har gjort några diagram som åskådliggör den utveckling som skett i frekvensen av *torde* i jämförelse med alternativa uttryckssätt.

Vad gäller den andra forskningsfrågan om användningen av *ha*-bortfallet i olika kontexter, har jag sökt med lämpliga CQP-uttryck förekomster av lemgrammet *torde* (*verb*) följt av lemgrammet *ha* (*verb*) eller ej, följt av ett supinum. Jag har utfört dessa sökningar i alla samma korpusar som jag gick noggrant manuellt igenom i samband med den första forskningsfrågan om användningen av

torde i olika kontexter. Sedan har jag gått igenom varje träff, mening för mening. Jag har därmed kunnat gallra bort oäkta fall och notera dubletter. Sedan har jag identifierat en korpus som har flera förekomster av *ha*-bortfallet än de andra och analyserat denna noggrannare. Därefter har jag också gjort några grövre sökningar i alla de fritt tillgängliga korpusarna i Korps Moderna-läge. Då har jag dock inte kunnat på grund av materialets massiva natur gå igenom resultaten manuellt och gallra bort sådana felaktiga träffar som säkert därmed ingår i resultaten.

3.2 Allmänt om materialet

Mitt material består av olika korpusar av nusvenska – samt vid behov korpusar med äldre texter – som finns tillgängliga i Språkbankens korpusverktyg Korp.³² Språkbankens korpussamling består av texter som till största delen är annoterade. En del är skyddade och därmed i praktiken otillgängliga, men detta torde inte ha stor betydelse för min forskning; skyddade korpusar behöver helt enkelt inte användas. Språkbanken grundades år 1975 och tillhör Göteborgs universitets humanistiska fakultet. Inrättningen ålades att samla, utveckla och tillgängliggöra språkligt material. Språkbanken säger sig användas ”flitigt av forskare i och utanför Sverige, både i empirisk forskning och i undervisning i svenska, nordiska språk och kontrastiv eller typologisk lingvistik”. (Borin.) Språkbankens korpusar används oftast med hjälp av Korp, som är ett gränssnitt för korpussökningar. Språkbanken erbjuder också andra verktyg som Karp och Sparv. Jag använder mig i min analys dels av resultaten från alla fritt tillgängliga korpusar i Korps Moderna-läge, dels av resultaten från sju utvalda korpusar. De senare har jag gått manuellt igenom och behandlar dem noggrannare i analysen.

Det finns tre huvudsakliga sätt att söka i materialet i Korp: Enkel sökning, Utökad sökning och Avancerad sökning. I enkel sökning kan man söka en ordform i de korpusar som man vill eller man kan söka ett lemgram i stället för ett ord. Då är alla böjningsformerna av ordet inkluderade. Lemgram kan vara polysemiska men homonymi noteras såtillvida att homonymiska ord som tillhör olika ordklasser behandlas som separata lemgram, t.ex. *fast* (adjektiv), *fast* (adverb) och *fast* (konjunktion) är olika lemgram. Lemgram är också skiftlägesoberoende, det vill säga versaler och gemener räknas som sak samma (Hammarstedt & al. 2017: 3). I statistikfliken kan man visserligen delvis se analys av frekvenser baserade på olika stavningar, t.ex. *torde* vs. *Torde*. Lemgram i Språkbankens korpusar har annoterats maskinellt, vilket innebär att annoteringar kan vara felaktiga (Hjortstam 2018: 22).

32 Korp finns tillgänglig i: <<https://spraakbanken.gu.se/korp>>.

I utökad sökning kan man t.ex. definiera vilka ord eller snarare token som följer varandra. Det går att tillåta alternativa token genom knappen ”Eller”. Det går även att tillåta ”Vilket ord som helst” mellan andra ord och tillåta detta ord, vad det än må vara, upprepas n gånger mellan andra ord. Man kan ange flera krav för samma token: man kan t.ex. kräva att tokenet är ett visst grafiskt ord³³ och (”och”-knappen) samtidigt är en form av ett visst lemgram eller en viss böjningsform³⁴. Det går även att använda reguljära uttryck för noggrannare och komplexare sökningar (Se kapitlet 2.4). I avancerad sökning kan man göra noggrannare sökningar med CQP-språk, som är ett formellt (icke-mänskligt) språk liknande programmeringsspråk. (Se kapitel 2.5.)

3.3 De sju utvalda korpusarna

I det följande rapporterar jag om sju korpusar som jag gått igenom mening för mening. I alla korpusar har jag sökt med lemgrammet *torde* (*verb*), vilket i CQP-form blir `[lex contains "torde\\.\\.vb\\.1"]`. I analysdelen presenterar jag resultaten av denna sökning, men här nedan strävar jag efter att ge en allmän inblick på dessa korpusar.

De korpusar som jag har valt är:

- i. Poeter.se
- ii. Skönlitteratur 2000–2013
- iii. Samhällsvetenskap
- iv. av Läkartidningen den årgång från åren 2002-2006 vars relativa *torde*-frekvens är median i frekvensordning (vilket är Läkartidningen 2003)³⁵
- v. av GP (Göteborgsposten) den årgång från åren 2002-2013 vars relativa *torde*-frekvens är median i frekvensordning (egentligen finns det två medianer eftersom antalet delkorpusar är i detta fall ett jämnt tal; av medianerna har jag valt den nyare, det vill säga GP 2012)
- vi. 8 sidor (en lättläst tidning)
- vii. Barnlitteratur 1988–2013 (i praktiken innehållande främst böcker från 2000-talet)

33 Dvs. en sträng utan alternativ, vilka man kunde specificera bl.a. genom reguljära uttryck.

34 Detta händer med optionen ”msd”. En lista av användbara morfosyntaktiska förkortningar finns i <https://spraakbanken.gu.se/korp/markup/msdtags.html>. Genom att trycka på ”i”-knappen får man en enkel form av samma lista, där man kan välja grafiskt en lämplig morfosyntaktisk beskrivning, om den finns med på listan.

35 Förfarandet har valts eftersom materialet i Läkartidningen-korpusarna är så massivt. Detsamma gäller även GP-korpusarna.

Korpusen *Poeter.se* innehåller material från webbsidan <poeter.se>, där registrerade användare kan publicera dikter och andra texter som de har skrivit och har så att säga i sin bordslåda. Det går också att kommentera andras alster, antingen direkt under texten eller – som det rekommenderas – på forumet. Icke-registrerade och icke-inloggade användare kan läsa texter och de kommentarer som skrivits direkt nedanför texterna, men de kan inte skriva på webbsidorna eller läsa forumet. Projektet började år 2003 och har varit menat för främst poesi, men numera finns det en uppsjö av olika slags prosatexter på sidorna. Varje text ska märkas som tillhörande någon av fyra huvudkategorierna: *Bunden vers*, *Fri vers*, *Prosa* och *Andra genrer*. (Poeter.se 2013, Poeter.se utan datum: passim på sidorna.)

Språkbankens korpus med namnet *Poeter.se* har senast uppdaterats den 16.10.2019³⁶. Enligt diagrammet i korpusväljaren hos Korp stammar texterna från åren 2003–2017. Korpusen erbjuder möjligheten att visa utökad kontext via en länk till den exakta sidan hos domänen poeter.se där texten har publicerats. I några få fall är denna sida inte längre tillgänglig, eventuellt på grund av att den användare som skrivit texten har raderat den eller avslutat sitt konto. Korpusen innehåller 106 196 502 token.

Skönlitteratur 2000–2013 är en delkorpus av korpusen Skön- och facklitteratur, som i sin tur är en delkorpus av korpusen Finlandssvenska texter. Det är alltså fråga om finlandssvensk skönlitteratur, men med finns verk som biografier och andra genrer som kommer nära facklitteratur. Detta gäller att noteras i analysen. Korpusen *Skönlitteratur 2000–2013* omfattar 2 501 260 token. Enligt diagrammet i korpusväljaren hos Korp stammar texterna från åren 2002–2013; därtill är en del av texterna odaterade. Det odaterade materialet omfattar 8 478 token. Korpusen stödjer inte utökad kontextvisning.

Korpusen *Samhällsvetenskap* är en delkorpus av korpusen *Akademiska texter*. Denna har i sin tur två delkorpor: den ena är just *Samhällsvetenskap* och den andra heter *Humaniora*. Bägge är ungefär lika stora och har en liknande frekvens av lemmogrammet *torde* (*verb*). Korpusen *Samhällsvetenskap* innehåller 10 855 954 token. Texterna är daterade 1997–2012. Det ingår inget odaterat material. Texterna innehåller dock icke-obetydliga mängder texter, antagligen citat, som torde vara äldre, vilket är snabbt att ställa fast på grund av en föråldrad ortografi. För att få en ungefärlig bild på fenomenet har jag sökt med det reguljära uttrycket *. *fv.** i korpusen. På det sättet har jag kartlagt att bokstavskombinationen <fv>, som är typisk för äldre ortografi,

36 Informationen hittas genom att hålla muspekaren ovanpå korpusens namn.

förekommer 501 gånger i korpusen. En del är dock meningar eller fraser på danska och en stor del namn som naturligtvis förekommer med <fv> även i moderna texter. Detta gäller att iakttas i analysen. Korpusen stödjer inte utökad kontextvisning.

Korpusen innehåller textattributet ”typ”, eller `_.text_type` i CQP-frågor, som har två möjliga värden: "PhD" och "Licentiat". I praktiken betyder detta att alla texter är annoterade antingen som doktorsavhandlingar eller som licentiatavhandlingar. För att få en djupare uppfattning av korpusens innehåll, som är torftigt dokumenterat hos Språkbanken, har jag utfört några sökningar med textattribut och skiljetecken.

I principen går det att söka helt enkelt med enbart textattribut, som t.ex. `[_.text_type = "Licentiat" & _.text_subject != "Medie- och kommunikationsvetenskap" & _.text_subject != "Social och ekonomisk geografi"]`. Då uppstår dock problemet att varje ord och nästan varje skiljetecken blir ett eget resultat och KWIC-resultatfliken blir enorm och oöverskådlig. Därför har jag använt mig av ett reguljärt uttryck `[\.\?!]` för att få fram alla eller åtminstone nästan alla meningar. Meningar är ju i varje fall grundenheten i träffarna hos Korp. Meningar i Korp slutar nästan alltid på punkt, utropstecken eller frågetecken. Så är i det nämnda reguljära uttrycket någon av dessa tre alternativ utsatta inom hakparenteser, `<.>`, `<?>` eller `<!>`, det token som söks. De omvända snedstrecken fungerar som avbrottstecken för att berätta för maskinen att punkten och frågetecknet avses här i sina vanliga betydelser och inte som specialtecken vilka de vanligtvis är i reguljära uttrycks syntax. Eftersom många meningar kan i sig innehålla dessa skiljetecken, främst punkter, utanför meningsslutet, har jag förutsatt i mina CQP-frågor att det ovan beskrivna tokenet inte får föregås av ett token som innehåller ett nummer. Detta har jag uttryckt med ett reguljärt uttryck `.*[0-9].*` vilket jag sedan förbjudit med den negerande `!=`-operatoren: `[word != ".*[0-9].*"]`. Så har jag kunnat gallra bort de fall där ett nummer följs av punkt mitt i en mening, vilket händer ofta i denna korpus. Jag har inte kunnat gallra bort de fall där t.ex. en förkortning följs av en punkt, men sådant vore onödigt i och med att det inte har betydelse i det hela, då jag bara har gjort dessa sökningar för att få en grov allmän bild av korpusens innehåll.

När jag sökte med `[word != ".*[0-9].*"] [word = "[\.\?!]" & _.text_type = "Licentiat"]` fick jag 45 153 resultat. Jag har också utrett vilka ämnen (textattributet `_.text_subject`) licentiatavhandlingarna har. Detta kan i sig vara av föga betydelse, men en

iakttagelse kan vara befogat att rapportera här. Största delen av licentiatavhandlingarna har ”Social och ekonomisk geografi” som s.k. ämne. Ämnet ”Medie- och kommunikationsvetenskap” förekommer ungefär tio gånger mera sällan. Det tredje ämnet bland licentiatavhandlingarna, ”Psykologi” gav bara 13 träffar. Det kan inte vara en hel licentiatavhandling med bara ungefär tretton meningar. Jag gick manuellt igenom meningarna och sökte i Google och upptäckte att det är huvudsakligen fråga om sammanfattningen i en Lindgrens licentiatuppsats (Lindgren 2011), som finns fritt tillgänglig i nätet i sin helhet. Det förefaller märkligt att uppsatsen inte inkluderats i korpusen helt och hållet.³⁷

Texter som har typen ”PhD” finns mycket flera än licentiatavhandlingar. Sökningen [word != ".*[0-9].*"][word = "[\.\?!]" & _.text_type = "PhD"] gav 442 173 träffar. Det kan alltså estimeras att korpusen *Samhällsvetenskap* innehåller ungefär 10 gånger flera doktorsavhandlingar än licentiatavhandlingar. Ämnen som ingår i doktorsavhandlingarna i korpusen är utöver desamma som förekommer i licentiatavhandlingar även åtminstone ”Sociologi”, ”Statsvetenskap”, ”Juridik”, ”Utbildningsvetenskap” samt ”Ekonomi och näringsliv”.

Korpusen *Läkartidningen 2003* är en delkorpus av korpusen *Läkartidningen*, som i sin tur är en delkorpus av korpusen *Medicinska texter*. *Läkartidningen* är en tidskrift som utges såväl i pappersform som på nätet och som innehåller bl.a. referentgranskade vetenskapliga artiklar (*Läkartidningen*). Bakgrundsorganisationen för tidskriften är Sveriges läkarförbund (ibidem). Korpusen omfattar 1 748 780 token enligt faktarutan som syns när man rör på muspekaren över korpusens namn i korpusväljaren. Detta är lite motstridigt i och med att faktarutan som syns när man rör på muspekaren över diagrammet i korpusväljaren ger tokenantalet för ”Material daterat 2003” som 1 749 785 i skrivande stund, alltså 1 005 gånger flera. Korpusen stödjer inte utökad kontextvisning, men genom att klicka på träffen får man i högra marginalen se bl.a. namnet på artikeln ur vilken meningen i fråga är tagen, och ibland kan man hitta artikeln gratis med Google.

37 Sökningen [word != ".*[0-9].*"][word = "[\.\?!]" & _.text_type = "Licentiat" & _.text_subject = "Social och ekonomisk geografi"] gav 41 426 träffar. Sökningen [word != ".*[0-9].*"][word = "[\.\?!]" & _.text_type = "Licentiat" & _.text_subject = "Medie- och kommunikationsvetenskap"] gav 3 714 träffar. Sökningen [word != ".*[0-9].*"][word = "[\.\?!]" & _.text_type = "Licentiat" & _.text_subject = "Psykologi"] gav 13 träffar, vilket är litet. Sökningen [word != ".*[0-9].*"][word = "[\.\?!]" & _.text_type = "Licentiat" & _.text_subject != "Psykologi" & _.text_subject != "Medie- och kommunikationsvetenskap" & _.text_subject != "Social och ekonomisk geografi"] gav inga träffar, dvs. inga andra ämnen förekom bland licentiatavhandlingar, om inte någon text saknar annotering, vilket vore teoretiskt möjligt.

Korpusen *GP 2012* är en delkorpus av korpusen *GP*, som i sin tur är en delkorpus av korpusen *Tidningstexter*. Det kan vara bra att notera att *Tidningstexter* är långt ifrån enda korpus i Korps *Moderna*-läge vilken ägnats åt tidningstexter. Det finns ändå en betydande samling texter i GP-korpusarna från åren 2001–2013 samt 1994. Jag har valt att noggrannare gå igenom årgången 2012, vilken omfattar 17 162 039 token. Det är fråga om tidningen *Göteborgs-Posten*, som utges i Göteborg med omnejd utan avbrott sedan 1859, men beroende på hur man räknar kan anses ha grundats redan 1813 (GP). Tidningen säger sig vara oberoende liberal, som en textruta i ledarartiklarna i nätversionen deklarerar (13.2.2021).

GP 2012 uppvisar samma motstridighet som i *Läkartidningen 2003*: faktarutan som syns när man rör på muspekaren över diagrammet i korpusväljaren ger en annan siffra för tokenantalet, i detta fall 17 192 005 token. Korpusen stödjer i princip utökad kontextvisning, men i praktiken ger detta föga nytta, då den visade kontexten är så knapp. Textattributen *artikelförfattare* och *artikelavdelning* syns i marginalen då man klickar på träffen.

Korpusen *8 sidor* är en delkorpus av korpusen *Tidningstexter*. *8 sidor* är en lättläst tidning som finansieras åtminstone delvis av staten och som utges av Myndigheten för tillgängliga medier, MTM (8 sidor 2016a). Tidningen är särskilt avsedd för ”elever i grundskolan, personer med utvecklingsstörning, invandrare och många andra människor som har svårt att läsa en vanlig dagstidning” (ibidem). Tidningen har utgetts i flera årtionden (ibidem). Innehållsligt fokuserar den på nyheter och stilistiskt säger sig tidningen inte ha några ”svåra och ovanliga ord” (8 sidor 2016b).

Korpusen *8 sidor* omfattar 2 832 875 token. Texterna stammar enligt diagrammet i korpusväljaren från åren 2003–2012. Korpusen stödjer inte utökad kontext.

Korpusen *Barnlitteratur 1988–2013* är en delkorpus av korpusen *Skön- och facklitteratur*, som i sin tur är en delkorpus av korpusen *Finlandssvenska texter*. Dateringen förknippas med motstridigheter: faktarutan i korpusväljaren beskriver att korpusen består av texter som publicerats 2000–2013. Enligt Språkbankens webbsida som listar alla de tillgängliga resurserna³⁸ består korpusen av barnböcker skrivna av finlandssvenska författare och utgivna åren 1988, 2002, 2005, 2006, 2008, 2009, 2010, 2011 och 2012. Jag gjorde några sökningar med textattributet *år* och kom fram till att den sistnämnda informationen torde stämma. Korpusen innehåller 191 406 token. Korpusen stödjer i princip utökad kontextvisning, men den visade kontexten är väldigt knapp.

38 <<https://spraakbanken.gu.se/resurser/barnlitteratur>>.

Jag har försökt att välja en mångsidig samling av samtida och ungefär lika stora korpusar, som tydligt representerar olika genrer. Tyvärr är en av dem, *Barnlitteratur 1988–2013*, betydligt mindre än de andra, men jag anser ändå att den som helt egen och klart separat genre kompletterar min studie värdefullt. Materialet i de valda korpusarna stammar huvudsakligen från 2000-talet, med somliga innehållande delvis materialet från slutet av 1900-talet. Korpusen *Poeter.se* omfattar material till och med år 2017 åtminstone, men samtliga andra består av material som stammar från senast år 2013.

Korpusarna kan kategoriseras på flera olika sätt, vilket illustreras i Tabell 3. Det finns tre korpusar som kan klassificeras som skönlitteratur: *Barnlitteratur 1988–2013*, *Skönlitteratur 2000–2013* och *Poeter.se*. Härav består de två förstnämnda huvudsakligen av prosa medan den sistnämnda består av lyrik med smärre mängder prosa mitt emellan. De två korpusarna med prosaistisk skönlitteratur är också de enda korpusarna som handlar om finlandssvenska. Det finns tre korpusar som består av tidningstexter: *8 sidor*, *GP 2012* och *Läkartidningen 2003*. Därifrån representerar den sistnämnda en vetenskaplig stil och innehåll, GP är en vanlig tidning, och 8 sidor är en lättläst tidning. Det finns två korpusar som kan klassificeras som vetenskapliga texter: *Läkartidningen 2003* och *Samhällsvetenskap*. Det finns två korpusar som uppvisar en enkel stil som är avsedd för speciella målgrupper: *8 sidor* och *Barnlitteratur 1988–2013*. De sju valda korpusarna representerar således sju genrer: **lyrik** (*Poeter.se*), **vuxenskönlitteratur** (*Skönlitteratur 2000–2013*), **vetenskapliga avhandlingar** (*Samhällsvetenskap*), **vetenskaplig tidning** (*Läkartidningen 2003*), **vanlig tidning** (*GP 2012*), **lättläst tidning** (*8 sidor*) och **barnlitteratur** (*Barnlitteratur 1988–2013*).

Tabell 3. De sju närmare genomgångna korpusarnas indelning och egenskaper

Korpusen ³⁹	Genreegenskaper		Regional härkomst	Storlek (n token)
<i>Samhällsvetenskap</i>	Vetenskap			10 855 954
<i>Läkartidningen 2003</i>		Tidningar		1 748 780?
<i>GP 2012</i>				17 162 039?
<i>8 sidor</i>	Enkel stil			2 832 875
<i>Barnlitteratur</i>		Skönlitteratur	Prosa	Finlandssvenska
<i>Skönlitteratur</i>			2 501 260	
<i>Poeter.se</i>			Lyrik huvudsakligen	

39 Jag har förkortat en del av namnen på korpusarna av utrymmesskäl.

4 Torde i korpusarna

I analysen i detta kapitel presenterar jag olika slags resultat, iakttagelser och synpunkter som framgått när jag gått igenom korpusar i Korp. Jag använder kursiv för att uttrycka mina sökuttryck som inte är i CQP-form. CQP-fraserna återger jag däremot i icke-proportionellt (monospace) typsnitt, som det är vedertaget att göra med uttryck som är på formella språk.

Korps resultatmeningar (KWIC) återger jag som de är och med vanligt typsnitt. Detta innebär att det finns flera märkligheter. T.ex. är skiljetecken oftast omgivna av mellanslag. Detta beror på att meningarna har tokeniserats, det vill säga gjorts läsbara för datorprogrammet som annoterat korpusmaterial (Hjortstam 2018: 17). Jag har inte använt anföringstecken kring exempelmeningarna, ty de innehåller ofta redan i sig anföringstecken, och det kunde åsamka förvirring. När jag återger relativa frekvenser, använder jag för korthetens skull uttrycket ppm (parts per million) för att syfta på andelen token per miljon token.

4.1 Torde i samtliga korpusarna

Jag har sökt i alla de korpusar som finns i läget *Moderna* – utom dem som är skyddade – och utfört en enkel sökning med lemgrammet *torde* (*verb*). Som CQP-fras blir det `[lex contains "torde\.\.vb\.1"]`. På det sättet fick jag också förekomster av böjningsformen *tör* med i resultaten. Mest frekvent förekom *torde* vid sökningstillfället i korpusen *Statens offentliga utredningar*, som innehåller 428 188 012 token. Detta gällde både lemgrammet *torde* (*verb*) och böjningsformen *torde* (urskild från böjningsformen *tör*). Det kan te sig lite förvirrande att det finns två korpusar i Korp som heter *Statens offentliga utredningar*: den ena är den egentliga *Statens offentliga utredningar* medan den andra är en delkorpus av korpusen *Riksdagens öppna data*. Den senare är ungefär hälften mindre. Den mest *torde*-frekventa korpusen var alltså den större och egentliga *Statens offentliga utredningar*, medan den mindre korpus med samma namn kom på plats 7. I den egentliga *Statens offentliga utredningar* förekom böjningsformen *torde* 438 086 gånger, vilket betyder 1 023,1 gånger per miljon token. Därav var 71 förekomster (0,2 ppm) av den grafiska formen <Torde>. Alla andra förekomster var av formen <torde>. Presensformen *tör* förekom också i den egentliga *Statens offentliga utredningar*. Det absoluta antalet var 451 träffar (1,05 ppm), varav 19 var förekomster av <Tör> och 432 av <tör>.

Flest förekomster av böjningsformen *tör* uppvisade korpusen *Äldre svenska romaner*, men om vi bortser från den och försöker att få en bild av nusvenskans situation, hade korpusen *Etnologiska frågelistor* den största andelen (5,7 ppm, absolut 328) och *Statens offentliga utredningar* det största absoluta antalet (451 förekomster, relativt 1,05 ppm). Andelen i samtliga tillgängliga korpusar var 0,2 ppm. Det gäller dock att notera att en stor del av resultaten med *tör* var felträffar. Det förekommer många fall där *tör* står där det borde stå t.ex. *stör*, *gör* eller *för*. Ett resultat ur korpusen *Flashback: Vetenskap & humaniora* till och med ger en hänvisning åt detta håll, se ex. (28)

(28) Hon brukade t.ex. säga " du tör " i stället för " du stör " . (*Flashback: Vetenskap & humaniora.*)

(29) Kolla madrassen om du tör ! (*Flashback: Hem, bostad & familj.*)

(30) Men de grävsta " påståendena " torde ha dokumenterade grundvalar . (*Åbo underrättelser 2012.*)

En mindre självklar kategori av felträffar är de talrika fallen där det förefaller att *tör* eller *tör sig* använts i stället för det korrekta *törs* som i exempel (29). Särskilt med tanke på att *töras* hör etymologiskt ihop med *tör* : *torde* kunde man resonera att det är fråga om förekomster av *tör* men i en lite avvikande fastän historiskt förståelig betydelse. Jag har ändå klassificerat dessa fall som felträffar. *Torde* kan i sin tur förekomma, där t.ex. *tordes* har egentligen menats, eller *borde*, som kanske är fallet i exempel (30).

Statens offentliga utredningar är betänkanden som vissa av staten utnämnda experter framställer i förberedningsprocessen av nya lagar. Betänkandena publiceras i en serie som heter just Statens offentliga utredningar. (Regeringskansliet 2019.) Annan information om korpusen är vanskelig att finna. När man söker i enbart en korpus och gör ett trenddiagram, får man se, hur många förekomster som finns för respektive år. Detta gäller ändå såtillvida korpusen innehåller data om textens ålder. Om *Statens offentliga utredningar* ser man att tidsintervallen sträcker sig fr.o.m. januari 1922 t.o.m. december 1996. Månaderna verkar visserligen inte separerade utan månaderna inom samma år förefaller ha exakt samma frekvens. *Torde* förekommer ända till tidsintervallens slut.

Jag har även gjort beräkningar av *torde* i förhållande till motsvarande, alternativa uttrycksätt. Detta är viktigt för att kunna urskilja i vilken mån frekvensen av *torde* beror på om det bakomliggande semantiska innehållet över huvud taget förekommer ofta eller nästan inte alls i materialet. Om *torde* inte förekommer i en korpus, har det inget bevisvärde för verbets frekvens om motsvarande

uttryckssätt inte heller förekommer i korpusen i fråga. Sammalunda ter sig en hög andel av verbet mindre betydelsefull om man märker att andra, motsvarande uttryckssätt är mycket vanligare i samma korpus. Jag har identifierat som alternativa uttryckssätt: *sannolikt*, *troligen*, *troligtvis*, *förmodligen* och *antagligen*. Alla är adverb, till skillnad från *torde*, som är ett verb, men detta har inte betydelse i denna frågeställning.

Jag sökte med frasen [(word = "[Tt]orde" | word = "[Ff]örmodligen" | word = "[Aa]ntagligen" | word = "[Tt]roligen" | word = "[Tt]roligtvis" | word = "[Ss]annolikt")] i samma korpusar som jag gått igenom för *torde* (verb)-lemgrammets del. Siffrorna i Diagram 1 baserar sig på data hämtade den 12.3.2021. Däremot baserar sig siffrorna i diagrammen 2 och 3 på data hämtade den 19.3.2021. Däremellan har Korp och dess korpusar förändrats lite grann, och t.ex. antalet fritt tillgängliga korpusar är i skrivande stund bara 226 medan det förut var 227. Sökfrasen var emellertid densamma. För att göra sökningen snabbare har jag alltså sökt för *torde*-verbets del bara de grafiska formerna <torde> och <Torde>, vilket gör att t.ex. den arkaiska presensformen *tör* faller bort från resultaten, till skillnad från de mera ovan rapporterade sökningar med lemmat *torde* (verb), vilka också omfattade bland annat *tör*.

Det vanligaste alternativet var adverbet <förmodligen>, alltså med gement f. Dess andel var 21 % av alternativen. På andra plats kom <torde> (20 %) med skillnad på bara 1 procentenhet. Därefter kom <antagligen> (17 %) och <troligen> (12 %). Beaktansvärt är att <sannolikt> representerades av endast 10 % av fallen medan det kommer fram nedan att detta adverb i vissa närmare genomgångna korpusar var det vanligaste alternativet. Den grafiska formen <Torde> med versalt t var det allra raraste alternativet. Detta är inte förvånande i och med att *torde* är ett verb och kunde således placeras i början på meningen främst i en frågesats, men sådana här epistemiska strukturer inte typiskt förekommer i frågesatser av semantiska skäl. Detta gäller däremot inte adverb. I alla fall var andelen av alternativ skrivna med versal i början på ordet enbart 12 %. För noggrannare siffror och illustration hänvisas till Diagram 1. Det kan ännu noteras att <torde> var vanligast i jämförelse med alla de alternativa uttryckssätten i korpusen *Statens offentliga utredningar* med <sannolikt> som näst vanligast.

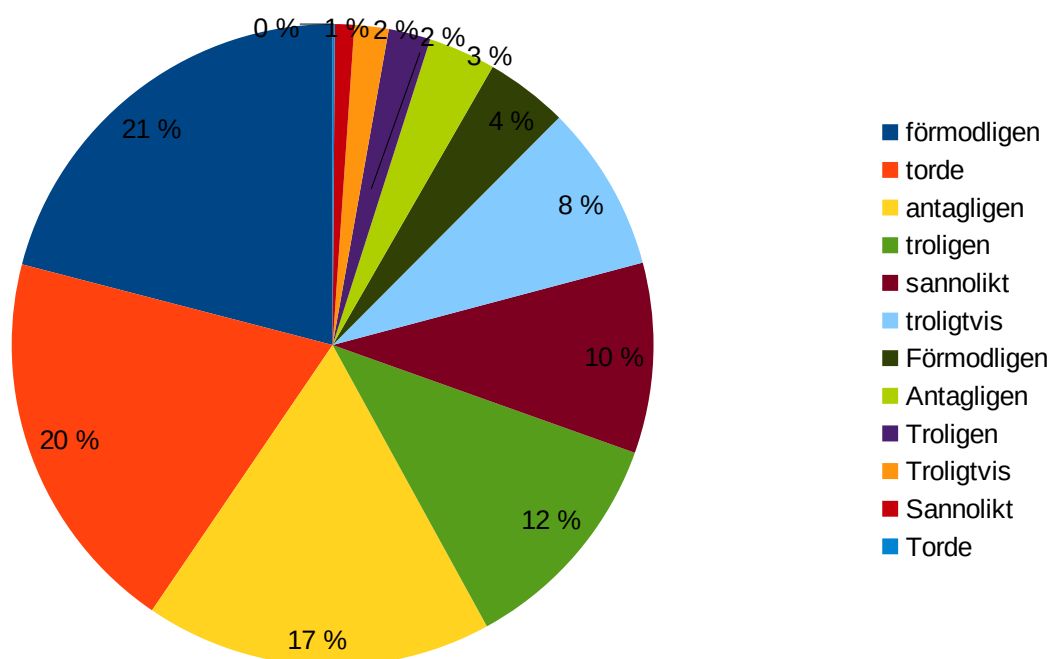


Diagram 1. Andelarna av *torde* och alternativa uttryckssätt i alla fritt tillgängliga korpusar (n %) i förhållande till varandra

Ur denna synvinkel föreligger det alltså inga grunder till att stämpla användningen av *torde* som något icke-naturligt, vilket grammatiken *Lukion ruotsin kielioppi* gör (se kapitel 1). Visserligen kan det skönjas ur kronologiskt perspektiv en sjunkande tendens i frekvensen av verbet. Jag har försökt att utreda om det verkligen förhåller sig så att *torde* var frekventare förr eller om de alternativa uttryckssätten också har blivit mera sällsynta genom åren. Diagram 2 illustrerar hur såväl *torde* som de alternativa uttryckssätten över huvud taget har minskat i andel i Korps korpusar genom tiderna. Detta kan vara orsakat av att innehållet i de nyare korpusarna har blivit annorlunda eller mångsidigare och semantiska grunder har gjort att sådana här epistemiska uttryck inte förekommer så tätt som i de äldre korpusarna. Ett undantag är åren 2015, 2016 och 2017, som bildar en enstaka, märklig, statistisk topp. Dessutom är frekvenserna främst små före 1920-talet, men materialet från den tiden är så föga att den inte kan befinnas statistiskt representativt.

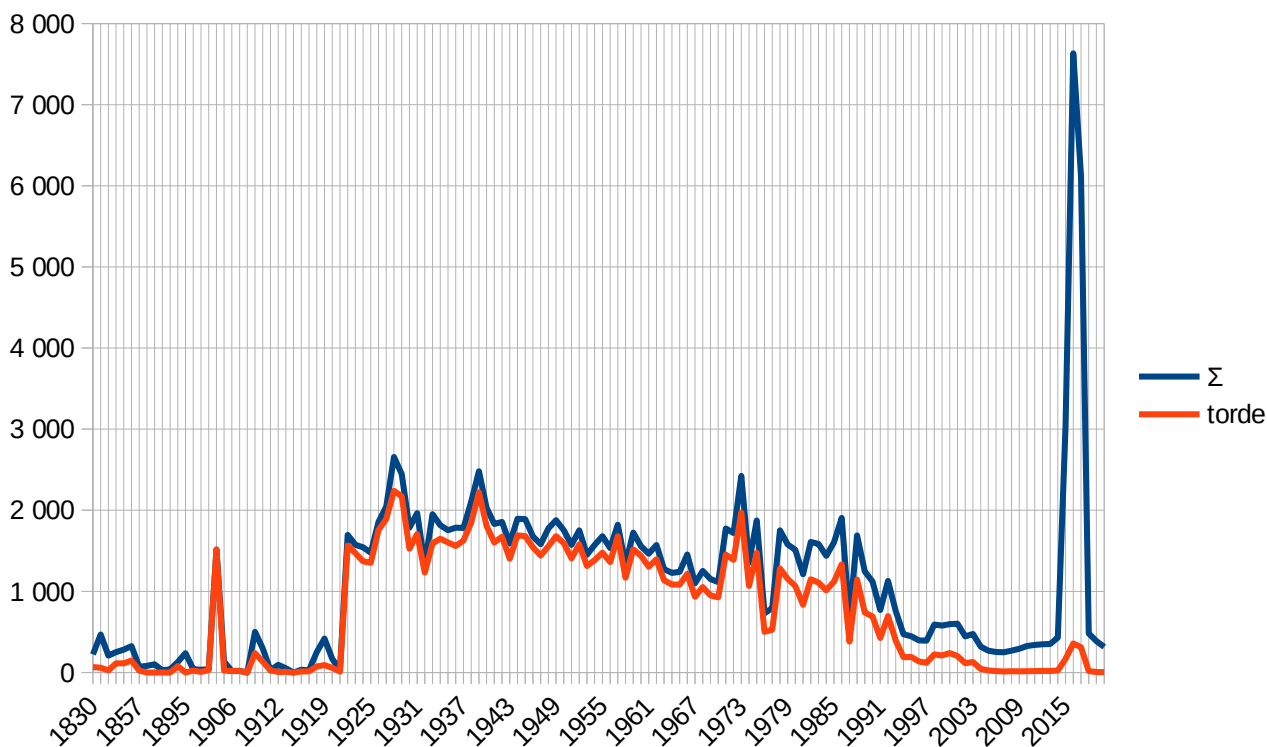


Diagram 2. Andelarna av *torde* samt motsvarande uttryckssätt (*n* ppm) per år

Det gäller dock att notera hur *torde* även sjunkit i förhållande till de alternativa uttryckssätten. Diagram 3 åskådliggör detta. Sjunkningen av *torde* har till största delen skett mellan åren 1978 och 2006, närmare sagt från 73 % år 1978 till 6 % år 2006, det vill säga 67 procentenheter. Därefter har situationen varit stabilare. Den ovan nämnda statistiska toppen under åren 2015, 2016 och 2017 som syntes i Diagram 2 syns inte längre i Diagram 3. De talspråkligare korpusarna stammar huvudsakligen från senare tider, vilket naturligtvis förvränger resultatet. I alla fall har sjunkningen stannat på grundval av materialet och frekvensen av verbet ser stabil ut. Före 1920-talet uppvisar materialet påfallande variation, antagligen på grund av att materialinsamlingen har skett så osystematiskt.

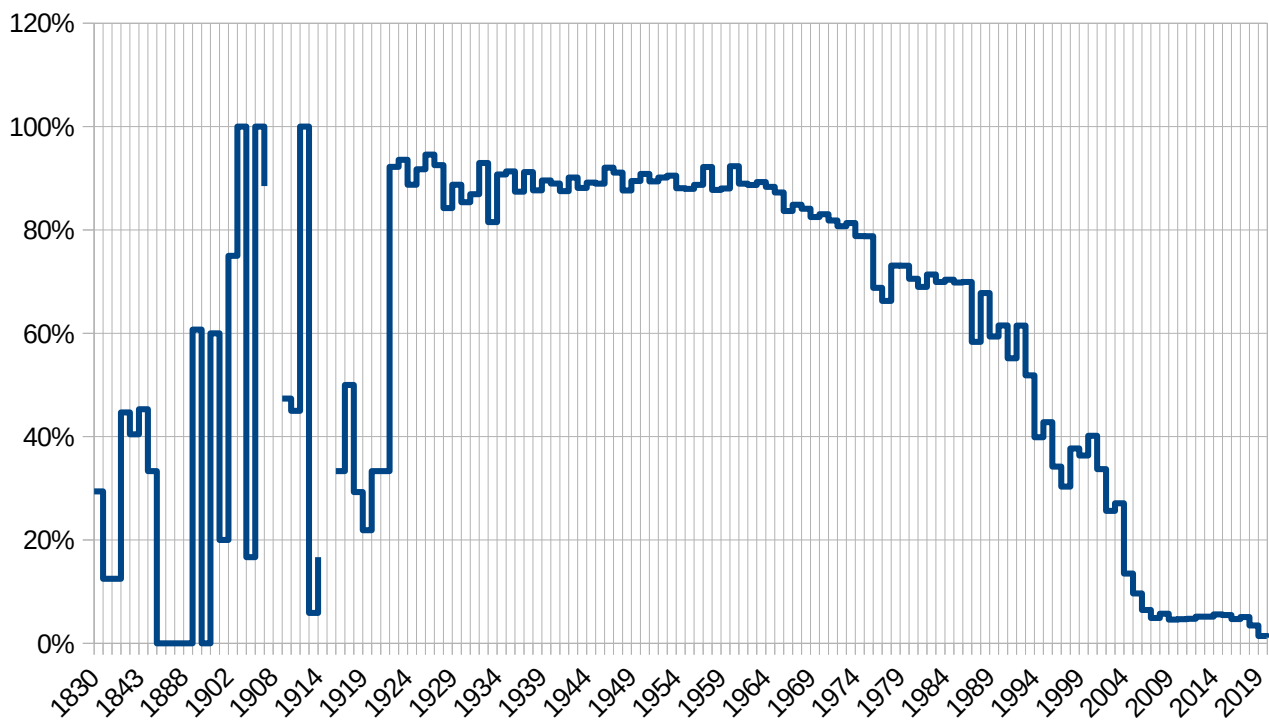


Diagram 3. Andelen av *torde* av motsvarande uttryckssätt (n %) per år

4.2 *Torde* i olika genrer i de utvalda korpusarna

Poeter.se

Sökning i korpusen *Poeter.se* med lemgrammet *torde* (*verb*) gav 667 resultat. Detta är 6,3 ppm eller träff per miljon token. Andelen av den grafiska formen <torde> var 5,8 ppm (616 stycken). Andelen av den grafiska formen <tör> var 0,2 ppm (23 st.). Andelen av <Torde> var 0,3 ppm (27 st.) och andelen av <TORDE> 0 ppm (men ändå 1 stycke!). Andra former av lemgrammet förekom inte. Sammanlagt var andelen av den grammatiska böjningsformen *torde* alltså 6,1 ppm (644 st.)

I min närmare analys avslöjades, att 4 träffar av böjningsformen *torde* och 1 träff av böjningsformen *tör* var äkta men utan större tvivel för gamla. Därtill var 1 träff av *torde* och 2 träffar av *tör* både för gamla och oäkta. En del av dessa för gamla träffar stammade t.o.m. från 1600-talet, och närmare bestämt från Skogekär Bergbos diktsamling *Wenerid*.

Av de kronologiskt passande träffarna var 26 förekomster av *torde* och 3 av *tör* oäkta eller tvivelaktiga. Jag identifierade 12 träffar av *torde* där *tordes* 'vågade' antagligen hade avsetts och 2

träffar av *tör* där *törs* 'vågar' antagligen hade avsetts. Se därför exempel (31). I en del fall är användningen av *torde* i stället för det korrekta *tordes* lätt att förklara med rimmen, som i ex. (32)⁴⁰. Det fanns 10 träffar där *borde* verkar ha avsetts. *Trodde* verkade ha avsetts i 2 träffar av *torde*. Dessutom föreföll *tör* ha 1 gång använts i stället för *för* (konjunktion). Således förblev det 2 tvivelaktiga träffar av *torde* och 1 av *tör*, som jag inte kunde tolka på detta sätt.

(31) men aldrig gjorde , sånt som jag ville men inte torde .

(32) Trots jag borde , jag inget torde .

Det förekom 32 dubletter med *torde* och 2 dubletter med *tör*. Alla dubletter verkade vara äkta och tillräckligt nya förekomster. Sammanlagt hade korpusen 63 felträffar (inklusive de för gamla och dubletterna) av *torde* och 9 felträffar av *tör* (inklusive de för gamla och dubletterna). Om man inte räknar dubletterna (men inkluderar de för gamla), hade man 31 felträffar av *torde* och 7 av *tör*. Således blir antalet riktiga förekomster av böjningsformen *torde* $644 - 31 = 613$ stycken (5,8 ppm), och för formen *tör* blir det $23 - 7 = 16$ stycken (0,2 ppm, det vill säga ungefär detsamma som andelen inklusive felträffarna).

Det fanns i *Poeter.se* en del fall där *torde* kombinerades med ett adverb med liknande betydelse som *knappast* (i negativa satser) eller *väl*. Exempelen (33), (34), (35) och (36) må belysa detta. Fallen med *väl* är mindre klara eftersom *väl* även kan ha andra betydelser. Det förefaller ändå att de använts här just för att uttrycka en viss grad av sannolikhet i propositionen, vilket modal verbet *torde* redan gör. Så är fallet utan tvekan i meningar med *knappast* eller *sannolikt*. Det är fråga om en så kallad tautologi⁴¹ eller pleonasm. En orsak till fenomenet kunde vara att skribenten själv är van vid ordet *torde* men tillägger ett adverb med liknande betydelse för att vara säker på att läsare med mindre brett ordförråd förstår. Fenomenet förefaller så pass ovanligt att man inte kan tala om en typisk kollokation.

Ett grammatiskt särfenomen förekommer i en mening (ex. (37)), där *torde* verkar ha en betydelse som kombinerar kommentar av propositionens sannolikhet med ett konditionalt inslag. Det ser alltså ut som en kombination av *torde* och *skulle*. Fenomenet kunde förklaras av att *torde* morfologiskt är ett preteritum, trots att det semantiskt sett är presens, och konditional betydelse i

40 (Ex. (31) kan se ut som ett likadant fall men den utökade kontexten av ex. (31) är fri vers som inte leker med slutrim på det sätt som den dikt ur vilken ex. (32) är taget, gör.)

41 SAOB:s definition av tautologi: "(ofrivillig) i den språkliga formen varierad upprepning av tankeinhåll l. ordförbindelse, "tårta på tårta"; jfr PLEONASM"

svenskan och i många andra språk brukar uttryckas med hjälp av former som morfologiskt är eller historiskt har varit dåtidsformer. I exempel (38) återges ännu ett fall där skribenten leker med språket och torde använda *torde* främst för just ordlekens skull, men jag har ändå klassificerat det som en äkta förekomst.

(33) gemensamma med tv torde väl vara att tv även

(34) Men om varje stånd ger 20 miljoner mindre Som ” kastas ” iväg Så torde kanske att nån kommer fram
Vara mindre ... än annars Så även om man ståndaktigt försöker Så är det 20 milj mindre varje gång I alla fall
... .

(35) Man kan nog gå på begravningar men man gör det mest för de anhörigas och för sin egen skull , den döde
torde knappast märka nåt inte heller hunden lär fatta vad det handlar om .

(36) Skratt och skrik som från en berg-och-dal-bana fyllde hela behandlingsrummet som sannolikt torde ha haft
ljudisolering .⁴²

(37) men med bättre passform torde de ha blivit nöjda

(38) Så önskar jag att det torde , bli så , att det aldrig mer , någonstans , stode ordet borde .

Det vanligaste av de alternativa epistemiska uttrycken var <förmodligen> med gement f (26 % av fallen), precis som det också förhöll sig i alla fritt tillgängliga korpusar sammanlagt. Däremot hade <torde> en tämligen låg ställning i korpusen med bara 5 %. Se noggrannare information i Diagram 4.

42 Det är också möjligt att skribenten menat *borde*, men detta är så oklart att jag klassificerat det som en äkta förekomst av *torde* i analysen ovan.

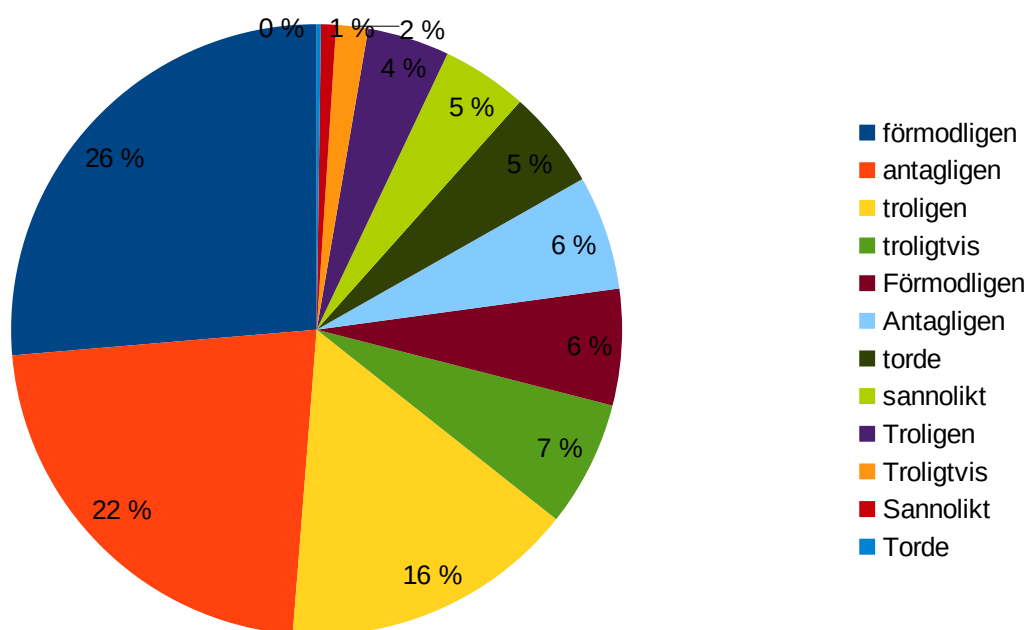


Diagram 4. Andelarna av *torde* och alternativa uttrycksätt (n %) i *Poeter.se*

Skönlitteratur 2000–2013

Korpusen *Skönlitteratur 2000–2013* gav 17 resultat för lemmgrammet *torde* (*verb*). Det betyder 6,8 ppm. Av resultaten stammar 5 från romaner baserat på Google-sökningar med namnet på verket. Andra var biografier, memoarer eller liknande. Namnet avslöjas i högra marginalen när man klickar på ordet *torde* i den beträffande resultatmeningen i KWIC-vyn. Två namn på skönlitterära verk var felskrivna i marginalen, nämligen *En morgontrött fågelskådarens bekännelser* (borde vara *fågelskådares*) och *Gräset är mörkare p andra sidan* (borde vara *på*, och inte *p*). Det fanns inga oäkta eller tvivelaktiga förekomster. Det verkade inte heller finnas med för gamla förekomster. Inga dubletter uppdagades heller. Samtliga förekomster av lemmgrammet var av den grafiska formen <torde>. Två exempel på innehållet i korpusen återges i ex. (39) och (40).

(39) Rektoreernas djungeltelegraf torde höra till de mer effektiva som finns .

(40) Det torde vara bortom allt tvivel att en präst måste äga mera än akademisk utbildning för att klara av de stora krav som ämbetet ställer .

Det vanligaste av de alternativa epistemiska uttrycken var <antagligen> med gement a (41 % av fallen). Däremot hade <torde> en tämligen låg ställning i korpusen med bara 3 %. Formen <förmodligen> kom denna gång först på tredje plats (15 %). Se noggrannare information i Diagram 5.

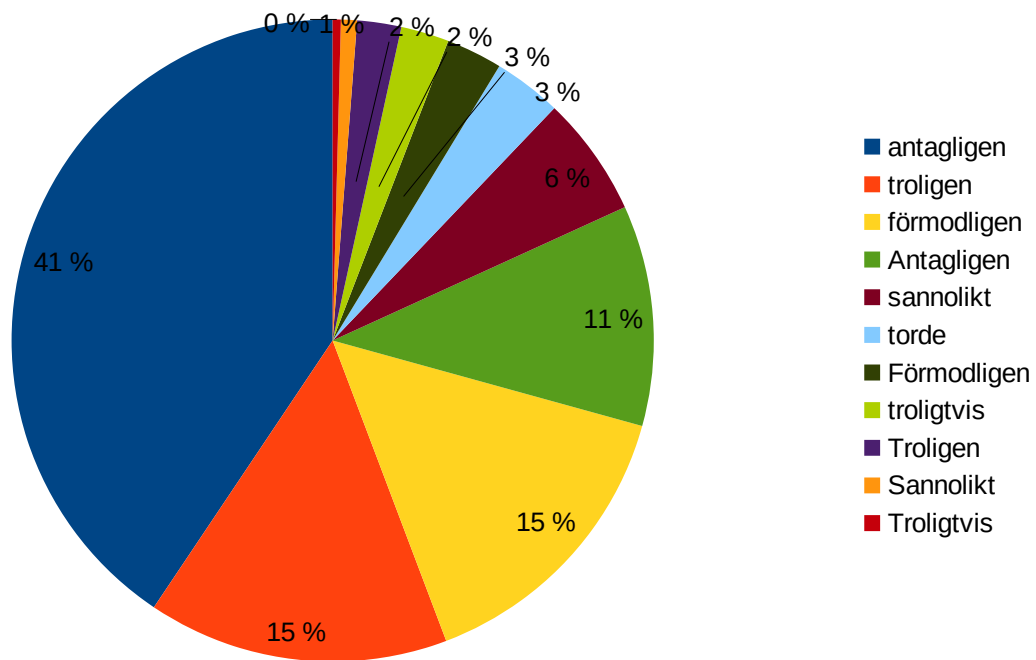


Diagram 5. Andelarna av *torde* och alternativa uttrycksätt (n %) i *Skönlitteratur 2000–2013*

Samhällsvetenskap

Korpusen *Samhällsvetenskap* gav 612 träffar med sökningen *torde* (*verb*). Det betyder 56,4 ppm. Av alla träffar här stammade 38 från licentiatavhandlingar (6,2 % av alla träffarna), eller sannolikt från ett och samma licentiatarbete vars ämne var ”Social och ekonomisk geografi”. Samtliga andra resultat av lemmgrammet *torde* (*verb*) härstammade från doktorsavhandlingar (”PhD”). Det fanns bara en dubblett. Det fanns 599 träffar av <torde> (55,2 ppm), 9 av <Tör> (0,8 ppm) och 4 av <tör> (0,4 ppm). Alla 13 träffar av böjningsformen *tör* var tydligt oäkta. Ett fall var en del av en felskrivning av ordet *organisatör*. I andra fall var det fråga om två ortnamn i Norden, dåtida Sverige, nämligen Yttre Tör och Övre Tör, vilka *torde* syfta på nutida Södertörn. Textens textattribut *ämne* var ”Ekonomi och näringsliv”, men det verkade handla om förläningar och andra historiska förhållanden.

Av träffarna av <torde> var fyra tvivelaktiga. Därav föreföll två vara fall där *borde* hade avsetts; se härpå exempel (41). En märklig mening med tvivelaktigt *torde* återges i exempel (42). En tydlig pleonasm förekom i exempel (43).

(41) Eftersom förskolan betraktas som en arena för social utveckling och detta även är en förmåga som anses vara viktig att förvärva i ett framtidsperspektiv så torde förskolan vara en plats där denna förmåga utvecklas för alla , vilket innebär att idén om att alla barn inte passar in i förskolans praktik är något som bör granskas kritiskt .

(42) Pappan dömdes ett uttryck för ” vanlig svartsjuka ” , vilket ” torde villkorligt i Irak .

(43) Att partierna byggt upp personliga relationer och ett ömsesidigt förtroende för varandra , efter dels de tidigare överläggningarna och dels den gemensamma ståndpunkten i kärnkraftsomröstningen 1980 , torde säkerligen ha inverkat på utgången av det avtal som denna gång kom att slutas.138 Redan månaden efter 1982 års val beslutade partistyrelsen att uppdra åt partiets ledning att ” aktivt utreda frågan om valteknisk samverkan ” (KDS 1982b) .

(44) Då därjämte den formella sidan af nutida vetenskapsmäns produktion alltid anses tillfredsställande , då den klart och logiskt ansluter sig till innehållet , men vältaligheten i sig icke längre skattas annat än såsom individuell talang att inom ramen af hufvudfordran på klarhet och koncishet röra sig med större eller mindre ledighet och elegans , torde varje hänsyn till denna för stiftaren så högt skattade hufvud- uppgift för hans stiftelse numera böra öfvergifvas .

Det hittades 10 träffar av <torde> som var äkta men tydligt för gamla. Som ovan konstaterats, finns det en hel del äldre texter med i hela korpusen, men tio stycken är ändå så många att det torde ha förvrängt resultatet uppåt i någon mån. En gammal mening återges i ex. (44). Träffar som skulle ha verkat både oäkta och för gamla hittades inte.

Det vanligaste av de alternativa epistemiska uttrycken var <sannolikt> med gement s (25 % av fallen). Formen <torde> var andra vanligaste med andelen 22 %. För mera information, se Diagram 6.

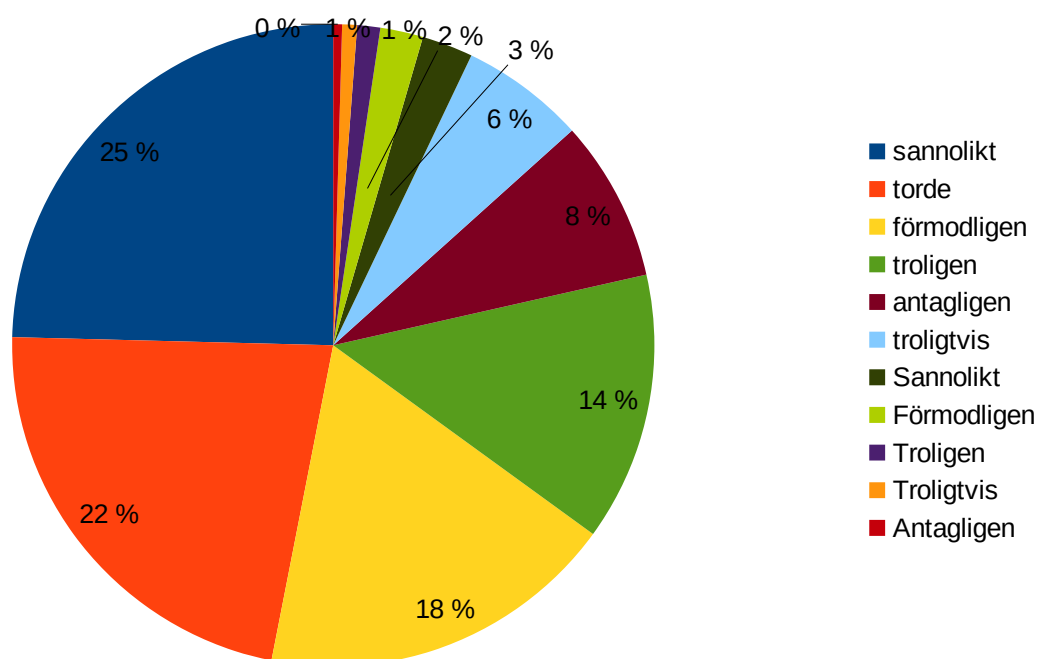


Diagram 6. Andelarna av *torde* och alternativa uttrycksätt (n %) i *Samhällsvetenskap*

Läkartidningen

Korpusen *Läkartidningen 2003* innehöll 203 träffar. Detta innebär 116,1 ppm, vilket är en relativt betydlig mängd i förhållande till många andra korpusar. Alla träffarna var av den grafiska formen <torde>, alltså inga *tör*-former och inga skrivna med versaler. Alla fall verkade äkta, och över huvud taget tedde sig språkbruket prydligt och normenligt. Ett tvivelaktigt fall presenteras i ex. (45), men jag hittade den ursprungliga artikeln med Google och i kontexten verkade det helt trovärdigt att *torde* hade avsetts och inte t.ex. *borde*, så jag räknade fallet som en äkta *torde*-förekomst. Det hittades en dublett – ex. (46) – men det kunde också resoneras att det ingalunda är fråga om någon dublett: jag hittade den ursprungliga artikeln med Google, och den ena förekomsten av meningen var i en sammanfattningsruta medan den andra förekomsten befann sig i brödtexten. Jag har ändå räknat det som en dublett. Det upptäcktes inga fall som skulle verka för gamla. Ett ytterligare exempel av träffarna i korpusen utgörs av ex. (47), vilket även visar hur brett ämnesspektrum korpusen har.

(45) De latinska suffixen - ivus och - orius (transitivus eller transitorius = ' övergående ') torde fogas snarast till verb , inte till substantiv (sidan 152) , och i fråga om termen leukemi menar jag nog att bokstaven k inte ersätter c , snarare ch (sidan 124)

(46) Om nociceptiva nervbanor med ursprung i muskuloskeletal vävnad under inflytande av vissa utlösande eller aggraverande belastningar kan utveckla en ökad retraherbarhet torde en sådan patofysiologisk process analogt

kunna ske även i sensoriska nervbanor omfattande andra kvaliteter än smärta och involvera andra kropps- och sinnesorgan såsom luftvägar , magetarm , urinblåsa och reproduktionsorgan .

(47) Vidare torde det faktum att giraffer har ett medelblodtryck kring 250 mm Hg bero på att deras CNS annars inte fick genomblödning , med sin placering nära 3 m över hjärtlöjd , och knappast på att deras njurar primärt skulle kräva ett sådant perfusionstryck .

Det vanligaste av de alternativa epistemiska uttrycken i *Läkartidningen 2003* var <sannolikt> med gement s (44 % av fallen). Formen <torde> var den näst vanligaste med andelen 19 %. *Sannolikt* var så övervägande att till och med den versala formen <Sannolikt> var vanligare (5 %) än de gemena formerna <antagligen> (2 %) och <troligtvis> (2 %). För mera information, se Diagram 7.

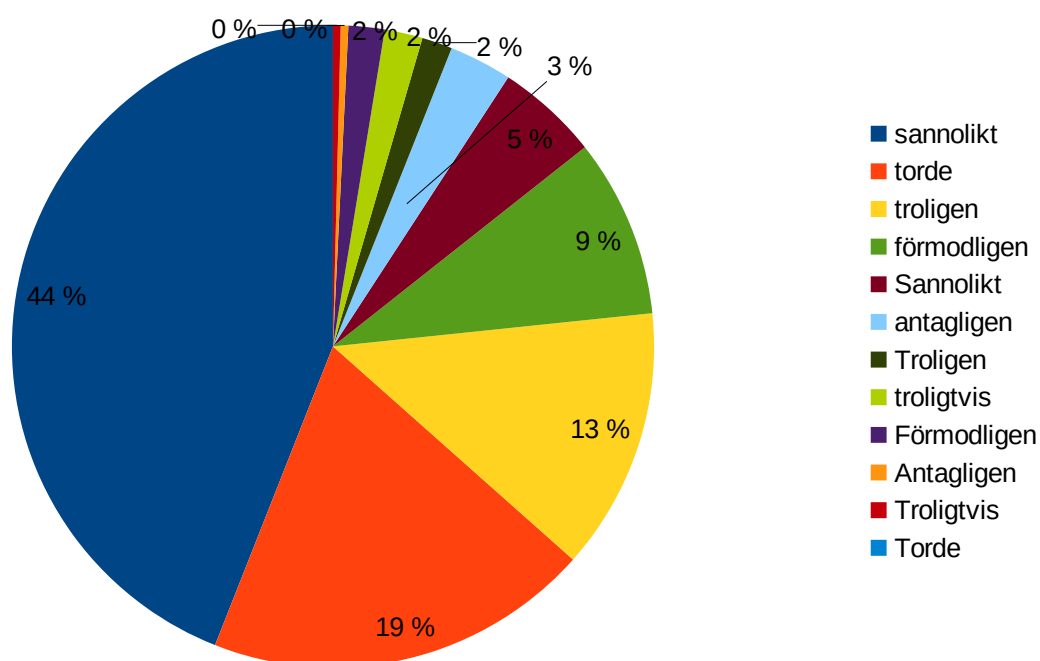


Diagram 7. Andelarna av *torde* och alternativa uttryckssätt (n %) i *Läkartidningen 2003*

Fast jag inte gått igenom andra årgångar av *Läkartidningen* träff för träff, kan jag ändå fastställa något allmänt om korpusen *Läkartidningen* ur ett kronologiskt perspektiv. Korpusen omfattar årgångarna fr.o.m. 1996 t.o.m. 2006. Det går att urskilja en mild sjunkande tendens bland tidningens tillgängliga årgångar. Åren 2000, 2002 och 2005 uppvisar en tillväxt av *torde*-frekvens i jämförelse med det föregångna året; i de andra fallen har frekvensen minskat. Den lägsta frekvensen är i den senaste årgången, det vill säga 2006, nämligen 94,5 ppm, men denna årgångs korpus är en av minsta bland *Läkartidningen*-korpusarna, något som kan öka slumpens inverkan på resultaten. Totalt var frekvensen av lemgrammet *torde* (*verb*) i *Läkartidningen*-korpusarna 126,8 ppm eller 2 669

stycken. Den grafiska formen <Torde> representerades av 3 träffar eller 0,1 ppm medan alla de andra träffarna var av formen <torde>.

När jag jämförde frekvensen av *torde* och de alternativa uttryckssätten i alla *Läkartidningen*-korporar, framgick det att fördelningen mellan de olika alternativen var ungefär densamma som i den ovan presenterade *Läkartidningen 2003*. Andelen av <torde> var denna gång 21 %. För noggrannare siffror, se Diagram 8. Över huvud taget kan man konstatera att frekvensen av epistemiska uttryck av det behandlade slaget var högst just i *Läkartidningen*-korporarna bland de genomgångna korporarna. Närmare bestämt var den absoluta frekvensen ca 596 ppm⁴³ medan den i alla fritt tillgängliga korporar var ca 304 ppm. Det ligger nära till hands att man är mera benägen till att använda sig av synonymer och variera ordförrådet när man måste upprepa samma semantiska innehåll flera gånger i texten, som man enligt resultaten har måst göra i *Läkartidningen*-korporarna. Detta kan förklara även den höga andelen av *torde*.

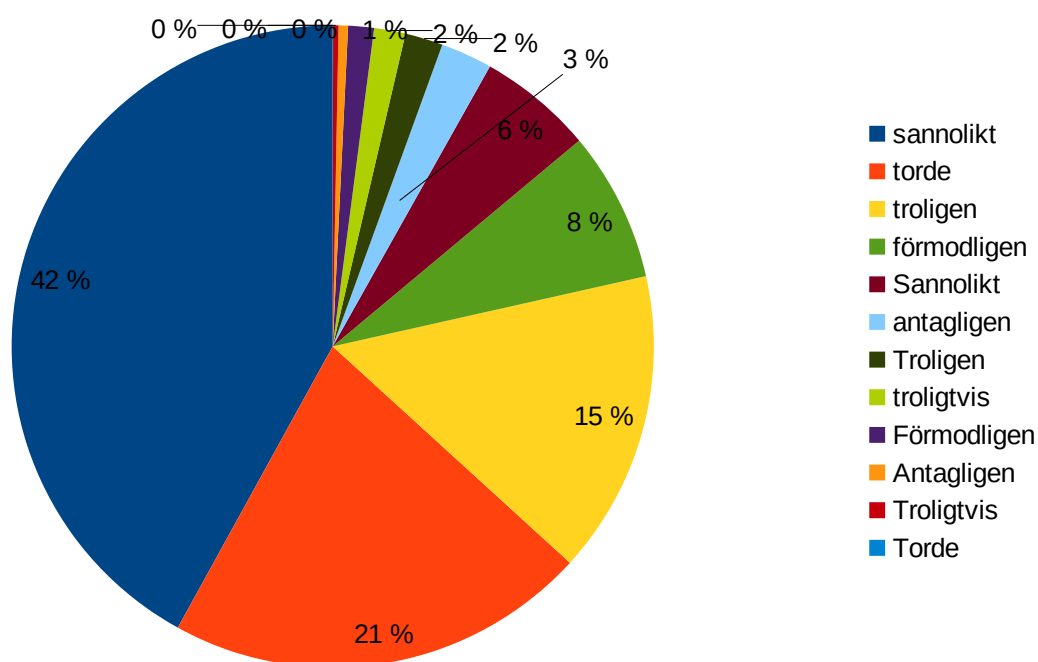


Diagram 8. Andelarna av *torde* och alternativa uttryckssätt (*n* %) i samtliga *Läkartidningen*-korporar

GP

43 I årgången 2003 var den ca 597 ppm, alltså skillnaden är liten. Den högsta frekvensen hade *Läkartidningen 1997* med ca 690 ppm.

Korpusen *GP 2012* gav 177 träffar av lemgrammet *torde* (*verb*). Detta innebär 10,3 ppm. Av träffarna var 175 (10,2 ppm) av den grafiska formen <torde>. Formen <Torde> representerades av 1 träff (0,1 ppm), som tillsammans med den knappa kontext som finns tillgänglig återges i ex. (48). Attributet *artikelavdelning* för denna förekomst är tomt i korpusen, men den verkar stamma från något slags koncist sammandrag eller anteckningsartade iakttagelser, kanske i form av en tabell. Man kan tänka sig att hjälpverbet *torde* hjälper en att nå den korthet som man eftersträvar i sådana textgenrer. Formen <tör> representerades likaså av 1 träff (0,1 ppm), men denna var sannolikt en oäkta förekomst, och jag har klassificerat den som ett tvivelaktigt fall, se ex. (49).

(48) Dryckerna : Torde tillfredsställa alla smakriktningar och plånböcker , även de tjockaste . Imponerande champagneavdelning [*Exemplet innehåller kontext*]

(49) tör och KO , konsu-

Jag hittade 11 träffar (ca 0,6 ppm) som uppvisade ålderdomlig språkdräkt, men jag har ändå klassificerat dem som äkta, egentliga träffar med pålitlig datering, eftersom de verkar härstamma från samtida kåserier. Då skulle det vara pastischerande stil som skulle förklara den ålderdomliga språkdräkten. I de flesta av sådana fall är textattributet *artikelavdelning* tomt, men skribenten identifieras som Anders Hansson (textattributet *artikelförfattare*), vilken på grundval av några sökningar i Google verkar vara en kåsör. Det kan dock vara befogat att komma ihåg att min tolkning kan vara fel. Den pastischerande stilen kunde naturligtvis också vara skälet till att verbet *torde* valts, men detta betyder inte att resultaten vore helt förvrängda på grund av detta: om man i nutid använder *torde* i pastischerande stil är det också nutida användning av *torde*.

Ett beaktansvärt fenomen, som gäller dessa pastischerande träffar, är att 8 (ca 0,5 ppm) av dessa är sådana, där *torde* utan större tvekan finner användning i betydelsen 'bör lämpligen, begärs', som nämns i en del ordböcker men sällan förekommer i moderna texter. Dessutom är det intressant att de flesta av dessa fall även har liknande innehåll: något försvunnet föremål bör återlämnas. Jag dokumenterar dessa fall noggrant eftersom de också är de enda exempel som jag har på *torde* i betydelsen 'bör lämpligen, begärs': ex. (50) och (51) representerar detta, likaså (52) med skillnaden att det inte är tillskrivet Anders Hansson utan Manne Ekman och har <w> i stället för <v>. Ex. (53) är också tillskrivet Manne Ekman och har ålderdomlig språkdräkt, men har <v>, inte <w>, och *torde* används i den normala, epistemiska, betydelsen. Det må tas upp ännu ett intressant exempel, men av annat slag: i ex. (54) förekommer *torde* i en frågesats eller något liknande, vilket är sällsynt. Visserligen har satsen inte frågesatsordföljd, utan frågetecknet är det enda som avslöjar satsens

frågesatskaraktär i skrift. Jag hittade i korpusen 7 dubletter, varav 5 var ovan beskrivna fall med tvivelaktig datering – därav hade 4 betydelsen 'bör lämpligen, begärs'.

- (50) Mellan Sillgatan och Otterhällan borttappades en gesällbok , som upphittaren torde mot vedergällning återlemna å annonskontoret .
- (51) Den som tillvaratagit desamma , torde mot vedergällning aflemna dem hos kopparslagare Pettersson vid Vallgatan .
- (52) I folkträngseln om aftonen den 30 sistlidne förlorades emellan stora och lilla torget en mössa , som upphittaren torde vara god mot wedergällning inlemna hos Mätaren Andersson .
- (53) ” Teckningarna äro en förnämlig och läcker ögonfägnad : sällan torde man finna ett så flott och festligt samspel mellan svart och vitt .
- (54) Den numera nästan bortglömde Sigvard Marjasin torde alltså bara vara ett undantag som bekräftar regeln ?

Det vanligaste av de alternativa epistemiska uttrycken i *GP 2012* var <förmodligen> med gement f (25 % av fallen). Formen <troligen> var den näst vanligaste med andelen 21 %. Formen <torde> kom först på sjätte plats med andelen 5 %. För mera information, se Diagram 9.

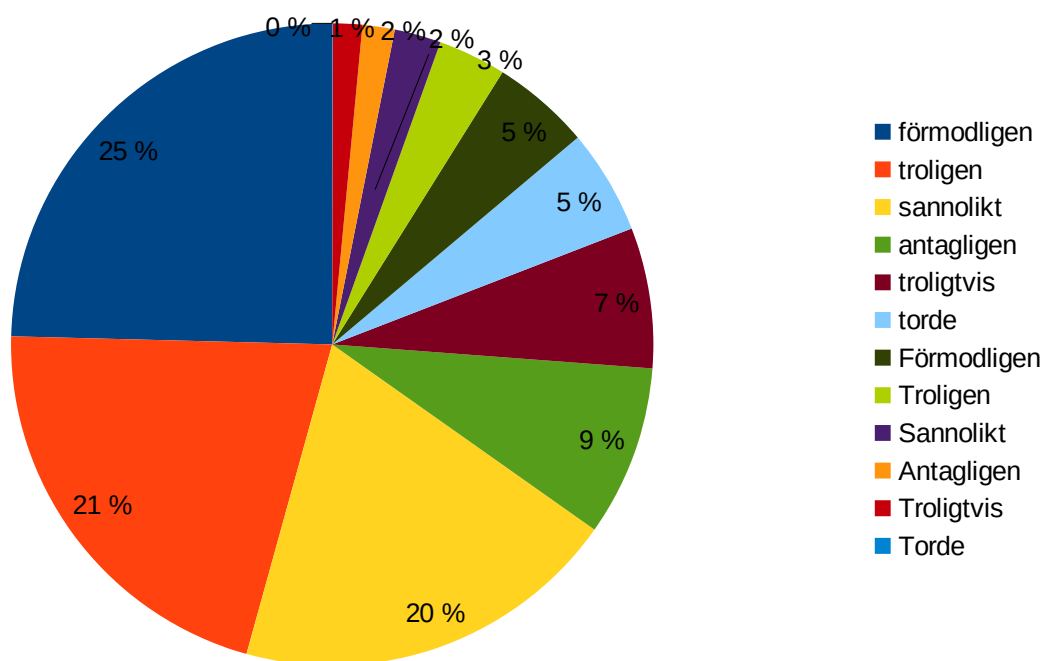


Diagram 9. Andelarna av *torde* och alternativa uttryckssätt (*n* %) i *GP 2012*

Fast jag inte har gått igenom alla delkorpusarna i korpusen *GP* träff för träff, kan jag ändå redogöra för somliga allmänna iakttagelser om de olika årgångarna samt delkorpusen *GP – Två dagar*, som består av odaterat material, bland annat för kronologisk jämförelse. Hela *GP*-korpusen omfattar

3 703 förekomster av *torde* (*verb*) (13,6 ppm). Därav representerar endast 12 förekomster (11,6 ppm) det odaterade *GP – Två dagar*, så den förvränger inte den kronologiska analysen betydligt. Det kan skönjas en milt sjunkande tendens i frekvensen av *torde*. De år då den relativa *torde*-frekvensen stigit i förhållande till föregående år är endast 2008, 2009 och 2012; dessutom har den relativa andelen förblivit densamma år 2007. Andelen av den grafiska formen <torde> i alla *GP*-korpusarna är 3 681 stycken (13,5 ppm), andelen av <tör> 11 (0 ppm), andelen av <Torde> 7 (0 ppm) och andelen av <TORDE> 4 (0 ppm). Av de 11 träffarna med *tör* är 4 otvivelaktiga förekomster av *tör* i betydelsen 'torde', medan 1 är underlig, men klart i verbal användning, troligast i betydelsen 'törs'; de andra är felskrivningar för bland annat <för>. De fyra äkta förekomsterna av *tör* 'torde' ingår i årgångarna 1994 (2 av dem), 2006 (1 av dem) och 2009 (1 av dem). En av dessa presenteras i ex. (55); den stammar från *GP 1994*.

(55) Hon tör inte ha svårt att finna nya läsare nu

När jag jämförde frekvensen av *torde* och de alternativa uttryckssätten i alla *GP*-korpusar, kom det fram att det fanns några tydliga skillnader i fördelningen mellan de olika alternativen jämfört med *GP 2012*, men andelen av *torde* var ungefär densamma som i *GP 2012*, fast lite större; formen <torde> kom nämligen på femte plats med 6,15 %. Vanligast var <sannolikt> med 22,1 %, och näst vanligast var <förmodligen> med 22,0 %. För noggrannare siffror, se Diagram 10.

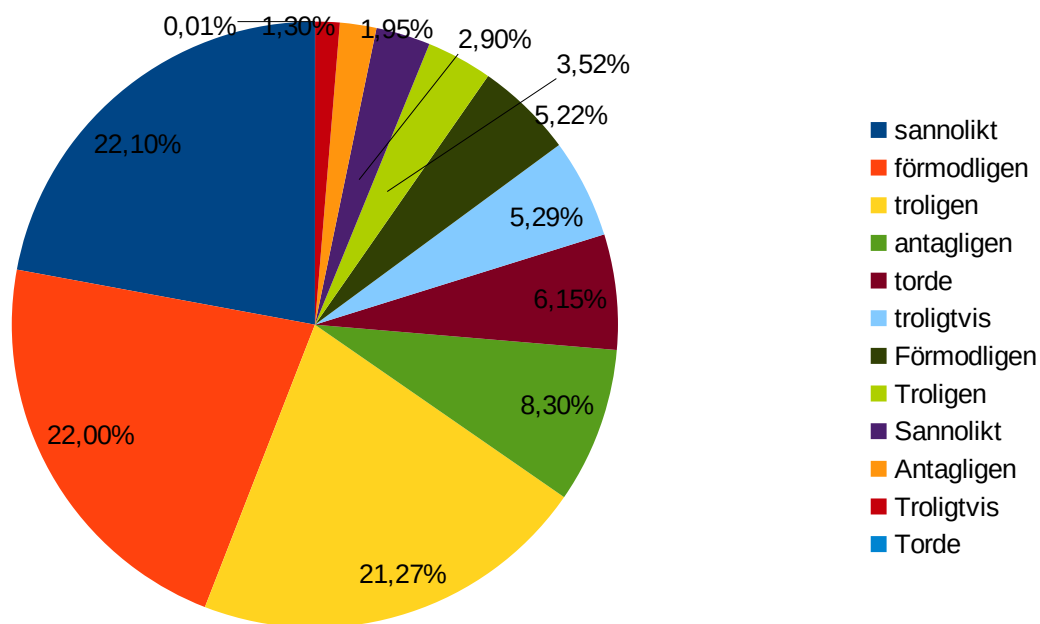


Diagram 10. Andelarna av *torde* och alternativa uttryckssätt (*n* %) i samtliga *GP*-korpusar

Korpusen *8 sidor* gav 0 träffar av lemgrammet *torde* (*verb*). För att vara säker på att det inte bara handlar om ett avvikande annotering sökte jag också med det reguljära uttrycket [Tt]orde, men resultatet blev detsamma. Det var förväntat att *torde*-frekvensen vore låg i en korpus som säger sig inte använda ovanliga ord. Resultatet *torde* också kunna förklaras med att en statlig tidning som strävar efter att framställa saker neutralt men även enkelt törhända inte vill ta ställning till hur sannolikt någonting är – vilket är det som görs med *torde* – utan snarare rapportera om klara fakta och referera människors åsikter i utkristalliserad form.

Det vanligaste alternativet till *torde* i *8 sidor* var <troligen> med gement t (59 % av fallen). Detta adverb var så dominerande att den versala formen <Troligen> var näst vanligast (34 %) och allt annat tillsammans var bara 7 %. Det är vanligt att man inom lättläst inte undviker upprepning av samma ord på det sätt som man gör i många andra genrer utan kanske snarare även ser det som fördel att vara stabil i ordvalet (jfr Österlund 2011: 15). För mera information, se Diagram 11.

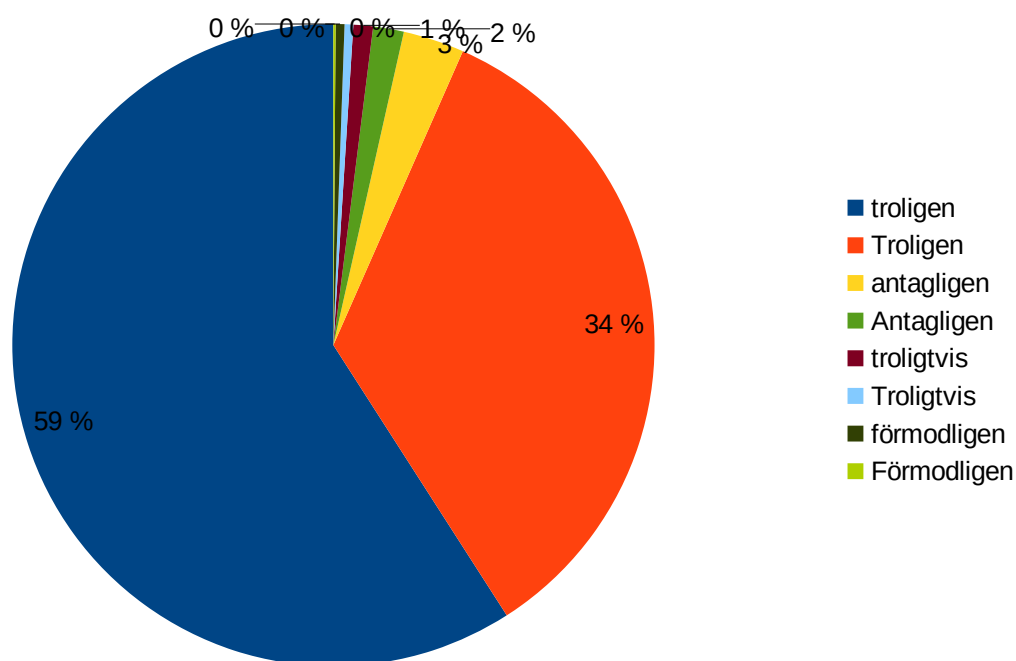


Diagram 11. Andelarna av de alternativa uttryckssätten (n %) i *8 sidor*

Även korpusen *Barnlitteratur 1988–2013* gav 0 träffar av lemgrammet *torde* (*verb*). För att vara säker på att det inte endast handlar om en avvikande annotering sökte jag också med det reguljära uttrycket [Tt]orde, men resultatet blev fortfarande detsamma. Barnlitteraturen brukar främst vara avsedd för barn som målgrupp, och det kan te sig ändamålsenligt för skribenterna att undvika högre stilnivåer och över huvud taget skriva tämligen enkelt, lite på samma sätt som i fallet med *lättläst* och *8 sidor*. Korpusen är också liten i jämförelse med alla de andra korpusarna som jag gått igenom, något som kan ha påverkat resultatet.

De vanligaste alternativen till *torde* var <Antagligen> och <antagligen> med lika många träffar (33 + 33 % av fallen). Detta adverb var alltså dominerande. Den i allmänhet så populära <förmodligen> förekom endast i 8 % av fallen. För mera information, se Diagram 12.

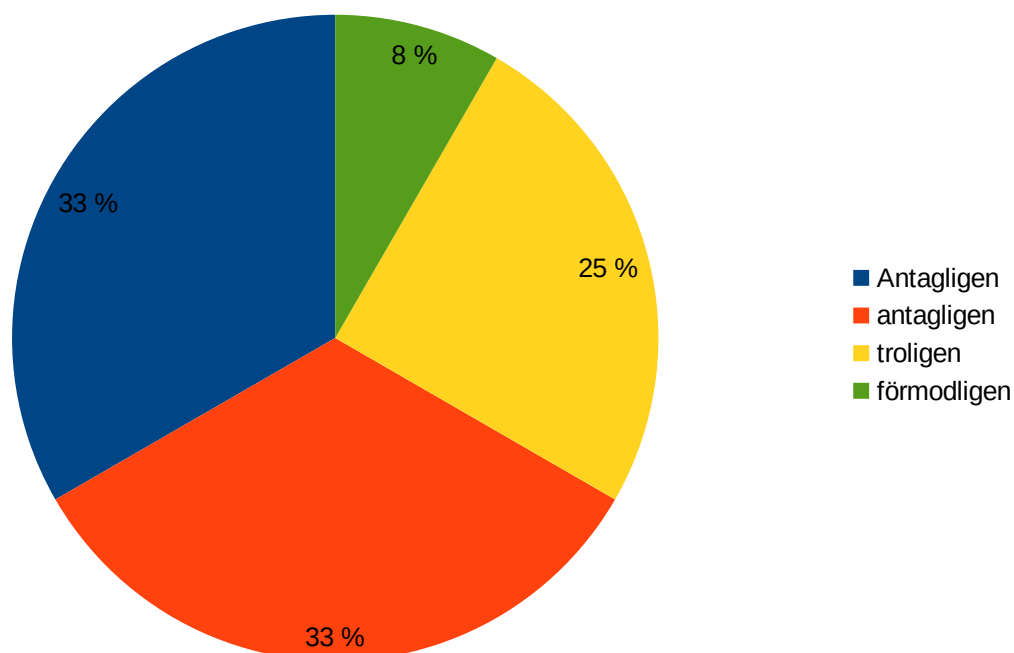


Diagram 12. Andelarna av de alternativa uttryckssätten (n %) i *Barnlitteratur 1988–2013*

4.3 *Ha* med *torde* + supinum

I det följande rapporterar jag om mina forskningar som gäller den andra forskningsfrågan om bortfallet av det perfektuella hjälpverbet *ha* i olika kontexter. Det har hittats så pass få förekomster av *ha*-bortfallet i de sju utvalda korpusarna att det ur statistiskt perspektiv är svårt att dra några

slutsatser på basis av endast det materialet. Nedan presenteras dock resultaten av mina sökningar i alla fritt tillgängliga korpusar. Där blir den bild som de sju utvalda korpusarna ger förstärkt.

Jag har sökt i de sju utvalda korpusarna med CQP-frasen `[lex contains "torde\\.\\.vb\\.1" []{0,2} [lex contains "ha\\.\\.vb\\.1" []{0,2} [msd = ".*SUP.*"]]` för att hitta de fall där *torde* + supinum har använts normalt med det perfektuella hjälpverbet. För att tillåta träffar där bestämningar ingår mellan *torde* och *ha* eller mellan *ha* och *torde* har jag tillagt ett optionellt tomt token `[]` på bägge sidorna av lemgrammet *ha* (*verb*). För att inte få alldeles för många felaktiga träffar, alltså så kallat skräp, har jag bestämt med upprepningsoperatorm att det tomma tokenet får uppträda 0, 1 eller 2 gånger på respektive plats: `[] {0,2}`.

På motsvarande sätt har jag sökt i samma korpusar med frasen `[lex contains "torde\\.\\.vb\\.1" [lex not contains "ha\\.\\.vb\\.1"]{0,2} [msd = ".*SUP.*"]]` för att hitta de fall där *ha*-bortfallet har skett. Här har jag velat tillåta 0–2 token mellan *torde* och supinum. Jag har emellertid inte kunnat nöja mig med ett tomt token som i den föregående sökfrasen, ty då skulle också fall där *ha* uppträder ha blivit med i resultaten. Därför har jag använt mig av en negation av lemgrammet *ha* (*verb*): `[lex not contains "ha\\.\\.vb\\.1"]`.

Korpusen *Poeter.se* innehöll flest förekomster av bortfallet, 17 stycken. Av dessa var 15 förekomster äkta, varav 1 var en dubblett. De felaktiga förekomsterna var här, som också i alla andra genomgångna korpusar, sådana där något verb – annat än *ha* – hade placerats i infinitiv före det som hade annoterats som supinum men i själva verket var particip perfekt, typiskt i användning som predikatsfyllnad. Det var alltså fråga om fel annotering i XML-texten. Ett exempel på en oäkta träff av det nämnda slaget återfinns i ex. (56). Samtidigt åskådliggör exemplet även hur långa meningar korpustexten kan innehålla, när meningsgränserna har identifierats maskinellt. Sökningen av fall med utsatt *ha* i *Poeter.se* gav 52 träffar, varav 2 var dubletter och ingen var oäkta.

(56) Årets skiva : " Sånger att älska till " med Sanna och eminent band med hennes man Daniel och jävla bra handklaver Låten " Torrknulla mig i rosetten " torde bli dökallmärkt på SR och SVT Sannas texter är lika saftiga som poetiska och tyder på en stor musikalisk begåvning tillika en verbal sådan Plus HUMOR Har bara en invändning : På vissa av Sannas sånger sjungz som nån som är tjock och inte vacker och såna sånger passar

inte så bra för just Sanna att sjunga eftersom hon är VACKER YPPIG SNÄLL och kanske inte lika tuff som hon låter alltid

(57) Att söka efter den enda , sanna formen torde vara dömt att misslyckas .

(58) Det torde ha varit ett beslut som Persson kunde motverkat men inte gjorde .

I korpusen *Skönlitteratur 2000–2013* hittades inga träffar med bortfall, men 4 träffar utan bortfall. Ingen av dessa var dubblett eller oäkta. Korpusen *Samhällsvetenskap* gav i sin tur 4 träffar av utelämnat *ha*, varav ingen var dubblett men 2 var oäkta, härpå ex. (57). Däremot hittades träffar av utsatt *ha* 87 stycken, alla äkta och inga dubletter. Ett intressant fenomen hittades i ex. (58), där det perfekta hjälp verbet ingick i konstruktionen med *torde* men hade däremot utelämnats i samma mening i motsvarande konstruktion med *kunde*. Detta är inte i linje med SAG:s omnämnande om att just *torde* och *måtte* tillåter utelämnandet av *ha* (Teleman & al. 1999, band 4: 286). Visst torde detta uttryck i SAG ha avsetts vara relevant i kontexten av epistemiska hjälpverb, och fast *kunde* kan fungera som epistemiskt hjälpverb, är det i detta fall potentiellt hjälpverb i stället. *Läkartidningen 2003* uppvisade inga förekomster av bortfall men 15 förekomster av utsatt *ha*, varav alla var äkta, inga dubletter. *GP 2012* gav 2 träffar av bortfall, varav 1 var äkta. Sökningen av utsatt *ha* gav 15 träffar, varav 2 var dubletter. Dubletterna var visserligen såtillvida gränsfall av en dubblett att attributet *artikelförfattare* hade olika värden hos båda dubletterna än hos de ursprungliga förekomsterna av respektive mening. *8 sidor* och *Barnlitteratur 1988–2013* innehöll över huvud taget inga förekomster av *torde*, som det ovan konstaterats.

Eftersom *Poeter.se* var den enda av de sju utvalda korpuserna som gav någorlunda betydelsefullt antal träffar, har jag undersökt denna korpus utförligare. För att testa min hypotes om att ju flera bestämmningar man har mellan *torde* och supinum, desto sannolikare utelämnar man det perfekta hjälp verbet, räknade jag hur stor del av träffarna för bortfall respektive utsatt *ha* som innehöll 0, 1 respektive 2 token mellan *torde* och supinum (vid sidan om eventuellt *ha*). En svag tendens kunde skönjas som stödjer min hypotes, men det är knappast statistiskt betydande. Därför gjorde jag några grövre sökningar i alla de fritt tillgängliga korpuserna. Dessa resultat har jag inte gått igenom manuellt eftersom materialet är så massivt, utan analysen baserar sig på maskinell behandling med tämligen komplexa reguljära uttryck. I Tabell 4 vidare nedan markeras dessa resultat därför med beteckningen *brutto*.

Jag använde följande CQP-fraser. För att få fram fallen med inga bestämmningar eller token mitt emellan sökte jag med `[lex contains "torde\\.\\.vb\\.1"] []{0} [lex contains`

"ha\\.\\.vb\\.1"] []{0} [msd = ".*SUP.*"] (utan bortfall) och [lex contains "torde\\.\\.vb\\.1"] [lex not contains "ha\\.\\.vb\\.1"]{0} [msd = ".*SUP.*"] (med bortfall). För att få fram fallen med ett token mitt emellan använde jag [lex contains "torde\\.\\.vb\\.1"] [lex not contains "ha\\.\\.vb\\.1"]{1} [msd = ".*SUP.*"] för meningar med bortfall. Däremot för motsvarande fall utan bortfall måste jag iaktta två olika möjligheter: bestämningen kunde teoretiskt finnas före eller efter *ha*. Därför använde jag den logiska operatoren <|>: [lex contains "torde\\.\\.vb\\.1"] (([] {1} [lex contains "ha\\.\\.vb\\.1"] [] {0}) | ([] {0} [lex contains "ha\\.\\.vb\\.1"] [] {1})) [msd = ".*SUP.*"]. När jag sökte fall med utsatt *ha* och två bestämningar blev det ännu längre: [lex contains "torde\\.\\.vb\\.1"] (([] {1} [lex contains "ha\\.\\.vb\\.1"] [] {1}) | ([] {0} [lex contains "ha\\.\\.vb\\.1"] [] {2}) | ([] {2} [lex contains "ha\\.\\.vb\\.1"] [] {0})) [msd = ".*SUP.*"]. Den motsvarande frasen för meningar med bortfall var: [lex contains "torde\\.\\.vb\\.1"] [lex not contains "ha\\.\\.vb\\.1"]{2} [msd = ".*SUP.*"].

Resultaten gällande *ha*-bortfall i *Poeter.se* och i alla fritt tillgängliga korpusar sammanfattas i Tabell 4. I *Poeter.se* syns en tendens att en eller två bestämningar har ökat sannolikheten att bortfallet äger rum. 59 % av träffarna med bortfall hade inga bestämningar mellan *torde* och supinum medan rentav 79 % av träffarna utan bortfall inte hade några bestämningar mitt emellan.

Det syns konsekvent i det senare och större materialet att bestämningar används flera då man utelämnar *ha*, och mindre då man inte utelämnar det. Beträffande fallen med bortfall, är en bestämning frekventare än noll bestämningar och två bestämningar frekventare än en bestämning. För fallen utan bortfall gällde det motsatta. 26 % av fallen med bortfall hade noll bestämningar, 53 % av fallen utan bortfall. Naturligtvis har dessa träffarna inte rensats utan kan innehålla flera felaktiga träffar. En snabb blick över träffarna hänvisar dock till att det inte skulle finnas många. Det gäller att notera de begränsade sökfraserna, vilka kan göra att alla relevanta resultat inte ingått. Sådana fall där flera än 2 token mellan *torde* och supinum föreligger har nämligen inte ingått i resultaten. På motsvarande sätt har det kunnat ingå fall där det finns t.ex. två token mellan *torde* och supinum men satsgräns uppträder mitt emellan så att det egentligen är fråga om två – i stället för en – olika satser, som inte har något att säga om min forskningsfråga.

Tabell 4. Antalet bestämmingar mellan *torde* och supinum i fall där perfektuellt hjälpverb utelämnats respektive i fall där det utsatts

Sökning	Antal träffar där andra token än lemgrammet <i>ha</i> (<i>verb</i>) förekommer mellan <i>torde</i> (<i>verb</i>) och supinumet <i>n</i> gånger			Antal träffar där supinum egentligen ej föreligger (fel annotering)
	<i>n</i> = 0	<i>n</i> = 1	<i>n</i> = 2	
Poeter.se utan <i>ha</i>	10 (59 %)	3 (18 %)	2 (12 %) ⁴⁴	2 (12 %)
Poeter.se med <i>ha</i>	41 (79 %)	9 (17 %)	2 (4 %)	0 (0 %)
Alla utan <i>ha</i> (brutto)	2 288 (26 %)	3 320 (37 %)	3 343 (37 %)	Ej granskat
Alla med <i>ha</i> (brutto)	15 828 (53 %)	9 193 (31 %)	5 037 (17 %)	Ej granskat

På ett allmänt plan kan man konstatera att av de sju utvalda korpusarna var bortfallet vanligast i *Poeter.se*, där 22 % av de genomgångna och rensade⁴⁵ träffarna med *torde* och tillhörande supinumfras uppvisade *ha*-bortfall. För *GP 2012* var motsvarande siffra 7 % och för *Samhällsvetenskap* 2 %. Två andra korpusar gav endast träffar med utsatt *ha* och två korpusar innehöll över huvud taget inga förekomster av *torde*. I samtliga fritt tillgängliga korpusar var den totala bruttoandelen av träffar med *ha* och 0–2 bestämmingar mellan *torde* och supinum 2,2 ppm och bruttoandelen av träffar med inget *ha* men 0–2 bestämmingar mellan *torde* och supinum 0,2 ppm. Den rensade andelen vore sannolikt mycket mindre. Dessa iakttagelser kan tyda på att *ha*-bortfallet inte är särskilt vanligt i samband med *torde*. Det är svårt att säga något allmänt om huruvida bortfallet t.ex. förekommer oftare i koncisa tidningstexter och tabeller i och med att kontexten utanför samma mening oftast inte är tillgänglig. För att på något sätt åskådliggöra bortfallet i kontext anger jag dock två exempelmeningar ur träffarna: ex. (59) och (60).

(59) Först genom Timur Lenks härjningar på 1400-talet torde platsen lagts helt öde . (Svenska Wikipedia (Januari 2017).)

(60) Men företeelsens karaktär av sup- och lekplats för lobbyister torde gått upp för allt fler , viket talar för min tomatteori . (Bloggmix 2009.)

44 Jag räknade hit ett fall där ett snedstreck (antagligen i disjunktiv betydelse) och <borde> hade lagts mellan <torde> och supinumet. Ur teknisk synpunkt utgör dessa verkligen två token, men i praktiken närmar sig detta ett fall där det inte skulle finnas något mellan <torde> och supinumet i och med att <borde> är bara ett alternativt ordval till <torde>, som tilläggs i satsen med hjälp av snedstreckets, och har ingen självständig, från rollen av <torde> separat, syntaktisk roll i satsen. Träffen i fråga hette: *som torde / borde blivit solrosor*

45 Dubbletter ej medräknade.

Ex. (59) har en saklig och vetenskaplig stil medan ex. (60) har en talspråkligare karaktär och det kanske leks med språket. Allt som allt visar resultaten att *torde* är ett mångsidigt verb i svenskan.

5 Diskussion

Den första forskningsfrågan angick användningen av *torde* i korpusarna i Språkbankens Korp-verktyg och i olika genrer. Lemgrammet *torde (verb)* förekom tämligen ofta i alla fritt tillgängliga korpusar i Korps Moderna-läge som helhet: det var näst vanligast av de alternativa epistemiska uttrycken. Andelen av den grafiska formen <torde> var 20 % medan vanligast var <förmodligen> med 21 %. Skillnaden är alltså liten mellan dessa två. Flest träffar av lemmat *torde (verb)* gav korpusen *Statens offentliga utredningar*, 438 086 stycken (1 023,1 ppm).

Moderna-läget innehåller emellertid trots namnet tämligen mycket material som härstammar från 1900-talets början och en del är till och med från 1800-talet. Det verkar finnas en tendens att frekvensen av *torde* i förhållande till de andra alternativa uttrycksätten har minskat under de senaste hundra åren, som Diagram 3 bevisade. Av bl.a. denna anledning har det varit viktigt att jag också undersökt förekomster och användning av verbet i sju utvalda korpusar vilka i princip omfattar endast samtida texter, det vill säga från ungefär 1990-talet och framåt.

Jag har gått igenom manuellt alla träffarna av *torde* i de sju utvalda korpusarna och gallrat bort träffar som var tydligt för gamla eller på något annat sätt oäkta eller vilseledande. Det fanns sammanlagt 1 676 träffar av lemmat *torde (verb)* i de sju korpusarna. Av dessa befanns 26 stycken äkta men för gamla och 3 stycken både för gamla och oäkta. Antalet oäkta förekomster var påfallande; det verkar att många svensktalare förväxlar *torde* och dess gamla presensform *tör* med *törs* eller *borde*. Eventuellt kunde sådana förväxlingar undvikas om större uppmärksamhet fästes vid *torde* och andra hjälpverb i skolväsendet.

Över huvud taget bedömdes 119 träffar (7,1 %) vara på något sätt felaktiga, alltså för gamla, oäkta, tvivelaktiga eller dubletter. Tydligt var dock flertalet äkta (rensade) träffar, närmare bestämt 1 557 (92,9 %). Andelen äkta träffar i förhållande till alla träffar var störst av de sju korpusarna i *Skönlitteratur 2000–2013* med 100,0 %, och minst i *Poeter.se* med 89,4 %. Den förra bestod av professionella, i kommersiellt syfte publicerade verk och gav 17 träffar av lemmat, alltså ganska få, något som kan förklara resultatet. *Poeter.se* i sin tur består av vanliga människors delvis experimentella texter, som utgetts så som de är; därför är det förståeligt att lemmat å ena sidan förekommer tämligen frekvent och att felaktig användning å andra sidan är frekvent också. Se även Diagram 13.

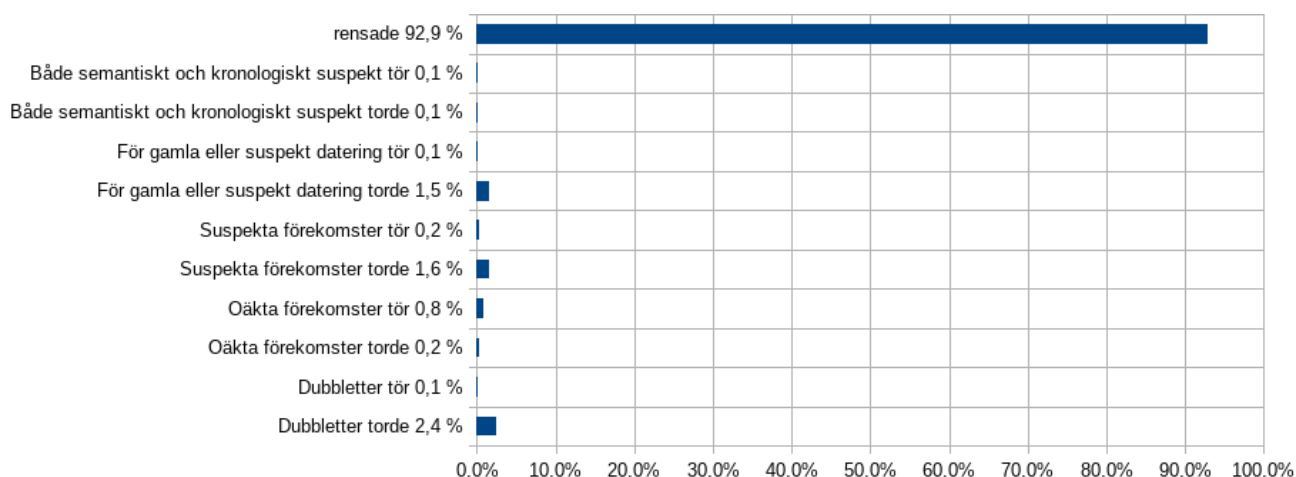


Diagram 13. Andelarna av rensade förekomster av lemmat *torde* (*verb*) och dubletter, oäkta, tvivelaktiga och för gamla förekomster i de sju utvalda korpusarna

Den morfologiska formen *tör* visade sig ha ännu färre äkta förekomster som jag hade förväntat. Det totala antalet *tör* var 37 stycken. Därav var högst 15 äkta moderna förekomster, exklusive dubletterna. Alla var i *Poeter.se*, där människor leker med språket, och det var ofta vanskligt för mig att bedöma om ordet hade använts i den egentliga ordboksbetydelsen eller ej. Däremot hittade jag formen *torde* i betydelsen 'bör lämpligen, begärs' endast i *GP*; visserligen tog jag kanske inte tillräckligt hänsyn till denna detalj i början av processen när jag gick igenom träffarna så en del av sådana fall har kunnat bli obemärkta av mig. Meningarna med *torde* i denna betydelse, vilken en del ordböcker nämner, verkade representera en pastischerande stil som försöker imitera äldre texter. *GP* är en tidning som har innehållit kåserier så det kan vara fråga om kåserier.

Äkta förekomster av den grafiska formen <Torde> med versalt t var ovanliga, vilket är naturligt i och med att *torde* är ett verb, och verb placeras först i meningen främst i binära frågesatser medan *torde* av semantiska skäl sällan förekommer i frågesatser. *GP 2012* innehöll en mening där <Torde> använts i något som verkade härstamma från en koncis, tabellarisk framställning, nämligen ex. (48). Eftersom *torde* gör meningarna kortare än om man använde motsvarande adverb kan man tänka sig att koncisa texter i tabeller och på andra ställen där det råder brist på utrymme vore en naturlig omgivning för *torde*; detta har jag dock inte kunnat bevisa i mitt material. Det fanns även några enstaka meningar där *torde* användes i någon form av frågesats.

I enkla genrer som **barnlitteratur** och **lättläst tidning** förekommer *torde* betydligt mera sällan än i vissa andra genrer. Det vore dock förhastat att dra slutsatsen att undvikandet av *torde* vore den enda

skillnaden mellan dessa genrer gällande epistemiskt modala uttryck. Det kom tydligt fram att det rådde en stor variation även i frekvensen av alternativa uttryck: i de mera akademiska genrerna **vetenskapliga avhandlingar** och **vetenskaplig tidning** samt i genren **vanlig tidning**⁴⁶ var <sannolikt> den vanligaste grafiska formen av de epistemiska uttryckssätt som jag kartlagt. I bägge de akademiska genrerna kom <torde> på andra plats. Däremot de enkla genrerna **barnlitteratur** och **lättläst tidning**, som inte alls uppvisade förekomster av *torde*, uppvisade inte heller några förekomster av *sannolikt*; i **barnlitteratur** var <Antagligen> och <antagligen> vanligast, och i **lättläst tidning** var <troligen> och därefter <Troligen> förhärskande. Det är vanligt att man inom lättläst inte undviker upprepning av samma ord så som man kanske gör i de flesta andra genrer (jfr Österlund 2011: 15). Detta kan förklara den tydliga preferensen av ett enstaka ord. **Vuxenskönlitteratur** är lite av ett mellanting med <antagligen> som vanligast men även <sannolikt> och <torde> representerade. Se även Tabell 5 om detta.

Tabell 5. De vanligaste alternativa uttryckssätten i respektive genre

Genre	Vanligaste	Näst vanligaste	% <i>torde</i>	% äkta (icke-bortgallrade) av <i>torde</i> -träffarna
Lyrik	förmodligen	antagligen	5 %	89,4 %
Vuxenskönlitteratur	antagligen	troligen	3 %	100,0 %
Vetenskapliga avhandlingar	sannolikt	torde	22 %	95,4 %
Vetenskaplig tidning	sannolikt	torde	21 %	99,5 %
Vanlig tidning	sannolikt	förmodligen	6 %	89,3 %
Lättläst tidning	troligen	Troligen	—	—
Barnlitteratur	Antagligen	antagligen	—	—
Samtliga	förmodligen	torde	20 %	92,9 %

Det gäller att minnas att materialet ändå är för litet för att kunna dra långtgående, statistiskt pålitliga slutsatser. Teoretiskt kunde resultaten ha förvrängts av att de båda korpusarna för skönlitteratur, *Skönlitteratur 2000–2013* och *Barnlitteratur 1988–2013*, var delkorpusar av *Finlandssvenska texter*. Man kan inte utesluta möjligheten att det skulle finnas någon skillnad mellan finlandssvensk och sverigesvensk användning av *torde* fastän jag inte är medveten om en sådan skillnad och det inte har framgått något som skulle hänvisa till sådant under forskningsprocessen. Finskans motsvarande bruk av moduset potentialis, alltså faktumet att det ifrågavarande semantiska

46 I den närmare genomgångna årgången *GP 2012* var dock <förmodligen> vanligast.

innehållet kan uttryckas genom verb och inte endast genom adverb, kunde teoretiskt påverka finlandssvenskan så att *torde* å ena sidan vore frekventare än i sverigesvenska. Å andra sidan kunde finskans potentialis även tvärtom göra att finlandssvenska talare vore medvetna om finskans potentialis och dess eventuella inverkan och därför medvetet undvika *torde* som de skulle uppfatta som finlandism. Utöver detta var de båda finlandssvenska korpusarna även tämligen små i förhållande till de flesta andra utvalda korpusarna. Det stora antalet *torde* i genren **vetenskaplig tidning** kan vara påverkad av att det över huvud taget fanns flest alternativa epistemiska uttryck i just denna genre och man har kunnat använda synonymer för att undvika upprepning; detta är dock knappast att betrakta som skevhet utan det är snarare en förklaring till resultatet. Vidare är det möjligt att alla texter i en korpus av en viss genre i verkligheten inte representerar den genre som korpusnamnet eller beskrivningen av korpusen förutsätter. I *Poeter.se* åtminstone förekommer texter som kunde eventuellt snarare klassificeras som prosa i stället för lyrik; gränsen är svår att dra.

Resultaten ger i alla fall en fingervisning som framtida forskare kan undersöka vidare. Åtminstone hypotesen att *torde* är vanligare på högre stilnivåer än på lägre stilnivåer kan anses någorlunda bekräftad. Resultaten nyanserar bilden på så sätt att *torde* verkar vara vanligare i facktexter och dylika än i skönlitterära texter fast båda kan representera högre stilnivåer. Dessutom framgick det att det också finns skillnader mellan stilnivåer och genrer i användningen av motsvarande adverb.

Den andra forskningsfrågan angick utelämnandet av det perfektuella hjälpverbet *ha* i olika kontexter. Fenomenet visade sig så ovanligt i de närmare undersökta korpusarna att det knappast kan sägas något statistiskt pålitligt om *ha*-bortfallet i olika genrer. Vanligast var den av de sju korpusarna i *Poeter.se*, där 22 % av de genomgångna och rensade⁴⁷ träffarna med *torde* och tillhörande supinumfras uppvisade *ha*-bortfall. För *GP 2012* var motsvarande siffra 7 % och för *Samhällsvetenskap* 2 %. Två andra korpusar gav endast träffar med utsatt *ha* och två korpusar innehöll över huvud taget inga förekomster av *torde*.

I samtliga fritt tillgängliga korpusar var den totala, orensade andelen av träffar med *ha* och 0–2 bestämningar mellan *torde* och supinum 2,2 ppm och den orensade andelen av träffar med inget *ha* men 0–2 bestämningar mellan *torde* och supinum 0,2 ppm. De gjorda iakttagelserna kan tyda på att *ha*-bortfallet inte är särskilt vanligt i samband med *torde*. Det skulle dock behövas vidare studier om detta och jämförelse med andra hjälpverb.

47 Dubbletter ej medräknade.

Jag hade två hypoteser om bortfallet. Hypotesen att bortfallet skulle förekomma oftare i koncisa tidningstexter och tabeller och dylika kunde varken bekräftas eller förkastas eftersom kontexten utanför samma mening är oftast inte tillgänglig. Detta står för framtida forskning att utreda. Den andra hypotesen om bortfallet, det vill säga att *ha* utelämnas desto sannolikare ju flera bestämmningar som finns mellan *torde* och supinum, fick däremot stöd i materialet. Det framgick att korrelationen mellan antalet bestämmningar och utelämnandet av *ha* var stark i materialet. Det gäller visserligen att notera att de CQP-sökfraser som jag använde exkluderade eventuella fall med tre eller flera bestämmningar mellan *torde* och supinum. Dessutom är det tänkbart att en sats kan innehålla noll bestämmningar mellan *torde* och supinum men vara annars så lång och tung att språkanvändaren ser behovet av att satsen blir förkortad genom *ha*-bortfall. En utförligare undersökning om ämnet återstår för framtida forskning.

6 Avslutning

I denna studie har jag undersökt det modala hjälp verbet *torde* i Språkbankens korpusar genom korpusverktyget Korp. Metoden har varit korpuslingvistik och särskilt frekvensanalys. Materialet har varit dels alla fritt tillgängliga korpusarna i Korps Moderna-läge, dels sju utvalda korpusar vilkas träffar jag har gått igenom noggrannare. Metoden har varit tämligen lyckad och gett värdefulla resultat. Om jag hade valt att gå manuellt igenom några andra korpusar än dem som jag valt och kanske gått igenom flera och större korpusar, kunde jag eventuellt ha fått annorlunda och pålitligare resultat. De skönlitterära korpusarna borde kanske ha valts så att de inte endast representerar finlandssvenska som de nu gjorde för att utesluta den regionala härkomstens inverkan på saken. Orsaken till mitt korpusval i detta avseende var övervägningar gällande korpusarnas storlek: jag ville ha ungefär lika stora korpusar. Jag borde emellertid ha reflekterat korpusutbudet noggrannare. Att välja korpusar är dock svårt.

Jag hade två forskningsfrågor:

1. Hur används *torde* i olika genrer och korpusar?
2. I vad mån och i vilka kontexter brukar det perfektuella hjälp verbet *ha* utelämnas i samband med *torde* och supinum av huvud verbet?

Jag hade en hypotes gällande den första forskningsfrågan: *torde* skulle användas mera på högre stilnivåer. Denna hypotes fick stöd i materialet. Dessutom gjorde jag en iakttagelse som åtminstone för mig själv var överraskande. *Torde* förekom betydligt oftare i facktexter och vetenskapliga texter än i skönlitterära texter även om dessa också kan representera högre stilnivåer. Detta är en aspekt som skulle kräva vidare forskning, särskilt med tanke på att bägge korpusarna om skönlitteratur bestod av finlandssvenska texter och var dessutom relativt små, vilket kan ha förvrängt resultatet. Det fanns två korpusar som inte alls gav träffar om *torde*: den ena representerade barnlitteratur och den andra lättläst.

För den andra forskningsfrågan hade jag två hypoteser. Den första var att *torde* skulle användas oftare i koncisa tidningstexter och tabellartade framställningar. Detta var vanskligt att ta ställning till i och med att kontexten över meningsgränser oftast inte är tillgänglig i de utvalda korpusarna. Personligt tror jag fortfarande att hypotesen kunde stämma. I detta avseende var min metod inte helt

lyckad, kanske inte heller korpusvalet. Om jag hade hittat och valt korpusar som ger en gedigen utökad kontextvisning, kunde jag kanske ha fått mera belysande resultat gällande den första hypotesen om *ha*-bortfallet. Den andra hypotesen om *ha*-bortfallet var att bortfallet skulle äga rum sannolikare om det finns bestämmningar och särskilt om det finns flera bestämmningar mellan *torde* och supinum. Denna hypotes fick starkt stöd i materialet.

I litteraturen behandlas presensformen *tör* samt användningen av formen *torde* i betydelsen 'bör lämpligen, begärs' antingen inte alls eller beskrivs som ålderdomligt eller dylikt. I mitt material verkade dessa två fenomen ha nästan helt fallit ur bruk. Materialet var emellertid så litet att detta i sig inte ger anledning till att revidera den allmänna uppfattningen. Däremot kan uppfattningen att *torde* som helhet vore ålderdomlig eller sällsynt inte anses stämma i ljuset av mina resultat utan snarare är användningen eller icke-användningen av *torde* en fråga om genre och stil. Det framgick att svensktalare överraskande ofta förväxlar *torde* och *tör* med t.ex. *borde* och *töras*. Även detta kan tyda på att behandlingen av *torde* i skolundervisningen inte är helt onödigt.

Utöver den första hypotesen om *ha*-bortfallet får framtida forskning utgöra vad som är förhållandet av bortfallet i samband med *torde* till bortfallet i samband med andra hjälpverb. Man kunde även forska närmare på *torde* i betydelsen 'bör lämpligen, begärs' samt presensformen *tör*. Ur perspektivet av metodtriangulering kunde resultaten ännu nyanseras och kompletteras genom t.ex. intervjuer eller test i vilka informanter reagerar på frågor vid datorn.

Litteratur

8 sidor 2016a. ”8 sidors historia”. Tillgänglig: <<https://8sidor.se/8-sidors-historia/>> [Hämtad 6.3.2021].

8 sidor 2016b. ”Om 8 Sidor”. Tillgänglig: <<https://8sidor.se/om-8-sidor/>> [Hämtad 6.3.2021].

Alander, Sofia 2019. *MÅNNE. En korpusbaserad studie av satsadverbialet måne i finlandssvenska och sverigesvenska tidningstexter*. Pro gradu-avhandling. Helsingfors: Helsingfors universitet.

Aristoteles 320-tal f.Kr. (1990). *Metafysiikka*. Finsk översättning av Jatakari, Tuija & al. Helsingfors: Gaudeamus.

Blom, Anna & al. 2017. *Fokus kurs 4*. Helsingfors: Otava.

Bolander, Maria 2001. *Funktionell svensk grammatik*. Stockholm: Liber.

Borin, Lars (datum okänt). Språkbanken. *Om oss*. Göteborgs universitet, institutionen för svenska språket. Tillgänglig: <<https://spraakbanken.gu.se/swe/om>> [Hämtad 24.11.2019].

Bray, Tim & Paoli, Jean & Sperberg-McQueen, C. M. & Maler, Eve & Yergeau, François 2008: *Extensible Markup Language (XML) 1.0 (Fifth Edition)*. Cambridge (Massachusetts): World Wide Web Consortium. Tillgänglig: <<https://www.w3.org/TR/xml/>> [Hämtad 13.6.2020].

Bylin, Maria 2013. *Aspektuella hjälpverb i svenskan*. Doktorsavhandling. Stockholm: Stockholms universitet. Tillgänglig: <<https://www.diva-portal.org/smash/get/diva2:615608/FULLTEXT01.pdf>> [Hämtad 21.7.2020].

Bäckström, Linnéa 2019. *Etableringen av ha-bortfall i svenskan. Från kontaktfenomen till inhemsk konstruktion*. Göteborg: Göteborgs universitet.

Ernby, Birgitta 2008. *Norstedts etymologiska ordbok*. Stockholm: Norstedts.

Evert, Stefan & CWB-utvecklingsgruppen 2005–2020. *The IMS Open Corpus Workbench (CWB) CQP Query Language Tutorial. CWB Version 3.4.24*. Tillgänglig: <http://cwb.sourceforge.net/files/CQP_Tutorial.pdf> [Hämtad 4.6.2020].

GP (datum okänt). *Historik*. Tillgänglig: <<https://info.gp.se/om-gp/historik/>> [Hämtad 13.2.2021].

Hammarstedt, Martin & al. 2017. *Korp 6 – Användarmanual*. Göteborg: Göteborgs universitet. Tillgänglig: <https://gupea.ub.gu.se/bitstream/2077/53096/1/gupea_2077_53096_1.pdf>

Harmo, Nuutti 2016. *Vore i finlandssvenskt och sverigesvenskt tidningsspråk*. Pro gradu-avhandling. Helsingfors: Helsingfors universitet.

Hellquist, Elof 1922. *Svensk etymologisk ordbok*. Lund: Gleerups. Tillgänglig: <<http://runeberg.org/svetym/>> [Hämtad 15.4.2021].

Hjortstam, Klas 2018. *Att söka i Korp med CQP och Regexp - en introduktion*. Göteborg: Meijerbergs institut för svensk ordforskning. Tillgänglig: <https://meijerbergs.hum.gu.se/digitalAssets/1694/1694112_vagledning.pdf> [Hämtad 4.6.2020].

Kielipankki (Språkbanken i Finland) & FIN-CLARIN & CSC – IT Center for Science 2015–2020. *Korp Web service (API)*. Tillgänglig: <<https://www.kielipankki.fi/support/korpapi/>> [Hämtad 4.6.2020].

Lagervall, Marika 2014. *Modala hjälpverb i språkhistorisk belysning*. Doktorsavhandling. Göteborg: Göteborgs universitet. Tillgänglig: <https://gupea.ub.gu.se/bitstream/2077/36007/1/gupea_2077_36007_1.pdf> [Hämtad 25.7.2020].

Lindgren, Ola 2011. *Terapeuters och patienters tankar om vad som är verksamt i individuell psykoterapi*. Licentiatavhandling. Karlstad: Karlstads universitet. Tillgänglig: <<https://www.diva-portal.org/smash/get/diva2:450417/FULLTEXT01.pdf>> [Hämtad 26.5.2020].

Läkartidningen (datum okänt). Tillgänglig: <<https://lakartidningen.se/lakartidningen/>> [Hämtad 28.1.2021].

Martola, Nina (red.) 1997 och regelbundna uppdateringar. *Stora finsk–svenska ordboken*. Webbversion. Helsingfors: Institutet för de inhemska språken. Teknikplattform: Göteborg: Göteborgs universitet, Inst. för svenska språket. Tillgänglig: <<http://finsk-svenska.svenska.gu.se/>> [Hämtad 27.5.2020].

McEnery, Tony & Hardie, Andrew 2012. *Corpus linguistics. Method, theory and practice*. Cambridge, New York: Cambridge University Press.

McEnery, Tony & Wilson, Andrew 1996, 2001. *Corpus Linguistics. An Introduction*. Edinburgh: Edinburgh University Press.

Menge, Hermann & al. 1986, 1993. *Taschenwörterbuch. Altgriechisch*. München: Langenscheidt.

Nagy, Zsolt 2018. *Regex Quick Syntax Reference. Understanding and Using Regular Expressions*. Berkeley (Kalifornien): Apress L. P.

Nield, Thomas 2019. *An Introduction to Regular Expressions*. Sebastopol (Kalifornien): O'Reilly Media.

Niiniluoto, Ilkka 1980: *Johdatus tieteenfilosofiaan. Käsitteen- ja teorianmuodostus*. Helsingfors: Otava.

Nikander, Marja-Leena & Jantunen, Aili 1985. *Lukion ruotsin kielioppi*. Helsingfors: Otava.

Pitkäranta, Reijo 2001 (2018). *Suomi–latina–suomi-sanakirja*. Helsingfors: Gaudeamus.

Poeter.se 30.12.2013. *Om Oss*. Tillgänglig: <<https://www.poeter.se/Om+Oss>> [Hämtad 20.4.2020].

Poeter.se (datum okänt). *Riktlinjer*. Tillgänglig: <<https://www.poeter.se/Riktlinjer>> [Hämtad 20.4.2020].

Regeringskansliet (datum okänt). *Statens offentliga utredningar*. Tillgänglig: <<https://www.regeringen.se/rattsliga-dokument/statens-offentliga-utredningar/>> [Hämtad 30.12.2019].

SAOB. Ordbok utgiven av Svenska Akademien. Tillgänglig genom: <<https://svenska.se/>> [Hämtad: 26.4.2021].

SAOL. Ordbok utgiven av Svenska Akademien. Tillgänglig genom: <<https://svenska.se/>> [Hämtad: 26.4.2021].

SO. Ordbok utgiven av Svenska Akademien. Tillgänglig genom: <<https://svenska.se/>> [Hämtad: 26.4.2021].

Teleman, Ulf & al. 1999. *Svenska akademiens grammatik*. Ett verk i 4 band. Stockholm: Svenska akademien.

Tieteen termipankki [Vetenskapstermbanken]. *Kielitiede:modaalisuus*. Tillgänglig: <<https://tieteentermipankki.fi/wiki/Kielitiede:modaalisuus>> [Hämtad 21.9.2020].

Tognini-Bonelli, Elena 2001. *Corpus Linguistics at Work*. Amsterdam: John Benjamins Publishing Company.

Österlund, Maria 2011. *Så här skriver du lättläst*. Helsingfors: LL-Center. Tillgänglig: <<https://www.ll-center.fi/Site/Data/2794/Files/attskrivalattlast-sidor.pdf>> [Hämtad 26.4.2021].