

De som (som) vi använder

En korpusstudie av optionellt 'som' i svenska objektsrelativsatser

Henriette Matz

Institutionen för lingvistik

C-uppsats 15 hp

Allmän språkvetenskap

Kandidatuppsats i lingvistik

Vårterminen 2009

Handledare Thomas Hörberg, huvudhandledare Maria Koptjevskaja Tamm,
rådgivare Kristina Nilsson



Stockholms
universitet

De som (som) vi använder:

En korpusstudie av optionellt 'som' i svenska objektsrelativsatser

Henriette Matz

Sammanfattning

Denna uppsats behandlar förekomsten av den optionella subjunktionen *som* i svenska objektsrelativsatser. Sökningar gjordes i en svenskspråkig korpus för att hitta objektsrelativsatser med och utan *som*. Resultaten undersöktes mot bakgrund av två hypoteser gällande processningseffektivitet och tillgänglighet hos referenter. I båda fallen styrktes dessa hypoteser. *Som* tycks vara mer frekvent i relativsatser där många ord skiljer matrissatsens korrelat från relativsatsens finita verb vilket stödjer principen Maximize On-line Processing som formulerats av Hawkins och som grundar sig i teorier kring processningseffektivitet. *Som* tycktes också vara mer frekvent i relativsatser där det inbäddade subjektet var i hög grad tillgängligt, i fråga om pronominell och animat status hos referenten.

Nyckelord

Relativsats, processning, tillgänglighet, svensk grammatik, syntax

1. Inledning	1
1.1 Introduktion.....	1
1.2 Syfte	2
1.3 Uppsatsens struktur	2
2. Bakgrund	3
2.1 Relativsatser	3
2.1.1 Optionellt som.....	5
2.1.2 Relativsatsens funktion i en mening	6
2.2 Forskningsbakgrund.....	7
2.2.1 Användarbaserade modeller	7
2.2.2 Processing som avgörande faktor för uteslutande av relativsatsinledare	8
2.2.3 Referenttillgänglighet som avgörande faktor för utelämnande av relativsatsinledare.....	14
2.2.5 Jämförelse av modellerna	16
2.2.6 Sammanfattning av avsnitt 2.2.....	17
3. Metod.....	18
3.1 Material.....	18
3.2 Procedur	19
3.2.1 TIGERSearch.....	19
3.2.2 Relativsatsstruktur.....	20
3.2.3. Söksträngar	21
3.2.3 Problem med sökningarna.....	23
3.2.3.1 Egna problem	23
3.2.3.2 Korpusrelaterade problem	25
3.2.3.3 Svensk Trädbank gentemot andra korpusar	26
3.3 Analysmetod	26
4. Resultat	27
4.1 Antal ord mellan korrelatet och relativsatsens finita verb.....	28
4.1.1 Exempel på resultatmeningar	28
4.1.2 Hawkins teorier och matrismönster	30
4.1.3 Chi2-test	31
4.2 Tillgänglighet	32
4.2.1 Pronomen.....	32
4.2.1.1 Exempel på resultatmeningar.....	33
4.2.1.2 Chi2-test	33

4.2.2 Animacitet	34
4.2.2.1 Exempel på resultatmeningar.....	34
4.2.2.2 Chi2-test	35
5. Diskussion	36
5.1 Processningseffektivitet	36
5.2 Tillgänglighet	36
5.3 Korpusstudie	37
6. Sammanfattning	38
7. Referenser	41
7.1 Tryckta källor	41
7.2 Otryckta källor	42
8. Appendix	43

1. Inledning

1.1 Introduktion

I Svenska Akademiens Grammatik (1999b:471) beskrivs en relativsats som en bisats som främst har en attributiv funktion. Mening (1) är ett exempel på en relativkonstruktion, där relativsatsen är understruken:

(1) Glassen som Kalle åt var god.

I (1) är *som Kalle åt* den bit som kallas för relativsats medan *glassen*, det ord relativsatsen syftar på, kallas för korrelat. På det här sättet kan vi i svenskan konstruera relativsatser med flera underordnade led; vi kan t.ex. säga *Kalle som äter, Kalle som äter glassen som kommer från affären, Kalle som äter glassen som kommer från affären som stängdes igår*. När det gäller *som* så kan denna subjunktion ibland utelämnas, närmare bestämt när den används i en objektsrelativsats. Mening (1) skulle också kunna uttryckas som i (2):

(2) Glassen Kalle åt var god.

Utelämnandet av *som* medför ingen betydelseskilnad, men tycks inte kunna förekomma i alla fall där subjunktionen *som* används i en objektsrelativsats. Mening (3) känns klumpig i sin konstruktion:

(3) Glassen Kalle Andersson och hans allra bästa kompis Lisa Nilsson åt var god.

(3) skiljer sig från (2) främst i att den har en väldigt lång nominalfras som subjekt i relativsatsen. Det verkar alltså kunna vara så att storleken på relativsatsens subjekt avgör hur smidig eller klumpig den uppfattas som. Mening (4) känns inte lika klumpig:

(4) Glassen som Kalle Andersson och hans allra bästa kompis Lisa Nilsson åt var god.

Att mening (4) känns mindre klumpig skulle då kunna bero på att det är svårare att utelämna *som* i en relativkonstruktion där subjektsnominalfrasen (*Kalle Andersson och hans allra bästa kompis Lisa Nilsson*) är längre.

En annan möjlighet är att det skulle kunna vara lättare att utesluta *som* när relativsatsens subjekt är animat, dvs. en människa eller ett djur.

1.2 Syfte

Jag ska i denna uppsats undersöka olika faktorer som kan tänkas påverka utelämnandet av subjunktionen *som* i svenska objektsrelativsatser.

Som kan teoretiskt sett utelämnas i alla restriktiva objektsrelativsatser, men ändå känns det i vissa fall klumpigare att göra det än i andra. Med hjälp av sökningar i en svenskspråkig korpus ska jag hitta och analysera restriktiva objektsrelativsatser med och utan *som*, och med utgångspunkt i framförallt två teorier undersöka vilka faktorer som kan tänkas ligga bakom utelämnandet av *som*. Dessa teorier berör tillgänglighet och processningseffektivitet i språket.

De två hypoteser som jag utifrån dessa teorier har formulerat för att tillämpa på utelämnandet av *som* i svenska objektsrelativsatser ser ut som följer:

- Subjunktionen *som* borde sättas ut oftare i meningar där många ord skiljer matrissatsens korrelat från relativsatsens finita verb, i motsats till meningar med färre ord mellan dessa enheter. Detta beror på att konstruktionen då enligt Hawkins (2004) effektivare kan processas.
- Subjunktionen *som* borde utelämnas oftare i meningar där relativsatsens subjekt är ett pronomen eller animat, i motsats till konstruktioner där det inbäddade subjektet inte är ett pronomen eller inanimat. Detta beror på att pronomen och animata referenter i högre grad är tillgängliga.

En liknande undersökning på relativsatser i svenska har inte genomförts tidigare. Mitt syfte är därför också att se hur långt jag kan komma med en undersökning av detta slag: hur bra kommer jag kunna använda mig av en korpus för att få reda på det jag vill om svenska relativsatser? Den korpus som ska användas är en utvärderingsversion, vilket innebär att den inte tidigare har använts för liknande syften som denna undersökning. Detta är ytterligare en anledning till att undersökningen har som delsyfte att se hur bra en korpusundersökning av denna typ kan genomföras.

1.3 Uppsatsens struktur

I avsnitt 2 av denna uppsats ges en bakgrund genom att den svenska konstruktion som undersöks presenteras och förklaras, nämligen relativsatskonstruktionen. En bakgrund kring tidigare forskning som gjorts om ämnet presenteras också, samt de teorier som kommer är av störst vikt för denna undersökning

Därefter presenteras i avsnitt 3 undersökningsmetoden. Detta innefattar bl.a. en introduktion till den korpus som används, eftersom den är grundläggande i undersökningen. I metoddelen

redogörs också för tillvägagångssätt, med exempel på och förklaringar av de söksträngar som används.

Efter detta redovisas resultaten av undersökningen i avsnitt 4. Slutligen diskuteras resultaten i avsnitt 5 och sammanfattas i avsnitt 6.

2. Bakgrund

2.1 Relativsatser

I denna beskrivning av svenska relativsatser utgår jag från Svenska Akademiens Grammatik (1999b: 487-519). De exempel som används är mina egna om inget annat anges.

Den svenska relativsatsen är en typ av bisats som har som sin främsta funktion att tillskriva egenskaper, eller attribut, till den konstituent i den överordnade *matrissatsen* som den är underordnad. Denna konstituent kallas för *korrelat*. Korrelatet och den underordnade relativsatsen utgör tillsammans en nominalfras.

En relativsats kan inledas av subjunktionen *som* eller en satsbas. Satsbasen kan utgöras av ett relativt pronomen, en prepositionsfras, ett adverb, eller adverbet *var* med en preposition. Jag ger exempel på satsbasens inledare nedan. Relativsatsen är här understruken, korrelatet är markerat i fetstil, och satsbasen markeras med kursiv stil.

(5) Det var **en bok** *vilken jag inte hade läst.*

vilken = relativt pronomen

(6) Hon tog hand om **en katt** *vars ägare var bortrest.*

vars = relativpronomen i genitiv

(7) I rummet fanns **en burk** *i vilken hon la sina pengar.*

i vilken = prepositionsfras bestående av preposition *i* och relativpronomenet *vilken* (här kan även relativpronomen i genitiv användas)

(8) Det finns **en plats** *dit jag vill åka.*

dit = adverb

(9) Vi hade **en granne** *varifrån det alltid luktade gott.*

varifrån = adverbet *var* med preposition

En attributiv relativkonstruktion med *som* kan se ut som i (10), subjunktionen *som* är markerad med kursiverad stil.

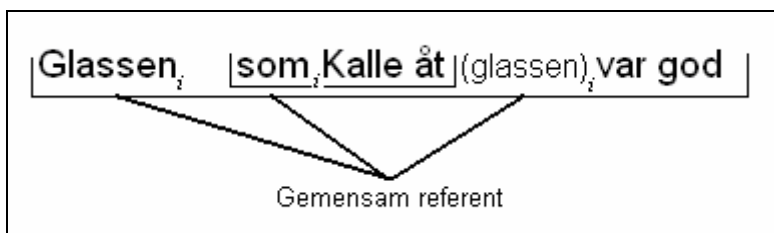
(10) **Glassen** *som* Kalle åt var god.

(10) kan benas ut i två olika delar: *Glassen* var god / *Kalle åt glassen*. De båda delarna har *glassen* som gemensam referent, *glassen* i den första delen är alltså koreferentiell med *glassen* i den andra. Referenten nämns dock med sitt rätta namn endast en gång, i den första satsen. Den senare satsen har med hjälp av subjunktionen *som* underordnats, så att den fungerar som en attributiv sats till det överordnade *glassen*, korrelatet, och har ett underförstått led som är koreferentiellt med den.

Det som händer i relativkonstruktioner som dessa kan förklaras på olika sätt. Chomsky och generativisterna menar att det koreferentiella elementet *glassen* har försvunnit från sin position i en av meningens två satser, och också lämnat efter sig ett spår (Chomsky 1995:35-36). Detta kan visas om vi markerar de enheter som har samma referent, *glassen*, med ett nedsänkt *i* som betyder *index*:

(11) **Glassen_i** *som_i Kalle åt_i* var god.

Relationen beskrivs i Figur 1.



Figur 1. Relativsats

Det koreferentiella elementet **glassen** i mening (11) har den syntaktiska funktionen objekt i meningen, men kan också ha andra funktioner, vilket visas i exempel (12) – (14):

(12) Jag känner **en katt** *som* alltid är snäll.

en katt har funktionen subjekt.

(13) Det är **ett nytt fenomen** *som* han strax ska beskriva.

nytt fenomen har funktionen direktobjekt.

(14) Det finns **en fågelfamilj** *som* jag brukar ge mat till.

en fågelfamilj har funktionen indirekt objekt.

Ett ofta förekommande begrepp när det gäller relativsatsar är *filler-gap dependencies*; relationerna mellan matrissatsens korrelat (*filler*) och den plats som det koreferentiella

elementet har försvunnit från (*gap*). Här kommer termerna 'korrelat' och 'gap' att användas. I sats (11) ovan utgörs korrelatet av *glassen*, och gapet är den indexerade platsen efter det finita verbet.

2.1.1 Optionellt som

Relativsatsinledaren kan i vissa fall uteslutas utan att en menings betydelse ändras. Exempel på detta syns i (15) och (16):

(15) **Glassen** som Kalle åt var god.

(16) **Glassen** Kalle åt var god.

Huruvida satsinledaren kan utelämnas eller ej beror på ett antal faktorer. SAG (1999b:489) beskriver detta som "När den relativa bisatsen är restriktiv och det underförstådda ledet i bisatsen inte är subjekt kan subjunktionen *som* utelämnas, dvs. den relativa bisatsen kan sakna inledare". Detta innebär att det är i meningar där det underförstådda ledet har funktionen direkt och indirekt objekt som subjunktionen kan utelämnas. I de fall där *vilken* är utbytbart mot *som* så kan också denna subjunktion uteslutas. En förutsättning för ett utelämnande är alltså att *som* kan stå som relativsatsinledare, även om det också kan vara relativpronomenet *vilken* som inleder relativbisatsen. Med tanke på att subjunktionen *som* används oftare än *vilken*, så är det rimligt att som i SAG:s definition ovan enbart nämna att *som* kan utelämnas. Det är också denna subjunktion jag kommer referera till som optionellt element i denna uppsats, trots att även *vilken* kan utelämnas på samma sätt.

Att en relativbisats är restriktiv innebär att den utgör obligatorisk information som begränsar (eng. *restricts*) den nominalfras som den fungerar som attribut till. En restriktiv objektsrelativsats kan se ut som i mening (17):

(17) Här är **den glassen** (*som*) Kalle vill ha.

I mening (18) däremot tillför relativbisatsen inte obligatorisk information:

(18) Igår på stan åt Kalle **en glass**, som han inte bjöd på.

Det är bara i restriktiva meningar som (17) som *som* kan uteslutas. I meningar som (18), där relativsatsen inte uttrycker begränsande information till det nominala subjektet *en glass*, blir det ogrammatiskt om *som* utelämnas. Här syns också att den icke-restriktiva meningens matrissats lättare kan stå själv, än den restriktiva meningens matrissats.

(19) ?Här är den glassen.

(20) Igår på stan åt Kalle en glass.

Exempel (19) är svårtydd utan relativsatsen; eftersom det inte anges vidare vilken glass det handlar om krävs en förklarande kontext. Alltså behövs i detta fall den restriktiva relativsatsen för att meningen ska bli komplett.

I en relativsats som fungerar som utbrytningskomplement kan satsinledaren *som* också i vissa fall uteslutas. Detta är möjligt om relativsatsens subjekt inte är underförstått, som i (21), eller om korrelatet inte angränsar till relativbisatsen, som i (22).

(21) Det var **Kalle** jag ville prata med.

(22) **Henne** är det vi pratar om.

Jag ska i denna uppsats undersöka restriktiva objektsrelativsatser som konstrueras med *som*. Relativsatser är en bred kategori, och jag kommer här undersöka både indirekta och direkta objektsrelativsatser, och konstruktioner där relativsatsen fungerar både som attribut och som utbrytningskomplement till korrelatet.

Ett exempel på en sådan sats är ovan nämnda exempelmening (1). Om *som* skulle uteslutas i denna mening, som vi har sett i mening (2) att den kan, så ser meningen istället ut som i exempel (23):

(23) **Glassen**; Kalle åt var god

Det koreferentiella elementet syns här dels i den överordnade matrissatsen som utgörs av *glassen*, och dels i det spår som man kan tänka sig att det lämnat efter sig på platsen efter det finita verbet. Om den underordnade satsen är väldigt lång som i (3) så ser hela relativsatsen med utsatta index ut som i (24):

(24) **Glassen**; Kalle Andersson och hans allra bästa kompis Lisa Nilsson åt var god.

Här syns att de båda indexerade enheterna skiljs åt av en lång sträng ord, vilket skulle kunna vara en orsak till att meningen känns klumpig.

2.1.2 Relativsatsens funktion i en mening

Utöver den attributiva funktion som en svensk relativsats kan ha, vilket presenterats ovan, så kan den också fungera som apposition till en sats och som komplement i en utbrytningskonstruktion.

Exempel på dessa andra typer av bisatser ges nedan:

(25) Det var en stor händelse, vilket jag gladde mig åt.

Relativsatsen *vilket jag gladde mig åt* fungerar som apposition till satsen *Det var en stor händelse*.

(26) Det var igår som jag kom hit.

Relativsatsen *som jag kom hit* fungerar som utbrytningskomplement till det utbrutna ledet *Det var igår*, och har en adverbial funktion.

I relativbisatser som fungerar som utbrytningskomplement kan endast subjunktionen *som* användas. Dessa konstruktioner har, precis som de attributiva bisatserna, ett underförstått led som har ett korrelat i matrissatsen. I dessa fall kan det underförstådda ledet dock ha såväl nominal som också adverbial eller predikativ funktion.

(27) Det är **prata** som hon är bra på.

Korrelatet **prata** är ett predikat.

(28) Det är **livligt** som hon pratar.

Korrelatet **livligt** är ett adverbial.

Både korrelatet och den relativa bisatsen har funktionen subjektspredikativ till verbet *vara*. I vissa fall kan dock verbet *bli* användas istället för *vara*; detta är möjligt om den aktion som bisatsen uttrycker är avgränsad.

(29) Det blev **Obama** som valdes till president i USA.

Det utbrytningskomplementet beskriver är ett ”faktiskt sakförhållande som talaren förutsätter bekant för lyssnaren” (SAG 1999d:517). I (29) förusätts det att talaren redan vet att någon har valts till president i USA; det som förmedlas är att det är Obama som blivit vald till president.

Endast attributiva relativsats och utbrytningskomplement är relevanta för min undersökning, eftersom konstruktioner där relativsatsen fungerar som apposition inte konstrueras med *som*.

2.2 Forskningsbakgrund

2.2.1 Användarbaserade modeller

Den övergripande teori som min undersökning baseras på handlar om språket som *användarbaserat* (eng. *usage-based*). Enligt denna modell är det själva språkanvändningen som formar hur ett språk ser ut. Detta skiljer sig från de generativa teorier som ser människans språkliga kapacitet som något medfött, och där andra faktorer såsom sociala eller mentala kognitiva processer inte anses ha någon påverkan på själva den språkliga strukturen (Chomsky 1975). Användarbaserade modeller för språk pekar istället på att det är mer allmänna kognitiva mekanismer, inlärningsprinciper och sociala faktorer som påverkar och formar vår språkliga kompetens (Tomasello 2003, Diessel 2007).

Detta innebär att man inom de användarbaserade modellerna menar att språkförståelse och språkproduktion påverkar grammatiken, på så sätt att de grammatiska strukturer som föredras ur förståelse- och produktionsperspektiv också är de som är mest frekventa inom och emellan språk. Att en struktur är mer effektiv ur ett psykologiskt perspektiv innebär då att denna struktur bör vara mer frekvent förekommande inom och emellan språk, i förhållande till alternativa mindre effektiva strukturer. Hawkins (2004:3) har formulerat detta samband med följande hypotes (*Performance-Grammar Correspondence Hypothesis*):

“Grammars have conventionalized syntactic structures in proportion to their degree of preference in performance, as evidenced by patterns of selection in corpora and by ease of processing in psycholinguistic experiments.”

Framförallt två tidigare teorier och deras resultat är relevanta för min undersökning. Dessa handlar om tillgängligheten hos språkliga enheter, och effektiviteten i processning.

2.2.2 Processning som avgörande faktor för uteslutande av relativsatsinledare

Begreppet parsning används inom psykologisk språkvetenskap som beteckning för den mentala process hos lyssnaren som skapar en syntaktisk representation av en mening (van Gompel & Pickering 2007:289). Denna procedur är grundläggande för förståelsen av en mening, eftersom den låter oss förstå vilka relationer satsens olika komponenter har till varandra. En mening parsas i den ordning som segmenten uppfattas av åhöraren. Efterhand som ord registreras av åhöraren placeras orden genom parsningsprocessen in i den syntaktiska struktur som behövs för förståelse av meningen. Med andra ord sker denna process genom att förståelsen och avkodningen av en mening ökar gradvis.

Det som behövs för att kunna genomföra en lyckad parsning av en mening är att rätt egenskaper tillskrivs rätt språkliga enheter så fort som möjligt. Ju fortare detta sker, desto fortare kan den övergripande syntaktiska strukturen konstrueras. Detta fokuserar Hawkins (2004, 2007) på när han förklarar att tolkningsgraden ska *maximeras*, dvs. den syntaktiska strukturen i en mening ska uttolkas och konstrueras så fort som möjligt.

Principen *Maximize On-line Processing (MaOP)* (2004:51) formulerar detta:

“The human processor prefers to maximize the set of properties that are assignable to each item X as X is processed, thereby increasing O(n-line) P(roperty) ratios. The maximization difference between competing orders and structures will be a function of the number of properties that are unassigned or misassigned to X in a structure/sequence S, compared with the number in an alternative.”

Den mänskliga processor som här beskrivs är förmågan att parsas en mening. Målet i processningen är att antalet egenskaper som tillskrivs meningen när enheten X parsas ska vara

så högt som möjligt. Att uttolkningsprocessen maximeras innebär att så få misassignments och unassignments som möjligt sker. Misassignments innebär att fel syntaktiska eller semantiska egenskaper tillskrivs en enhet X när den processas. Unassignments innebär att vissa egenskaper hos enheten X förblir oavkodade, eftersom de är beroende av egenskaper hos enheter som ännu inte påträffats, vilket skulle kunna undvikas med en annan ordföljd eller andra ordval. Det MaOP uttrycker är att processningen fungerar olika effektivt beroende på ordföljd. En viss ordföljd har bättre möjlighet än en annan att maximera den pågående processningen; dvs. den undviker i högre grad misassignments och unassignments. Denna ordföljd är den som kommer föredras.

I en svensk relativsats handlar detta om huruvida en konstruktion med eller utan *som* är den som i högsta grad maximerar processningen. Wasow, Jaeger & Orr (kommande) påpekar att en relativsatsinledare borde förekomma mer frekvent i komplexa än i mindre komplexa satser, eftersom denna inledare hjälper till att klargöra konstituenternas relationer till varandra.

Meningarna (30) - (33) är komplexa på det vis att två av dem är tvetydiga när de konstrueras utan *som*.

- (30) Jag gav Kalle glassen.
- (31) Jag gav Kalle (som) glassen åts av en sked.
- (32) Hunden Lisa sover.
- (33) Hunden (som) Lisa sover bredvid.

Om meningarna (31) och (33) konstrueras utan *som* kan de lätt bli tvetydiga. Innan meningen är färdigparsad kan *glassen* i (31) analyseras som direktobjekt, trots att det egentligen är subjekt i relativsatsen. Det som sker är då misassignments, (31) analyseras helt enkelt som (30) eftersom fel egenskaper tillskrivs konstituenterna. På samma sätt kan det ske misassignments under parsningen av (33); *Lisa* kan analyseras som huvudord i NP:n *hunden Lisa*, och meningen tolkas då som i (32), vilket inte sker om *som* satts ut. På detta sätt minskar *som* risken för misassignments. Temperley (2003) har undersökt just huruvida närvaron av relativsatsinledaren *that* i engelska relativsatser gör att tvetydigheten minskar, och hans resultat bekräftar att *that* hjälper till att undvika otydligheter.

Hawkins (2004) har konstruerat en matris för att praktiskt kunna beräkna det förhållande som finns mellan de syntaktiska egenskaper som tillskrivs en konstruktion allteftersom enheterna processas ($O(n\text{-line}) P(\text{roperties})$) och det totala antalet syntaktiska egenskaper som en mening har och som ska tillskrivas orden i meningen för en korrekt tolkning ($U(\text{ltimate}) P(\text{roperties})$). Det som beräknas kallas för OP-till-UP-ratio, och denna ratio beräknas vid varje punkt i en

mening. Alltså: vid varje ord i en mening anger denna OP-till-UP-ratio hur mycket av meningen som helhet som kunnat avkodas dithills.

Egenskaperna som ska tilldelas är av fyra sorter: Ordklass, frastyt så som den kunnat konstrueras så långt, mor-dotter-relationer mellan noderna (i min översättning kallad förankring), och så en kategori som i Hawkins (2004:56) inbegriper fenomen som kasus och theta-roller. Denna sista kategori syftar till att visa på ytterligare argumentstruktur-egenskaper som kan tilldelas ett ord och kommer här att kallas för *relationer*.

I Tabell 1 sätter jag in en svensk relativsats för att illustrera hur satsens UP-till-OP-ratio ser ut:

Tabell 1. En svensk relativsats införd i matris-mönstret efter Hawkins (2004:56-57).

OP-till-UP-ratio	(Glassen)	1.	2.	3.	(var god)
		som	Kalle	åt	
Ordklass	(Substantiv 1)	Subjunktion	Substantiv 2	Verb1	
Frastyt	(NP1, S1)	S2	NP2	VP1	
Förankring	(NP1(Subst1)) (S1(NP1))	S2(subj), NP1(S2)	NP2(subst2), S2(NP2)	VP1(v), S2(VP1)	
Relationer		S2=Attr-NP1		NP1=Obj-Verb1, NP2=Sub-Verb1	
Resultat		5	4	6	
		5/15=33%	9/15=60%	15/15=100%	

Jag beskrev ovan att egenskaperna som ska tilldelas enheterna i en mening är av fyra sorter. I varje kolumn är de egenskaper utskrivna som kan tillskrivas ordet när det processas. När det första ordet *som* uppfattas är det som kan tillskrivas ordet att det är en subjunktion, att det inleder en ny sats S2, att denna sats är inbäddad i den första nominalfrasen NP1 som utgörs av *glassen*, samt att S2 har en attributiv funktion inom NP1. Det är alltså fem egenskaper, av sammanlagt 15 i hela meningen, som är klargjorda vilket ger en OP-till-UP-ratio på 33%.

I den tredje kolumnen finns relativsatsens andra ord, *Kalle*, som genast kan avkodas som ett substantiv, huvud för en ny nominalfras NP2, och som är underordnad den andra satsen S2. Det är här fyra ytterligare egenskaper som har delats ut, vilket tillsammans med de tidigare fem ger en summa på nio sammanlagda egenskaper av 15 totala, alltså 60%.

Slutligen framkommer i meningens sista ord *åt* att detta är meningens första verb som är förankrat i meningens första verbfras, samt att denna verbfras innefattas i underordnade relativsatsen, S2. Dessutom kan de två nominalfrasernas theta-rollerna uttolkas, vilka är subjekt

till verbet för NP2, och objekt till verbet för NP1. Dessa sex ytterligare attribut innebär att 100% av meningen nu är avkodad, och att den är förstådd och uttolkad i sin helhet.

Ju tidigare information kan uttolkas, desto effektivare och därmed bättre är det för kommunikationen. Därför är det fördelaktigt att ha så höga procenttal så tidigt som möjligt i en mening, eftersom detta innebär att mer information har förståtts tidigare vilket i sin tur betyder att meningen parsas fortare, och att kommunikationen är maximalt effektiv.

Jag provar nu att sätta in motsvarande mening utan *som* i matrisen:

Tabell 2. En svensk relativsats utan subjunktion.

OP-till-UP-ratio	(Glassen)	1. Kalle	2. åt	(var god)
Ordklass	(Substantiv 1)	Substantiv 2	Verb1	
Frastyp	(NP1, S1)	NP2	VP1, S2	
Förankring	(NP1(Subst1) S1(NP1))	NP2(subst)	VP1(v), S2(VP1), S2(NP2), NP1(S2)	
Relationer			NP1=OBJ NP2=SUB S2=Attr-NP1	
Resultat		3	10	
		3/13=23%	13/13=100%	

I denna mening är det som kan tillskrivas relativsatsens första ord *Kalle* att det är meningens andra substantiv, samt att detta substantiv är förankrat i en överordnad nominalfras som också den är meningens andra. Vid detta ord har alltså tre av 13 egenskaper kunnat avkodas vilket ger 23%, jämfört med 33% vid ord 1 i relativsatsen med *som*.

När ord två uppfattas kan detta tolkas som meningens första verb som är förankrat i meningens första verbfras. Dessutom låter det finita verbet oss förstå att det handlar om en hel andra sats, S2, som innehåller denna verbfras och också den andra nominalfras som utgörs av satsens första ord *Kalle*. Här syns också att den andra satsen S2 i sin tur är underordnad den första nominalfras NP1 som utgörs av *glassen*. Dessutom gör det finita verbet att vi kan tillskriva *glassen* den syntaktiska rollen objekt och *Kalle* den syntaktiska rollen subjekt. Nu är alltså meningen komplett uttolkad med 100% av egenskaperna tillskrivna satsens enheter.

Nästa steg blir att jämföra skillnaderna vid varje ord och räkna ut ett gemensamt resultat, som tydligare kan visa huruvida någon av konstruktionerna är mer effektiv än den andra. Detta görs i Tabell 3, som är hämtad och översatt från Hawkins (2004:57).

Tabell 3. Jämförelse i effektivitet mellan relativsatser med och utan som.

Ord:	1.	2.	3.
Med <i>som</i>	33	60	100
Utan <i>som</i>	23	100	→
	+10	-40	= -30

I denna tabell visas hur mycket som kunnat uttolkas vid varje ord i de båda meningarna. Vid relativsatsens första ord har konstruktionen med *som* kunnat uttolka 33% av satsens egenskaper, medan satsen utan *som* avkodat 23%. Detta ger differensen $33 - 23 = +10$ mellan de båda satserna. Vid ord två har konstruktionen med *som* avkodat 60% medan konstruktionen utan *som* avkodat hela satsen, alltså 100%. Vid detta ord är differensen mellan de båda satserna alltså $60 - 100 = -40$. Om de båda differenserna sedan adderas så får vi $+10 + -40 = -30$. Denna summa kallar jag för preferenstal, eftersom det är ett tal som visar preferensen för att ha med *som*.

Ett preferenstal, som det tal -30 som räknades fram i Tabell 3, är högre ju högre preferensen för användande av *som* är. Ligger talet på minus så är preferensen snarare att inte använda *som*.

I Tabell 3 har alltså skillnaden mellan versionerna med och utan *som* jämförts vid varje ord, och utifrån detta har ett sammanlagt resultat räknats fram, som i detta exempel är -30. Detta resultat tyder på en preferens för att inte använda *som* i satser som denna.

I en sats med fler ord i relativsatsens subjektsnominalfras blir resultatet annorlunda.

Tabell 4. En svensk relativsats med fyraordig subjektsnominalfras.

OP-till-UP-ratio	(Glassen)	1. som	2. den	3. väldigt	4. unge	5. Kalle	6. åt	(var god)
Ordklass	(Substantiv 1)	Subjunktion	Bestämd artikel	Adverb	Adjektiv	Substantiv 2	Verbl	
Frastyp	(NP1, S1)	S2	NP2				VP1	
Förankring	(NP1 (Subst1), S1 (NP1))	S2 (subj), NP1 (S2)	NP2 (best. art), S2 (NP2)	NP2 (Adv)	NP2 (Adj)	NP2 (subst2)	VP1 (v), S2 (VP1)	
Relationer		S2=Attr-NP1					NP1=Obj-Verbl, NP2=Sub-Verbl	
Resultat		5	4	2	2	2	6	
		5/21=24%	9/21=43%	11/21=52%	13/21=62%	15/21=71%	21/21=100%	

Här kan vi se att den mänskliga processorn vid relativsatsens första ord *som*, precis som i tabell 1, har avkodat 5 av satsens egenskaper. Men eftersom antalet totala egenskaper som ska delas ut

nu har ökat till 21, i och med de extra orden i nominalfrasen, så innebär de fem utdelade egenskaperna att 24% har avkodats vid satsens första ord.

Satsens andra ord utgörs här av den bestämda artikeln *den*, som under processningen avkodas som förankrad i en andra nominalfras NP2, samtidigt som en andra sats S2 som innehåller denna andra nominalfras kan hittas. Dessa fyra egenskaper innebär att vi nu tillsammans med de tidigare fem sammanlagt lyckats utdela nio av satsens 21 egenskaper, vilket innebär att 43% nu är uttolkat.

Ord tre, fyra och fem är adjektiv, adverb och substantiv som alla tillhör den andra nominalfras som konstruerades vid ord 2. Hos dessa kan vi hitta två egenskaper vardera. Detta ger att elva egenskaper har avkodats vid ord 3, 13 vid det fjärde ordet, och 15 vid ord 5. När vi når det femte ordet har därmed sammanlagt 71% av satsens syntaktiska egenskaper kunnat delas ut. De återstående 29% hittar vi i relativsatsens sjätte ord, det finita verbet, precis som i Tabell 1 och 2.

En motsvarande relativsats utan *som* skulle se ut som i Tabell 5:

Tabell 5. En svensk relativsats utan subjunktion med fyraordig subjektsnominalfras.

OP-till-UP-ratio	(Glassen)	1. den	2. våldigt	3. unge	4. Kalle	5. åt	(var god)
Ordklass	(Substantiv 1)	Bestämd artikel	Adverb	Adjektiv	Substantiv 2	Verb1	
Frastyp	(NP1, S1)	NP2				VP1, S2	
Förankring	(NP1 (Subst1), S1 (NP1))	NP2 (best. art)	NP2 (Adv)	NP2 (Adj)	NP2 (subst2)	VP1 (v), S2 (VP1), S2 (NP2), NP1 (S2)	
Relationer						NP1=OBJ NP2=SUB S2=Attr-NP1	
Resultat		3	2	2	2	10	
		3/19= 16%	5/19= 26%	7/19= 37%	9/19= 47%	19/19= 100%	

I denna sats kan vid relativsatsens första ord *den* uttolkas att detta är en bestämd artikel som är förankrad i en överordnad nominalfras, den andra i meningen. Det är alltså tre egenskaper som kan utdelas, vilket innebär att det är 16% av satsens totalt 19 egenskaper.

Satsens andra, tredje och fjärde ord är här, som i tabell 3, adverb adjektiv och substantiv. Hos dessa kan vi även nu hitta två egenskaper vardera, vilket innebär att fem egenskaper har avkodats vid ord 2, sju vid det tredje ordet, och nio vid ord 4. Processandet av det femte ordet innebär därmed att sammanlagt 47% av satsen har kunnat avkodas. Slutligen får vi med det femte ordet, det finita verbet *åt*, 100% av alla egenskaper klargjorda.

Jag för åter in dessa värden i en tabell för att räkna ut ett preferenstal för de båda satserna.

Tabell 6. Jämförelse i effektivitet mellan relativsatser med och utan som.

Ord:	1.	2.	3.	4.	5.	6.
Med <i>som</i>	24	43	52	62	71	100
Utan <i>som</i>	16	26	37	47	100	→
	+8	+17	+15	+15	-29	≡ +26

Det sammanlagda preferenstalet hamnar här på +26, vilket tyder på en preferens för att använda *som* i konstruktionen. Tabell 7 sammanfattar skillnaden i preferenstal i en relativsats med ett eller fyra ord som skiljer korrelatet från relativsatsens finita verb.

Tabell 7. Sammanfattning av preferenstal för relativkonstruktioner med ett och fyra ord mellan korrelatet och det finita verbet.

Antal ord:	1.	4.
Preferenstal:	-30	26

Tabell 7 visar att preferenstalet är högre ju fler ord som finns mellan korrelatet och relativsatsens finita verb. Med andra ord innebär detta att preferensen för att använda relativsatsinledaren *som* är högre när det är fler ord som skiljer dessa enheter åt.

Det som Hawkins MaOP och hans matrismönster för att beräkna OP-till-UP-ratio beskriver är alltså att fler ord mellan korrelatet och relativsatsens finita verb ökar sannolikheten för att en relativsatsinledare *som* ska användas. Detta beror på att unassignments då i högre grad undviks än om *som* inte sätts ut.

Även Race & MacDonald (2003) pekar på längd som påverkande faktor för huruvida en relativsatsinledare utelämnas eller ej.

2.2.3 Referenttillgänglighet som avgörande faktor för utelämnande av relativsatsinledare

Wasow & Jaeger (2008) drar i en undersökning slutsatsen att det framför allt är *tillgängligheten* (*accessibility*) hos relativsatsens subjektreferent som styr huruvida relativsatsinledaren utelämnas eller ej. Begreppet tillgänglighet förklaras närmare nedan, och kan även det ses som relaterat till processning: en referents tillgänglighet styr hur effektivt denna referent kan processas. I denna uppsats används dock begreppet processning för att beskriva Hawkins teorier som presenterades ovan i avsnitt 2.2.2.

Ett exempel på en mening som användes i undersökningen ges i (34) (2008:7). Relativsatsens subjekt har här markerats med fet understrykning.

- (34) I don't remember what they call it... some kind of **word** they use when you get a positive indication of drugs.

I mening (34) är det pronomenet *they*, på svenska *de*, som undersökts för tillgänglighet beroende på ett antal faktorer som jag tar upp nedan i texten.

Som utgångspunkt för undersökningen användes en tillgänglighetsskala (*accessibility scale*), där referenter högre upp på skalan är mer tillgängliga än referenter längre ner (Ariel 1990, citerat i Wasow & Jaeger 2008:170).

(35) Pronoun > Demonstrative > First Name > Definite NP > Indefinite NP

Den genomförda undersökningen visade att relativsatsinledaren oftast utelämnas när relativsatsens subjekt är ett pronomen, mindre ofta när det är ett demonstrativt pronomen, ännu mindre ofta när det är ett egennamn, och så vidare efter skalan i (35). Detta innebär också att sannolikheten för utelämnande av inledare i relativsatsen i (34) borde vara stor, eftersom meningen har pronomen som relativsatssubjekt. I denna mening stämmer detta; ingen relativsatsinledare finns utsatt. I undersökningen i stort pekade också resultaten på att pronomen som relativsatssubjekt tenderar att minska sannolikheten att en relativsatsinledare används. Författarnas antagande är att utelämnandet av ett lättillgängligt subjekt sparar tid och ansträngning för talaren, vilket leder till effektivare kommunikation. Språkproduktionsmässigt är det mindre komplext att planera ett uttalande med subjektsreferenter som står högre på tillgänglighetsskalan. Ju mer komplex formuleringen av ett subjekt är desto längre tid antas denna formuleringsprocess ta, och därför används i högre grad en relativsatsinledare för att behålla turen i samtalet.

Wasow & Jaeger syftar med begreppet tillgänglighet framförallt på något som kallas konceptuell tillgänglighet (*conceptual accessibility*), och som hos Bock & Warren (1985:50) beskrivs som "the ease with which the mental representation of some potential referent can be activated in or retrieved from memory". Tillsammans med skalan i (35) innebär detta att ett ords mentala representation lättare kan aktiveras eller tas fram från minnet, ju högre upp på skalan ordet befinner sig. Ju mer tillgängligt ett relativsatssubjekt är, desto större är sannolikheten att relativsatsinledaren utelämnas.

Wasow & Jaeger gör också skillnad mellan härledd (eng. *derived*) och inneboende (eng. *inherent*) tillgänglighet hos en referent. Den härledda tillgängligheten beror på faktorer utanför själva referenten, nämligen hur given eller ny referenten är i den kontext som används. Skalan i (35) syftar just på härledd tillgänglighet; ju högre upp på skalan desto mer framträdande eller närvarande är referenten i den givna kontexten. Till inneboende tillgänglighet hör egenskaper som är kontextoberoende. Ett sådant fenomen som Wasow & Jaeger nämner, dock utan att hitta bevis för att detta verkligen påverkar utelämnandet av en relativsatsinledare, är animacitet.

Prat-Sala & Branigan (2000) diskuterar begreppet konceptuell tillgänglighet ytterligare, och tar bland annat upp just animacitet som en avgörande faktor för tillgänglighet. Enkelt beskrivet så

står en referent högre upp i *animacitetshierarkin* ju mer mänsklig den är. Denna hierarki för att gradera animacitet har av Givón (2001:56) formulerats som i (36).

(36) Female > Human > Animate > Spatial > Temporal > Entity

Högst upp i hierarkin står alltså mänskliga referenter, därefter animata, rumsliga, temporala och slutligen entiteter. Utöver att en referent med animata egenskaper i sig själv är mer tillgänglig än inanimat referent, så finns det enligt Prat-Sala & Branigan också en tendens för mer animata referenter att inneha subjektspositionen i en mening.

Dessa faktorer tagna tillsammans innebär att ett ords tillgänglighet kan tänkas bero på faktorer som dess position på skalorna för tillgänglighet och animacitet. Ju mer animat en referent är desto större chans är det att den innehar subjektspositionen, och dessutom är sannolikheten större, eftersom animata referenter tycks vara mer tillgängliga än inanimata referenter, att relativsatsens inledare utelämnas.

2.2.5 Jämförelse av modellerna

En studie av Race & MacDonald (2003) styrker tillgänglighet som avgörande faktor vid uteslutande av relativiserare. Det konstateras bl.a. att resultatet av en genomförd undersökning visar att "When the embedded subject was a common noun, participants tended to use "that" significantly more often than when the embedded subject was a highly accessible, frequent pronoun" (2003:4). Detta anges som det fenomen som undersökningen visade ha störst relevans för utelämnande av den engelska relativsatsinledaren *that*.

Två andra faktorer som presenteras som avgörande för utelämnande av den engelska relativsatsinledaren *that* uppges vara längden på relativsatsens subjektsnominalfras, samt längden på relativsatsen efter subjektsnominalfrasen. I båda dessa fall uppges det vara sannolikare att *that* finns med om subjektsnominalfrasen eller relativsatsen efter denna är längre. Taget tillsammans innebär detta att likheter finns med Hawkins; det är avståndet mellan korrelat och finit verb som påverkar om relativsatsinledaren ska sättas ut eller ej. Ju fler ord som skiljer dessa åt, dvs. ju längre subjektsnominalfrasen och efterföljande ord i relativsatsen tillsammans är, desto större är sannolikheten att relativsatsinledaren kommer användas.

Hawkins princip MaOP (2004) gör ingen skillnad mellan om relativsatssubjektet är ett pronomen, ett egennamn, eller ett substantiv. Enligt denna princip är det som är relevant för en effektiv avkodning av en sats antalet ord som finns mellan *glassen* och *åt*. Detta innebär också att det är likgiltigt huruvida det handlar om en lång subjektsnominalfras, eller en kort nominalfras med ett antal adverb. SAG (1999b:489) menar dock att det är ovanligt att *some* utelämnas framför ett adverbial; här skiljer sig alltså Hawkins teori från både Wasow & Jaeger (2008) och SAG.

Å andra sidan finns också likheter mellan Hawkins och Wasow & Jaegers perspektiv. Ur Hawkins synvinkel finns en tendens att, som det verkar, använda *som* när den underliggande satsen är lång. Detta skulle bero på att satsen då av olika anledningar blir lättare att processa. Enligt Wasow & Jaeger är hypotesen att ett pronomen, som ju består av ett enda ord, som relativsatsens subjekt ökar sannolikheten för att *som* kan utelämnas. Här överlappar alltså de båda teorierna: Hawkins bryr sig visserligen inte om vilken ordklass de ord har som skiljer korrelatet från det finita verbet, men menar att det är bättre ju färre ord det är. Tillgänglighetshierarkin har på sin högsta position enordiga pronomen, och längst ner befinner sig nominalfraser som ju kan innehålla flera ord. Här skulle man alltså kunna se Hawkins MaOP som en kompletterande förklaring till varför pronomen är mer tillgängliga – enligt denne på grund av sin korthet. Att nominalfraser befinner sig längst ner på tillgänglighetsskalan skulle då kunna förklaras av att de tenderar att vara längre.

2.2.6 Sammanfattning av avsnitt 2.2

De teorier som presenterats här berör tillgänglighet och processningseffektivitet som bakomliggande orsaker till varför en relativsatsinledare kan uteslutas i vissa fall men inte i andra.

Gällande tillgänglighet är den gemensamma hypotesen att relativsatsinledaren är sannolikare att uteslutas när relativsatsens subjekt är mer tillgängligt. Vad som menas med tillgänglighet skiljer sig dock lite åt; här har jag tagit upp teorier som fokuserar på Ariels tillgänglighetsskala och Givóns animacitetshierarki.

Hawkins princip MaOP handlar däremot om processning. När många ord skiljer matrissatsens korrelat från relativsatsens finita verb lönar det sig att sätta ut *som*, eftersom detta underlättar processandet av meningen. *Som* kan därmed väntas vara mer frekvent i meningar med stort avstånd, dvs. många ord, mellan matrissatsens korrelat och relativsatsens finita verb, än i meningar med ett litet avstånd.

Sammanfattningsvis kan sägas att det som enligt dessa undersökningar har relevans för utelämnandet av en relativsatsinledare är:

- antal ord mellan korrelatet och relativsatsens finita verb
- tillgänglighet baserat på tillgänglighetsskalan och animacitetshierarki

Det är dessa hypoteser som ska testas i denna undersökning, som genomförs med hjälp av sökningar i en svenskspråkig korpus. Denna undersökning ska nu presenteras och förklaras närmare.

3. Metod

Syftet med den undersökning som genomförs är att testa ett antal hypoteser kring vad som har betydelse för utelämnandet av subjunktionen *som* som relativsatsinledare i objektsrelativsatser. Detta utförs genom sökningar i en svenskspråkig korpus.

För att hitta ett statistiskt underlag för hypoteserna formulerades söksträngar för att hitta och utvärdera följande fenomen:

- Hur många ord som skiljer matrissatsens korrelat från det finita verbet.
- Hur tillgänglig nominalfrasens subjektsreferent är ifråga om pronomen och animacitet.

Alla dessa fenomen har i tidigare undersökningar gjorda på engelskspråkiga korpusar visat sig ha relevans för utelämnandet av en relativsatsinledare. Härefter presenteras tillvägagångssätt och vad som används för att undersöka dessa fenomen också i svenska.

3.1 Material

För mitt arbete med att hitta relevanta meningar för undersökningen används den svenskspråkiga korpusen Svensk Trädbank, version 1.0.¹ Denna består av SUC och Talbanken. SUC, som är den större av korpusarna, skapades som ett samarbete mellan institutionerna för lingvistik på Umeå och Stockholms Universitet. Den består av svenska texter från 1990-talet och innehåller en miljon morfosyntaktiskt analyserade svenska ord.

Talbanken är från 1970-talet och innehåller både talat och skrivet språk. I Svensk Trädbank 1.0 är det dock bara den professionella skriftliga delen som tagits med, ca 85000 ord. Eftersom SUC är den större av de båda korpusarna och dessutom är den standard som finns för svenska korpusar, så följer Svensk Trädbank den standard som finns i SUC.

Svensk Trädbank innehåller sammanlagt runt 1850000 ord. Version 1.0 är en utvärderingsversion, vilket innebär att den är ny och ännu inte har hunnit användas för så många korpusbaserade undersökningar. Därför går det inte att vara helt säker på hur en undersökning baserad på denna korpus artar sig. Dessutom innebär det att det kan finnas felaktigheter i taggning och annotering som inte har upptäckts ännu. Jag har dock ändå valt att använda mig av Svensk Trädbank framförallt eftersom den syntaktiska trädstrukturen som korpusen är

¹ <http://spraakbanken.gu.se/stb/om/eng/index.html>

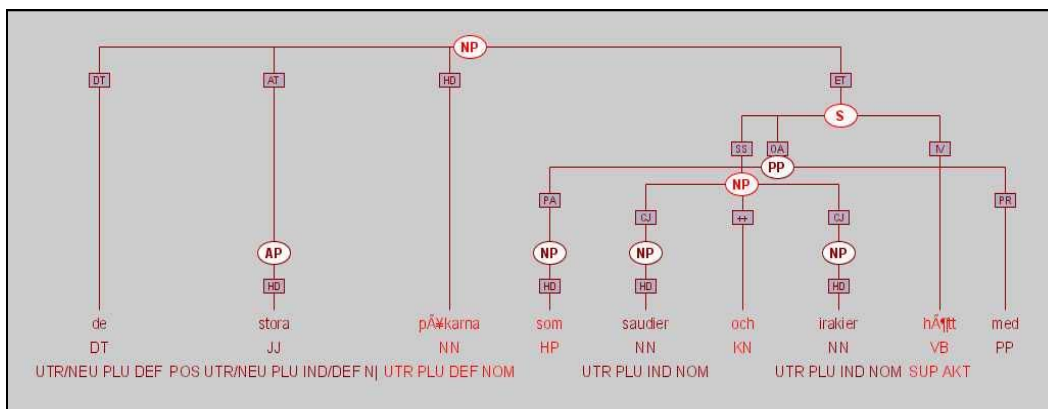
uppbyggd i möjliggör en sökning ovanifrån i trädet, så att jag inte har behövt specificera ordföljden linjärt. Att söka ovanifrån innebär att det går att söka från noder högre upp i meningens syntaktiska trädstruktur än själva orden. Exempelvis går det att välja att söka efter en hel nominalfras, istället för att specificera att det som söks är ett substantiv, ett egennamn eller ett pronomen. På detta sätt ökar möjligheterna till variation inom de språkliga enheter som träffas. En sökning ord för ord hade försvårat mitt arbete eftersom de relativsatser utan *som* som jag vill hitta kan innehålla mycket emellan korrelatet och det finita verbet.

3.2 Procedur

3.2.1 TIGERSearch

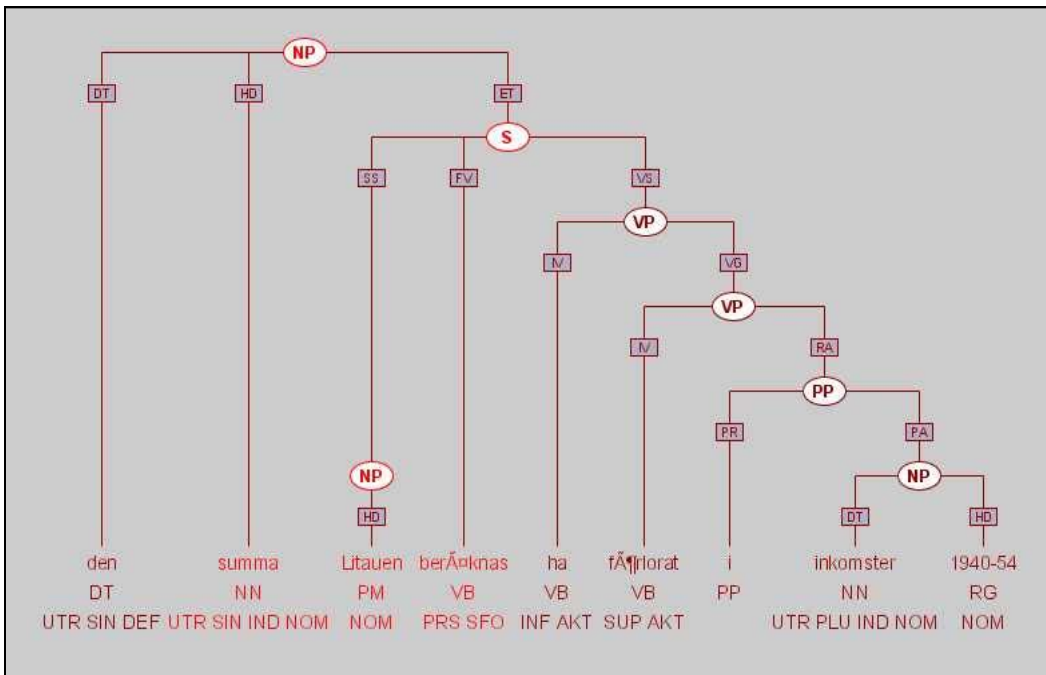
Korpusen är kodad i TIGER-XML, och det verktyg som jag har använt för att söka i textsamlingen är TIGERSearch, version 2.1. TIGERSearch är en fritt tillgänglig programvara som går att ladda ner gratis från internet.²

Ett sökresultat i TIGERSearch kan se ut som i Figur 2 och Figur 3:



Figur 2. Exempel på sökresultat i TIGERSearch, objektrelativsats med *som*: "de stora påkarna som saudier och irakier hött med".

² <http://www.ims.uni-stuttgart.de/projekte/TIGER/TIGERSearch/>



Figur 3. Resultat av sökning efter objektsrelativsats utan som: ”den summa Litauen beräknas ha förlorat i inkomster 1940-54”.

Trädets noder är avbildade som runda enheter med frasbeteckning utskrivna inuti. För de terminala noderna som utgörs av ord eller skiljetecken finns ordklass och egenskaper såsom numerus och genus utsatt. Allt detta går att använda vid formuleringen av söksträngar (König Lezius & Voorman 2003).

De noder som i trädigrammen syns som runda enheter kallas i TIGERSearch för *Categories*, dvs. kategorier. Vid sökningar i korpusen skrivs dessa element ut som *cat*. De ord som finns i de terminala noderna är taggade för *Parts of speech*, eller satsdelar. I sökningar skrivs detta ut som *pos*.

Utöver fraser och ordklasser är också relationer mellan enheterna utsatta, i form av fyrkantiga boxar på de linjer som sammanbinder konstituenterna. Dessa boxar markerar med vilken relation en nod dominerar en annan nod. Exempelvis syns att den högsta nominalfrasen i Figur 3 dominerar substantivet *påkarna* genom en box där det står *HD*. *HD* står för *Head*, och innebär att substantivet utgör den här nominalfrasens huvudkonstituent. Boxen signalerar alltså att nominalfrasen *NP* genom en huvudrelation dominerar substantivet *påkarna*. Jag använde mig dock inte av denna funktion eftersom jag var rädd att av misstag utesluta satser.

3.2.2 Relativsatsstruktur

De syntaktiska strukturer vi ser i Figur 2 och 3 ovan är likadana, med undantag för att meningen i Figur 2 innehåller subjunktionen *som*. Den övergripande struktur som trots allt är gemensam ser ut som i (37):

$$(37) \quad [NP1[N S[(som) NP2 VB]]]$$

Den första nominalfras *NP1*, som inleder frasstrukturen i (37), motsvaras i Figur 2 och 3 av den översta noden med *NP* utskrivet inuti cirkeln. (37) visar att det först i denna första nominalfras ska finnas ett substantiv, vilket är det ord i trädet under vilken det står *NN* som i Svensk Trädbanks taggningsmodell betyder just substantiv. Detta är korrelerat. Den första nominalfrasen ska också dominera en sats *S*, vilket är relativsatsen. I träden finns mycket riktigt en nod *S* som är direkt dominerad av den första nominalfrasen. De båda träden skiljer sig åt med avseende på om *som* finns med i strukturen eller inte; i (37) är *som* satt inom parentes för att visa att detta element är optionellt.

Den underliggande relativsatsen *S* dominerar i sin tur konstruktionens andra nominalfras *NP2* som i träden utgörs av den nod *NP* som direkt domineras av *S*. Denna nominalfras är subjektet i den relativa konstruktionen. Det sista element som är specificerat i strukturen i (37) är det verb som följer på subjektsnominalfrasen. I träden kommer mycket riktigt ett verb efter den andra nominalfrasen *NP2*. Detta verb är markerat med *VB* under själva ordet, vilket betyder just verb.

3.2.3. Söksträngar

Exempel på satser som har strukturen i (37) och Figur 2 och 3 ges nedan:

(38)	Igår såg han hunden	<u>(som)</u> <u>Anna</u> <u>hatade.</u>
	[NP1 [N	S[(som) NP2 VB]]]
(39)	Det är en fråga	<u>(som)</u> <u>styrelsen redan</u> <u>har diskuterat.</u>
	NP1 [N	S[(som) NP2 VB]]]
(40)	Det är ju det	<u>(som)</u> <u>jag</u> <u>säger.</u>
	NP1 [N	S[(som) NP2 VB]]]
(41)	Det var tydligen Maria	<u>(som)</u> <u>de</u> <u>hade sett.</u>
	NP1 [N	S[(som) NP2 VB]]]

Enheternas fras/ordklass-kategorier är utskrivna under orden. En viktig poäng med dessa exempel är att korrelerat utgörs av olika typer av nominalfraser: i mening (38) är det ett bestämt substantiv, i mening (39) ett obestämt substantiv, i mening (40) ett deklarativt pronomen, och slutligen i mening (41) ett egennamn. Söksträngarna har också utformats för att kunna hitta meningar som dessa, med olika typer av ord i korreleratets nominalfras.

Det var meningar som (38) - (41) som skulle hittas genom de söksträngar som användes. Dessa söksträngar syns i (42) och (43).

```
(42) #np:[cat="NP"] &
      #np > #n1:[pos="PM"|pos="NN"|pos="PN"] &
      #np > #s:[cat="S"] &
```

```

#s > #np2:[cat="NP"] &
#s > #v:[pos="VB"] &
#sb:[word="som"] . #np2 &
#np2 > #n

(43) #np:[cat="NP"] &
#np > #n1:[pos="PM"|pos="NN"|pos="PN"] &
#np > #s:[cat="S"] &
#s > #np2:[cat="NP"] &
#s > #v:[pos="VB"] &
#np2 > [pos!="HP"] &
#np2 > #n2 &
#n1 . #np2

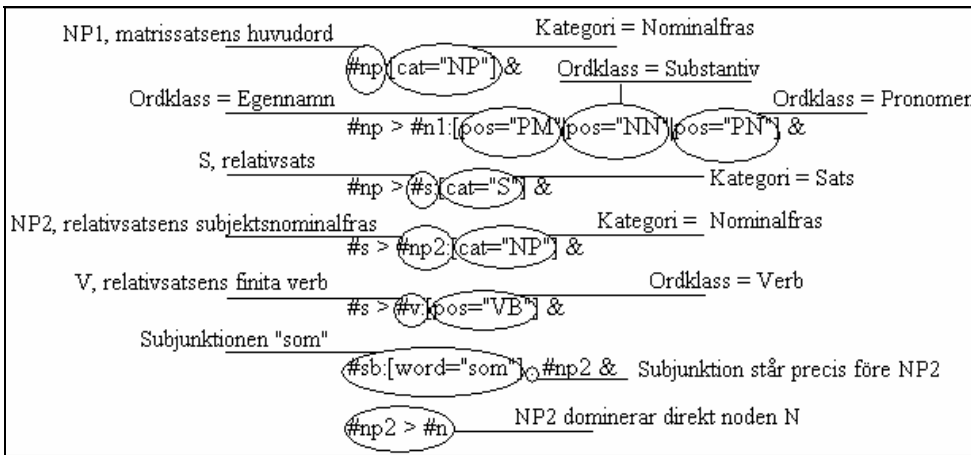
```

Söksträngen i (42) är formulerad på så sätt att den ska hitta en nominalfras (*#np*) som direkt dominerar ett nomen (*#n1*) med ordklassfunktionen egennamn, pronomen eller substantiv, precis som i exempelmeningarna (38) – (41) ovan. Nominalfrasen dominerar också en sats (*#s*) som i sin tur direkt dominerar en andra nominalfras (*#np2*) och ett verb (*#v*). En subjunktion *som* (*#sb*) ska stå precis framför den andra nominalfrasen. Denna andra nominalfras dominerar också en enhet *#n* som inte är specificerad för några egenskaper. Eftersom den domineras av en nominalfras kan den ha alla funktioner som en nominalfras kan ha. Detta antas i sökningen vara relativsatsens subjekt. Poängen med att inte specificera denna nod för några egna egenskaper var att på så sätt möjliggöra variation i de element som förekom där. Denna sökning resulterade i relativsatser med ett utsatt *som*, bland annat den trädstruktur som syns i Figur 2.

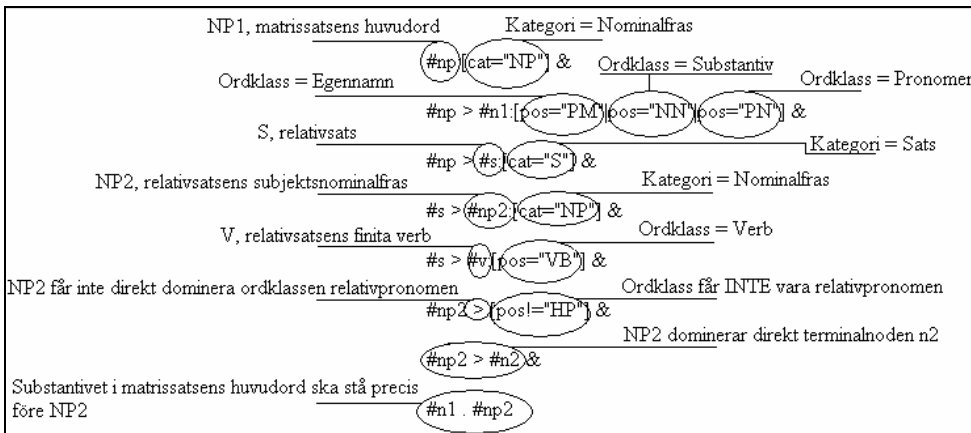
Söksträng (43) har samma struktur som sträng (42), med det undantaget att en subjunktion *som* inte är utsatt. Dessutom uttrycks att den andra nominalfrasen inte får dominera en terminalnod med den morfologiska egenskapen 'HP', vilket är vad *som* taggats för i denna korpus. *HP* betyder i Svensk Trädbanks terminologi ett 'wh-pronoun', eller på svenska ett relativpronomen. Detta innefattar också det som i SAGs terminologi kallas för subjunktion, nämligen *som*.

Den linjära ordningen i sträng (43) är specificerad så att det första substantivet (*#n1*) ska stå precis framför den andra nominalfrasen (*#np2*). På detta sätt lyckades jag hitta relativsatser utan *som*, såsom trädet i Figur 3 ovan. Min avsikt med att inte närmare specificera vad som skulle finnas i terminalnoden *#n* var att denna bit av satsen skulle kunna ha varierande längd, vilket också syntes i resultatet.

Figur 4 och 5 ger en tydligare bild av vilka delar söksträngarna innehåller.



Figur 4. Förklaring av element i söksträng (42)



Figur 5. Förklaring av element i söksträng (43).

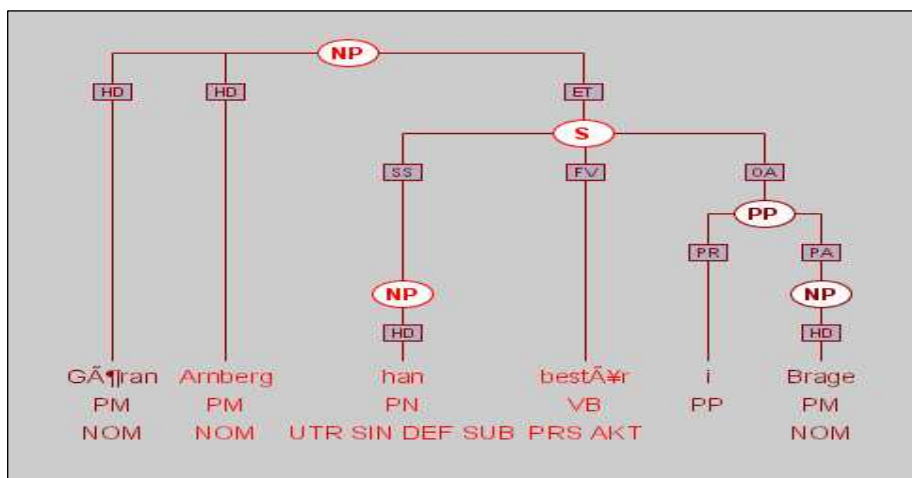
En lista över de förkortningar och tecken som används i TIGERSearch finns i Appendix.

3.2.3 Problem med sökningarna

3.2.3.1 Egna problem

En liknande korpusundersökning av relativsätser i svenska har inte gjorts tidigare, vilket innebar att söksträngarna fick formuleras helt efter egen förmåga. De slutgiltiga strängarna hittade de önskade relativkonstruktionerna, men bättre och mer precisa formuleringar kan förmodligen utformas om så önskas. Jag har dock sett denna undersökning som ett första försök att genomföra en studie av detta slag i svenskan, och därför inte ägnat den större delen av min tid åt att fullända sökningarna.

Båda söksträngarna hittade utöver de önskade resultaten också ett antal meningar som inte hade den sökta konstruktionen. Ett exempel på en sådan mening som hittades vid sökning med sträng (43) visas i Figur 6:



Figur 6. Utsnitt ur felaktigt resultat vid sökning med söksträng (43): "Tränare och spelare kommer och går, men Göran Arnberg han består i Brage."

Här syns en nominalfras NP som precis som söksträngen uttrycker dominerar ett substantiv, som här utgörs av egennamnet *Göran Arnberg*. Nominalfrasen dominerar också en sats som dominerar ytterligare en nominalfras och ett verb. Denna andra nominalfras utgörs av pronomenet *han*. Substantivet *Göran Arnberg* och den andra nominalfrasen står i direkt linjär anslutning till varandra. Problemet här är att denna mening inte har vanlig svensk ordföljd, utan snarare en versartad sådan där pronomenet *han* har samma referent som *Göran Arnberg*. Meningar som dessa hittades alltså även om de inte motsvarade den sökta strukturen.

Sökresultaten kontrollerades och felaktiga resultat, såsom det avbildade i Figur 6, rensades ut manuellt. I de allra flesta fall var det inga svårigheter att se om en sats var en relativsats eller inte.

En svårighet som uppkom var däremot att avgöra huruvida relativsatser med *som* var restriktiva eller ej. Mening (44) nedan, som är hämtad från resultatet av söksträng (42), är en relativsats som är klart icke-restriktiv.

- (44) "... där han bland annat analyserar Bellmans konst med utgångspunkt från **begreppet 'humor'**, som han menar vara ett grundbegrepp..."

Den understrukna texten visar på den relevanta delen av meningen. Här är korrelatet i matrissatsen "begreppet 'humor'", vilket är vad subjunktionen *som* syftar tillbaka på. Relativsatsen är inte restriktiv eftersom den inte bidrar till identifikation av referenten *begreppet 'humor'*. Relativsatsen är här inte ett obligatoriskt element som specificerar bara matrissatsens korrelat, utan tillför snarare ny betydelse till meningen som helhet. Eftersom meningen inte är restriktiv så kan *som* inte utelämnas. Denna sorts meningar uteslöts därför.

I exempel (45), också denna mening hämtad från sökresultaten av söksträng (42), är det däremot svårare att avgöra huruvida satsen är restriktiv eller inte.

(45) Sandvik har **koncerngemensamma regler** som vi följer.

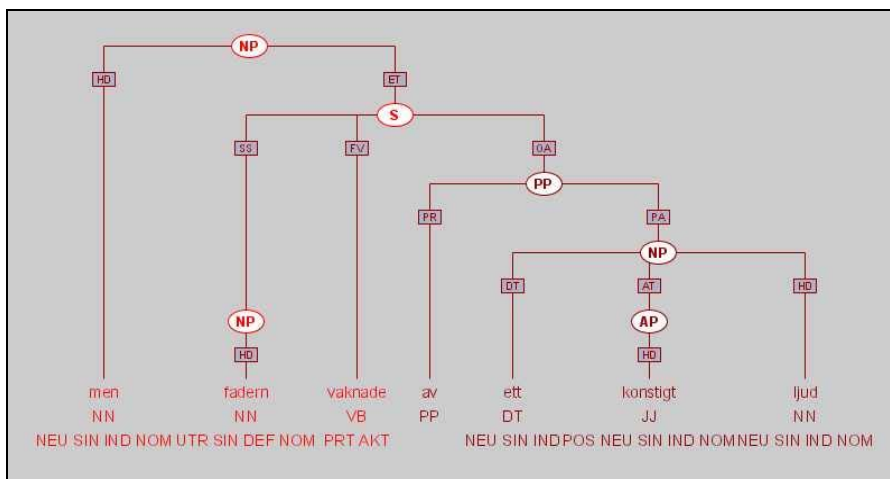
Här syftar relativsatsens *som* på de ”koncerngemensamma regler” som står i matrissatsen. Det är dock inte klart huruvida relativsatsen har en restriktiv betydelse eller ej; i vissa sammanhang skulle *som* kunna uteslutas, medan det också skulle kunna låta ogrammatiskt i ett annat sammanhang, mycket beroende på betoning. För säkerhets skull har jag valt att utesluta dessa ambigüosa meningar.

Strängarna är formulerade så att de söker efter meningar i Svensk Trädbank där en sats direkt dominerar ett verb som är en terminalnod. Detta utesluter alltså att detta verb kan domineras av en verbfras-nod. I TIGERSearch är det inte så, i kontrast till lingvistikens i övrigt, att ett verb alltid domineras av en verbfras. Verbfraser är i TIGERSearch istället en nod som dominerar mer än en enhet, i övriga fall kan ett verb domineras direkt av satsen.

Verbfraser förekommer i TIGERSearch främst i perfektkonstruktioner där verb som *har gjort* utgör en verbfras. Av denna anledning har perfektkonstruktioner inte inkluderats i sökningarna.

3.2.3.2 Korpusrelaterade problem

Eftersom den version av Svensk Trädbank som jag använt mig av är en utvärderingsversion så var det väntat att en del annoteringsfel skulle förekomma. I Figur 7 syns ett exempel på ett feltaggat träd:



Figur 7. Exempel på feltaggning i Svensk Trädbank: ”men fadern vaknade av ett konstigt ljud”.

Det första ordet i den överordnade nominalfrasen, ”men”, har taggats som ett substantiv trots att det här rör sig om en subjunktion. Fel som detta utgjorde dock ingen större del av resultaten, men är en av orsakerna till att meningar behövde rensas bort manuellt.

3.2.3.3 Svensk Trädbank gentemot andra korpusar

Sökningarna i Trädbanken med TIGERSearch lyckades inte hitta riktigt alla önskade meningar, och samtidigt utesluta alla oönskade meningar; som redan nämnts uteslöts konstruktioner med verb som stod i perfekt-tempus. Trots vissa problem med att formulera söksträngar valdes Trädbanken som den mest lämpade korpusen för undersökningen. Detta har sin grund i det faktum att korpusens trädstruktur möjliggjorde definitioner av övergripande fraser snarare än enskilda ord. En korpus som PAROLE är inte syntaktiskt annoterad, varför endast terminalnoder dvs. enheter som utgörs av ord kan specificeras, och inte som i TIGERSearch noder på en högre nivå i frasstrukturträdet. Om det är variationer i längd på själva relativsatsen som söks så blir detta ett problem, eftersom den linjära ordningen gör att varje möjligt element då måste specificeras. I en trädstruktur som i Svensk Trädbank går det att formulera en nominalfras, utan att närmare gå in på vilka terminalnoder denna fras ska dominera. Detta gör att de element som ingår i nominalfrasen kan variera på alla sätt som ord i nominalfraser i allmänhet kan variera. En annan möjlighet än variationer i subjektsnominalfrasens längd är att eventuella adverb finns mellan denna subjektsnominalfras och det finita verbet. Dessa adverbiella element som kan finnas i relativsatsen inkluderas också i mina sökningar i Svensk Trädbank, eftersom antalet element som kan förekomma mellan subjektsnominalfrasen och det finita verbet inte är specificerat.

Det var just detta faktum att syftet var att få en varierande längd på relativsatsen som gjorde att jag valde att använda mig av Svensk Trädbank i min undersökning.

3.3 Analysmetod

De meningar som korpussökningarna har fått fram har samlats i tabeller i MS Excel, varefter de manuellt har gått igenom och rensats från felaktiga resultat. Det som återstår efter denna process är en tabell där enheterna NP1, eventuellt *som*, NP2, eventuella adverbiella element, och V finns uppställda. Efter rensningen kvarstår sammanlagt 854 meningar med *som* och 1023 meningar utan *som*.

I mening (46) ges ett exempel på en relativkonstruktion från resultaten av söksträng (43) med olika frastyper utmärkta:

(46)	Den roman	<u>vi</u>	<u>just</u>	<u>läser</u>
	NP1	NP2	tidsadverbial	VB

En av faktorerna som undersöktes i detta arbete är hur många ord som skiljer matrissatsens korrelat från relativsatsens finita verb. I (46) är antalet ord som skiljer korrelatet *den roman* från det finita verbet *läser* två: pronomenet *vi* och tidsadverbialet *just*.

Listor har sammanställts med information om antalet ord som skiljer korrelatet, eller *som*, från det finita verbet i satserna. Det har skapats listor både med avseende på antalet ord totalt, antalet ord i subjekts-NP:n, och antalet ord med en adverbialfunktion. Resultatmeningarna har också taggats för huruvida subjektsnominalfrasen i relativsatsen är pronominellt eller ej, samt hur animat samma subjekt är.

Endast meningar med ett ord mellan korrelatet och relativsatsens finita verb har undersökts för förekomst av pronomen. Detta enda ord är i dessa fall subjektsnominalfrasens pronomen. Anledningen till detta är en strävan att isolera ett eventuellt pronomen som påverkande faktor för utelämnandet av *som*. Om relativsatsen med endast ett ord mellan korrelatet och relativsatsens finita verb undersöks kan det uteslutas att andra möjliga faktorer såsom längden på subjektsnominalfrasen haft påverkan för utelämnandet av *som*.

Som utgångspunkt för animacitet har Givóns (2001) animacitetshierarki använts. Den uppskattning av animacitet som gjorts följde dock inte hierarkin helt utan är snarare en grov uppskattning. Det som skilts på är animata referenter (människor), mindre animata referenter (djur), kollektiv (såsom *styrelsen* eller *regeringen*) och inanimata referenter (föremål). De tre första kategorierna har alla räknats som animata, även om de enligt Givón skulle innehålla olika positioner i animacitetshierarkin.

Resultaten har sedan analyserats för att i bästa mån kunna svara på de hypoteser som ställts upp, vilka presenterades i avsnitt 2.1.

4. Resultat

De resultat som framkommit är här uppdelade efter de hypoteser som de besvarar. Först kommer de att beskrivas mot bakgrund av Hawkins princip kring det antal ord som förekommer mellan korrelatet och relativsatsens finita verb. Därefter presenteras hur resultaten förhåller sig när det gäller förekomsten av pronomen och animata referenter i relativsatsens subjekt.

För undersökningen kring processningslättnad visas ett någorlunda linjärt samband mellan antalet ord och frekvensen av utsatta *som*. När resultatmeningar beräknas enligt Hawkins modell visar detta också på att sambandet är så gott som linjärt.

För undersökningen kring tillgänglighet hos subjektsreferenten i relativsatsen syns ett samband mellan pronominella subjekt och frekvensen av *som*. För animacitet visar resultaten också på ett

samband mellan animata subjektsreferenter och utsatta *som*, om än inte lika tydligt som för pronomina subjektsreferenter.

4.1 Antal ord mellan korrelatet och relativsatsens finita verb

En av de hypoteser som ligger till grund för denna undersökning baseras på Hawkins MaOP, och presenterades i avsnitt 2.1. Här redovisas hur resultatet av den genomförda undersökningen förhåller sig till denna hypotes.

Tabell 8 nedan visar antalet ord som skilde korrelatet från relativsatsens finita verb i resultatmeningarna.

Tabell 8. Det faktiska antal ord som i relativsats med och utan *som* skilde korrelatet från relativsatsens finita verb.

Antal ord:	1-2	3-4	5-6	7-8	9+	Summa
Antal meningar med <i>som</i>	745	80	15	9	5	854
Antal meningar utan <i>som</i>	995	22	4	2	0	1023
Summa	1740	102	19	11	5	1877
Andel med <i>som</i>	43%	78%	79%	82%	100%	

Det procenttal som anges under varje kolumn anger hur stor del av meningarna med ett visst antal ord mellan korrelatet och relativsatsens finita verb som innehåller *som*.

4.1.1 Exempel på resultatmeningar

Exempel på resultatmeningar med olika antal ord mellan korrelatet och relativsatsens finita verb med och utan *som* ges i (47) - (50) nedan. Det finita verbet har här markerats med fet understrykning.

(47) Jag vill förena vardag med **det jag gör som artist.**

Ett ord, pronomenet *jag*, skiljer här korrelatet *det* från det finita verbet *gör*.

(48) I går kunde polisen gripa ytterligare en medlem i **den liga man nu är på spåren.**

Två ord, pronomenet *man* och tidsadverbialen *nu* skiljer här korrelatet *den liga* från det finita verbet *är*.

(49) Han sticker inte under stol med **den starka oro som han själv känner inför framtiden.**

Två ord, pronomenet *han* och pronomenet *själv*, finns mellan korrelatet *den starka oro* och relativsatsens finita verb *känner*.

- (50) Skulle Shamir verkligen gå i land med detta, innebär det med all sannolikhet att han tvingas bilda **en regering** som han inte ens själv tycker om.

Fyra ord, pronomenen *han* och *själv* samt adverbialfrasen *inte ens* skiljer här korrelatet *en regering* från det finita verbet *tycker*.

Dessa exempelmeningar ger en bild av hur meningarna som hittades kunde se ut, och hur de mättes för antalet ord mellan korrelat och finit verb.

Endast fem meningar med 9 ord eller fler mellan korrelat och finit verb i relativsatsen hittades, varav alla konstruerade med *som*. Dessa meningar presenteras i (51) – (55) nedan.

- (51) "Det hemliga ljuset" är **en lyrikhändelse** som var och en med behov av ord att ta in i sitt eget öde bör unna sig.

14 ord skiljer subjunktionen *som* från det finita verbet *bör*: nominalfrasen *var och en* med tillhörande prepositionsfras *med behov av*, som i sin tur har en inbäddad verbfras *att ta in i sitt eget öde*, där *i sitt eget öde* utgör ännu en prepositionsfras.

- (52) Det var också **det rådet** som den äldre mannen i den inledande historien i kapitel 1 ville följa.

10 ord skiljer subjunktionen *som* från det finita verbet *ville*: nominalfrasen *den äldre mannen* med tillhörande prepositionsfras *i den inledande historien i kapitel 1*, där *den inledande historien* utgör ännu en nominalfras i vilken prepositionsfrasen *i kapitel 1* är inbäddad.

- (53) **Det avtal mellan de nordiska länderna om handräckning i skatteärenden** som Danmarks regering tillsammans med Färöarnas landsstyre och Grönlands landsstyre samt Finlands, Islands, Norges och Sveriges regeringar undertecknade den 7 december 1989 skall i originaltexternas lydelse gälla som lag här i landet.

16 ord skiljer subjunktionen *som* från det finita verbet *ville*: den sammanhängande nominalfrasen *Danmarks regering tillsammans med Färöarnas landsstyre och Grönlands landsstyre samt Finlands, Islands, Norges och Sveriges regeringar*.

- (54) Vi kan således än en gång konstatera **den stora betydelse för nätverksbyggandet** som linking-pin-funktionen i form av Centralen med dess personal hela tiden kom att få.

10 ord skiljer subjunktionen *som* från det finita verbet *kom*: nominalfrasen *linking-pin-funktionen i form av Centralen med dess personal*, där *i form av Centralen* och *med dess personal* utgör två inbäddade prepositionsfraser. Här finns också adverbfrasen *hela tiden* precis innan det finita verbet.

(55) "Lag" står för **det förslag som arbetsgruppen för livsmedelspolitik som departementsfolk och referensgrupp från partierna presenterade** i höstas.

9 ord skiljer subjunktionen *som* från det finita verbet *presenterade*: nominalfrasen *arbetsgruppen för livsmedelspolitik som departementsfolk och referensgrupp från partierna*. Vad meningens andra *som* har för funktion här är något oklart; möjligen är det en felskrivning som egentligen ska vara konjunktionen *och*. Huvudstrukturen i meningen förändras dock inte av detta.

Här har exempel på resultatmeningar som användes för att mäta avstånd mellan korrelat/*som* och meningens finita verb visats. Meningarna med mer än 9 ord har alla med undantag för en hela nominalfraser som skiljer *som* från finit verb. Undantaget är mening (54) där adverbfrasen *hela tiden* står före det finita verbet.

4.1.2 Hawkins teorier och matrismönster

Här visas hur undersökningens 1877 resultatmeningar förhåller sig till Hawkins (2004) MaOP och den matris för att beräkna OP-till-UP-ratio som han skapat, vilket presenterats i avsnitt 2.4.4.

Resultatmeningar från undersökningen med olika antal ord mellan korrelat/*som* och finit verb i relativsatsen sattes in i denna matris, och preferenstal räknades ut. I Tabell 9 nedan ges en sammanställning av preferenstalen för konstruktioner med olika antal ord.

Tabell 9. Sammanfattning av preferenstal för relativkonstruktioner med ett, tre, fem, sju och tio ord mellan korrelatet och det finita verbet.

Antal ord:	1.	3.	5.	7.	10.
Preferenstal:	-30	+11	+18	+40	+34

Med resultaten som redovisas i Tabell 9 bekräftas Hawkins hypotes att preferensen för att använda *som* i en konstruktion ökar gradvis allt eftersom antalet ord mellan korrelat och relativsatsens finita verb ökar.

När endast ett ord skiljer matrissatsens korrelat från relativsatsens finita verb ligger preferenstalet på minus, vilket innebär att en preferens finns för att inte använda *som*. I en konstruktion med tre ord mellan korrelat och relativsatsens finita verb är preferenstalet 11, vilket tyder på en preferens för att använda *som*. När antalet ord ökar ökar också preferenstalet, med undantag för konstruktionen med tio ord som har ett lägre tal än konstruktionen med sju ord. Bortsett från detta undantag finns en tendens att i högre grad använda *som* när fler ord skiljer matrissatsens korrelat från relativsatsens finita verb. Denna preferens kan dock inte sägas vara helt linjär, men tyder ändå på ett tydligt mönster.

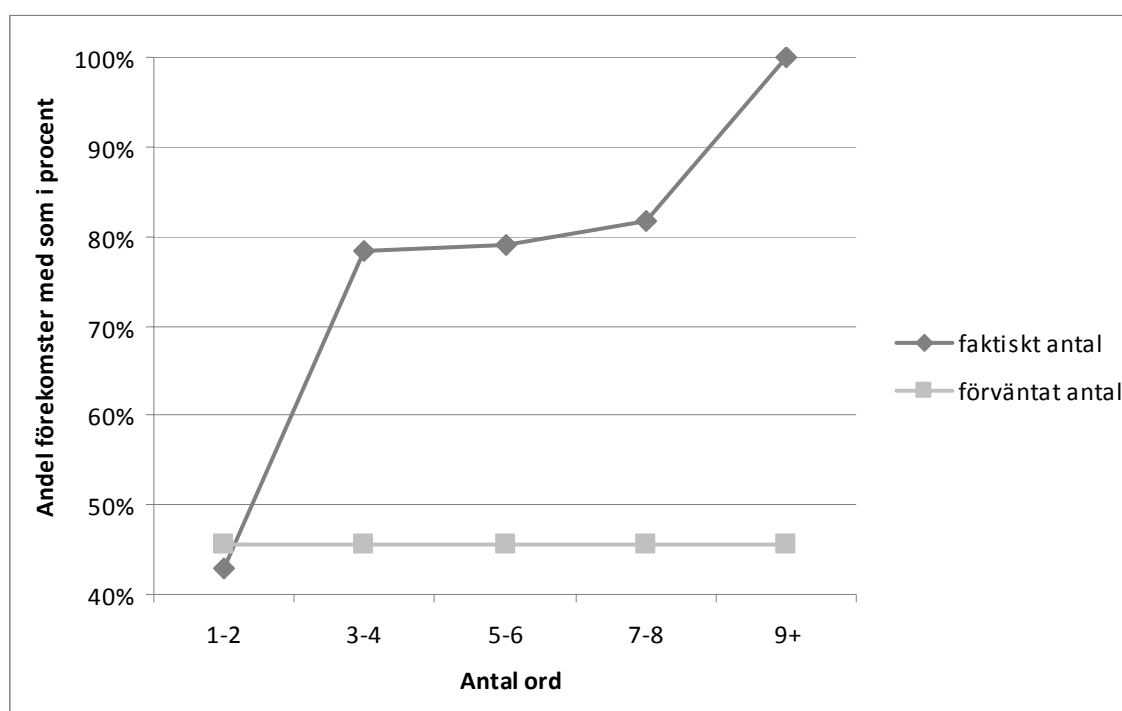
4.1.3 Chi²-test

Ett chi²-test används för att visa hur skevheten i den faktiska distributionen jämfört med hur en förväntad inte skev distribution av meningar skulle se ut. Den förväntade distributionen av antalet ord mellan korrelerat/som och relativsatsens finita verb ser ut som i Tabell 10.

Tabell 10. Det antal förväntade ord som i relativsats med och utan som skilde korrelerat från relativsatsens finita verb.

Antal ord:	1-2	3-4	5-6	7-8	9+	Summa
Antal meningar med <i>som</i>	792	46	9	5	2	854
Antal meningar utan <i>som</i>	948	56	10	6	3	1023
Summa	1740	102	19	11	5	1877
Faktiskt antal	45%	45%	45%	45%	45%	

En jämförelse mellan det faktiska antalet förekomster med *som*, som visas i Tabell 8, och det förväntade antalet förekomster med *som* ger det resultat som syns i Figur 8.



Figur 8. Jämförelse av faktiskt och förväntat antal meningar med som i meningar med olika antal ord mellan korrelerat och finit verb.

Det som visas i detta diagram är att det faktiska antalet förekomster av meningar med *som* understiger det förväntade antalet meningar med 1-2 ord som skiljer korrelerat från relativsatsens finita verb. När antalet ord blir högre än två så överstiger det faktiska antalet meningar däremot det förväntade antalet. Detta visar på en skevhet i resultatet som verkar vara för stor för att bero på slumpen. Ett chi²-test visar att så också är fallet:

$$\chi^2(4) = 70,05, p > 0.0001$$

Detta innebär att Hawkins hypotes visat sig sannolik; eftersom skevheten inte kan bero på slumpen måste det finnas andra orsaker som påverkar utelämnandet av *som*, och då ger Hawkins MaOP en rimlig förklaring.

Den första hypotes som här undersökts har alltså blivit bekräftad.

4.2 Tillgänglighet

Enligt den andra hypotes som denna undersökning utgått från så kan en faktor som påverkar utelämnandet av *som* vara tillgängligheten hos relativsatsens subjekt. Denna hypotes har redovisats närmare i avsnitt 2.2.2 ovan.

4.2.1 Pronomen

Här visas hur resultatmeningarna förhåller sig till den hypotes om tillgänglighet utifrån pronominell status hos nominalfrasens subjekt som formulerats i avsnitt 1.2.

Tabell 11 visar det antal meningar som hade ett pronomen som subjektsnominalfras i relativsatsen. Endast meningar med ett ord mellan korrelatet och relativsatsens finita verb har undersökts.

Tabell 11. Antalet faktiska meningar med pronomen respektive lexikal enhet som huvudord i relativsatsens subjektsnominalfras.

Faktisk distribution				
	Pronomen	Lexikal	Summa	Andel pronomen
Med <i>som</i>	278	216	494	56%
Utan <i>som</i>	864	79	943	92%
Summa	1142	295	1437	
Andel med <i>som</i>	24%	73%		

Det procentuella värde som visas längst ner i varje kolumn anger den andel av meningarna som har *som*. Tabellen visar att det i meningar med pronomen i relativsatsens subjekt är 24% av meningarna som har ett *som* utsatt. I meningar som inte har pronomen i relativsatsens subjekt är motsvarande andel 73%. Kolumnen längst till höger visar också att 56% av meningarna med *som* som hittades hade ett pronomen som subjekt i relativsatsen. I meningarna utan *som* var det 92% som hade ett pronomen som subjekt.

Detta bekräftar hypotesen att *som* oftare utelämnas när relativsatsens subjekt är ett pronomen.

4.2.1.1 Exempel på resultatmeningar

I (56) – (58) nedan ges exempel på meningar med olika typer av enheter i relativsatsens subjekt. Här har relativsatsens subjekt markerats med fet understrykning.

(56) De hade **en AMU-utbildning till kylmontör** som jag var mycket intresserad av.

Relativsatsens subjekt utgörs av det personliga pronomenet *jag*, som enligt Ariels tillgänglighetshierarki är den mest tillgängliga enheten.

(57) Toppolitikerna och toppfunktionärerna levde avskilda från folket på **ett sätt** som Wetzl menar gjorde dem främmande för verkligheten.

Här är det egennamnet *Wetzl* som utgör relativsatsens subjekt, vilket enligt tillgänglighetshierarkin är en mindre tillgänglig enhet än pronomen.

(58) Målsättningen har tillkommit utifrån **den verklighet** som sjukvårdsdistriktet har att arbeta i.

I denna mening är relativsatsens subjekt substantivet *sjukvårdarna*, vilket inte är ett pronomen och därför enligt Ariels (1990) tillgänglighetshierarki mindre tillgängligt än pronomen.

4.2.1.2 Chi²-test

I Tabell 12 visas den förväntade distributionen av pronomen i subjektsnominalfrasen i meningar med och utan *som*.

Tabell 12. Antalet förväntade meningar med pronomen eller lexikal enhet som huvudord i relativsatsens subjektsnominalfras.

Förväntad distribution				
	Pronomen	Lexikal	Summa	Andel pronomen
Med <i>som</i>	393	101	494	79%
Utan <i>som</i>	749	194	943	79%
Summa	1142	295	1437	
Andel med <i>som</i>	34%	34%		

I Tabell 12 visas att 34% av det totala antalet meningar med pronomen i subjektsposition kunde väntas vara konstruerade med *som* om fördelningen varit rent slumpmässig, oberoende av hur många ord som fanns mellan korrelatet och det finala verbet i relativsatsen.

Också för pronomen har ett chi²-test gjorts för att undersöka om den snedfördelning som finns mellan meningar med och utan *som* sannolikt beror på slumpen.

$$\chi^2(2) = 282,95 \quad p > 0.0001$$

För relativsatser med pronomen som subjekt är alltså den skevhet i distribution mellan konstruktioner med och utan *som* för stor för att vara slumpmässigt uppkommen. Detta implicerar ett samband mellan användandet av pronomen och utelämnandet av *som*; också hypotes 2 har alltså bekräftats med avseende på pronominellt subjekt.

4.2.2 Animacitet

Resultaten undersöktes också för animacitetshierarkin som definierad av Givón (2001), dock med vissa anpassningar (se avsnitt 3.3), där utgångspunkten var att animacitet har betydelse för huruvida *som* kan utelämnas eller ej. Här visas hur resultatet förhåller sig till denna hypotes.

I Tabell 13 visas antalet meningar där relativsatsens subjektsnominalfras var mer eller mindre animat, i motsats till att inte vara animat alls.

Tabell 13. Antal faktiska meningar med animat referent som subjektsnominalfrasen i relativsatsen.

Faktisk distribution				
	Animat	Inanimat	Summa	Andel animata subjekt
Med <i>som</i>	666	188	854	78%
Utan <i>som</i>	902	121	1023	88%
Summa	1568	309	1877	
Andel med <i>som</i>	42%	61%		

Tabellen visar att 42% av meningarna med animata subjektsreferenter konstrueras med *som*. Bland meningarna med inanimata subjektsreferenter var det 61% som hade *som* utsatt. Tabellen visar också att det av de resultatmeningar som hade *som* utsatt var 78% som hade animat referent. I meningarna utan *som* var motsvarande andel 88%.

Resultaten visar att det i resultatmeningarna fanns en viss preferens för att inte använda *som* när relativsatsens subjekt hade en animat referent, vilket innebär att denna del av undersökningens andra hypotes kan anses delvis bekräftad.

4.2.2.1 Exempel på resultatmeningar

De exempel som ges här visar att variation förekommer i huruvida konstruktioner med och utan *som* har en animat subjektsreferent. Exempelmeningar med relativsatser vars subjektsreferent är animat eller inanimat visas i (59) – (61). I dessa exempel har relativsatssubjektet strukits under med fet understrykning.

- (59) Och för en månad sedan fick hon beskedet att hon och Sveriges Television får **den eftertraktade första intervjun som kronprinsessan ger.**

I denna mening, som är konstruerad med *som*, är relativsatsens subjektsreferent den i allra högsta grad mänskliga och därmed animata *kronprinsessan*.

- (60) **Det regionen framför allt vill satsa på är högteknologi**, eftersom man har lång erfarenhet av att arbeta med ko, stål och kemiska produkter.

Denna konstruktion utan *som* har det kollektiva subjektsreferenten *regionen*, som har räknats som animat.

- (61) Avsikten är att **under den tid förbifarten ej har byggts ut i sin helhet** skall delen Drumstorp - väg 1008 användas som industrigata.

Här har relativsatsens, som konstruerats utan subjunktionen *som*, den inanimata referenten *förbifarten* som subjekt. Värt att notera är att verbet *byggts* står i passiv form, vilket innebär att *förbifarten* i en icke-passiv konstruktion står som objekt, och inte subjekt. Också detta exempel kan alltså tolkas som att subjekt tenderar att vara animata.

4.2.2.2 Chi²-test

För att kunna avgöra om denna preferens är signifikant uppskattas först de förväntade värdena för distribution av *som* i relativsatsen med animat eller inanimat subjektsreferent.

Tabell 14. Antalet förväntade meningar med animat referent som subjektsnominalfras i relativsatsen.

Förväntad distribution				
	Animat subjekt	Inanimat subjekt	Summa	Andel animata subjekt
Med <i>som</i>	713	141	854	84%
Utan <i>som</i>	855	168	1023	84%
Summa	1568	309	1877	
Andel med <i>som</i>	45%	45%		

Den förväntade distributionen av *som* i meningar med animata och inanimata subjektsreferenter, med utgångspunkt i en slumpmässig distribution, visas i Tabell 14. Denna tabell åskådliggör att 45% av antalet meningar med animat referent kan förväntas vara konstruerade med *som*. Ett chi²-test ger följande resultat:

$$\chi^2(2) = 62,47 \quad p > 0.0001$$

Enligt detta test finns en viss preferens för att utesluta *som* när relativsatsens subjekt har en animat referent. Utifrån detta kan även hypotesen gällande animacitet anses vara bekräftad, om än inte lika starkt som för pronomen.

5. Diskussion

Jag har i denna uppsats undersökt förekomsten av *som* i objektsrelativsatser i svenska. Jag har utgått från två hypoteser som presenterats i avsnitt 2.2. Detta beror på att pronomen och animata referenter i högre grad är tillgängliga.

De resultat som mina analyser av 1877 träffar i Svensk Trädbank har gett visar att båda hypoteser har bekräftats, om än i olika hög grad.

5.1 Processingeffektivitet

När resultaten undersökts med avseende på processingeffektivitet med utgångspunkt från Hawkins matris (2004), så bekräftades den första hypotesen: subjunktionen *som* förekom oftare i meningar med många ord mellan matrissatsens korrelat och relativsatsens finita verb. Denna teori tycks alltså stämma även för svenska relativsatser.

De 1877 meningar som sammanlagt hittades visar visserligen på denna preferens, men med en större databas skulle en mer omfattande undersökning kunna genomföras. Resultaten visar på tendenser, vilka dock skulle kunna stärkas i en större undersökning. I denna undersökning hittades endast 5 meningar med 10 och fler ord mellan korrelat och relativsatsens finita verb. En mer omfattande undersökning skulle kunna ge säkrare resultat för meningar med många ord mellan korrelat och finit verb.

5.2 Tillgänglighet

Den andra hypotesen bekräftades i hög grad för pronomina referenter. Tillgänglighetshierarkin som formulerad av Ariel ger en sannolik förklaring till varför *som* är mindre frekvent i relativsatser med pronomint subjekt. Meningarna undersöktes visserligen inte för hela skalan utan endast för pronomen, men hypotesen kan ändå anses bekräftad för distinktionen mellan pronomina och lexikala referenter. I framtida forskning skulle också större distinktioner mellan olika nivåer på skalan kunna undersökas. En större uppdelning av skalan skulle kunna visa på ett tydligare linjärt samband, om ett sådant finns att visa, vilket skulle stärka hypotesen om tillgänglighet enligt tillgänglighetshierarkin ytterligare.

Med avseende på animacitet var resultatet likartat; det visade på en viss preferens att använda *som* där relativsatsens subjektsreferent är inanimat, men här pekade resultatet inte lika entydigt på denna preferens. Detta skulle kunna bero på ett antal faktorer. Dels är det i svenskan, och

därmed också i resultatmeningarna, vanligt med konstruktionen ”det (som) jag gör/har/är” etc. Detta *det* är inanimat, men förekommer ofta utan *som*. Detta skulle kunna bero på att konstruktionen är så pass vanlig att den kan förstås som en relativsats även utan *som*; den skulle då helt enkelt förstås som en helhet utan att orden behöver parsas var för sig. Eftersom denna konstruktion är vanlig utan *som*, innebär det att det i resultatmeningarna återfinns flera meningar som konstruerats utan *som*, och där subjektsreferenten (*det*) är inanimat. Därmed ökar andelen konstruktioner utan *som* där subjektsreferenten är inanimat, vilket ju går emot hypotesen. Detta skulle alltså kunna vara en bidragande faktor till varför hypotesen inte bekräftats. En undersökning där detta tas med i beräkningen och konstruktioner av denna typ sorteras bort vore hjälpsamt för att tydligare kunna se samband mellan animacitet och närvaro av *som*.

Även här skulle alltså en mer omfattande och detaljerad undersökning kunna genomföras, som tydligare visar på samband mellan animacitet och utelämnande av *som*. I en undersökning där också fler nivåer i animacitetshierarkin tas med skulle en tydligare bild av detta samband kunna ges.

För att undersöka den andra hypotesen skulle en studie gjord på talspråk vara av intresse. Eftersom dessa berör tillgänglighet utan att, som Hawkins, fokusera på skriftspråket så är det relevant att se på tillgängligheten också i vardagligt tal, eftersom det blir ännu tydligare hur tillgängligt en referent är när denna måste hittas genast, utan betänketid. Det är inte omöjligt att resultaten då skulle bli något annorlunda eftersom det i talspråk finns fenomen som intonation och gestikulation som kan underlätta för en mening att blir riktigt tolkad. Möjligen skulle det alltså kunna vara ännu lättare att utesluta *som* i talspråk, eftersom annat än bara ordföljd då hjälper till att förklara sammanhanget. För att bekräfta detta krävs dock som sagt en ny undersökning, gjord på talspråk.

5.3 Korpusstudie

Gällande den korpusundersökning som genomförts kan sägas att den har visat sig genomförbar och framgångsrik. Att utföra en studie av svenska relativsats i Svensk Trädbank 1.0 har varit både effektivt och resultatrikt, även om en mer omfattande korpus som underlag skulle kunna ge tydligare resultat. Mer förfinade söksträngar skulle också kunna bidra till en mer heltäckande undersökning.

6. Sammanfattning

Denna uppsats har behandlat möjligheten att utesluta subjunktionen som i en svensk objektsrelativsats. Sökningar har gjorts i en svenskspråkig korpus varifrån resultat har sammanställts och analyserats, med avseende på två faktorer: antal ord mellan korrelat och relativsatsens finita verb, och tillgänglighet.

För undersökningen kring antal ord mellan matrissatsens korrelat och relativsatsens finita verb syntes en linjär fördelning av hur stor andel meningar med ett visst antal ord som var konstruerade med *som*. Tabell 8 visade att meningar med ett till två ord mellan korrelat och relativsatsens finita verb i något högre grad var konstruerade utan *som*; endast 43% hade *som* utsatt. I konstruktioner med tre till fyra ord mellan matrissatsens korrelat och det finita verbet i relativsatsen var 78% konstruerade med *som*, vilket är en övervägande del. Andelen relativsatser som var konstruerade med *som* ökade sedan gradvis, för att i meningar med nio ord och mer mellan korrelat och finit verb i relativsatsen vara 100%; här hittades alltså inga meningar som inte hade *som* utsatt.

När meningar sattes in i den av Hawkins konstruerade matris som syftar till att tydliggöra hur preferenser för användande av *som* ser ut, så bekräftade resultaten den tendens som visats i Tabell 8. Resultaten, som visats i Tabell 10, visade också de på en ökande preferens för användande av *som* där fler ord skiljde matrissatsens korrelat från relativsatsens finita verb.

Ett chi²-test genomfördes för att kontrollera huruvida den skevhet i distributionen mellan meningar med och utan *som* kunde bero på slumpen. Chi²-testet visar att denna skevhet är för stor för att bero på slumpen. Om så inte är fallet så kan andra faktorer antas ligga bakom de olikheter som finns i Tabell 8. Den förklaring som antalet längd på avståndet mellan korrelat och finit verb ger kan anses vara godtagbar, med tanke på den linjäritet som resultaten både i Tabell 8 och 10 uppvisar.

Med detta kan Hawkins teori anses vara bekräftad. Min undersökning har visat att det finns stöd för dennes antaganden kring processning och effektivitet, med avseende på svenska objektsrelativsatser.

Meningarna undersöktes också för tillgänglighet. En referent ansågs i hög grad tillgänglig om den var pronominell och animat, och den konstituent som undersöktes var det inbäddade subjektet. Om subjektsreferenten var i hög grad tillgänglig ansågs det sannolikare att *som* inte skulle förekomma.

Denna hypotes bekräftades i hög grad för pronominella referenter. Tabell 11 visar att endast 24% av relativsatserna med pronominellt subjekt var konstruerade med *som*, gentemot 73% i

relativsatser med lexikalt subjekt. Denna skevhet testades också mot ett chi²-test, som bekräftade att snedfördelningen är för stor för att vara slumpmässigt framkallad. Istället ger tillgänglighetshierarkin som formulerad av Ariel en sannolik förklaring till varför *som* är mindre frekvent i relativsatser med pronominellt subjekt. Meningarna undersöktes visserligen inte för hela skalan utan endast för pronomen, men hypotesen kan ändå anses bekräftad för distinktionen mellan pronominella och lexikala referenter.

Med avseende på animacitet var resultatet likartat. Skillnaden i fördelning mellan meningar med och utan *som* är här inte lika stor som med avseende på pronomen. För meningar med animat subjektsreferent var det något mindre än hälften, 42%, som var konstruerade med *som*. För meningar med inanimat subjektsreferent var det istället något mer än hälften, 61%, som var konstruerade med *som*. Detta tyder på en viss preferens att använda *som* där relativsatsens subjektsreferent är animat än när referenten är inanimat. Det chi²-test som gjordes för animacitetsresultaten visade på en viss signifikans för detta, dock inte lika stark som för pronominella referenter.

Sammanfattningsvis kan sägas att resultatmeningarna gav bevis framförallt för hypotesen att *som* lättare kan utelämnas om relativsatsens subjektsreferent är pronominell. Ett visst stöd tycks också finnas för att *som* lättare kan utelämnas när subjektsreferenten är animat. Detta ger i sin tur stöd åt teorin att pronominella referenter är mer tillgängliga än lexikala, samt till viss del att animata referenter är mer tillgängliga än inanimata.

Meningarna har undersökts för processningseffektivitet och tillgänglighet. Både Hawkins och Wasow & Jaegers teorier har styrkts: både antalet ord mellan matrissatsens korrelat och relativsatsens finita verb, samt huruvida relativsatsens subjektsreferent är pronominell eller inte syns i denna undersökning vara av betydelse för huruvida *som* kan utelämnas eller ej. Dessa teorier är förenliga eftersom ett pronomen, som enligt Wasow & Jaeger ökar möjligheten att utelämnas *som*, är enordigt, och Hawkins teori förutspår att få ord mellan korrelat och relativsatsens finita verb ökar sannolikheten för utelämnande av *som*.

Resultaten har för dessa hypoteser visat sig vara signifikanta med avseende på ett chi²-test, medan resultaten för animacitet var något mindre signifikanta.

Sammanfattningsvis kan sägas att stöd för Hawkins princip MaOP har hittats i svenska objektsrelativsatser. På liknande sätt har det visat sig troligt att teorier kring tillgänglighet med avseende på pronomen kan bekräftas för svenska objektsrelativsatser.

Denna undersökning har syftat till att undersöka om de svenska relativkonstruktioner, som exempelmening (3), som intuitivt uppfattas som klumpiga också är ovanligare. Resultatet av undersökningen stödjer denna tes, eftersom jag fann att majoriteten av relativsatser med mer än sju ord konstrueras med *som*. I relativsatser med mer än åtta ord är konstruktioner utan *som* helt

obefintliga. Dessutom har resultat hittats som visar på signifikans för huruvida det inbäddade subjektet är ett pronomen eller ej.

Gällande den korpusundersökning som genomförts kan sägas att den har visat sig genomförbar och framgångsrik. Att utföra en studie av svenska relativsatser i Svensk Trädbank 1.0 har varit både effektivt och resultatrikt.

7. Referenser

7.1 Tryckta källor

- Ariel M., 1990. *Accessing noun-phrase antecedents*. London: Routledge
- Bock J.K. & Warren R.K., 1985. Conceptual accessibility and syntactic structure in sentence formulation. *Cognition*, 21: 47-67
- Chomsky N., 1975. *Reflections on Language*. New York: Pantheon Books
- Chomsky N., 1995. *The Minimalist Program*. Cambridge, MA: The MIT Press
- Diessel H., 2007. Frequency effects in language acquisition, language use, and diachronic change. *New Ideas in Psychology*. 25: 108-127
- Hawkins J.A., 2002. Symmetries and asymmetries: their grammar, typology and parsing. *Linguistics*, 28: 95-149
- Hawkins J.A., 2004. *Efficiency and Complexity in Grammars*. New York: Oxford University Press Inc.
- Hawkins J.A., 2007. Processing typology and why psychologists need to know about it. *New Ideas in Psychology*, 25 (2): 87-107
- König E., Lezius W., Voorman H., 2003. *TIGERSearch 2.1 User's manual*. Stuttgart: University of Stuttgart.
- Prat-Sala M. & Branigan H.P., 2000. Discourse Constraints on Syntactic Processing in Language Production: A Cross-Linguistic Study in English and Spanish. *Journal of Memory and Language*, 42: 168-182
- Race D.S. & MacDonald M.C., 2003. The use of "that" in the Production and Comprehension of Object Relative Clauses. *26th Annual Meeting of the Cognitive Science Society*
- Svenska Akademiens grammatik, 1999b. *Volym 4: Satser och meningar*. Red Ulf Teleman m fl. Stockholm: Norstedts Ordbok
- Temperley D., 2003. Ambiguity Avoidance in English Relative Clauses. *Language*, 79: 464-84
- Tomasello M., 2003. *Constructing a Language: A Usage-Based Theory of Language Acquisition*. Harvard University Press.

van Gompel R.P.G. & Pickering M.J., 2007. Syntactic parsing. *The Oxford Handbook of Psycholinguistics*. Ed. Gaskell G. New York: Oxford University Press Inc.

Wasow T., Jaeger F. & Orr D., kommande

Wasow T., & Jaeger F., 2008. Processing as a Source of Accessibility Effects on Variation. *Proceedings of the 31st BLS*. 169-180

7.2 Elektroniska källor

Swedish Treebank. [Online] Tillgänglig på: <http://spraakbanken.gu.se/stb/om/eng/index.html>

[Besökt 09-05-27]

TIGERSearch [Online] Tillgänglig på <http://www.ims.uni-stuttgart.de/projekte/TIGER/TIGERSearch>

[Besökt 09-05-27]

8. Appendix

SUC Morphological Features		
Tag	Value	Feature
AKT	Active	Voice
DEF	Definite	Definiteness
GEN	Genitive	Case
IND	Indefinite	Definiteness
INF	Infinitive	Verb form
IMP	Imperative	Verb form
KOM	Comparative	Degree
KON	Subjunctive	Mood
NEU	Neutre	Gender
NOM	Nominative	Case
MAS	Masculine	Gender
OBJ	Object	Pronoun form
PLU	Plural	Number
POS	Positive	Degree
PRF	Perfect	Perfect form
PRT	Preterite	Verb form
PRS	Present	Verb form
SFO	S-form	Voice
SIN	Singular	Number
SMS	Compound	Case
SUB	Subject	Pronoun form
SUP	Supinum	Verb form
SUV	Superlative	Degree
UTR	Common	Gender
	Supplementary Tag	
AN	Abbreviation	

	SUC Part-of-Speech Categories	
Tag	Meaning	
AB	Adverb	
DT	Determiner	
HA	WH-adverb	
HD	WH-determiner	
HP	WH-pronoun	
HS	WH-possessive	
IE	Infinitival marker	
IN	Interjection	
JJ	Adjective	
KN	Coordinating conjunction	
NN	Noun	
PC	Participle	
PL	Particle	
PM	Proper Noun	
PN	Pronoun	
PP	Preposition	
PS	Possessive pronoun	
RG	Cardinal number	
RO	Ordinal number	
SN	Subordinating conjunction	
VB	Verb	
	Supplementary Tag	
UO	Foreign word	
	Delimiters	
MAD	Major delimiter	
MID	Minor delimiter	
PAD	Pairwise delimiter	
Tag	Meaning	
AP	Adjective phrase	
AVP	Adverb phrase	
HD	Multi-word head	
NP	Noun phrase	
PP	Prepositional phrase	

ROOT	Root node	
S	Sentence/Clause	
VP	Verb phrase	
XP	Other phrase	

Stockholms universitet
106 91 Stockholm
Telefon: 08-16 20 00
www.su.se



Stockholms
universitet