



UPPSALA
UNIVERSITET

Institutionen för lingvistik och filologi
Språkteknologiprogrammet
Examensarbete i datorlingvistik

9 mars 2006

Valensbeskrivningar av högfrekventa svenska verbalsubstantiv

Ylva Erixon

Handledare:
Eva Ejerhed, Uppsala universitet
Anna Sågvall Hein, Uppsala universitet

Sammandrag

Syftet med detta examensarbete är att ta fram underlag för valensbeskrivningar av ett fyrtiotal svenska högfrekventa verbalsubstantiv. Underlaget består i data om vilka och hur många prepositioner som förekommer i direkt efterföljande led till substantivet. I arbetet ingår att ta fram högfrekventa verbalsubstantiv med efterföljande led från en korpus (Scarriekorpusen) utifrån ett antal suffix

Resultatet visas i form av tabeller för hur vanligt det är att det efterföljande leDET är en preposition, vilka prepositionerna är samt hur frekvent de enskilda prepositionerna förekommer. I arbetet ingår också en jämförelse med verbens valens samt en jämförelse med data från en ordbok (Nationalencyklopedins ordbok).

Uppsatsen avslutas med en fördjupning av lemmat *utveckling* med genomgång av de olika prepositionerna m.a.p om de är valensled eller inte samt vilken relation som verkar råda mellan substantivet och prepositionsfrasen.

Innehåll

Sammandrag	ii
Tack	vi
1 Inledning	1
1.1 Syfte	1
1.2 Avgränsningar	2
1.3 Disposition	2
2 Lingvistisk bakgrund	3
2.1 Verbalsubstantiv	3
2.2 Valens	5
2.2.1 Substantivvalens	5
2.2.1.1 Relation till verbet	6
3 Språkteknologisk bakgrund	7
3.1 Scarrie	7
3.2 Prepositionsfrasanknytning	8
3.3 ComLex	8
3.4 Prepositionsfrasklassificering	9
3.5 Valensmodell för engelska substantiv	10
3.6 Parole	10
3.7 Sammanfattning	11
4 Metod	12
4.1 Urval av verbalsubstantiv	12
4.2 Analysmetod av verbalsubstantiven	13
5 Övergripande resultat och analys	16
5.1 Andel prepositioner i högerkontexten	16
5.2 Vilka prepositioner som finns i högerkontexten	16
5.3 Koppling till verben	17
5.4 Jämförelse med NEO	21
5.5 Slutsatser	26
6 Djupare analys	28
6.1 Inledning	28
6.2 Syfte och metod	28
6.3 Resultat - översikt	29
6.4 Utveckling av	29

6.5	<i>Utveckling i</i>	30
6.6	<i>Utveckling på</i>	31
6.7	<i>Utveckling för</i>	31
6.8	<i>Utveckling inom</i>	32
6.9	<i>Utveckling under</i>	32
6.10	<i>Utveckling mot</i>	33
6.11	<i>Utveckling från</i>	33
6.12	<i>Utveckling med</i>	34
6.13	<i>Utveckling till</i>	34
6.14	Sammanfattning	35
7	Avslutning	36
7.1	Något om de övriga substantivens valens	36
7.2	Sammanfattning	36
7.3	Reflektioner	37
7.4	Vidareutveckling	37
	Litteraturförteckning	39

Figurer

2.1	Exempel från avledningslexikonet <i>derinfl.lexicon</i> för avledningsändel- sen <i>-ande</i>	4
4.1	Exempel på konkordans för <i>utveckling</i>	14

Tabeller

2.1	De 17 substantivändelser från SUC 1.0-korpusen som har en frekvens på över 1000	4
3.1	Exempel på en ordformsingång för lemmat <i>utveckling</i> i <i>ScarrieLex</i>	7
4.1	De 41 st mest frekventa verbalsubstantiven på <i>-ing/-ning, -are, -ande/-ende</i>	13
4.2	Statistik för <i>utveckling</i> och dess efterföljande prepositioner	15
5.1	Verbalsubstantiv med procentandel prepositioner i högerkontexten	17
5.2	Verbalsubstantiv med respektive dominerande prepositioner	18
5.3	Verbalsubstantiv med efterföljande prepositioner som utgör minst 1.00 % av allt som kommer i omedelbar högerkontext	19
5.4	Relativt god överensstämmelse mellan prepositionerna i NEO och <i>Scarrie</i>	22
5.5	NEO har prepositioner som inte <i>Scarrie</i> har	22
5.6	<i>Scarrie</i> har prepositioner som inte NEO har.	24
5.7	NEO har prepositioner som inte <i>Scarrie</i> har och <i>Scarrie</i> har prepositioner som inte NEO har	25
5.8	<i>Scarrie</i> s prepositioner för de ord som saknas i NEO's lista över substan- tivvalenser	26
6.1	Resultat av hur prepositionerna fördelade sig för substantivet <i>utveckling</i>	29

Tack

Jag vill tacka mina handledare Eva Ejerhed och Anna Sågvall Hein för all hjälp de gett mig under arbetets gång. Tack för tips, idéer, genomläsningar, lösningar och övrig handledning. Jag vill tacka främst Per Starbäck men även alla andra som hjälpt mig med de tekniska bitarna såsom latex-problem och andra datorproblem. Jag vill även tacka Sven-Göran Malmgren för att jag fick ta del av hans lista på valenser i Nationalencyklopedins ordbok. Tack även alla som läst igenom någon eller några versioner av uppsatsen och gett feedback.

Till sist vill jag tacka alla i min närhet som stöttat och trott på mig under hela arbetsprocessen.

1 Inledning

Substantivet kan enligt Liljestränd (1993) ”ses som den viktigaste ordklassen i språket” (Liljestränd, 1993, s. 69). Olika substantiv är i olika stor grad benägna att knyta efterställda attribut till sig, och verbalsubstantiv är enligt Malmgren (1994) en av de grupper som är mest benägna att göra det. Exempel på verbalsubstantiv med efterföljande prepositionsfras:

- (1) Det sker en ständig **utveckling** av nya produkter
- (2) Det pågick en **utredning** om det ökade våldet
- (3) Man flyttade **produktionen** till Örebro

Prepositionsfrasen i (1) och (2) är attribut till substantivet medan den i (3) är en bestämning till verbet. De led som är attribut till substantivet kallas substantivets valensled. Om man lägger in beskrivningar av vilka attribut ett substantiv kan ta som valensled i ett datamaskinellt lexikon kan man ta ställning till om en PP kan föras till ett substantiv eller inte. Man kan alltså finna en lösning på det vanliga PP-attachmentproblemet¹. Dessa valensbeskrivningar kan sedan exempelvis användas i en parser i ett maskinöversättningssystem så att man får fler korrekta översättningar. Prepositioner är något av det svåraste att hantera i maskinöversättningssammanhang. Om man använder en defaultöversättning av en preposition kan man råka ut för felöversättningar. T.ex. så kan

ledaren för delegationen
felaktigt översättas till
the leader for the delegation
i stället för
the leader of the delegation.

Om man i stället beskriver substantivet och den efterställda prepositionsfrasen som en relation kan man sedan översätta relationen och på så vis få rätt preposition. T.ex. så är *ledaren för delegationen* en objektsrelation och likaså *the leader of the delegation*.

1.1 Syfte

Syftet med detta arbete är att ta fram underlag för valensbeskrivningar av ett antal högfrekventa verbalsubstantiv. Underlaget ska bestå i data om vilka verbalsubstantiv som är vanligast förekommande samt vilka prepositioner de kan följas av.

Valensbeskrivningarna ska sedan tala om vilken relation som råder mellan verbalsubstantivet och dess direkt efterföljande prepositionsfras samt vilken eller vilka

¹se avsnitt 3.2 samt t.ex. *The Semantic Score Approach to the Disambiguation of PP Attachment Problem*, Chao-Lin Liu, Jing-Shin Chang and Keh-Yih Su, Department of Electrical Engineering, National Tsing Hua University, Hsinchu, Taiwan, R.O.C

prepositioner som kan ingå i prepositionsfrasen. Exempel på relationer är objekt och lokation. I arbetet ingår en fördjupning av ett av verbalsubstantiven där data studeras mer noggrant och förslag på relationer ges.

Syftet med valensbeskrivningarna är att de ska kunna implementeras som regler i en parser där regelanropen finns i ett datamaskinellt lexikon för varje enskilt lexem. Analyserna som genereras av reglerna ska sedan kunna tillämpas i översättningssyfte.

I arbetets syfte ingår även en jämförelse mellan substantivens och verbens valens för att se vilka samband som finns mellan dem. Även en jämförelse med data från en ordbok ingår i syfte att se vad man kan vinna på att ta ut data från en korpus i stället för att endast använda sig av en ordbok.

1.2 Avgränsningar

Den första avgränsning jag gör är att välja verbalsubstantiv framför övriga substantiv. Detta görs för att begränsa antalet förekomster i undersökningsmängden. Man kan även utgå från verbens valens när man beskriver verbalsubstantivets valens. (Det görs en jämförelse med verben i avsnitt 5.3.)

Nästa avgränsning jag gör är att välja Scarriekorpusen som källa till mitt material. Mer om Scarriekorpusen finns att läsa i kapitel 3.

Verbalsubstantiven kan följas av annat än prepositionsfraser, t.ex. olika typer av bisatser. De förekommer dock i mindre utsträckning och det är svårare att få fram data om dessa och därför tar jag inte med dem.

En annan avgränsning som jag gör är att endast titta på ett fåtal substantivbildande suffix när orden väljs ut. Jag väljer ut tre suffix: *-ing/-ning*, *-ande/-ende* och *-are*. Mer om detta finns att läsa i kapitel 4.

När orden väljs ut så sammanför jag alla böjningsformer till ett lemma. Jag analyserar inte orden på lexemnivå.

Jag gör ingen skillnad på nomen agentis och nomen actionis vad gäller vilka prepositioner de efterföljs av.

Jag gör en noggrann analys av ett av orden och gör en mera generell analys av de andra på grund av det omfattande materialet samt begränsningen av tid.

När analyserna av substantiven och prepositionerna görs så går jag endast igenom en delmängd av alla förekomster på grund av det stora antalet, se vidare kapitel 6.

1.3 Disposition

I kapitel 2 ges en lingvistisk bakgrund till arbetet med betoning på verbalsubstantiv och valens. I kapitel 3 ges en språkteknologisk bakgrund. Här beskrivs något om vad som tidigare gjorts inom ramen för substantivvalens. I kapitel 4 beskriver jag hur jag har gått tillväga för att få fram valensbeskrivningarna. I kapitel 5 redogör jag för de data jag kommit fram till genom att ta ut statistik för 41 st verbalsubstantiv. Här gör jag även en jämförelse med verben och data från en ordbok. I kapitel 6 gör jag en djupare analys av ett verbalsubstantiv. I kapitel 7 görs en generalisering av de övriga verbalsubstantivens valensrelationer. Kapitlet innehåller även en sammanfattning, reflektioner, samt idéer till hur man kan bygga vidare på de resultat jag har kommit fram till.

2 Lingvistisk bakgrund

Jag kommer i detta kapitel att ta upp definitioner på verbalsubstantiv och valens.

2.1 Verbalsubstantiv

Verbalsubstantiv är substantiv avledda av verb. De flesta bildas genom att man lägger till ett suffix till stammen, t.ex. *utbildning* från *utbilda*. De kan också bildas utan suffix, så kallat nollsuffix, t.ex. *bad* från *bada*.

Verbalsubstantiven kan indelas i två grupper: **nomen actionis**, som betecknar aktionen/handlingen, t.ex. *simning*, och **nomen agentis** som betecknar aktören, t.ex. *simmare*. Nomen agentis kan också beteckna bland annat ett instrument och nomen actionis kan också beteckna bland annat kollektiv agent, föremål, plats eller tid. Dessa båda grupper av substantiv kan man, enligt Malmgren (1994), bilda av nästan alla verb, ofta med hjälp av suffixen **-ning** och **-are**.

Ibland är det svårt att avgöra om verbet eller substantivet är det ursprungliga ordet, t.ex. *oroa* och *oro*. För att ta reda på detta behövs etymologiska studier och det är ett arbete som inte ryms inom ramen för detta examensarbete. De tveksamma fallen kommer jag därför att förbigå i min analys och i stället endast behandla de klara fallen.

Jag utesluter även avledda sammansättningar, t.ex. *väskryckare* och *tvååring*.

De substantivbildande suffixen är i olika grad produktiva. Fritt produktivt enligt Svenska Akademiens grammatik (SAG), (Teleman m.fl., 1999), är framför allt **-ande**. Produktiva för verbalsubstantiv är även **-(n)ing** och **-are**. Man kan även ta reda på hur produktiva suffixen är genom att undersöka en korpus. SUC 1.0-korpusen (Stockholm Umeå Corpus, se Ejerhed och Källgren (1997)) är en svensk korpus som innehåller 1 miljon löpord. Utifrån den har man i Umeå skapat ett avledningslexikon, *derinfl.lexicon*. Lexikonet innehåller uppgifter om avledningar och deras frekvens för olika morfosyntaktiska beskrivningar av hela ord, se figur 2.1.

Jag gick igenom detta lexikon för hand¹ och letade upp alla substantivförekomster och räknade ut vilka suffix som var vanligast för substantiven. 17 st av avledningsändelserna hade en frekvens på över 1000, se tabell 2.1. Den klart dominerande är *-ning*.

De tre suffix jag har valt ut för min undersökning är **-ing/-ning**, **-ande/-ende** och **-are**.² Liljestränd (1993) beskriver dessa suffix:

¹ Detta hade med fördel kunnat göras maskinellt i stället.

² Läs mer om hur jag valde ut dessa i kapitel 4.

("<Xande>")	(405 AB POS "Xande") (536 JJ POS UTR/NEU SIN/PLU IND/DEF NOM "Xande") (1602 NN NEU SIN IND NOM "Xande") (37 NN UTR PLU IND NOM "Xande") (149 NN UTR SIN IND NOM "Xande") (4815 PC PRS UTR/NEU SIN/PLU IND/DEF NOM "Xande"))
("<Xanden>")	(329 NN NEU PLU IND NOM "Xande") (1 NN UTR PLU IND NOM "Xande") (49 NN UTR SIN DEF NOM "Xande") (6 NN UTR SIN IND NOM "Xande"))
("<Xandena>")	(73 NN NEU PLU DEF NOM "Xande") (1 NN UTR PLU DEF NOM "Xande"))
("<Xandens>")	(20 NN NEU PLU IND GEN "Xande") (2 NN UTR SIN DEF GEN "Xande"))
("<Xandes>")	(6 AB "Xande") (20 NN NEU SIN IND GEN "Xande") (1 NN UTR PLU IND GEN "Xande") (2 NN UTR SIN IND GEN "Xande") (3 PC PRS UTR/NEU SIN/PLU IND/DEF GEN "Xande"))
("<Xandet>")	(791 NN NEU SIN DEF NOM "Xande"))
("<Xandets>")	(52 NN NEU SIN DEF GEN "Xande"))

Figur 2.1: Exempel från avledningslexikonet *derinfl.lexicon* för avledningsändelsen *-ande*. X står för en sekvens med minst en vokal. Till vänster står lexemet med en viss avledning. Till höger står de förekomster av olika ordklasser och böjningar av ordet som finns i SUC. Här kan man t.ex. se att *Xande* förekommer 405 gånger som adverb (AB), 536 gånger som adjektiv (JJ), 1788 (1602 + 37 + 149) gånger som substantiv (NN) och 4815 gånger som particip (PC).

Ändelse	Antal substantiv-förekomster
X-ning	28 770
X-het	6 241
X-ion	5 935
X-ing	5 895
X-are	5 220
X-nad	3 329
X-else	3 228
X-ande	3 141
X-al	2 995
X-ik	2 914
X-i	2 446
X-ator	1 746
X-em	1 519
X-skap	1 405
X-är	1 423
X-ska	1 193
X-itet	1 006

Tabell 2.1: De 17 substantivändelser från SUC 1.0-korpusen som har en frekvens på över 1000. X står för en sekvens med minst en vokal.

-ande/-ende : betecknar personer i både fast och tillfällig verksamhet, t.ex. *studerande, gående*.³

-are : ett av de mest produktiva suffixen, förekommer i ord för person, djur, redskap, yrkesutövare osv, t.ex. *läsare, murare*.

-ing/-ning : bildar substantiv som betecknar verksamhet, t.ex. *simning, samling*

Värt att nämna här är att ord med **-ande** utan böjningsändelser till övervägande del tillhör andra ordklasser än substantiv, dvs particip, adjektiv eller verb.

2.2 Valens

Det finns olika definitioner av vad valens är.

Enligt SAG definieras valens som de semantiska eller syntaktiska krav som ett ord (särskilt substantiv, adjektiv och verb) har på sin omgivning. Det finns två olika typer av valens, semantisk och syntaktisk. Den semantiska valensen är det valensbärande ordets krav på semantisk roll hos de omgivande referenterna, t.ex. agens, mottagare, föremål. Syntaktisk valens är hur de omgivande referenterna anges just vid detta ord och om de är obligatoriska eller inte.

Malmgren (1994) definierar valens som de bestämmningar som är nära knutna till sina huvudord. De kan också kallas fyllnadsbestämningar. Det finns optionella, som går att stryka, och obligatoriska, som inte går att stryka. Han menar också att ett ords valens är de upplysningar om ordets fraseologi som bör finnas i en bra ordbok.

Andersson (1993) skiljer mellan valensled och fria led samt mellan obligatoriska och optionella valensled. Han menar också att det är svårt att dra gränsen mellan dessa. Valensled är obligatoriska bestämmningar i synnerhet till verb, de är semantiskt fast beroende av sitt huvudord.

Nikula (1986) definierar valens som ”egenskapen hos verbet (eller hos ord ur andra ordklasser) att ’öppna luckor’ som kan eller måste fyllas med fyllnadsled”. T.ex. har vissa verb luckor för subjekt och direkt objekt, medan andra verb dessutom har en lucka för indirekt objekt.

Gemensamt för definitionerna är att valensled är de led som är nära knutna till sitt huvudord, obligatoriska eller optionella. Valensen är alltså en beskrivning av vilka valensled ett visst ord måste eller kan ta.

Den definition jag kommer att använda för valensled är de led som i första hand kan tolkas som att de modifierar huvudordet och inte är bestämmningar till hela satsen.

2.2.1 Substantivvalens

SAG skiljer på semantisk och syntaktisk valens. Det är den syntaktiska valensen som står i centrum för detta examensarbete.

I en nominalfras kan de efterställda prepositionsfraserna vara mer eller mindre bundna i förhållande till frasens huvudord. De är sällan obligatoriska. I (4) är prepositionsfrasen mer bunden till substantivet än i (5). Exempelen är hämtade ur Scarriekorpusen.

³De ord på *-ande/-ende* jag har i min undersökning (*förhållande* och *uttalande*) betecknar dock inga personer.

- (4) ... förbättra **behandlingen** av de blödarsjuka.
- (5) ... för att kunna styra **utvecklingen** i hamnen.

Enligt Malmgren (1994) och Teleman m.fl. (1999) är de största grupperna av substantiv som har bundna valensled substantiv avledda av adjektiv (nomen qualitatis) och verb (nomen actionis). Även substantiv som inte är avledda av verb eller adjektiv kan ta valens, t.ex. relationella substantiv såsom *bror, syster, farbror*.

Enligt Nikula (1986) är det endast de fraser där prepositionen bestäms av substantivet som är valensbundna, de är fyllnadsled. De fraser där prepositionen har en "egen betydelse" är av mer adverbial karaktär. De är tilläggsled.

En fråga jag måste ta ställning till i mitt arbete är hur jag ska hantera de som är mindre bundna till substantivet, om de ska räknas som valensled eller som någon form av tilläggsled. Jag måste avgöra om jag ska göra någon skillnad mellan tilläggsled och fyllnadsled. Jag återkommer till detta i kapitel 3.

2.2.1.1 Relation till verbet

Om en viss preposition förekommer som verbets eller adjektivets valens används i allmänhet samma preposition för substantivets valens, t.ex.

- (6) Han *deltog* sällan **i** förhandlingarna
- (7) Hans *deltagande* **i** förhandlingarna var sporadiskt

När verbet konstrueras med direkt objekt utan preposition, dvs. transitiva verb, används ofta *av* vid verbalsubstantivet, t.ex.

- (8) Hon *undersökte* skalbaggar noga
- (9) Hennes *undersökning* **av** skalbaggar var noggrant genomförd

Men även här finns undantag, t.ex. när substantivet inte slutar på *-ande* eller *-ning* eller när kopplingen till verbet respektive adjektivet inte är så stark. T.ex. *intervju med, mordet på*. *Av* är en svårtolkad preposition eftersom den även kan fungera som agentmarkör och då inte inleder ett valensbundet led. (Malmgren, 1994)

3 Språkteknologisk bakgrund

I detta kapitel tar jag upp några exempel på vad som redan har gjorts inom ramen för valens inom språkteknologin, det vill säga vad man har gjort för att skilja på obligatoriska och valfria efterled och vilka system för att koda valenser som finns. För det första bör det påpekas att det har gjorts mycket mer vad gäller verbvalenser än substantivvalenser och därför är det svårt att hitta litteratur om och exempel på implementeringar av substantivvalenser. En presentation av det jag hittat ges i detta kapitel.

I de olika undersökningarna jag presenterar använder man sig av olika terminologi. Jag väljer att använda mig av Nikulas termer fyllnadsled och tilläggsled. Definitionerna för dem är att fyllnadsled är de fraser där prepositionen bestäms av substantivet medan tilläggsled är av adverbial karaktär och har en ”egen betydelse”.

3.1 Scarrie

Scarriekorpusen¹ är den korpus utifrån vilken jag hämtar mitt undersökningsmaterial, dvs verbalsubstantiven, deras efterföljande prepositionsfraser och konkordanser för dessa ord.

Scarriekorpusen innehåller alla artiklar skrivna 1995 och 1996 från Svenska Dagbladet och Uppsala Nya Tidning. Det är över 220 000 artiklar, mer än 70 miljoner löpord, och mer än 1.6 miljoner ordtyper. Scarriekorpusen finns numera även i en taggad version² och det är ur den jag hämtar mitt material.

Utifrån Scarriekorpusen har man skapat ett lexikon, ScarrieLex. ScarrieLex innehåller 265 552 ordtyper och 4 899 fraser. (Olsson, 2004) En ordformsingång innehåller bl.a. information om lemma (*svlemma*), ordklass (*svpos*) och morfosyntaktiska särdrag (*svmorphcode*), se tabell 3.1. Det är här det är tänkt att valensinformationen ska läggas till i form av en kod.

svlemma	svlexeme	svpos	svinflection	svstem	svpattern	svaffix	svmorphcode	svfreq	svstyle	svreplacement	svderiv
utveckling.nn	1	NOUN	STOL	utveckling	\$	ar	NNUIB	15	N		
utveckling.nn	1	NOUN	STOL	utveckling	\$	ens	NNUSDG	99	N		
utveckling.nn	1	NOUN	STOL	utveckling	\$	en	NNUSDB	9053	N		
utveckling.nn	1	NOUN	STOL	utveckling	\$	s	NNUSIG	24	N		
utveckling.nn	1	NOUN	STOL	utveckling	\$		NNUSIB	90890	N		SMS

Tabell 3.1: Exempel på en ordformsingång för lemmat *utveckling* i ScarrieLex.

¹Se t.ex. <http://www.hlcentral.org/projects/detail.php?acronym=SCARRIE> eller Povlsen m.fl. (1999)

²Ohlander (2005) har i sitt examensarbete skapat denna taggade version

3.2 Prepositionsfrasanknytning

När ett substantiv följs av en preposition är det inte alltid som prepositionsfrasen modifierar substantivet. Vissa fall är klara, man vet direkt vad de modifierar. Andra fall är mer svåravgörbara. I det klassiska exemplet

(10) Jag såg mannen med teleskopet

finns det två tolkningar av prepositionsfrasen, antingen att den modifierar verbet eller att den modifierar substantivet. Problemet brukar kallas för PP-attachment-problemet. I min analys kommer jag att kunna stöta på fall där prepositionsfrasen modifierar något annat än substantivet och fall som är svåra att avgöra.

Hindle och Rooth (1993) ville hitta ett sätt att lösa problemen med prepositionsfrastillhörigheten. De tog fram några kriterier för att ta ut de lätta fallen. De tittade på treordskombinationer där första ordet är ett verb, andra ordet ett nomen och tredje ordet en preposition. Verbet eller prepositionen kunde vara utelämnat. Deras kriterier för de lätta fallen var

- Om nominet är ett pronomen tillhör prepositionen verbet, t.ex. *spare it from*
- Om verbet är i passiv form tillhör prepositionen verbet (förutom om prepositionen är *by*)
- Prepositionen tillhör substantivet om det inte finns något verb att koppla prepositionen till

Därefter använde man sig av olika statistiska gissningsmetoder. De hoppades kunna ge varje fall en lösning men upptäckte att det fanns fall där detta inte var möjligt, t.ex.

(11) ... known to frequent the same bars in one neighborhood.

Om man ofta besöker en bar och baren ligger på ett ställe så sker besökandehändelsen på samma ställe. Fall som detta, då en tolkning alltid validerar den andra tolkningen, kallade Hindle och Rooth (1993) för systematiskt tvetydiga. I min undersökning väljer jag att i stället skapa en defaultregel så att de tvetydiga fallen får en valensrelation:

Om PP:n går att koppla till substantivet så görs det

3.3 ComLex

ComLex-projektet var ett projekt vars mål var att skapa ett lexikon med syntaktiska särdrag, ett valenslexikon, för engelska ord i språkteknologiskt syfte (Grishman m.fl., 1994). Enligt Meyers m.fl. (1996) är det viktigt att skilja på adjunkter, (adjuncts), som är optionella valensled, och komplement, (complements), som är obligatoriska valensled. Detta var viktigt eftersom komplementen skulle tas med i lexikonet medan adjunkterna inte skulle det. Om man gör felaktiga klassificeringar av komplement och adjunkter kan det bli fel i parsningarna. Man satte därför upp nödvändiga kriterier för verbens komplement.

- 1. Obligatoriska** - komplement behövs för att satsen ska bli grammatiskt korrekt.
- 2. Passivitet** - ett led kan vara subjekt i en passiv sats endast om det är ett komplement.

3. **Speciella roller** - ett komplement har rollen som *theme* (tema), *source* (källa), *goal* (mål), *patient* (patient), *recipient* (mottagare), *experiencer* (upplevare), *proposition* (proposition), *question* (fråga) etc.
4. **Underförstådd betydelse** - ett komplement är optionellt men underförstått om det är utelämnat
5. **Selektionsrestriktioner** - om verbet ålägger selektionsrestriktioner på ett led så är ledet ett komplement.

Resultatet av projektet blev ett lexikon innehållande 38 000 engelska uppslag. Varje lexikoningång är organiserad som en hierarkisk mängd av typade särdragsvärdelistor. Särdragen indelas i komplement och andra särdrag. Andra särdrag kan t.ex. för substantiv vara titel och skala och för adjektiv kan det vara komparerbarhet. Exempel på annat verbsärdrag är VSAY vilket förekommer hos alla verb som kan förekomma med direkta citat. Substantiv har 9 möjliga komplementtyper och 9 möjliga särdrag (Macleod m.fl., 1994b). Exempel på komplementtyper för substantiv är *noun-for-to-inf* och *noun-be-wh-s* (Macleod m.fl., 1994a). Komplementtypen har sedan underkategorier som säger exakt hur komplementen kan se ut. I dessa fall *noun-for-to-inf* och *noun-to-inf* samt *noun-be-wh-s* och *noun-be-wh-to-inf*. Komplementtyperna är inte baserade på prepositioner utan på om de kan ta bisatskomplement med eller utan kopulaverb. T.ex. så kan man inte säga *the conflict whether we can leave*. Däremot kan man säga *the idea that he would steal*. Min undersökning bygger däremot på prepositioner vilket gör att jag inte kan använda mig av ComLex komplementtyper.

3.4 Prepositionsfrasklassificering

Merlo (2003) och medarbetare har byggt en prepositionsfrasklassificerare som skiljer på fyra typer av prepositionsfraskfunktioner.

- substantivargument
- substantivadjunkter
- verbargument
- verbadjunkter

De definierar argument som komplement som fyller sin roll i relation till sitt huvud. Adjunkter definierar de som komplement som tillför en speciell egenskap till sitt huvud eller sats. Ett komplement är ett argument om dess tolkning uteslutande beror på huvudet, medan det är en adjunkt om tolkningen är relativt konstant om den sätts ihop med olika huvuden. De tog fram sex kriterier för prepositionsfrasklassificeraren för att skilja på argument och adjunkter.

1. **Huvudberoende** Argument är beroende av sina huvud eftersom de utgör en del av frasen. Adjunkter är inte det. PP-argument kan bara förekomma tillsammans med det specifika verbala eller nominala huvudet. PP-adjunkter kan förekomma med många huvuden.
2. **Optionalitet** Många PP-argument som hör till verb gör meningarna ogrammatiska om man tar bort dem. Adjunkter är optionella.
- 3 + 4. **Iterativitet och ordning** Två argument kan inte upprepas och de måste ligga nära sitt huvud. Inget av dessa kriterier gäller för adjunkter.

5. Kopulär parafrasering NP-argument kan inte parafraseras med en kopulär relativsats medan adjunkter kan det, t.ex:

adjunkt: *en man från Paris* - *en man som var från Paris*

argument: *försäljningen av kläder* - **försäljningen som var av kläder*

Adjunkter kan förekomma efter kopulaverb.

6. Verbalsubstantiv PP som följer ett verbalsubstantiv är vanligen ett argument. För att hitta verbalsubstantiven undersökte man morfologin.

Kriterierna ovan användes av prepositionsfrasklassificeraren för att känna igen de fyra olika typerna av prepositionsfraser. Substantivargument visade sig vara lättare att identifiera än verbargument.

Syftet med prepositionsfrasklassificeraren var att identifiera olika prepositionsfrasklassifikationer i ett parsträd. De skapade två korpora bestående av exempel på PP-sekvenser i Penn Tree-bank. Den första korpusen bestod av enkla PP och den andra bestod av två PP efter varandra. Klassificeraren lyckades sätta rätt benämning på 74 % av fallen. De jämförde även med en klassificerare som bara skilde på två typer, argument och adjunkter, men det visade sig att den som skilde på fyra typer var bättre.

Min uppgift är att undersöka om kriterium nr 6 verkligen gäller för svenska verbalsubstantiv. Skulle kriteriet gälla så skulle alla prepositionsfraser i mitt undersökningsmaterial vara valensbundna. Man skulle i sådana fall kunna göra en automatisering som tillskrev prepositionsfrasen efter ett verbalsubstantiv en valensrelation till substantivet.

3.5 Valensmodell för engelska substantiv

Herbst (1987) har utarbetat en valensmodell för engelska substantiv. Han skiljer på komplement och perifera element. I ett PP-komplement styrs prepositionsvalet av substantivet. De perifera elementen är mer konstanta.

Herbst (1987) konstaterar att verb och verbalsubstantiv oftast har samma komplement.

Herbst (1987) nämner två sätt att skilja på komplement och perifera element:

1. Perifera element är inte bundna av några restriktioner förutom att de måste betyda något, medan komplement bara förekommer med speciella styrande element.
2. Formerna på perifera element är oberoende av de styrande elementen medan de styrande elementen bestämmer formerna på komplementet.

Om man tar bort ett element och meningen blir ogrammatisk eller förändrar betydelsen så är det alltså ett obligatoriskt komplement.

Denna modell stämmer inte helt för svenska substantiv eftersom det inte finns några klara gränser mellan tilläggsled och fyllnadsled.

3.6 Parole

PAROLE-projektet var ett EU-projekt som var inriktat på att bygga upp ett europeiskt nätverk av språkliga resurser. För varje språk skapades en korpus och ett lexikon. Det

svenska lexikonet består av ett morfologiskt lager och ett syntaktiskt. I det syntaktiska lagret finns information om de enskilda orden och deras syntaktiska egenskaper. För substantiven finns information om obligatoriska komplement och optionella framförställda och efterställda attribut. Bland de efterställda attributen finns prepositionsfraser med och information om vilken eller vilka prepositioner det kan vara samt om rektionen är animat eller inanimat. Substantiv som har obligatoriska komplement är t.ex. *aning*, *anhängare* och *avsaknad*. En beskrivning av det svenska PAROLE-lexikonet finns i [PAROLE].

3.7 Sammanfattning

Hindle och Rooth (1993) och Merlo (2003) tar upp problematiken med PP-attachments. Detta problem är relevant för mig eftersom jag ska avgöra om en PP tillhör substantivet eller inte. Merlo (2003) menar att en PP som följer ett verbalsubstantiv troligtvis är ett fyllnadsled. Jag ska undersöka om fallet är så på svenska. Hindle och Rooth (1993) kunde inte hitta en lösning som var tillämplig på alla meningar. De kallar de fall där en tolkning alltid validerar den andra tolkningen för systematiskt tvetydiga. Jag väljer i stället att använda mig av min defaultregel och låter på så vis de tvetydiga PP:na inkluderas bland de icke tvetydiga valensleden.

Merlo (2003) och Herbst (1987) har satt upp kriterier för tilläggsled och fyllnadsled. Även i ComLex-projektet har man satt upp kriterier, dock endast för verben. Dessa kriterier liknar varandra, bl.a. att fyllnadsled är beroende av sina huvuden och mer eller mindre obligatoriska. Jag väljer att följa min defaultregel som säger att om PP:n går att koppla till substantivet så ska det göras och gör således ingen skillnad på tilläggsled och fyllnadsled utan låter allt vara valensled.

Inte i någon av undersökningarna finns namn på relationerna mellan prepositionen och den efterföljande NP:n. Endast i PAROLE nämner man något om substantivet i den efterföljande NP:n. Man talar om ifall den är animat eller inanimat. Jag får alltså ingen ledning här för hur jag ska definiera relationerna i mitt material. I stället får jag själv definiera vilka relationer valensleden tycks uttrycka.

4 Metod

Detta kapitel beskriver hur uppgiften har genomförts, hur verbalsubstantiven valdes ut och på vilket sätt de analyserades.

4.1 Urval av verbalsubstantiv

Det jag hade som utgångspunkt när jag valde ut mina verbalsubstantiv var den taggade Scarriekorpusen samt en frekvenslista över alla ord från Scarriekorpusen som taggats som substantiv.¹ Det första jag gjorde var att jag tog fram verbalsubstantiv ur Scarriekorpusen samt valde ut vilka jag skulle analysera. Det finns i huvudsak två sätt att ta ut verbalsubstantiv från en frekvenslista med substantiv. Antingen kan man gå igenom listan för hand och ta ut de mest frekventa verbalsubstantiven eller så kan man utgå från ändelser och maskinellt plocka ut lexem som är taggade som substantiv och som slutar på en viss ändelse. Jag valde det senare.

För att veta vilka ändelser jag skulle leta efter använde jag mig av det avledningslexikon (derinf.lexicon) som beskrivs i 2.1 (se tabell 2.1). I ett första skede valde jag ut de sex mest frekventa suffixen; **-ing/-ning**, **-het**, **-ion**, **-are**, **-else** och **-ande/-ende**. Som synes är inte dessa suffix de sex översta i tabell 2.1. Det beror på att **-ande** och **-ende** är varianter av samma suffix, likaså **-ing** och **-ning**. Att **-nad** inte finns med bland dessa beror på att det aldrig läggs till verb.

Utifrån dessa ändelser tog jag fram de löpord som hade en frekvens på minst 500. Detta gjordes genom att söka i frekvenslistan över alla ord som var taggade som substantiv och som slutade på någon av dessa ändelser inklusive alla böjningsändelser². Eftersom orden i frekvenslistan inte är lemmatiserade så gick jag igenom listan för hand för att få fram vilka som hade en lemmafrekvens på minst 1500. Jag gick inte vidare till att analysera vilka olika lexem som lemmarna representerar. Jag fick även rensa bort de ord som inte var verbalsubstantiv. Kriteriet som jag hade för att ett substantiv var ett verbalsubstantiv var att det skulle finnas en belagd form av ett verb som substantivet är avlett ifrån. Detta gäller oavsett om det skett en betydelseglidning eller inte. Detta uppskattas till att ha tagit ungefär 35 arbetstimmar. Resultatet blev 197 lemman vilket var alldeles för många för att hinna analysera. För att begränsa omfånget valde jag ut de tre inhemska produktiva avledningsändelserna

- -ing/-ning
- -are

¹Ohlander (2005) har skapat denna frekvenslista och taggat Scarriekorpusen.

²De böjningsändelser som är relevanta är singular definit och indefinit form samt plural definit och indefinit form. Genitiv behövs inte eftersom ett substantiv i genitiv alltid fungerar som ett framförställt attribut till huvudordet i en annan nominalfras (Teleman m.fl., 1999) och det tar alltså inte några efterställda attribut.

- -ande/-ende

Jag begränsade mig också till att endast ta med de lemmor som hade en frekvens på minst 5000. Resultatet av detta blev 41 st lemmor och de kan ses i tabell 4.1. Det är dessa ord jag vidare analyserar.

Verbal-substantiv	Antal förekomster
utveckling	19 419
utbildning	13 221
författare	12 650
förändring	11 847
utredning	11 668
ledning	11 331
undersökning	11 199
ledare	11 026
utställning ^a	10 539
försäljning	10 371
lösning	10 185
förhållande	10 148
förening	9 982
forskare	9 453
lärare	9 346
spelare	9 345
förhandling	9 249
förutsättning	9 229
läkare	8 647
mening	8 623
ägare	8 380
forskning	8 227
tävling	7 481
uppfattning	7 324
förklaring	6 947
satsning	6 808
föreställning	6 780
ökning	6 676
besparing	6 264
omsättning	6 242
uttalande	5 755
ordning	5 589
åklagare	5 474
investering	5 432
ersättning	5 407
behandling	5 381
läsare	5 242
deltagare	5 194
invandrare	5 158
bedömning	5 136
handling	5 075

Tabell 4.1: De 41 st mest frekventa verbalsubstantiven på *-ing/-ning*, *-are*, *-ande/-ende*

^a *Utställning* är bildat på partikel verbet *ställa ut*, inte *utställa*

4.2 Analysmetod av verbalsubstantiven

För att kunna analysera verbalsubstantiven har jag valt att använda konkordanslistor. Jag skapade sådana för varje enskilt ord sorterade på högerkontextens tagg. För detta ändamål användes ett skript (`ylver/EXJOB/PERL/SKRIPT/konkstat.py`) som Tina Ohlander skapat för sitt examensarbete (se Ohlander (2005)). Detta skript ger förutom konkordanslistor även statistik för vilka och hur många prepositioner som förekommer i högerkontexten. För varje enskild preposition finns information om

hur stor andel den är av alla ord och hur stor andel den är av alla prepositioner. Exempel på hur konkordanslistorna kan se ut finns i figur 4.1 och ett exempel på hur statistiktabeln ser ut finns i tabell 4.2.

AS inledningen_NCUSN@DS till_SPS **utvecklingen**_NCUSN@DS av_SPS bl_XF a_XF datoriseringen_N
UAS så_RG0S mycket_RGPS för_SPS **utvecklingen**_NCUSN@DS av_SPS offshore_XF racing_XF som_CC
USNIS betydelse_NCUSN@IS för_SPS **utvecklingen**_NCUSN@DS av_SPS snabbareoch_XF mindre_AQC0
S testverksamhet_NCUSN@IS för_SPS **utveckling**_NCUSN@IS av_SPS t_XF ex_XF vacciner_NCNP@IS o
spåra_V@N0AS denna_DF@US@S **utveckling**_NCUSN@IS bakom_SPS att_CSS vi_PF@UPS@S tycks_V@I
IPAS en_DI@US@S lång_AQPUSNIS **utveckling**_NCUSN@IS bakom_SPS sig_PF@000@S _FE </s>
och_CCS neurobiologiska_AQP0SNDS **utvecklingen**_NCUSN@DS bakom_SPS vår_PS@US0@S personlige
NCUSN@DS angående_SPS hur_RH0S **utvecklingen**_NCUSN@DS beträffande_SPS lantbruk_NCNSN@IS ska
<s> En_DI@US@S ojämn_AQPUSNIS **utveckling**_NCUSN@IS beträffande_SPS mäns_NCUPG@IS respektive_C
e_V@IIAS för_SPS sämre_AQC00N0S **utveckling**_NCUSN@IS bland_SPS räntekänsliga_AQP0PN0S aktier_N
lädjande_RG0S är_V@IPAS inte_RG0S **utvecklingen**_NCUSN@DS bland_SPS utomnordiska_AQP0PN0S medbo

Figur 4.1: Exempel på konkordans för *utveckling*. Sökordet är i fetstil för att tydliggöra, i konkordanslistorna är det inte det.

Jag skapade liknande statistiktabeln och konkordanslistor för alla 41 orden.³ Därefter tog jag ut de mest frekventa prepositionerna för varje lemma. Jag valde att ta med de prepositioner som hade en frekvens på minst 1.00 % av allt i högerkontexten. Jag sammanställde detta i en tabell för att kunna göra jämförelser. Först analyserades vilka prepositionerna var och hur frekvent de förekom. Sedan gjordes en jämförelse med verbets valens för att se vilka samband som finns mellan verbvalenserna och substantivvalenserna. Därefter gjordes en jämförelse med Nationalencyklopedins ordbok (NEO) för att se vad man kan vinna på att ta ut data från en korpus i stället för att endast använda sig av en ordbok. Mer om detta finns i kapitel 5.

Nästa steg var att gå igenom konkordanserna för vart och ett av orden för att se om de efterställda prepositionsfraserna var valensled till substantivet eller inte samt att analysera de grammatiska relationerna som råder mellan substantivet och dess efterföljande prepositionsfras. På grund av det omfattande materialet och att jag hade begränsat med tid gick jag endast igenom ett ord grundligt, se kapitel 6. I kapitel 7 säger jag något kort om de övriga 40 ordens valensled.

³Statistiktabeln finns i /home/stp99/y/lver/EXJOB/METOD/STATISTIK/ där filen alla_stat innehåller alla statistiktabeln. De enskilda statistiktabeln finns i XXX_stat där XXX står för substantivet. Tabellerna är inte justerade utan ser ut som de gjorde efter att ha genererats av skriptet.

Statistik för <i>utveckling</i> och dess efterföljande prepositioner			
Stickord:	utveckling utvecklingen utvecklingar utvecklingarna		
Konkordansrader:	19709 (totalt antal stickordsförekomster)		
Antal prepositioner	Procent av totalt		
7271	36.89%		
lemma	antal	procent av prep	procent av alla ord
av	2143	29.47 %	10.87 %
i	2014	27.70 %	10.22 %
på	699	9.61 %	3.55 %
för	415	5.71 %	2.11 %
inom	366	5.03 %	1.86 %
under	363	4.99 %	1.84 %
mot	342	4.70 %	1.74 %
från	202	2.78 %	1.02 %
med	174	2.39 %	0.88 %
till	109	1.50 %	0.55 %
genom	66	0.91 %	0.33 %
sedan	58	0.80 %	0.29 %
efter	48	0.66 %	0.24 %
hos	34	0.47 %	0.17 %
åt	32	0.44 %	0.16 %
vid	32	0.44 %	0.16 %
kring	22	0.30 %	0.11 %
över	22	0.30 %	0.11 %
mellan	18	0.25 %	0.09 %
bland	16	0.22 %	0.08 %
om	15	0.21 %	0.08 %
utan	15	0.21 %	0.08 %
inför	8	0.11 %	0.04 %
framför	7	0.10 %	0.04 %
ur	7	0.10 %	0.04 %
enligt	6	0.08 %	0.03 %
bakom	4	0.06 %	0.02 %
trots	4	0.06 %	0.02 %
alltsedan	3	0.04 %	0.02 %
liksom	3	0.04 %	0.02 %
utanför	3	0.04 %	0.02 %
beträffande	2	0.03 %	0.01 %
runt	2	0.03 %	0.01 %
utifrån	2	0.03 %	0.01 %
via	2	0.03 %	0.01 %
å	1	0.01 %	0.01 %
af	1	0.01 %	0.01 %
alltifrån	1	0.01 %	0.01 %
emellan	1	0.01 %	0.01 %
före	1	0.01 %	0.01 %
förrän	1	0.01 %	0.01 %
ifrån	1	0.01 %	0.01 %
inklusive	1	0.01 %	0.01 %
inuti	1	0.01 %	0.01 %
nära	1	0.01 %	0.01 %
oavsett	1	0.01 %	0.01 %
per	1	0.01 %	0.01 %
tills	1	0.01 %	0.01 %

Tabell 4.2: Statistik för *utveckling* och dess efterföljande prepositioner

5 Övergripande resultat och analys

I detta kapitel redogörs för och analyseras de statistiska resultat jag har fått fram för vart och ett av de 41 verbalsubstantiven. Resultatet redovisas i form av en tabell med de prepositioner som förekom med minst 1.00 % av allt som kom först i högerkontexten. Dessförinnan redovisas för varje verbalsubstantiv hur stor andel av allt som förekommer först i högerkontexten som är prepositioner. Därefter görs en jämförelse med verbens valens. Till sist kommer ett avsnitt som redovisar en jämförelse med Nationalencyklopedins Ordbok (NEO) som har gjorts för verbalsubstantiven.

5.1 Andel prepositioner i högerkontexten

De 41 st verbalsubstantiven skiljer sig kraftigt åt vad gäller hur ofta de följs av preposition till skillnad från andra ordklasser (inkl skiljetecken). I tabell 5.1 kan man se att andelen prepositioner är mellan 12.22 % och 66.50 % för de olika substantiven.

En tendens man kan se är att nomen *agentis* har hamnat långt ner i listan, de följs mer sällan av prepositioner än nomen *actionis*. Enligt Teleman m.fl. (1999) är det nomen *actionis* (och substantiv bildade på adjektiv) som oftast följs av bundna valensled. Även intuitionen säger att det känns mer naturligt att t.ex. *ökning* och *satsning* följs av preposition än *läsare* och *åklagare*.

Jag har undersökt om de som har hög andel prepositioner i högerledet också har fler dominerande prepositioner än de som ligger lägre ner. I tabell 5.2 kan man se resultatet. Om man tittar på de som har en preposition över 50 % så finns de representerade främst i den övre halvan men även *åklagare* som ligger längst ned har en dominerande preposition. Endast fyra av dessa har ingen annan preposition med minst 10 %. De andra dominerande prepositionerna är minst dubbelt så vanliga som den näst mest frekventa prepositionen. Många av de övriga substantiven har också en dominerande preposition även om den inte förekommer med minst 50 %. Det är alltså inte givet att bara för att substantivet har hög andel prepositioner så är det en av dem som är dominerande.

5.2 Vilka prepositioner som finns i högerkontexten

I tabell 5.3 kan man se vilka prepositioner som utgjorde minst 1.00 % av alla ord som stod omedelbart till höger om substantivet. Det man måste komma ihåg är att dessa inte är analyserade och alltså är det inte säkert att alla dessa prepositioner verkligen är valensled till substantivet. De kan modifiera verbet eller hela satsen.

Det första man lägger märke till är att det handlar om 15 st prepositioner och att vissa är mer och vissa mindre vanliga. Märkbart är att *i* förekommer för alla verbalsubstantiv. Vanliga prepositioner är också *på* som förekommer hos 32 substantiv,

Verbal-substantiv	Andel prepositioner i högerledet
ökning	66.50 %
satsning	60.27 %
förhållande	57.73 %
ersättning	56.33 %
försäljning	50.54 %
förutsättning	48.28 %
förklaring	47.90 %
behandling	44.10 %
förhandling	44.06 %
uttalande	42.35 %
ledning	42.26 %
lärare	38.99 %
förändring	38.28 %
omsättning	37.90 %
investering	37.34 %
utveckling	36.89 %
besparing	34.19 %
uppfattning	33.43 %
lösning	32.75 %
ledare	32.19 %
föreställning	30.49 %
utställning	30.48 %
bedömning	28.90 %
tävling	28.60 %
deltagare	28.45 %
ordning	27.21 %
ägare	26.94 %
undersökning	26.32 %
forskning	26.25 %
lärare	25.55 %
utbildning	25.30 %
utredning	24.30 %
forskare	23.92 %
spelare	22.56 %
invandrare	17.67 %
förening	17.59 %
mening	17.40 %
handling	16.80 %
läkare	16.34 %
läsare	15.16 %
författare	12.88 %
åklagare	12.22 %

Tabell 5.1: Verbalsubstantiv med procentandel prepositioner i högerkontexten

för hos 27 st, *med* hos 23 och *av* hos 21. *Efter* förekommer bara hos ett substantiv och med ganska låg frekvens. *Kring*, *mellan*, *mot* och *under* förekommer hos två eller tre substantiv.

Många prepositioner förekommer med ganska låg frekvens. Endast *Av*, *för*, *på* och *till* förekommer hos ett eller två substantiv med minst 25 %. De förekommer hos ganska många substantiv. *Efter*, *inom*, *kring*, *mot* och *under* förekommer aldrig mer än 5 %. De förekommer hos ganska få substantiv.

5.3 Koppling till verben

I avsnitt 2.2.1.1 redogör jag för Malmgrens teorier angående kopplingen mellan verbens och verbalsubstantivens valens. Teorierna går ut på att samma preposition som används för verbet också ofta används för verbalsubstantivet. Om preposition saknas för verbet används ofta *av*. Undantag är om verbalsubstantivet inte slutar på *-ande* eller *-ning* eller inte har så stor koppling till verbet. Jag undersöker nu om dessa teorier stämmer med mina resultat från Scarriekorpusen. I sådana fall skulle man kunna skapa regler som automatiskt talade om vilken eller vilka prepositioner som substantivet

Verbal-substantiv	Andel prepositioner i högerledet	Prepositioner med minst 50 % av alla prepositioner		Prepositioner med minst 10 % av alla prepositioner																
ökning	66.50 %			av	40.56 %	med	33.82 %													
satsning	60.27 %	på	77.20 %																	
förhållande	57.73 %	till	64.05 %	mellan	14.19 %															
ersättning	56.33 %			för	43.70 %	till	15.12 %	från	11.74 %											
försäljning	50.54 %	av	59.38 %	i	11.84 %															
förutsättning	48.28 %	för	92.09 %																	
förklaring	47.90 %	till	86.42 %																	
behandling	44.10 %	av	56.09 %	med	13.66 %															
förhandling	44.06 %			med	31.80 %	om	26.50 %	mellan	13.75 %	i	13.22 %									
uttalande	42.35 %			om	26.84 %	från	21.62 %	i	18.30 %	av	10.36 %									
ledning	42.26 %			av	34.07 %	för	20.32 %	i	17.08 %											
förändring	38.28 %			i	43.05 %	av	28.51 %													
omsättning	37.90 %	på	50.15 %	i	13.81 %															
investering	37.34 %	i	58.78 %	på	16.30 %															
utveckling	36.89 %			av	29.47 %	i	27.70 %													
besparing	34.19 %			på	41.95 %	i	24.61 %	inom	10.75 %											
uppfattning	33.43 %	om	70.49 %	i	11.71 %															
lösning	32.75 %			på	49.84 %	för	16.38 %	av	10.07 %											
ledare	32.19 %	för	55.71 %	i	21.55 %															
föreställning	30.49 %			om	44.04 %	i	11.96 %	av	11.15 %	på	10.87 %									
utställning	30.48 %			på	22.49 %	i	20.48 %	med	17.00 %	om	16.66 %	av	13.27 %							
bedömning	28.90 %	av	69.35 %	i	41.26 %	på	13.11 %	om	12.88 %	för	10.86 %									
tävling	28.60 %			från	19.21 %															
deltagare	28.45 %	i	52.41 %	på	49.38 %	i	23.36 %													
ordning	27.21 %			till	48.54 %	i	20.36 %	av	13.88 %											
ägare	26.94 %			av	41.36 %	i	12.92 %													
undersökning	26.32 %			om	26.27 %	kring	15.72 %	i	15.36 %											
forskning	26.25 %			i	38.99 %	på	19.11 %	vid	16.84 %											
lärare	25.55 %			i	28.16 %	av	16.17 %	för	13.17 %											
utbildning	25.30 %			om	40.93 %	av	19.84 %													
utredning	24.30 %			vid	31.54 %	i	23.65 %	på	11.17 %											
forskare	23.92 %			i	41.48 %	från	11.02 %													
spelare	22.56 %			i	41.60 %	från	15.86 %	på	11.34 %											
invandrare	17.67 %			för	26.26 %	i	25.15 %	med	19.90 %											
förening	17.59 %			med	37.66 %	om	22.71 %	i	18.10 %											
mening	17.40 %			i	32.45 %	från	11.16 %													
handling	16.80 %			i	26.05 %	på	21.18 %	vid	15.15 %											
läkare	16.34 %			i	21.67 %	av	20.30 %	med	13.08 %	på	10.96 %									
läsare	15.16 %			till	37.70 %	i	15.50 %													
författare	12.88 %			för	11.47 %															
åklagare	12.22 %	i	52.35 %																	

Tabell 5.2: Verbalsubstantiv med respektive dominerande prepositioner

skulle följas av utifrån verbets prepositioner eller saknaden av sådana. När jag talar om vilka prepositioner verbet har utgår jag ifrån Nationalencyklopedins Ordboks nätversion¹. Det hade varit mer fördelaktigt att hitta prepositionerna i en korpus men det kräver extra arbete eftersom man måste gå in och analysera vilka prepositioner som hör till verbet och vilka som inte gör det.

Jag börjar med den första teorin och undersöker de verb som följs av en preposition. Efter att ha jämfört verben och verbalsubstantiven delar jag upp verbalsubstantiven i fem grupper utifrån hur väl de stämmer överens med Malmgrens teori.

A. Verbet konstrueras med endast en preposition och den förekommer med högst procentsats hos substantivet

- investera/investering i
- satsa/satsning på
- delta(ga)/deltagare i
- förhålla (sig)/förhållande till

¹www.ne.se

Verbalsubstantiv	av	efter	för	från	i	inom	kring	med	mellan	mot	om	på	till	under	vid
bedömning	20.04			1.12	2.23						2.25				
behandling	24.74		2.44		3.71			6.03		1.39		1.76			
besparing			3.02		8.41	3.68						14.34			
deltagare				5.46	14.91							2.48	1.43		1.24
ersättning	2.13		24.61	6.62	4.80			1.11				2.17	8.52		2.38
forskare				2.36	5.66	1.30		1.17				2.67			7.55
forskning			1.13		4.03	1.97	4.13				6.90	2.37			1.47
förening	1.39		4.62		4.42			3.50							
föreställning	3.40		1.25		3.65			2.34			13.43	3.31			
författare					2.00			1.24					4.86		
förhandling			1.10		5.82			14.01	6.06		11.68	1.56			
förhållande			1.37		4.47			2.76	8.19			1.70	36.98		
förklaring	1.34				1.81						1.03		41.39		
försäljning	30.01		1.01		5.98			3.58				2.75	2.55	1.86	
förutsättning			44.46		1.59										
förändring	10.91		2.20		16.48	1.56						2.24			
handling				1.88	5.45						1.02	1.30	1.53		
invandrare				2.80	7.35			1.67				2.00	1.19		
investering			3.60		21.94							6.08	1.06		
ledare			17.93		6.94							1.14			
ledning	14.40	1.53	8.59		7.22			3.19				1.85	1.04		
läkare			1.30		4.25							3.46			2.47
lärare			1.12	1.19	9.96			1.16				4.88			4.30
läsare	3.08				3.28			1.98			1.06	1.66	1.17		
lösning	3.30		5.36		2.51			1.30				16.32	1.25		
mening					3.15			6.55			3.95				
omsättning	2.26		1.68		5.23			3.05				19.01	1.19	2.32	
ordning			1.55		6.36			1.41				13.44			
satsning			2.23		5.57							46.53			
spelare			1.14	2.49	9.36			1.70				2.21	1.44		
tävling			3.10		11.80			1.82				3.68	3.75		
undersökning	10.89			2.19	3.40						2.29	2.23			
uppfattning	2.15				3.91							23.56			
utbildning	4.09		3.47		7.12	1.05		1.18				2.45	2.02		1.62
utredning	4.82		1.25		1.81		1.15					9.95			
utställning	4.05				6.24			5.18				5.08	6.86		
uttalande	4.39			9.15	7.75							11.37	2.85	1.22	
utveckling	10.87		2.11	1.02	10.22	1.86				1.74		3.55		1.84	
åklagare			1.40		6.40							1.19			
ägare	3.74				5.48			1.10					13.08		
ökning	26.97		1.61	1.54	5.35			22.50				5.15			

Tabell 5.3: Verbalsubstantiv med efterföljande prepositioner som utgör minst 1.00 % av allt som kommer i omedelbar högerkontext

B. Verbet konstrueras med flera olika prepositioner. En av dem förekommer med högst frekvens hos substantivet

- ersätta/ersättning för (även *med* finns med)
- forska/forskning om (*i* har tredje högsta frekvensen)
- förhandla/förhandling med (*om* har näst högsta frekvens)
- tävla/tävling i (alla verbets prepositioner (*på*, *om*, *för* och *med*) finns hos substantivet)

C. Alla prepositioner som verbet konstrueras med finns hos substantivet, dock inte med högst frekvens

- behandla/behandling med för
- handla/handling om
- omsätta/omsättning i
- utbildas²/utbildning för till
- åklagas³/åklagare för

²passivformen av verbet används här

³passivformen av verbet används här

- invandra/invandrare till från

D. Minst en av verbets prepositioner förekommer med substantivet

- förena/förening med
- läsa/läsare till med
- ordna/ordning med i
- spela/spelare med
- öka/ökning med
- forska/forskare i

E. Ingen av verbets prepositioner förekommer med substantivet

- förklara/förklaring
- läka/läkare
- lära/lärare
- lösa/lösning

Det verkar finnas alldeles för många fall då verbets preposition inte förekommer hos substantiven för att man skulle kunna göra en automatisering. I 11 fall av 24 förekommer verbets alla prepositioner. I 9 fall förekom en eller flera av prepositionerna och i 4 fall förekom ingen.

Jag går nu igenom de som kan ha objekt utan preposition, dvs transitiva verb. Enligt Malmgren används ofta *av* vid substantiven. Efter att ha jämfört verben och verbalsubstantiven delar jag in substantiven i 3 grupper utifrån hur frekvent *av* förekommer efter respektive substantiv. (Observera att en del substantiv redan förekommit i de fem första grupperna.)

1. Verbet kan konstrueras med objekt utan preposition och *av* förekommer med högst procentsats hos substantivet. (Procentsatserna inom parentes är hur frekvent *av* är för substantivet.)

- bedöma/bedömning (20.04 %)
- behandla/ behandling (24.74 %)
- försälja/försäljning (30.01 %)
- leda/ledning (14.40 %)
- läsa/läsare⁴ (3.08 %)
- undersöka/undersökning (10.89 %)
- utveckla/utveckling (10.87 %)
- öka/ökning (26.97 %)

2. Verbet kan konstrueras med objekt utan preposition och *av* förekommer med minst 1 % av allt i högerkonstexten hos substantivet, det är dock ej den mest frekventa

- ersätta/ersättning (2.13 %)
- förena/förening (1.39 %)
- förklara/förklaring (1.34 %)
- föreställa/föreställning (3.40 %)
- förändra/förändring (10.91 %)
- lösa/lösning (3.30 %)

⁴*av* ligger egentligen på andra plats efter *i* men det är så lite som skiljer mellan dem, därför har jag valt att ta med det här.

- omsätta/omsättning (2.26 %)
- ställa ut/utställning (4.05 %)
- uppfatta/uppfattning (2.15 %)
- utbilda/utbildning (4.09 %)
- utreda/utredning (4.82 %)
- uttala/uttalande (4.39 %)
- äga/ägare (3.74 %)

3. Verbet kan konstrueras med objekt utan preposition och *av* förekommer aldrig med minst 1 % av allt i högerkontexten hos substantivet.

- bespara/besparing
- författa/författare
- förutsätta/förutsättning
- investera/investering
- leda/ledare
- läka/läkare
- lära/lärare
- ordna/ordning
- satsa/satsning
- spela/spelare
- åklaga/åklagare

Av de transitiva verben hade alltså 8 substantiv *av* med högst procentsats, 13 substantiv hade *av* med minst 1 % och 11 stycken hade inte *av* med minst 1 %. Av de 11 substantiven slutade hälften, 6 stycken, inte på *-ande* eller *-ing* (vilket är 6 av 8 stycken *-are*-ord). Man kan alltså inte generalisera genom att skapa en regel som tillskriver *av* till substantiv bildade på transitiva verb. Man verkar i stället vara tvungen att titta på de enskilda orden. Malmgrens teorier stämmer alltså inte fullt ut.

Mena/mening ligger helt utanför de båda grupperna eftersom verbet *mena* med betydelsen ”ha som syfte” varken konstrueras med enkelt objekt eller preposition utan med *att*-sats eller adverbial.

5.4 Jämförelse med NEO

Via Sven-Göran Malmgren har jag fått tillgång till en lista på alla substantiv i Nationalencyklopedins Ordbok (NEO) med uppgifter om valenser. Av de 41 verbalsubstantiven i min undersökning finns endast 32 st med i listan. *Forskare, invandrare, läkare, spelare, utbildning, utställning, uttalande, åklagare* och *ägare* finns inte med. Av dessa ord ligger alla utom *uttalande* i den nedre halvan i tabell 5.1. *Utställning* ligger i mitten.

Det finns ingen fullständig överlappning mellan NEO och Scarrie. Ibland föreslår NEO prepositioner som inte finns med minst 1 % i Scarrie, ibland har Scarrie prepositioner med hög frekvens som inte föreslås i NEO. Ibland stämmer de ganska bra överens. Jag har delat upp orden i grupper beroende på hur mycket som stämmer överens. Vid en första anblick ser det ut som att NEO har missat många prepositioner men det behöver inte vara hela sanningen eftersom man inte vet i vilken mån prepositionerna i Scarrie verkligen är valensbundna.

Jag har endast tagit med de prepositioner som finns med i tabell 5.3, dvs de med minst 1 % av allt i högerkontexten. När jag jämför Scarrie med NEO lägger jag störst vikt vid de prepositioner som har högst frekvens, framförallt de med minst 10 % av alla prepositioner. Jag börjar med en grupp där NEO och Scarrie stämmer ganska väl överens, se tabell 5.4.

Verbal-substantiv	NEO	Scarrie		
		prep	alla ord	alla prep.
mening	~en (i ngt) (ngns) ~ (om ngn/ngt) ~en (med ngt)	med	6.55 %	37.66 %
		om	3.95 %	22.71 %
		i	3.15 %	18.10 %
satsning	en ~ (på ngt)	på	46.53 %	77.20 %
		i	5.57 %	9.24 %
		för	2.23 %	3.70 %
bedömning	~ (av ngn/ngt)	av	20.04 %	69.35 %
		om	2.25 %	7.79 %
		i	2.23 %	7.73 %
		från	1.12 %	3.86 %
utredning	~ (av/om ngt)	om	9.95 %	40.93 %
		av	4.82 %	19.84 %
		i	1.81 %	7.47 %
		för	1.25 %	5.15 %
		kring	1.15 %	4.74 %
förhållande	ett ~ (mellan ngra) ett ~ (till ngn) ett ~ (med ngn)	till	36.98 %	64.05 %
		mellan	8.19 %	14.19 %
		i	4.47 %	7.74 %
		med	2.76 %	4.79 %
		på	1.70 %	2.94 %
		för	1.37 %	2.37 %

Tabell 5.4: Jämförelse mellan NEO och Scarrie m.a.p prepositioner i valensledet. Relativt god överensstämmelse

I nästa grupp föreslår NEO fler prepositioner än Scarrie, se tabell 5.5.

Verbal-substantiv	NEO	Scarrie			Kommentar
		prep	alla ord	alla prep.	
forskning	~ (av ngn) (i/kring/om ngt) ~ar (efter ngn/ngt)	om	6.90 %	26.27 %	<i>efter</i> och <i>av</i> finns inte med i Scarrie
		kring	4.13 %	15.72 %	
		i	4.03 %	15.36 %	
		på	2.37 %	9.03 %	
		inom	1.97 %	7.50 %	
		vid	1.47 %	5.61 %	
		för	1.13 %	4.31 %	
förklaring	en ~ (av ngn) (av/på/till ngt)	till	41.39 %	86.42 %	<i>på</i> finns inte med i Scarrie
		i	1.81 %	3.78 %	
		av	1.34 %	2.80 %	
		om	1.03 %	2.14 %	
förutsättning	(under) ~ (av ngt); en ~ (för ngt)	för	44.46 %	92.09 %	<i>av</i> finns inte med i Scarrie
		i	1.59 %	3.30 %	

Tabell 5.5: Jämförelse mellan NEO och Scarrie m.a.p prepositioner i valensledet. NEO har prepositioner som inte Scarrie har

I tabell 5.6 genererar Scarrie fler prepositioner än NEO. Här gör jag en intressant upptäckt. Påfallande ofta är det prepositionen *i* som inte föreslås i NEO, i tio av sexton fall. En gång av dessa har *i* dominerande frekvens i Scarrie (*förändring*) och i de övriga fallen lite lägre frekvens. Slutsatser man skulle kunna dra av detta är att antingen är *i* inte valensbundet vid dessa fall eller så är de tilläggsled. Vid tre ord föreslår även NEO *i* och hos två av dessa har Scarrie hög frekvens.

Verbal-substantiv	NEO	Scarrie			Kommentar
		prep	alla ord	alla prep.	
deltagare	en ~ (i ngt)	i från på till vid	14.91 % 5.46 % 2.48 % 1.43 % 1.24 %	52.41 % 19.21 % 8.70 % 1.43 % 4.35 %	NEO har inte med <i>från</i>
ersättning	~ (för ngt/ngn) ~ (av/från ngn) (för ngt)	för till från i vid på av med	24.61 % 8.52 % 6.62 % 4.80 % 2.38 % 2.17 % 2.13 % 1.11 %	43.70 % 15.12 % 11.74 % 8.53 % 4.23 % 3.85 % 3.79 % 1.97 %	NEO har inte med <i>till</i>
förhandling	~ar (mellan ngra) (om ngt); ~ar (med ngn) (om ngt)	med om mellan i på för	14.01 % 11.68 % 6.06 % 5.82 % 1.56 % 1.10 %	31.80 % 26.50 % 13.75 % 13.22 % 3.54 % 2.50 %	NEO har inte med <i>i</i>
behandling	~ (av ngn/ngt)	av med i för på mot	24.74 % 6.03 % 3.71 % 2.44 % 1.76 % 1.39 %	56.09 % 13.66 % 8.41 % 5.54 % 3.98 % 3.16 %	NEO har inte med <i>med</i>
handling ⁵	~en (i ngt)	i från till på om	5.45 % 1.88 % 1.53 % 1.30 % 1.02 %	32.45 % 11.16 % 9.09 % 7.71 % 6.10 %	NEO har inte med <i>från</i>
omsättning	en ~ (på ngt) ~ (av ngt)	på i med under av för till	19.01 % 5.23 % 3.05 % 2.32 % 2.26 % 1.68 % 1.19 %	50.15 % 13.81 % 8.05 % 6.13 % 5.97 % 4.42 % 3.13 %	NEO har inte med <i>i</i>
förening	en ~ (av ngra) (i) ~ (med ngt) en ~ (för ngt)	för i med av	4.62 % 4.42 % 3.50 % 1.39 %	26.26 % 25.25 % 19.90 % 7.90 %	NEO har inte med <i>i</i>
läsare	~ (av ngt)	i av med på till	3.28 % 3.08 % 1.98 % 1.66 % 1.17 %	21.67 % 20.30 % 13.08 % 10.96 % 7.72 %	NEO har inte med <i>i</i> , <i>med</i> och <i>på</i>

⁵handling kommer från *handla* med betydelsen "utveckla" "ha som ämne eller innehåll". Dessa betydelser är inte transitiva, de tar inte direkt objekt

Verbal-substantiv	NEO	Scarrie			Kommentar
		prep	alla ord	alla prep.	
ökning	~ (av/i ngt)	av med i på för från	26.97 % 22.50 % 5.35 % 5.15 % 1.61 % 1.54 %	40.56 % 33.83 % 8.05 % 7.74 % 2.43 % 2.32 %	NEO har inte med <i>med</i>
lösning	~en (av/på ngt)	på för av i med till	16.32 % 5.36 % 3.30 % 2.51 % 1.30 % 1.25 %	49.84 % 16.38 % 10.07 % 7.66 % 3.98 % 3.83 %	NEO har inte med <i>för</i>
försäljning	~ (av ngt)	av i med på till under för	30.01 % 5.98 % 3.58 % 2.75 % 2.55 % 1.86 % 1.01 %	59.38 % 11.84 % 7.08 % 5.44 % 5.05 % 3.67 % 1.99 %	NEO har inte med <i>i</i>
ledning	under ~ (av ngn)	av för i med på efter till	14.40 % 8.59 % 7.22 % 3.19 % 1.85 % 1.53 % 1.04 %	34.07 % 20.32 % 17.08 % 7.55 % 4.38 % 3.62 % 2.46 %	NEO har inte med <i>för</i> och <i>i</i>
undersökning	~ (av ngn/ngt)	av i om på från	10.89 % 3.40 % 2.29 % 2.23 % 2.19 %	41.36 % 12.92 % 8.69 % 8.49 % 8.32 %	NEO har inte med <i>i</i>
föreställning	(göra ngn) ~ar (för ngt) en ~ (av ngt) en ~ (om ngn/ngt)	om i av på med för	13.43 % 3.65 % 3.40 % 3.31 % 2.34 % 1.25 %	44.04 % 11.96 % 11.15 % 10.87 % 7.67 % 4.10 %	NEO har inte med <i>i</i> och <i>på</i>
förändring	~ (av ngn/ngt)	i av på för inom	16.48 % 10.91 % 2.24 % 2.20 % 1.56 %	43.05 % 28.51 % 5.86 % 5.75 % 4.07 %	NEO har inte med <i>i</i>
uppfattning	~ (av ngn/ngt) en (ADJ) ~ (om ngn/ngt)	om i av	23.56 % 3.91 % 2.15 %	70.49 % 11.71 % 6.42 %	NEO har inte med <i>i</i>

Tabell 5.6: Jämförelse mellan NEO och Scarrie m.a.p prepositioner i valensledet. Scarrie har prepositioner som inte NEO har.

I tabell 5.7 har NEO prepositioner som inte Scarrie har och Scarrie har prepositioner som inte NEO har. Här är det som i tabell 5.6 många fall där NEO inte har med *i*, fem av nio. Men även *på* saknas vid fyra fall. Mest anmärkningsvärt är *besparing* där NEO inte har några av Scarries prepositioner. Den preposition som Scarrie saknar är oftast *av* men även *mellan*, *till*, *åt* och *om*. De substantiv där *av* finns med i NEO men inte i Scarrie tillhör alla grupp 3 i avsnitt 5.3 (transitiva verb där *av* sällan

eller aldrig förekommer). Man förväntar sig alltså ett *av* efter substantivet. Det man ser här är alltså att verkligheten inte alltid är som man förväntar sig. I NEO förefaller det som att man har tagit med *av* utifrån hypotesen att ett verb med direkt objekt utan preposition får *av* vid substantivering. I Scarrie kan vi se att *av* inte förekommer så frekvent som man hade förväntat sig

Verbal-substantiv	NEO	Scarrie			Kommentar
		prep	alla ord	alla prep.	
investering	en ~ (av ngn) (i ngt)	i på för till	21.94 % 6.08 % 3.60 % 1.06 %	58.78 % 16.30 9.64 2.84	NEO har inte <i>på</i> och Scarrie har inte <i>av</i>
tävling	en ~ (med ngn/ngt) (i/om ngt) en ~ (mellan ngra) (i/om ngt)	i på om för med	11.80 % 3.75 % 3.68 % 3.10 % 1.82 %	41.26 % 13.11 % 12.88 % 10.86 % 6.35 %	NEO har inte <i>på</i> och <i>för</i> och Scarrie har inte <i>mellan</i>
utveckling	~ (av ngn/ngt) (till ngn/ngt) ~ (av ngt)	av i på för inom under mot från	10.87 % 10.22 % 3.55 % 2.11 % 1.86 % 1.84 % 1.74 % 1.02 %	29.47 % 27.70 % 9.61 % 5.71 % 5.03 % 4.99 % 4.70 % 2.78 %	NEO har inte <i>i</i> och Scarrie har inte <i>till</i>
ordning	~ (på ngt) ~ (med ngn/ngt) ~en (mellan ngra)	på i för med	13.44 % 6.36 % 1.55 % 1.41 %	49.38 % 23.36 % 5.71 % 5.19 %	NEO har inte <i>i</i> och Scarrie har inte <i>mellan</i>
lärare	~ (i ngt); ~ (för/åt ngn)	i på vid från med för	9.96 % 4.88 % 4.30 % 1.19 % 1.16 % 1.12 %	38.99 % 19.11 % 16.84 % 4.67 % 4.54 % 4.38 %	NEO har inte <i>på</i> och <i>vid</i> och Scarrie har inte <i>åt</i>
besparing	~ (av ngt)	på i inom för	14.34 % 8.41 % 3.68 % 3.02 %	41.95 % 24.61 % 10.75 % 8.83 %	NEO har inte <i>på</i> , <i>i</i> och <i>inom</i> Scarrie har inte <i>av</i>
författare	~n (av/till ngt)	till i med	4.86 % 2.00 % 1.24 %	37.70 % 15.50 % 9.65 %	NEO har inte <i>i</i> Scarrie har inte <i>av</i>
ledare	en ~ (av ngn) (om ngn/ngt) ~(för ngt)	för i på	17.93 % 6.94 % 1.14 %	55.71 % 21.55 % 3.53 %	NEO har inte <i>i</i> Scarrie har inte <i>av</i> och <i>om</i>

Tabell 5.7: Jämförelse mellan NEO och Scarrie m.a.p prepositioner i valensledet. NEO har prepositioner som inte Scarrie har och Scarrie har prepositioner som inte NEO har

Till sist har jag tagit med de ord som inte finns med i NEO, tabell 5.8.

Verbal-substantiv	NEO	Scarrie		
		prep	alla ord	alla prep.
utbildning	—	i	7.12 %	28.16 %
		av	4.09 %	16.17 %
		för	3.47 %	13.70 %
		på	2.45 %	9.69 %
		till	2.02 %	7.97 %
		vid	1.62 %	6.40 %
		med	1.18 %	4.65 %
		inom	1.05 %	4.13 %
åklagare	—	i	6.40 %	52.35 %
		för	1.40 %	11.47 %
		på	1.19 %	9.71 %
invandrare	—	i	7.35 %	41.60 %
		från	2.80 %	15.86 %
		på	2.00 %	11.34 %
		med	1.67 %	9.45 %
		till	1.19 %	6.72 %
spelare	—	i	9.36 %	41.48 %
		från	2.49 %	11.02 %
		på	2.21 %	9.78 %
		med	1.70 %	7.53 %
		till	1.44 %	6.38 %
forskare	—	för	1.14 %	5.05 %
		vid	7.55 %	31.54 %
		i	5.66 %	23.65 %
		på	2.67 %	23.65 %
		från	2.36 %	9.86 %
		inom	1.30 %	5.45 %
läkare	—	med	1.17 %	4.89 %
		i	4.25 %	26.05 %
		på	3.46 %	21.18 %
		vid	2.47 %	15.15 %
utställning	—	för	1.30 %	7.95 %
		på	6.86 %	22.49 %
		i	6.24 %	20.48 %
		med	5.18 %	17.00 %
		om	5.08 %	16.66 %
uttalande	—	av	4.05 %	13.27 %
		om	11.37 %	26.84 %
		från	9.15 %	21.62 %
		i	7.75 %	18.30 %
		av	4.39 %	10.36 %
		på	2.85 %	6.72 %
ägare	—	till	1.22 %	2.87 %
		till	13.08 %	48.54 %
		i	5.48 %	20.36 %
		av	3.74 %	13.88 %
		med	1.10 %	4.09 %

Tabell 5.8: Scarries prepositioner för de ord som saknas i NEO's lista över substantivvalenser

5.5 Slutsatser

Antalet prepositioner varierar kraftigt mellan olika substantiv. Även vilka prepositionerna är varierar mellan olika substantiv. Man kan inte fullt ut använda sig av verbens

valens för att förutsäga substantivens valens eftersom det finns för många undantag.
Verkligheten ser inte alltid ut som man förväntar sig. Det finns många fler prepositioner i Scarrie än det finns föreslagna i NEO.

6 Djupare analys

6.1 Inledning

Detta kapitel är en djupdykning i det mest frekventa av de valda verbalsubstantiven, *utveckling*. *Utveckling* med dess olika böjningar förekom 19 419 gånger i Scarrie-korpusen. Dess efterled består till en dryg tredjedel av prepositioner (36.89 %). 8 prepositioner förekommer med minst 1 % av allt i högerkontexten, *av* och *i* är vanligast. Vad gäller kopplingen till verbet tillhör *utveckling* grupp 1, dvs verbet kan ha objekt utan preposition och substantivet har högst frekvens på *av*. NEO föreslår *av* och *till* som prepositioner. Men *till* förekommer först på tionde plats i Scarrie. *I* som förekommer på andra plats finns inte med i NEO. Tittar man i tabell 4.2 kan man se att *utveckling* följs av 48 olika prepositioner. *Förrän* och *liksom* är dock feltaggade, de är inga prepositioner. Många prepositioner förekommer ett fåtal gånger. Fördjupningen innehåller min analys av de tio vanligaste prepositionerna

6.2 Syfte och metod

Fördjupningens syfte var att ta reda på om de efterföljande prepositionsfraserna verkligen är valensled till substantivet och vilka relationer de uttrycker. Detta måste göras manuellt och därför måste undersökningsmängden minskas. Metoden var att ta ut var N:te förekomst i konkordanslistorna som genererats av ett skript (konkstat.py, se sid 13, avsnitt 4.2) så att undersökningsmängden blev som minst 10 % av alla konkordansrader och som mest 50 %. Frekvensen för de olika substantiven varierar kraftigt vilket gjorde att antalet konkordansrader hamnade mellan 54 och 214.

Därefter gick jag igenom varje rad för att se om prepositionsfrasen i just denna mening var valensled till substantivet. Jag hade två saker att ta hänsyn till för varje konkordansrad:

1. Är det valensbundet?
2. Vilken relation är det?

När jag bestämde om det var valensled eller inte så utgick jag från min defaultregel som säger att om PP:n går att koppla till substantivet så ska det ske. Här fick jag även ta ställning till om jag ska skilja på fyllnadsled och tilläggsled. Jag valde att inte göra någon skillnad utan kallade allt för valensled utifrån defaultregeln. När jag bestämde om en PP gick att koppla till substantivet så använde jag mig av min språkliga intuition. Jag kunde rensa bort de fall där prepositionen tillhörde ett flerordsuttryck eller där det var uppenbart att den tillhörde verbet. I exempel som

(12) Orienteringen är under stark utveckling i Länna med en omfattande

är det uppenbart att PP:n tillhör hela satsen. I de tveksamma fallen som

(13) ett studentcentrum för miljö och utveckling i Uppsala

där man inte vet om PP:n (*i Uppsala*) ska kopplas till *studentcentrum, miljö och utveckling* eller *utveckling* använde jag mig av defaultregeln så att PP:n kopplas till *utveckling*.

Resultatet redovisas först i en översiktstabell och därefter mer utförligt preposition för preposition.

Eftersom syftet med valensrelationerna är att de ska kunna skrivas som regler och användas i maskinöversättningssyfte ville jag få en uppfattning om hur prepositionen kan översättas till engelska. Jag gick därför igenom ett antal länkade meningar med *utveckling* + respektive preposition i en parallellkorpus med texter länkade mellan svenska och engelska.¹ Jag har tittat på max 40 förekomster per genomgång. Läs mer om de engelska översättningarna under respektive preposition.

6.3 Resultat - översikt

Tabell 6.1 är en översikt över resultatet jag fått fram av analysen.

Preposition	Valensbundna (procent av alla)	Ej valensbundna				Osäker på	Summa
		verb	sats	fler	felt		
av	213 (99.5 %)		1				214
i	182 (90.5 %)	5	6	8			201
på	88 (88.9 %)	2	4			5	99
för	91 (88.3 %)	2		9 ^a	1		103
inom	89 (98.9 %)		1	1			91
under	88 (97.8 %)		1		1		90
mot	83 (97.6 %)	1		1			85
från	62 (92.5 %)	1	3	1			67
med	34 (58.6 %)	9	10	5			58
till	37 (68.5 %)	9	4	4			54

Tabell 6.1: Resultat av hur prepositionerna fördelade sig på valensbundna, ej valensbundna samt de jag var osäker på för substantivet *utveckling*

^a varav 6 st var *för att*

verb = prepositionen tillhör verbet

sats = prepositionen tillhör hela satsen

fler = prepositionen tillhör ett flerordsuttryck

felt = feltaggade

6.4 Utveckling av

Var tionde förekomst av *utveckling av* togs ut (214 av 2143) och av dessa var alla utom en valensbundna; den hörde till hela satsen.

¹ Se <http://logos.uio.no/cgi-bin/opus/opuscqp.pl>, där jag gick in under EUROPARL

(14) i sin utveckling av närmast feodala inkomst-

Av de substantiv som är huvudord i den nominalfras som följer *av* var 123 st abstrakta, 39 st konkreta, 23 st kunde vara både abstrakta och konkreta och 16 st var jag osäker på om de är abstrakta eller konkreta. 10 gånger såg man inte vilket substantivet var och 2 gånger var det inget substantiv. Relationen är i alla utom ett fall en objektsrelation. Exempel (min analys står inom parentes, valensrelationen står i fetstil):

(15) Utvecklingen av budgetunderskottet är viktigare (*abstrakt, objektsrelation*)

(16) svårt att se en dynamisk utveckling av den svenska arbetsmarknaden (*abstrakt, objektsrelation*)

(17) mobiltelefonförsäljningen ingår, pågår utvecklingen av alla de telefoner som (*konkret, objektsrelation*)

(18) Samtidigt ska vi fortsätta utvecklingen av nya system så att vi (*konkret + abstrakt, objektsrelation*)

Vad gäller de engelska länkarna var ungefär hälften *development of*. Övriga fall fördelade sig på en sammansättning med *development*, omskrivning med verb eller så användes helt andra ord

6.5 Utveckling i

Var tionde förekomst av *utveckling i* togs ut (201 av 2014) och av dessa var drygt 90 % valensbunda. Av de övriga fallen tillhörde prepositionen verbet, satsen eller en flerordskonstruktion.

Av de substantiv som är huvudord i den nominalfras som följer *i* var 136 st konkreta och 42 st abstrakta. Av de abstrakta var 6 substantiv temporala. En gång följdes *i* inte av ett substantiv (*utvecklingen i stort*). Två gånger såg man inte vilket substantivet var och i ett fall var jag osäker (*i praxis*).

Vad gäller valensrelationen så uttrycker de temporala en temporal relation. Av de övriga kan man inte göra några helt klara uttalanden. Man kan ana en skillnad mellan de konkreta och abstrakta substantiven. De abstrakta verkar vara en subjektsrelation. De konkreta lutar delvis åt en lokationsrelation men kan även vara subjektsrelation. Det skiljer lite mellan olika substantiv. För att avgöra om det är en subjektsrelation kan man skriva om rektionen till en genitiv. T.ex så kan *utvecklingen i matchen* skrivas om till *matchens utveckling*. Exempel:

(19) som i modern tid vände utvecklingen i Frankrikes största vinodlingsområd (*konkret, lokationsrelation*)

(20) av den politiska och ekonomiska utvecklingen i landet (*konkret, subjektsrelation*)

(21) fick också hjälp nedåt av utvecklingen i USA (*konkret, subjektsrelation*)

(22) lova att fortlöpande informera om utvecklingen i matchen (*abstrakt, subjektsrelation*)

(23) utvecklingen i exportindustrin försvagas så (*abstrakt, subjektsrelation*)

(24) Den allra bästa utvecklingen i maj hade S-E-bankens nya (*abstrakt, temporal, temporal relation*)

Exempel 20 och 21 kan man också ge tolkningen lokationsrelation.

Knappt hälften av de engelska länkarna var *development in*. Vanligast av de övriga var *development of*. Övriga fall var omskrivningar med verb eller andra ord eller enkelt substantiv utan preposition. Även några fellänkningar samt fall där prepositionen inte var valensled på svenska förekom.

6.6 Utveckling på

Var sjunde förekomst av *utveckling på* togs ut (99 av 699). Av dessa var ca 90 % valensbundna, 6 % ej valensbundna och 5 % var jag osäker på. Av de som inte var valensbundna tillhörde två verbet och fyra tillhörde satsen.

Av de substantiv som är huvudord i den nominalfras som följer *på* var 10 st konkreta, 64 st abstrakta och 12 st som jag inte kan avgöra om de är konkreta eller abstrakta. Av de abstrakta var 8 substantiv temporala. En gång såg man inte vilket substantiv det var och en gång följdes *på* av annat än substantiv;

(25) kraft att använda för kommunens utveckling på att dra fram gamla, redan

De temporala uttrycker en temporal relation. De övriga verkar uttrycka en subjeksrelation men några konkreta substantiv verkar även kunna uttrycka en lokationsrelation. Exempel:

(26) en snabbare eller långsammare utveckling på Arlanda flygplats (*konkret, lokationsrelation*)

(27) kan bidra till en bättre utveckling på miljövårdsområdet (*abstrakt, subjeksrelation*)

(28) vad tror du om utvecklingen på den amerikanska marknaden (*abstrakt, subjeksrelation*)

(29) utvecklingen på 80-talet och in i 90-talet (*abstrakt, temporal, temporalrelation*)

(30) och den svindlande tekniska utvecklingen på området har verkligen (*abstrakt eller konkret*)

I exempel 30 beror relationen på om *området* är abstrakt eller konkret.

Det var flera prepositioner som förekom i den engelska länkade texten. Vanligast är *in*. Andra som fanns med är *of*, *at*, *on* samt omskrivningar med verb eller andra ord.

6.7 Utveckling för

Var fjärde förekomst av *utveckling för* togs ut (103 av 415). Av dessa var knappt 90 % valensbundna. Av de som inte var valensbundna tillhörde de flesta en flerordskonstruktion, oftast *för att*.

Av de substantiv som är huvudord i den nominalfras som följer *för* var 20 st konkreta, 59 st abstrakta, samt 6 st som jag är osäker på om de är konkreta eller abstrakta. Av de abstrakta är 4 stycken temporala ord. 5 gånger ser man inte vilket substantivet är och 1 gång är det inte ett substantiv (*utveckling för dem*).

De temporala uttrycker en temporal relation. De övriga verkar uttrycka en patientrelation. Exempel:

- (31) kraftig ekonomisk och social utveckling för invandrarna (*konkret*, **patientrelation**)
- (32) att det blir en mycket bra utveckling för museet (*konkret*, **patientrelation**)
- (33) Denna förädling innebär samma utveckling för denna del av tjänstenäringen (*abstrakt*, **patientrelation**)
- (34) och många har hoppats på samma utveckling för stroke-vården (*abstrakt*, **patientrelation**)
- (35) är sällsynt svårt att gestalta utvecklingen för detta år (*abstrakt*, *temporal*, **temporalrelation**)

Bland de länkade texterna var många fall inte valensbundna på svenska, för övrigt förekom *of*, *for*, *in*, *on* samt omskrivningar med verb eller andra ord.

6.8 Utveckling inom

Var fjärde förekomst av *utveckling inom* togs ut (91 av 366). De var alla valensled utom en där prepositionsfrasen tillhörde hela satsen och en som ingick i en flerordskonstruktion (*inom ramen för*).

Av de substantiv som är huvudord i den nominalfras som följer *inom* var 80 st abstrakta, 1 var konkret och 5 st kunde vara både abstrakta och konkreta. 3 gånger såg man inte vilket substantivet var.

Relationen verkar vara subjektsrelation utom i de fall substantivet kan vara både abstrakt och konkret där det snarast lutar åt lokationsrelation ². Exempel:

- (36) att det är med och driver utvecklingen inom IT-området (*abstrakt*, **subjektsrelation**)
- (37) de bland annat på den dåliga utvecklingen inom byggindustrin som en av (*abstrakt*, **subjektsrelation**)
- (38) Utvecklingen inom det datorstödda musikskapandet (*abstrakt*, **subjektsrelation**)
- (39) utan hela Mälardalen och bevaka utvecklingen inom EU (*konkret* + *abstrakt*, **lokationsrelation**)

Vad gäller översättningarna var *in* den klart vanligaste prepositionen. Andra prepositioner som jag hittade var *of*, *within* och *at*. Även omskrivningar med andra ord samt enkelt substantiv förekom.

6.9 Utveckling under

Var fjärde förekomst av *utveckling under* togs ut (90 av 363). Alla utom två var valensbundna.

Av de substantiv som är huvudord i den nominalfras som följer *under* är alla tidsord, eller syftar på en tidsperiod, t.ex. *skede*. 6 gånger följs *under* av räkneord som är årtal. Relationen är alltså en temporal relation. Exempel:

²Substantivet är i alla dessa fall *EU*

- (40) så brukar utvecklingen under 80- och 90-talen samman (*abstrakt, temporal, temporal relation*)
- (41) första kvartal jämfört med utvecklingen under sista kvartalet i fjol (*abstrakt, temporal, temporal relation*)
- (42) flyktingar nu att konstatera att utvecklingen under den rödgröna majoriteten (*abstrakt, temporal, temporal relation*)
- (43) Utvecklingen under veckan gav alltså ingen (*abstrakt, temporal, temporal relation*)

Bland de engelska länkade meningarna var det främst andra omskrivningar som var vanliga, även några fellänkningar samt några prepositioner som inte var valensled på svenska förekom. Bland prepositionerna var *of, over* och *in* vanligast. Några enstaka fall av *during* fanns också.

6.10 Utveckling mot

Var fjärde förekomst av *utveckling mot* togs ut (85 av 342). Alla utom två st var valensbundna.

Av de substantiv som är huvudord i den nominalfras som följer *mot* är 3 st konkreta, 47 st abstrakta och 15 st kan vara både konkreta och abstrakta. Av de abstrakta var en temporal. 14 gånger ser man inte substantivet och 4 gånger är det inget substantiv (tre gånger följs *mot* av en att-sats och en gång av ett pronomen). Relationen verkar vara en målrelation. Exempel:

- (44) Den snabba utvecklingen mot alltmer trafiksäkra bilar (*konkret, målrelation*)
- (45) Den utvecklingen mot större självstyre lokalt (*abstrakt, målrelation*)
- (46) Utvecklingen mot ett utökat samarbete går (*abstrakt, målrelation*)
- (47) har lett till att den normala utvecklingen mot stabilitet i nya områden (*abstrakt, målrelation*)
- (48) vi bidra till norra Europas utveckling mot en dynamisk region med (*konkret och abstrakt, målrelation*)

För det mesta användes andra ord i de engelska texterna. Om det var en preposition så användes *towards*.

6.11 Utveckling från

Var tredje förekomst av *utveckling från* togs ut (67 av 202). 93 % var valensbundna, de övriga var inte valensbundna.

Av de substantiv som var huvudord i den nominalfras som följer *från* var 19 st konkreta, 30 st abstrakta och 4 st som både kan vara konkret och abstrakt. Av de abstrakta var 9 stycken temporala ord. 5 gånger ser man inte vilket substantivet är och 4 gånger är det inget substantiv utan ett årtal. Relationen verkar vara en ursprungsrelation och för de temporala orden tidsursprung. Exempel:

- (49) som styr bananflugans snabba utveckling från ägg och larv till färdig insekt (*konkret, ursprungsrelation*)
- (50) Utvecklingen från jägarfolk till ett modernare (*konkret, ursprungsrelation*)
- (51) Gruppens utveckling från mer typisk dödsmetall till (*abstrakt, ursprungsrelation*)
- (52) populärmusikutbud se den politiska utvecklingen från kommunistisk diktaturtill (*abstrakt, ursprungsrelation*)
- (53) får ingen klar bild av utvecklingen från 1600-tal till 1900-tal (*abstrakt, temporal, tidsursprungsrelation*)

Påfallande ofta finner man mönstret *från . . . till*, ibland kan man dock endast gissa sig till att ett *till* förekommer senare i satsen.

Jag hittade endast 15 fall i den engelska text jag gick igenom och av dessa var en fellänkad och tre stycken var inte valensled på svenska. Av de andra fallen var de flesta omskrivningar med andra ord.

6.12 Utveckling med

Var tredje förekomst av *utveckling med* togs ut (58 av 174). Drygt hälften av dem var valensbundna, de övriga var inte valensbundna. De flesta av dem tillhörde verbet eller satsen.

Av de substantiv som är huvudord i den nominalfras som följer *med* är 24 st abstrakta och 3 st kan vara både abstrakta och konkreta. Av de abstrakta är en temporal. 6 gånger ser man inte vilket substantivet är och 1 gång är det inte ett substantiv utan en att-sats. Relationen verkar vara beledsagande. Exempel:

- (54) på en fortsatt demokratisk utveckling med större deltagande av (*abstrakt, beledsagande*)
- (55) och inte har hunnit i kapp utvecklingen med utbildning, administration (*abstrakt, beledsagande*)
- (56) Västeuropa är man inne i en utveckling med sänkt normalarbetsvecka (*abstrakt, temporal, beledsagande*)

De allra flesta fall av de länkade texterna var inte valensled på svenska. Endast ett fåtal var det.

6.13 Utveckling till

Varannan förekomst av *utveckling till* togs ut (54 av 109). Knappt 70 % var valensbundna, resten var ej valensbundna. De tillhörde verbet, satsen eller en flerordskonstruktion.

Av de substantiv som är huvudord i den nominalfras som följs av *till* är 15 st konkreta, 9 st abstrakta och 7 st kan vara både konkreta och abstrakta. 4 gånger ser man inte vilket substantivet är och 2 gånger är det inte ett substantiv (*utveckling till det bättre*). Relationen verkar vara en målrelation. Exempel:

- (57) gick i skola och började sin utveckling till elitgymnast (*konkret*, **målrelation**)
- (58) Att verka för Paldiskis utveckling till en internationell hamn (*konkret*, **målrelation**)
- (59) Utvecklingen till en nordisk elmarknad (*abstrakt*, **målrelation**)
- (60) varna för de yngre högskolornas utveckling till universitet ter sig för (*konkret + abstrakt*, **målrelation**)

Det var endast ett fåtal förekomster som var valensbundna i de länkade texterna. Av dessa var de flesta omskrivningar med andra ord.

6.14 Sammanfattning

De relationer jag kommit fram till är

- objekt: av
- lokation: i, på, inom
- subjekt: i, på, inom
- mål: mot, till
- ursprung: från
- temporal: i, på, under
- indirekt objekt: för
- tidsursprung: från
- beledsagande: med

Några av relationerna har fler än en preposition. Vad gäller målrelationen så är *mot* vanligare än *till*. Det är fler abstrakta substantiv vid *mot* än vid *till*. *Till* går i flera fall att byta mot *mot* men inte tvärtom. När det gäller *i*, *på* och *inom* verkar det som att det efterställda substantivet styr vilken preposition det ska vara. *På* används oftare vid abstrakta substantiv medan *i* används oftare vid konkreta. *Inom* används nästan bara vid abstrakta substantiv. När det gäller temporalrelationerna hör prepositionen främst ihop med det efterföljande substantivet.

Om man tittar på de engelska länkarna verkar *in* vara den vanligaste prepositionen. Ofta hittar man även omskrivningar med andra ord. Detta visar att man i de flesta fall inte kan översätta prepositionerna med en defaultöversättning utan man behöver först beskriva substantivet och prepositionen i en relation för att sedan översätta relationen. Nästa steg vore alltså att hitta en engelsk översättning till relationerna ovan.

7 Avslutning

I detta kapitel säger jag något kortfattat om de övriga verbalsubstantivens valensrelationer. Därefter följer sammanfattning, reflektioner, samt idéer till hur man man bygga vidare på mina resultat.

7.1 Något om de övriga substantivens valens

I förekommer hos alla substantiv men oftast med låg frekvens vilket kan tolkas som att det i många fall kan handla om tilläggsled snarare än fyllnadsled eller att de inte är valensled över huvud taget. Det förekommer ju inte heller hos lika många substantiv i NEO som i mitt material.

Mot kan ha två betydelser, riktning eller emot. I min undersökning finns det med hos två substantiv, *utveckling* och *behandling* där de representerar var sin betydelse.

Malmgren (1994) nämner att *av* även kan tolkas som agentmarkör vilket inte var fallet i min undersökning av *utveckling*. Tittar man däremot i tabell 5.3 kan man misstänka att *av* efter *ersättning* oftast är agentmarkör. För *bedömning*, *behandling*, *förklaring*, *lösning*, *undersökning*, *utbildning*, *utredning* och *uttalande* kan det vara både agentmarkör och objektsrelation. I de övriga fallen är det antagligen en objektsrelation.

I min undersökning av *utveckling* inleder de 8 mest frekventa prepositionerna knappt 90 % eller mer valensbundna led. De två följande prepositionerna har lägre andel valensbundna led. Det man kan tänka sig är alltså att de flesta av prepositionerna i tabell 5.3 har hög andel valensbundna led. De prepositioner som ingick i flerordsuttryck för *utveckling* var främst *för*, *i*, *med* och *till*. I princip kan alla prepositioner inleda ett eller flera flerordsuttryck och det är svårt att förutsäga vilka av prepositionerna i tabell 5.3 som gör det.

7.2 Sammanfattning

I detta arbete togs underlag för valensbeskrivningar av 41 svenska verbalsubstantiv fram. Det gjordes genom att ta ut de mest frekventa verbalsubstantiven på **-ing/-ning**, **-ande/-ende** och **-are** ur Scarriekorpusen, vilket resulterade i 41 st substantiv. Underlaget består av data för vilka och hur många prepositioner som följer omedelbart efter substantivet.

För att avgöra om prepositionsfrasen är valensled till substantivet skapades även konkordanslistor. I detta arbete definieras valensled som de led som i första hand kan tolkas som att de modifierar huvudordet och inte är bestämmningar till hela satsen.

Resultatet visar att de 41 substantiven skiljer sig åt vad gäller hur ofta de följs av preposition samt vilka prepositioner de följs av. Vid en jämförelse mellan verbens

och substantivens valens visade det sig att endast i knappt hälften av fallen förekom verbets alla prepositioner hos substantivet. I 4 av 24 fall då verbet åtföljs av en preposition förekom ingen av dessa vid substantivet. Det visade sig också att det inte var så vanligt, som bl.a Malmgren (1994) nämner, att substantiv bildade på transitiva verb ofta efterföljs av *av*.

Scarries prepositioner jämfördes även med de föreslagna prepositionerna i NEO. Resultatet visade att det inte fanns någon fullständig överlappning mellan dem. För det första så har Scarrie många fler prepositioner än NEO. För de andra så har de ofta olika prepositioner. Dock vet man inte om alla Scarries prepositioner verkligen är valensbundna.

Vid genomgången av det vanligaste verbalsubstantivet, *utveckling*, framkom att de allra flesta prepositionsfraserna var valensled till substantivet. I några fall tillhörde prepositionen verbet, satsen eller en flerordskonstruktion. Valensrelationerna var av nio olika typer varav några representerades av flera olika prepositioner.

7.3 Reflektioner

Uppgiften var mycket mer omfattande än jag hade kunnat ana, det finns väldigt många verbalsubstantiv i Scarriekorpusen och dessutom ett stort antal förekomster av varje vilket gjorde det omöjligt att hinna gå igenom mer än ett fåtal.

En annan överraskning var den stora variation av prepositioner som följde varje verbalsubstantiv i korpusen. Om man tittar i en ordbok finner man en eller ett par prepositioner till varje substantiv men här fanns ett mycket större antal prepositioner representerade.

Det är också svårt att ta ställning till vilka led som är valensled och vilka som inte är det eftersom gränserna är flytande, speciellt mellan tilläggsled och icke valensled. Det är även svårt att göra exakta uttalanden om valensrelationerna. Man får i båda fallen använda sin intuition vilken ibland kan svika. På grund av att jag använde mig av min defaultregel fick jag troligen även med vissa fall som tillhörde verbet eller satsen.

På grund av språkets egenskap av att icke vara något beständigt och fast går det aldrig att göra helt orubbliga definitioner. Jag som språkanvändare och språkvare har därför min egen tolkning av vilka led som är valensled och vilka relationer de uttrycker. Min lösning av denna uppgift skulle antagligen aldrig bli helt densamma om någon annan hade utfört den men i stora drag skulle man dra samma slutsatser.

7.4 Vidareutveckling

Syftet med examensarbetet var att ta fram underlag för valensbeskrivningar och så har också gjorts. Valensbeskrivningarna skulle sedan kunna implementeras som regler i en parser. Utifrån de data jag har kommit fram till för *utveckling* skulle man kunna utveckla regler för de olika valensrelationerna. Många av de fall där prepositionen tillhör en flerordskonstruktion finns med som egna enheter i ScarrieLex vilket gör att de enkelt går att särskilja från de övriga prepositionsfraserna.

Även data angående vilka verbalsubstantiv som är vanliga och vilka prepositioner de följs av kan användas. När man t.ex. översätter från engelska till svenska kan man

använda sig av data för att veta vilka prepositioner som mest troligen följer substantiven. Även inom språkgranskning kan det vara bra att veta vilka prepositioner som är de mest troliga för ett visst substantiv. Man kan använda sig av olika sannolikheter för hur vanligt det är att ett verbalsubstantiv följs av en preposition i allmänhet och en viss preposition i synnerhet.

Litteraturförteckning

- Andersson, Erik. *Grammatik från grunden*. Hallgren & Fallgren, Stockholm, 1993.
- Ejerhed, Eva och Källgren, Gunnel. Stockholm Umeå Corpus version 1.0, suc 1.0, 1997.
- Grishman, Ralph, Macleod, Catherine, och Meyers, Adam. *Comlex Syntax: Building a Computational Lexicon*, 1994. Computer Science Department, New York University.
- Herbst, Thomas. A valency model for nouns in english. I: *Journal of linguistics*, nummer 24, ss 265–301. University of Augsburg, 1987.
- Hindle, Donald och Rooth, Mats. *Structural ambiguity and lexical relations*, 1993.
- Liljestrand, Birger. *Så bildas orden - Handbok i ordbildning*. Studentlitteratur, Lund, 1993.
- Macleod, Catherine, Grishman, Ralph, och Meyer, Adam. *Creating a Common Syntactic Dictionary of English*. <http://nlp.cs.nyu.edu/comlex/shared-resources.ps>, 1994a. Computer Science Department, New York University.
- Macleod, Catherine, Grishman, Ralph, och Meyers, Adam. *The Comlex Syntax Project: the First Year*. <http://nlp.cs.nyu.edu/comlex/arpa94.ps>, 1994b. Computer Science Department, New York University.
- Malmgren, Sven-Göran. *Svensk lexikologi*. Studentlitteratur, Lund, 1994.
- Merlo, Paola. *Generalised PP-Attachment Disambiguation using Corpus-based Linguistic Diagnosis*. <http://acl.ldc.upenn.edu/eacl2003/papers/main/p31.pdf>, 2003. Linguistic Department, University of Genova.
- Meyers, Adam, Macleod, Catherine, och Grishman, Ralph. *Standardization of the Complement Adjunct Distinction*, 1996. New York University.
- Nikula, Henrik. *Dependensgrammatik*. Liber, Malmö, 1986.
- Ohlander, Kristina. *Lingvistisk annotering av tidningstext och extraktion av verbala konstruktionsegenskaper*. Examensarbete, Institutionen för lingvistik och filologi, Uppsala universitet, 2005. http://stp.ling.uu.se/~matsd/thesis/arch/2005_ohlander.pdf.
- Olsson, Leif-Jöran. *Utveckling av lexikala resurser för ett språkgranskningsystem för svenska*. Examensarbete, Institutionen för lingvistik och filologi, Uppsala universitet, 2004. http://stp.lingfil.uu.se/exarb/arch/2004_olsson.pdf.

PAROLE. The Swedish PAROLE Lexicon - a language engineering resource with access to morphological and syntactic information in Swedish. <http://spraakdata.gu.se/parole/lexikon/swedish.parole.lexikon.html>.

Povlsen, Claus, Sågvall Hein, Anna, och de Smedt, Koenraad. Final Project Report. http://www.hltcentral.org/usr_docs/project-source/SCARRIE/ScarrieFinal.doc, 1999. slutrapport för Scarrie-projektet, projekt nr LE3-4239.

Teleman, Ulf, Hellberg, Staffan, och Andersson, Erik. *Svenska Akademiens grammatik*. Norstedts, Stockholm, 1999.