



GÖTEBORGS UNIVERSITET
INST FÖR SVENSKA SPRÅKET

GU-ISS-2012-02

Ett svenskt konstruktikon
Utgångspunkter och preliminära ramar

Benjamin Lyngfelt
Markus Forsberg



Forskningsrapporter från institutionen för svenska språket, Göteborgs universitet
Research Reports from the Department of Swedish

ISSN 1401-5919

www.svenska.gu.se/publikationer/GU-ISS

1. Inledning

Projektet *Ett svenskt konstruktikon*¹ handlar om samspelet mellan grammatik och lexikon och går konkret ut på att utveckla en svensk konstruktionsdatabas, ett s.k. konstruktikon, som ett komplement till befintliga lexikografiska resurser. Förebilden är det engelskspråkiga *constructicon* som utformats i Berkeley, USA, i anslutning till Berkeleys FrameNet. På samma sätt är det svenska konstruktikonet ett tillägg till FrameNets svenska motsvarighet: Svenskt frasnät (SweFN).

På sikt är målet att bygga upp en storskalig elektronisk resurs, som är användbar för språkteknologiska, lingvistiska och pedagogiska syften. Vi vill särskilt täcka in sådana konstruktioner som är problematiska ur inlärarsynpunkt och har ambitionen att utveckla metoder för automatisk igenkänning av konstruktioner. Till att börja med inriktas dock arbetet mot bygga upp en pilotdatabas i liten skala, för att prova fram ett lämpligt beskrivningsformat, utveckla tekniska resurser och visa vilka möjligheter en sådan här resurs kan erbjuda.

Nedan kommer först ett avsnitt om konstruktioner (avsnitt 2), och därefter presenteras projektets bakgrund i konstruktionsgrammatik, FrameNet och Språkbanken (avsnitt 3). I avsnitt 4 behandlas planerna för själva konstruktikonet.

¹ Deltagare i projektet är Lars Borin, Markus Forsberg, Benjamin Lyngfelt (projektledare), Julia Prentice, Rudolf Rydstedt, Emma Sköldberg och Sofia Tingsell. BL:s forskning finansieras av Svenska Akademien, med ekonomiskt stöd från Knut och Alice Wallenbergs stiftelse. MF finansieras av Vetenskapsrådet, genom projektet SweFN++.

2. Konstruktioner

Språket beskrivs huvudsakligen i ordböcker och grammatikor, och det finns en hel del språkliga strukturer som lätt hamnar mellan stolarna. Ordböckerna ska beskriva ordförrådet, grammatikor inriktar sig främst på att fånga generella regelbundenheter, men var beskrivs det som är varken eller? Mönster som är för generella för att kunna knytas till enskilda ord, men för specifika för att betraktas som generella regler?

Sådana mönster brukar kallas konstruktioner. De illustreras ibland i ordböckernas språkbruksexempel eller tas upp som specialfall eller undantag i grammatikböcker, men i båda fallen som en perifer del av beskrivningen. Det skrivs också en hel del språkvetenskapliga artiklar om intressanta konstruktioner, men typiskt om enskilda konstruktioner. Samlade, storskaliga beskrivningar saknas i stort sett – än så länge.

2.1. Några exempel

Det finns konstruktioner av vitt skilda slag. En typ som har uppmärksamats en del i svensk litteratur är reaktiva konstruktioner. Hit hör t.ex. den s.k. X-OCH-X-KONSTRUKTIONEN (Lindström & Linell 2007, Norén & Linell 2007), som illustreras i (1):

- (1) a. Jag tänkte lägga in lite ny musik på mobilen (eller mja, **ny och ny** vet jag inte .. ny för mig som inte hänger med så mycket, haha)
 b. Vad har ni i lön??? [...]
Lön och lön Jag har 9506 + 3400 + resterande från soc

Mönstret är att ett uttryck (*ny* resp. *lön* ovan) från en föregående mening eller yttrande upprepas två gånger med *och* emellan, med funktionen att nyansera eller omförhandla uttryckets innebörd. X kan vara i stort sett vilket uttryck som helst. Däremot är *och* obligatoriskt, men används inte som vanligt och konstruktionen fungerar inte som vanlig samordning (jfr *äpplen och päron*, *brott och straff*).

En annan reaktiv konstruktion med liknande funktion är AUX AUX (*Aux* från eng. *auxiliary*, alltså hjälpverb). Den går ut på att en sats

inleds med samma hjälpverb två gånger, och konstruktionen ”relativiserar eller graderar modaliteten i den aktuella situationen och/eller kontrasterar den mot något pragmatiskt relevantare” (Linell & Norén 2009:90). Ett par exempel ges i (2):

- (2) a. Men nu är det ju så att jag måste till skolan 15-17 eller **måste måste** jag väl inte, men det skulle ju vara bra.
 b. – Jupp .. gillar det inte men **kan kan** jag.

En annan omskriven konstruktionstyp rör argumentstruktur. Det handlar typiskt om hur verb kombineras med olika typer av argument, och särskilt intressant är kanske kombinationer som inte ingår i verbens ”vanliga” valensmönster. I (3) illustreras en konstruktion som kan kallas REFLEXIV RESULTATIVKONSTRUKTION (Jansson 2006, Lyngfelt 2007):

- (3) a. Har du **ölat dig odödlig**?
 b. Rui Costa [...] **tunnlade sig fri** till höger i straffområdet och serverade bollen på silverbricka till Ricardo sa Pinto.
 c. Kan man **måla sig frisk**?

Strukturen är ungefär [verba sig resultat], med ett adjektiv som uttrycker resultatet av verbets aktivitet och ett reflexivt objekt som inte fungerar som ett vanligt objekt. Verbet *öla* i (a) är påtagligt intransitivt, och *dig* är snarare ett argument till *odödlig* än till *öla*. Däremot är *tunnla* i (b) och *måla* i (c) transitiva, ’tunnla någon’ resp. ’måla något’, men *sig* uttrycker varken den tunnlade eller det målade, vilka tvärtom är undertryckta i konstruktionen. I stället hör reflexiven semantiskt snarare ihop med resultaten *fri* resp. *frisk* även här.

Ytterligare en konstruktion där objektet inte fungerar som objekt är INDIREKT KAUSATIV *BORT*-KONSTRUKTION (Sjögreen 2007):

- (4) a. Kan man **äta bort sin huvudvärk**?
 b. Folkpartiet vill **undervisa bort extremhögern**.

Här är det varken huvudvärken som ska ätas (4a) eller extremhögern som ska undervisas (4b), men genom ätandet/undervisningen är det meningen att huvudvärken/extremhögern ska försvinna (*bort*).

I en argumentkonstruktion av lite annat slag är det i stället subjektet som inte fyller sin vanliga roll utan snarare fungerar som argument i en underordnad predikation. Detta fenomen kallas ibland (subjekts-) lyftning och förekommer med ett litet antal aktiva verb (t.ex. *verka*,

tyckas) och ett stort antal passiva (se Lyngfelt 2011). Det uppträder också med några verb som typiskt inte fungerar som lyftningsverb:

- (5) a. Himalayas glaciärer **hotar** att smälta.
 b. Reparationerna och utbyggnaden av Rudbeckskolan **riskerar** att skjutas på framtiden.
 c. Äldre kvinna **misstänks** ha drabbats av vanvård.

Glaciärerna hotar ingen (5a), reparationerna tar inga risker (5b) och kvinnan är inte misstänkt (5c). Däremot finns det ett hot om, en risk för resp. en misstanke om en viss händelse (se Andersson 2006).

Ett tydligt exempel på hur argumentkonstruktioner kan ha egen betydelse är rörelsekonstruktioner, som mycket väl kan uttrycka rörelse utan att innehålla rörelseverb (se t.ex. Olofsson 2010, 2011), vilket illustreras i (6):²

- (6) a. Hon bytte vatten i hinkarna och **slamrade** in i försäljningsrummet.
 b. Han **måste** in i lägenheten.
 c. [...] några par **skrattade** iväg nerför trapporna och ut på grusplanen.
 d. Han **tyskade** iväg henne.

Gemensamt för alla dessa argumentkonstruktioner är att de kan sägas manipulera argumentstrukturen på olika sätt. De kan i och för sig också förekomma som ordinarie valensmönster (t.ex. *supa sig full* och *äta sig mätt*), men ett tydligt tecken på att de utgör konstruktioner i sig själva är att de som synes också kan användas där de avviker från och ibland t.o.m. bryter mot verbens ordinarie valens.

En konstruktion på lite annan språklig nivå är *I ADJEKTIVASTE LAGET*, som ungefär betyder att ett 'fenomen har mer än önskvärt av den egenskap adjektivet uttrycker' (se exemplen i 7a-b). Strukturellt utgörs den av en prepositionsfras med *i*, vars rektion består av substantivet *laget* och en adjektivfras i superlativ. Konstruktionen skiljer sig från en vanlig prepositionsfras med samma utseende dels genom sin speciella betydelse, dels genom att den inte kan konstrueras med artikel (jfr 7c).

- (7) a. Presentationen var **i längsta laget**.
 b. Tåget går **i tidigaste laget**.
 c. Jag hamnade alltid i (det) sämsta laget.

² Man kan också tänka sig en lexikal lösning, där rörelsebetydelsen projiceras från adverbena *in* resp. *iväg*, i stället för från en grammatisk rörelsekonstruktion (jfr Rydstedt u.u.).

En intressant konstruktion på satsnivå är *VAD GÖR X Y?*, vars engelska motsvarighet har beskrivits av Kay & Fillmore (1999). Konstruktionen ser ut som en fråga, men har en annan innebörd än frågans bokstavliga betydelse och fungerar snarast som ett uttryck för indignation (8a). Ibland kan dock de båda betydelseerna flyta samman (8b). Diskrepansen mellan konstruktionen och den bokstavliga frågebetydelsen har för övrigt gett upphov till ett stort antal vitsar (8c).

- (8) a. Vad gör du i min säng?
 b. Vad gör polisen på min blogg?
 c. Vad gör den här flugan i min soppa?

Bland andra konstruktioner kan exempelvis snabbt nämnas hur bl.a. uttryck för frekvens och hastighet har två varianter, antingen konstruktion med *i* och bestämd form (*två gånger i veckan, 1500 varv i minuten*) eller med *per* och obestämd form (*två gånger per vecka, 1500 varv per minut*). Vidare konstrueras tidsuttryck ofta med prepositionsfras (*i förrgår, i somras, på måndag, på 70-talet, om en vecka*), men i svenskan bara med nominalfras vid datumuttryck (*sista ansökningsdag är 20 januari; jag är född (den) 23 december*). Det sistnämnda har visat sig problematiskt för andraspråksinlärare (Prentice 2011); jfr t.ex. eng. *I was born on the 23rd of December*.

2.2. Summering

Konstruktionerna ovan kännetecknas av olika morfologiska, lexikala, syntaktiska, prosodiska, semantiska och pragmatiska egenskaper, i en eller annan kombination. Vissa delar är mer eller mindre fasta, andra är mer variabla. Konstruktionerna låter sig inte enkelt ordnas i språkliga nivåer, eftersom samma konstruktion ofta har särdrag från olika nivåer, t.ex. morfologi, syntax och pragmatik. Det de alla har gemensamt är att de på ett eller annat sätt är konventionaliserade förbindelser mellan form och innehåll. De är alltså tecken i saussureansk mening.

Just konstruktionerna i avsnitt 2.1 utmärks också av att de på ett eller annat sätt möjliggör avvikelser från mer generella mönster. Detta är dock ingen förutsättning, utan både mer generella och mer specifika strukturer kan också betraktas som konstruktioner, t.ex. *PÅSTÅENDE-SATS*, *VERBFRAS* och *VERBALSUBSTANTIV* å ena sidan – och *VORE*, *MINST* och *HAMNA PÅ EFTERKÄLKEN* å den andra. Inom konstruktionsorienterade språkteorier betraktas ofta hela språket som ett hierarkiskt

ordnat nätverk av konstruktioner, där mer specifika strukturer är undergrupper till mer generella (se avsnitt 3.1).

Generella konstruktioner brukar dock som sagt beskrivas i grammatikor och de mest specifika i ordböcker, så det är de däremellan som är i störst behov av en samlad beskrivning i ett konstruktikon.

3. Bakgrund

Konstruktikonet är ett möte mellan å ena sidan konstruktionsgrammatik och å andra sidan datoriserad lexikografi så som den yttrar sig i dels Berkeleys FrameNet, dels språkbankstraditionen i Göteborg. De båda sistnämnda förenas i projektet Svenskt Frasnät++, till vilket konstruktikonet utgör en påbyggnad.

3.1. Konstruktionsgrammatik

Språket betraktas ofta som ett system med två delar: ett lexikon och en grammatik.³ Grammatiken ses då typiskt som en uppsättning generella regler, varför enskildheter hänförs till lexikonet. Det sistnämnda innebär att fasta fraser och idiom räknas till lexikonet, fast de består av flera ord. Det innebär också att avvikelser från grammatikens regler gärna förklaras som lexikalt styrda undantag – alternativt hänförs till misstag, kreativitet eller helt enkelt avfärdas som perifert och ointressant.

Inom konstruktionsgrammatik (Fillmore, Kay & O'Connor 1988, Fillmore & Kay 1995, Goldberg 1995, 2006, Croft 2001, Boas & Sag u.u., Hoffman & Trousdale u.u. m.fl.) tar man avstånd från denna sorts distinktion mellan lexikon och grammatik, liksom från uppdelningen av språkliga fenomen som centrala eller perifera. Fenomen som är generella, specifika och alla grader däremellan antas vara lika naturliga delar av språket, och just denna varierande fin-/grovkornighet hos språkliga mönster betraktas som en utmärkande egenskap hos språk. Så i stället för uppdelningar i olika språkliga nivåer och moduler ses alla förbindelser mellan form och innehåll som samma sorts grundläggande entitet: tecken, och alla konventionaliserade sådana förbindelser betraktas som konstruktioner. Följaktligen ses språkssystemet som ett ordnat inventarium av konstruktioner, av varierande generalitetsgrad.

³ Vi bortser för tillfället från huruvida fonologi ska ses som en del av grammatiken eller som en egen del, liksom från frågan om hur språkssystemet förhåller sig till språkanvändningen.

Konstruktionsgrammatik har ambitionen att redogöra för hela språket, från det lexikalt specifika till det mycket generella. Emellertid är modellen främst känd för att ha tillämpats på fenomen i gränslandet mellan lexikon och grammatik och uppfattas ibland just som en gränslandsteori. Vid närmare påseende visar sig dock detta gränsland vara tämligen stort. Fenomen som typiskt räknas till lexikonet, t.ex. idiom, visar sig ofta ha både viss grammatisk struktur och tillåta viss variation (se Sköldberg 2004) – och grammatiska regler är aldrig helt generella, utan regleras alltid av mer eller mindre specifika begränsningar.

Vidare finns det en metodologisk poäng i att utgå från ”gränslandsfenomen”. Modeller som utgår från antingen grammatik eller lexikon får typiskt problem med att hantera drag av det omvända slaget. Genom att inrikta sig på strukturer med både grammatiska och lexikala särdrag tvingas man utveckla ett beskrivningsformat som integrerar information från båda håll, och som rimligen är tillämpligt även på strukturer som främst har särdrag av det ena slaget. Genom att såväl grammatiska som lexikala strukturer betraktas som konstruktioner kan man använda samma grundläggande beskrivningsmodell över hela skalan (Fillmore, Kay & O’Connor 1988).

Konstruktioner kan ha olika grad av schematicitet, enligt tabell 1 (jfr Fried u.u.).⁴

Tabell 1. *Konstruktioner med olika grad av schematicitet*

Schematicitetsgrad	Exempel
fyllda och fixerade	<i>så att säga, light, röda hund</i>
fyllda och delvis flexibla	<i>lägga[tempus] (alla) papperen/korten på bordet, sista minuten-resa[num/spec]</i>
delvis schematiska	<i>i ADJ-aste laget,</i> MINUTTAL <i>i/över</i> TIMTAL
helt schematiska	S[NP VP], VERB[tempus], VERBA <i>sig</i> RESULTAT

Helt fyllda konstruktioner brukar räknas till lexikonet, helt schematiska till grammatiken och de delvis schematiska hanteras på olika sätt, eller

⁴ VERBA *sig* RESULTAT är helt schematisk, eftersom *sig* inte enbart avser ordformen *sig* utan representerar reflexiva pronomen i allmänhet (man kunde alltså lika gärna skriva REFL, men *sig* bedömdes mer lättbegripligt).

inte alls, beroende på utgångspunkt. En typisk konstruktionsgrammatisk beskrivning är en särdragsanalys, där man i enklaste fallet redovisar konstruktionens form respektive innehåll, och i mer komplexa analyser urskiljer fonologiska, morfologiska, syntaktiska, pragmatiska osv. särdrag – vilket kan göras oavsett var på skalan en konstruktion befinner sig.

Det bör i sammanhanget nämnas att konstruktionsgrammatik utgör en ganska heterogen tradition. Dels finns det flera teoribildningar som är konstruktions- eller teckenbaserade, t.ex. kognitiv grammatik (Langacker 1987, 1991), dels finns det flera varianter av konstruktionsgrammatik:

- *Berkeley Construction Grammar* (bl.a. Fillmore & Kay 1995; för en introduktion, se t.ex. Fried & Östman 2005), den ursprungliga huvudfåran som andra varianter bygger på⁵
- *Cognitive Construction Grammar* (Goldberg 1995, 2006), som är mer kognitivt och mindre formalistiskt orienterad
- *Embodied Construction Grammar* (Bergen & Chang 2005), som är mer inriktad mot datorimplementering
- *Radical Construction Grammar* (Croft 2001), med mer fokus på språktypologi
- *Sign-Based Construction Grammar*; SBCG (Sag 2011, Boas & Sag u.u.), en hybrid mellan konstruktionsgrammatik och den likaledes teckenbaserade Head-Driven Phrase Structure Grammar (HPSG; Pollard & Sag 1994), mer formalistisk än övriga varianter.

Teoretiskt ligger de ganska nära varandra, delar många grundantaganden och är i huvudsak kompatibla. Varianterna skiljer sig åt främst vad gäller metaspråk, alltså beskrivningsformat, och i viss mån prioriteringar: man fäster olika vikt vid formell precision, implementerbarhet, integrering med kognitionsforskning osv. Det finns också en skillnad i hur man förhåller sig till lexikon.

Konstruktionsgrammatik är på sätt och vis både lexikalistisk och anti-lexikalistisk. Lexikalistisk i så måtto att även syntaktiska mönster beskrivs som tecken, i likhet med hur man brukar beskriva ord, och ett språks inventarium av konstruktioner antas vara strukturerat ungefär som man brukar tänka sig ett ordförråd. Icke-lexikalistisk i det avseendet att en del mönster som man tidigare brukat koppla till ordförrådet nu

⁵ Har även kallats *Standard*, *Vanilla*, *Vintage* och *Unification-Based Construction Grammar* – eller helt enkelt *Construction Grammar*. Specificerande bestämmingar används bara när man behöver kontrastera inriktningen mot andra varianter.

kan beskrivas mer schematiskt som konstruktioner, i synnerhet när det gäller argumentstruktur (se t.ex. Goldberg 1995). Det är på den senare punkten varianterna skiljer sig åt. En del fäster större vikt vid lexikon, eller *lexikala konstruktioner*, medan andra knyter mer information till syntaktiska konstruktioner. Mest uttalat lexikalistisk är SBCG, och minst lexikalistisk är Radical Construction Grammar.

Eftersom både Berkeleys *constructicon* och det svenska konstruktikonet utvecklas som tillägg till lexikografiska resurser (FrameNet resp. SweFN) ligger det nära till hands med en mer lexikalistisk hållning. Vad som visar sig lämpligast i praktiken återstår dock att se.

Vidare har olika sorters konstruktionsgrammatik som sagt delvis olika metaspråk. Samtliga beskriver konstruktioner som kombinationer av formella och innehållsliga särdrag, men beskrivningarna är mer eller mindre detaljerade och formaliserade. För att möjliggöra beskrivningar i stor skala strävar vi i arbetet med konstruktikonet mot ett så enkelt analysformat som möjligt (se vidare avsnitt 4).

3.2. FrameNet och Språkbanken

FrameNet är en lexikografisk databas som bygger på s.k. ramsemantik (eng. *Frame Semantics*, bl.a. Fillmore 1982).⁶ Principen är att lexikal betydelse beskrivs utifrån semantiska ramar – också kallade *scener*, *situationer* m.m. – där berörda deltagande element ingår. En ram som KOMMERSIELL TRANSAKTION kan t.ex. tänkas inkludera en köpare, en säljare, vara/varor, pris och betalning(-smedel). Denna rambetydelse kan instansieras av flera olika ord: t.ex. verb som *köpa*, *sälja* och *betala*, substantiv som *köpare*, *säljare* och *köp* och t.o.m. fasta fraser som *till salu*. Även adjektiv som *billig* och *dyr* knyter an till ramen. Å andra sidan kan samma ord uppträda i flera olika ramar; verbet *sparka* ingår i en helt annan situation (ram) i betydelsen 'avskeda' än i betydelsen 'träffa med foten'.

I FrameNet struktureras ordförrådet både utifrån ramar och utifrån lexikala enheter (ord). Betydelser definieras i ramarna, där man också karakteriserar ramelementen (deltagare som köpare, säljare, pris osv.), räknar upp vilka lexikala enheter som instansierar ramen samt redovisar ett antal exempel, som är annoterade (uppmärkta) så att det framgår vilka uttryck som instansierar ramen resp. de olika ramelementen. Annoteringen fungerar dels som en pedagogisk illustration, dels som ett

⁶ FrameNet är tillgängligt online på <<http://framenet.icsi.berkeley.edu/>> och beskrivs närmare i Ruppenhofer m.fl. (2010).

viktigt led i språkteknologiska tillämpningar där taggningen t.ex. kan möjliggöra automatisk identifiering av olika uttryck och funktioner.

Väljer man i stället lexikala enheter som ingång redovisas vilka olika ramar orden (eller deras underbetydelser) ingår i, vilka av ramelementen orden uppträder tillsammans med och hur dessa ramelement realiseras. Med ett verb som *köpa* uttrycks t.ex. köparen som nominalfras (och subjekt), men med *sälja* i en adverbial prepositionsfras (*till ngn*) eller inte alls. Sådan information redovisas både som uttrycksmönster, ung. valensramar, och som specifik realisering i enskilda exempel.

Ramarna i FrameNet är ordnade i ett hierarkiskt nätverk, där mer specifika ramar är underordnade mer generella. Även sådan övergripande information är fritt tillgänglig, så att man för varje ram kan se vilka andra ramar den står i relation till.

FrameNet är en resurs med flera förtjänster, såsom att genom ramarna kunna relatera ord med besläktad betydelse och att genom ramelementen fånga argumentstruktur och valensmönster. Men det finns förstås också begränsningar. En viktig sådan är att ramarna endast avser lexikal betydelse. Vid annotering av autentisk text stöter man snart på strukturer vars innebörd inte går att fånga lexikalt: konstruktioner.

Följaktligen har man i Berkeley börjat utveckla ett *constructicon* som komplement till FrameNets lexikala resurser (Fillmore 2008, 2009, Fillmore m.fl. u.u.). Formatet är likartat, och det nya tillägget ska kunna integreras med FrameNet i övrigt. Än så länge är det dock ganska litet till omfånget och innehåller drygt 70 påbörjade konstruktionsbeskrivningar, varav ett 50-tal är någotsånär fullständiga.

I stället för ramelement anges konstruktionselement, definierade efter deltagarroll eller frastyp. En konstruktikonpost innehåller följande fält:

- *Namn*: konstruktionens namn, oftast i lingvistiska termer
- *Definition*: fritextdefinition där alla konstruktionselement ingår explicit
- *Ram och Arv*: koppling till relevant ram och arvsrelation till annan konstruktion redovisas om relevant
- *Strukturbeskrivning*: schematisk redogörelse för konstruktionens form⁷
- *Konstruktionselement*: (lokala) definitioner av ingående konstruktionselement, inklusive CEE (*construction evoking element*) dvs. lexikalt specifika konstruktionselement

⁷ Detta fält syns inte i gränssnittet för den nätbaserade betaversionen, men redovisas i t.ex. Fillmore et al. (u.u.).

- *Exempel*: annoterade exempel, med alla konstruktionselement taggade.
- *Referens*: hänvisning till lingvistiska arbeten där konstruktionen behandlas

Berkeley's FrameNet (och *constructicon*) beskriver enbart engelska, men det finns numera FrameNets även för flera andra språk, åtminstone under uppbyggnad. En ambition är att de olika språkens FrameNet-resurser i görligaste mån ska vara kompatibla så att ramarna kan vara till nytta exempelvis i översättningssystem.

Sveriges motsvarighet kallas Svensk frasnät (SweFN), uppbyggt inom projektet *Swedish FrameNet++* (Borin m.fl. 2010).⁸ Plustecknen indikerar att detta inte är någon isolerad produkt, utan SweFN integreras med andra material vid Språkbanken i Göteborg, såväl korpusar som lexikala resurser. Språkbankens material har växt fram under decennier av korpus- och ordboksarbete, och en stor del av de resurserna är nu delvis integrerade. Kärnan i detta system är det elektroniska lexikonet Saldo (*Swedish Associative Thesaurus*; Borin, Forsberg & Lönngrén 2008, Borin & Forsberg 2009),⁹ vars lexikonenheter mappas till motsvarande enheter i andra material, inklusive SweFN. Rambeskrivningarna i SweFN har följande format:

- *Namn*: samma namn som motsvarande ram i FrameNet (med undantag för några nya ramar)
- *Domän* och *Semantisk typ*: övergripande ontologisk information; saknas specifika uppgifter anges Gen (eng. *general*, dvs. allmän)
- *Definition*: samma som i FrameNet; redovisas inte i betagränssnittet
- *Ramelement*: uppdelade i kärn- och periferielement; samma definitioner som i FrameNet
- *Sammansättningar*: fält för sms-mönster resp. -exempel; denna del av analyserna är ännu inte systematiskt genomförd
- *LU*: de lexikala enheter som instansierar ramen, redovisade ordklassvis som Saldo-enheter
- *Nya LU*: ett fält där man kan föreslå nya LU:s, som ännu inte ingår i Saldo
- *Exempel*: annoterade exempel, med LU och ramelement taggade
- *Kommentar*: ev. kommentar till analysen, ofta avseende relationer till besläktade ramar

⁸ SweFN är tillgängligt online på <<http://spraakbanken.gu.se/swefn/>>.

⁹ Se även <<http://spraakbanken.gu.se/eng/saldo/>>.

- *Praktisk info*: fält där det anges vem som har skapat beskrivningen, när den skapades och när den senast modifierades.

Det svenska konstruktikonet blir ett direkt tillägg till SweFN, till stor del med Berkeleys engelska *constructicon* som förebild men också anpassat för integrering med Språkbankens övriga material via Saldo.

4. Konstruktikonet

Konstruktikonet (Cxn)¹⁰ är en databas över svenska konstruktioner, med det deskriptiva syftet att ge en samlad redogörelse för konstruktionsspecifika uttrycksmönster och det funktionella syftet att utgöra en resurs dels för språkteknologiska applikationer, dels för lingvistiska och pedagogiska tillämpningar. Språkteknologin kräver formell exakthet, och lingvistikens fordrar deskriptiv adekvathet. Pedagogiken förutsätter begriplighet, men avsikten är inte att Cxn ska användas direkt som pedagogiskt redskap utan endast kunna utgöra ett underlag för sådana. Den direkta målgruppen är alltså experter på språk och/eller språkteknologi, inte allmänheten.

Databasens omfattning blir med nödvändighet en fråga om ekonomiska resurser, men ambitionen är att på sikt bygga upp ett Cxn i stor skala. I princip är alla typer av konstruktioner intressanta, men i praktiken lär fokus ligga på relativt specifika strukturer i gränlandet mellan lexikon och grammatik. Detta dels därför att det främst är där det saknas samlade beskrivningar, dels därför att Cxn utvecklas i anslutning till en lexikografisk resurs.

Nedan behandlas först konstruktionsbeskrivningarnas format och innehåll (4.1), konkordansverktyget Korp (4.2) och därefter frågor som rör själva konstruktikonets struktur (4.3). Betydelsen för berörda ämnesområden kommenteras i avsnitt 4.4.

4.1. Beskrivningsformat

Liksom i Berkeleys *constructicon* redovisas konstruktionerna med fritextdefinitioner, strukturbeskrivningar, konstruktionselement och annoterade exempel. Avsikten är att hålla analyserna relativt enkla, för att öka tillgängligheten, minska tidsåtgången och inte ställa alltför höga krav på analytikerna (arbetet ska delvis kunna utföras av studenter).

¹⁰ För att underlätta igenkänning bibehålls vedertagna engelska förkortningar: Cxn (konstruktikon), Cx (konstruktion), Ce (konstruktionselement), Cxt (konstrukt, det uttryck som instansierar konstruktionen).

Detta går delvis emot praxis i grammatikforskningen, där djup och detaljrikedom ofta prioriteras framför enkelhet:

[...] the meaning of linguistic expressions cannot be characterized by means of short, dictionary type definitions. (Langacker 1987:489)

Langacker har i och för sig rätt i att betydelse är komplex och mångfacetterad, men de approximationer som presenteras i ordböcker har ändå visat sig värdefulla i många sammanhang. Förhoppningen är att ungefär motsvarande komplexitetsgrad ska visa sig fruktbar även i ett konstruktikon. Djupare analyser är både utrymmes- och framför allt tidsödande och därför svåra att genomföra i stor skala.

Följaktligen prioriteras enkelhet och (potentiell) storskalighet i möjligaste mån. Samtidigt tillkommer ändå viss komplexitet jämfört med lexikala definitioner, eftersom man för syntaktiska konstruktioner även måste redovisa konstituentstruktur.

I figur 1 nedan presenteras (något förenklat) ett preliminärt format för konstruktionsbeskrivningarna tillsammans med en exempelkonstruktion. I bilaga 1 visas samma exempel som det ser ut i en nu tillgänglig utvecklingsversion av konstruktikonet,¹¹ vars preliminära gränssnitt har utformats av Jonatan Uppström (systemutvecklare vid Språkbanken). Den underliggande analysen i xml-format redovisas i bilaga 2.

¹¹ <<http://spraakbanken.gu.se/swe/resurs/konstruktikon/utvecklingsversion>>

Fält	Typ av information	Exempelanlys
id	konstruktionens namn	indirekt_kausativ_bort
type	Cx, Cat eller Role	Cx
cat	frastyp	vbm
inheritance	eventuell arvsrelation	
evokes	ev. aktualiserad ram i FrameNet	Causation_scenario
definition	fritextdefinition där konstruktionselementen ingår och markeras <taggas>	Någon utför en action som indirekt leder (eller antas leda) till att en företeelse försvinner
structure	morfosyntaktisk struktur, ordklass, frastyp, böjningsinfo och/eller lexem (lu i Saldo)	vb bort..1 NP
cee	ev. lexikalt specifika led i cx, angivna som Saldo-enheter	bort..1
coll	ev. led som är påfallande vanliga i konstruktionen	spela..2, spela..6
construction_elements	konstruktionselement (led som ingår i eller förutsätts av cx); ange roll, kategori, böjningsinfo och/eller lexem	<role name="Activity" cat="vb"/> <role name="Result" lu="bort..1"/> <role name="Theme" cat="NP"/> <role name="Agent" cat="NP"/>
examples	autentiska exempelmeningar där cx-elementen taggas; försök välja exempel som tillsammans speglar konstruktionens variationsmönster	Kan man äta bort sin huvudvärk? Folkpartiet vill undervisa bort extremhöger. Han har ändå ingenting att göra nu, sedan sommarvikarien spelat bort kommunens kassa på bingo.
comment	eventuella kommentarer till analysen	Strukturanalysen är förenklad. En semantiskt mer rättvisande analys är: vb [bort NP]
reference	hänvisning till analysens källa	Sjögreen, Christian d.y. (2007). Går det att äta bort sin huvudvärk? - en studie av den indirekt kausativa bort-konstruktionen i svenskan. (D-uppsats) Inst f svenska språket, Göteborgs universitet

FIGUR 1: Preliminärt beskrivningsformat med exempelanlys

id: Konstruktionens namn ska framför allt identifiera konstruktionen i databasen (och integrerade material), varför huvudkravet är att det ska vara unikt inom konstruktikonet. I övrigt kan det se ut i princip hur som helst, förutom att formatering, diakritiska tecken o.d. ska undvikas. Det är förstås önskvärt att namnet är belysande för konstruktionen, då det spar tid om samma namn kan användas även i gränssnittet, men det är

inte givet om konstruktionen bäst presenteras utåt med t.ex. en språkvetenskaplig term eller ett illustrerande exempel – kanske både och.

type: Typbenämningen på konstruktioner är Cx. Andra typer i databasen är Role (deltagarroll) och Cat (grammatisk kategori). Även dessa utgör egna poster enligt samma grundformat, men har förenklade beskrivningar där endast vissa fält är ifyllda.

cat: Här anges formkategori i Saldo-format, t.ex. *vb* för verb och *nn* för substantiv. För flerordsfraser tillfogas ett *m* (för multi-word expression), t.ex. *vbm*. Traditionella frasbeteckningar (XP) är neutrala med avseende på en- eller flerordsstatus; *VP* markerar alltså att både *vb* och *vbm* är möjliga (detta ingår *inte* i Saldo-formatet, utan har lagts till för konstruktikonets räkning).

inheritance: Här anges om en konstruktion ärver egenskaper från en överordnad, mer generell konstruktion. Fältet är alltså endast tillämpligt för hierarkiskt ordnade grupper av konstruktioner.

evokes: Om konstruktionens betydelse aktualiserar en rambetydelse anges detta här. Det är dock inte alltid någon sådan koppling finns, dels eftersom FrameNet inte är fullständigt, men framför allt därför att ramar hör samman med lexikal betydelse vilket inte alla konstruktioner matchar.

definition: Definitionen skrivs i fritext, med alla konstruktionselement (Ce) inkluderade och taggade, i enlighet med definitionerna i FrameNet och Berkeleys *constructicon*, fast på svenska (se vidare nedan).

structure: Strukturbeskrivningen är en enkel schematisk redogörelse för konstruktionens konstituentstruktur, där ordklass och specifika lexem anges enligt Saldo-format och frastyp enligt vedertagen standard (NP, PP osv.)

cee: Ev. lexikalt specifika konstruktionselement anges här, med Saldo-beteckning. Detta fält är främst till för att underlätta språkteknologiska tillämpningar, t.ex. (semi-) automatisk identifiering av konstruktioner i text.

coll: Ev. kollostruktionella led, dvs. uttryck som inte är fixerade men ändå särskilt vanliga i konstruktionen, anges med Saldo-beteckning. Ibland finns det flera sådana led, ibland inget. Observera att det är

lexem som anges, inte lemman (därav två olika *spela* i exemplet ovan). Dessa angivelser bygger i regel inte på fullödiga kollostruktionella analyser (jfr Stefanowitsch & Gries 2003), utan avser helt enkelt led som uppfattats som påtagligt vanliga.

construction_elements: Konstruktionselementen (Ce) är uppdelade i interna och externa, där de interna ingår i konstruktionen och de externa i dess valens. Varje Ce karakteriseras med ett slags särdragsanalys, där roll, kategori, böjningsinformation och/eller lexem redovisas, beroende på vad som är tillämpligt. Alla sådana egenskaper definieras unikt någon annanstans i databasen (eller anslutna resurser) – lexem och böjning i Saldo, roller och kategorier som egna poster i konstruktikonet. Den mest framträdande av dessa egenskaper märks *name* och är den tagg som Ce märks upp med i definition och exempel.

examples: För varje konstruktion redovisas några autentiska, annoterade exempel, typiskt hämtade ur Språkbankens korpusar. Exempelen är en viktig del av beskrivningen, dels som illustration, dels som underlag för språkteknologiska ändamål. Vid annoteringen taggas konstruktionen som helhet med sitt id, och varje konstruktionselement taggas för sig. Dessutom markeras CEE. Exempelen bör vara representativa och varierade och helst spegla typiska variationsmöjligheter.

comment: Här kan man lägga in vilken information som helst som inte passar in i grundformatet. Det kan gälla sådant som bruklighet (t.ex. om konstruktionen är särskilt vanlig i passiv), analyskommentarer, relationer till andra konstruktioner som bör beskrivas osv.

reference: Om analysen har en tydlig källa, där den som är intresserad av konstruktionen kan läsa mer, läggs referensen in här. Uppsatser som skrivs i anslutning till projektet publiceras typiskt online, så att man enkelt ska kunna komma åt dem i fulltext direkt från databasens gränssnitt.¹²

För att harmoniera med FrameNet och Berkeleys *constructicon* bibehålls många engelska beteckningar, men konstruktionernas namn och definitioner är ändå på svenska. Dels blir det mer lättarbetat så, dels är avsikten att definitionerna så småningom ska kunna användas i något

¹² För en förteckning över publikationer och uppsatser i anslutning till projektet, se: <http://www.svenska.gu.se/forskning/forskningsprojekt/konstruktionsgrammatik/>

gränssnitt mot svenska (normal-) användare. Samtidigt har projektet även en tvärspråklig dimension. Dels diskuteras samarbete med både Berkeley och ett planerat brasilianskt konstruktikonprojekt, dels är flera tänkbara lingvistiska och datalingvistiska tillämpningar tvärspråkliga.

I FrameNet går det tvärspråkliga samarbetet via ramarna. De lexikala enheterna är språkspecifika, men genom att man antar samma ramar för olika språk kan språkspecifika FrameNet-resurser ändå relateras till varandra. I konstruktikonet, däremot, är de centrala enheterna inte ramar utan konstruktioner, och dessa är till stor del språkspecifika. Alltså behövs någon form av gemensamt metaspråk, rimligtvis engelska med formella inslag.

Till att börja med utformas konstruktikonet dock främst med avseende på svenska. En tvärspråklig representation kan relativt enkelt läggas till i efterhand, och formatet för denna utvecklas i samråd mellan projektgrupperna i Göteborg, Berkeley och Juiz de Fora (Brasilien).

4.2. Korp

En viktig utgångspunkt är att konstruktionsbeskrivningarna bygger på analys av faktiskt språkbruk, här främst i form av korpusstudier med Språkbankens konkordansverktyg Korp. I Korp kan man söka i 74 korpusar av olika slag, som sammanlagt omfattar ca en miljard löpord, genom samma sökgränssnitt. I de flesta fall begränsas sökningarna till en del av korpusarna, eftersom materialet annars är för stort för allt annat än mycket specifika söksträngar. (Se Borin, Forsberg & Roxendal 2012.)

Materialet i Korp är automatiskt annoterat för sådant som ordklass, lemma, böjningskategori, dependensrelation samt olika egenskaper hos den text respektive exempel är hämtat ur. Man kan också specificera led i sökningen till delar av ord (t.ex. ord som slutar på *-chock*). Samma led kan ges flera specifikationer, additivt eller alternativt. Om man additivt specificerar att ett led dels ska vara adjektiv, dels ska sluta på *-bar*, fångar man adjektiv som *försumbar*, *ärbar* och *delbar* – men utesluter å ena sidan andra ordklasser (substantivet *ämbur* och verbformen *jobbar*) och å andra sidan adjektiv med annat slut (*glad*, *surrealistisk*, *hutilös*).

Med en alternativ specificering kan man t.ex. ange att ett visst led ska vara antingen adjektiv eller adverb, såsom illustreras med en sökning på [*så ADJEKTIV/ADVERB som möjligt*] i figur 2:

Sökresultaten kan sedan redovisas i flera olika format. Med defaultinställningen KWIC får man helt enkelt upp en lista med exempel, mening för mening. I figur 2 redovisas i stället statistik, där det bl.a. framgår att de särklassigt vanligaste adjektiven och adverbena i [*så ADJ/ADV som möjligt*] är *mycket*, *snart*, *fort* och *snabbt*. Man kan också utläsa i vilka korpusar beläggen finns, vilket kan indikera om en konstruktion är knuten till vissa genrer eller en viss tidsperiod.

Kombinationen av storlek, olika sorters korpusar, rik annotering och flexibelt sökgränssnitt gör Korp till ett utmärkt redskap för empiriska studier av konstruktioner. Som exempel på en konstruktikonanalys baserad på sökningar i Korp, se Börjesson (2012).

4.3. Avgränsningar och nätverksstruktur

Även om i princip alla konventionaliserade språkliga mönster kan betraktas som konstruktioner inriktar vi oss som sagt i första hand på sådana som är delvis schematiska (se tabell 1 i avsnitt 2.1 ovan). Visst kan man betrakta ord som *typ* och *hen* eller fasta fraser som *kalla fötter* och *det vill säga* som konstruktioner, men frågan är vad man kan fånga med det som man inte lika gärna kan hantera med en lexikal beskrivning.¹³ Vi prioriterar alltså mönster som är delvis produktiva eller åtminstone tillåter ett visst mått av variation.

Vidare är en viktig avgränsning att konstruktioner utgör förbindelser mellan form och innehåll. Detta innebär å ena sidan att t.ex. 'aspekt' inte kan vara en konstruktion, eftersom det enbart är en funktion. Det är först när funktionen kombineras med en form som det blir en konstruktion, t.ex. den imperfektiva aspektkonstruktionen *HÅLLA PÅ OCH/ATT VERBA*. Å andra sidan innebär det att ett formmönster bara utgör en konstruktion om det samtidigt hänger ihop med en viss betydelse. Detta kan illustreras med följande exempelsekvenser:

- (9) a. spela bort arvet, slarva bort nycklarna, lämna bort hunden, springa bort 10 kg
 b. skrika sig hes, växa sig stor, känna sig trött, byta sig ledig, hyra sig i form

I (9a) exemplifieras s.k. INDIREKT KAUSATIV *BORT-KONSTRUKTION* (Sjögreen d.y. 2007, se ex. (4) i avsnitt 1), som har strukturen [V bort NP], betyder ungefär 'verbhandlingen orsakar att NP försvinner' och utmärks av en indirekt relation mellan verb och objekt. Det är så att

¹³ I princip kan ord ses som lexikala konstruktioner.

säga inte arvet som spelas eller 10 kg som springs. Däremot är det faktiskt så att hunden lämnas, dvs. *hunden* fungerar som ett vanligt objekt till *lämna*. Därmed är det inte fråga om en *indirekt* kausativrelation. Följaktligen passar inte *lämna bort hunden* in i konstruktionen, utan hör i stället till en mer ordinär konstruktion med partikeladverbial.¹⁴

I (9b) ser vi i stället exempel på en REFLEXIV RESULTATIVKONSTRUKTION (Jansson 2006, Lyngfelt 2007), där predikativet uttrycker ett resultat av verbhandlingen – bli hes genom att skriva eller stor genom att växa. Här är det *känna sig trött* som inte passar in, eftersom det inte betyder att man blir trött genom att känna. Sedan kan man diskutera om *hyra sig i form* hör dit. Exemplet (*Hyr dig i form!*) kommer från en annons, där den som hyr en tv erbjuds en träningscykel på köpet; det finns alltså en indirekt koppling där hyrandet ska leda till att man kommer i form.

Det finns dock också en skillnad i form, nämligen att resultatet uttrycks i en prepositionsfras i stället för en adjektivfras som i de andra exemplen. I Goldbergs (1995) analys av motsvarande konstruktion i engelskan definieras resultatet som en adjektivfras, varför strukturer med liknande betydelse men prepositionsfras hänförs till en annan konstruktion. I Lyngfelt (2007) räknas däremot även prepositionsfraser in, förutsatt att de har rent resultativ betydelse.¹⁵ Jansson (2006) avgränsar i praktiken sin undersökning till adjektiv, men talar samtidigt om exempelvis *hyra sig i form* som en resultativkonstruktion och avstår från att explicit ta ställning i frågan.

Exemplen i (9) illustrerar också några andra avgränsningsfrågor. Hör t.ex. *skrika sig hes* och andra resultativer i (9b) till samma konstruktion som *måla stugan röd*? Jansson (2006) räknar dem som samma konstruktion, medan Lyngfelt (2007) antar en särskild reflexiv konstruktion. När det gäller (9a) är frågan om man ska anta en särskild *bort*-konstruktion eller om man hellre ska utgå från en mer generell INDIREKT KAUSATIV PARTIKEL-KONSTRUKTION. Här bygger avgränsningen till just *bort* från början på en praktisk avgränsning i den uppsats där konstruktionen först beskrevs (Sjögreen d.y. 2007). Först när man

¹⁴ En mer subtil avgränsningsfråga gäller *slarva bort nycklarna*. I analysen i figur 1 ovan specificeras det handlande verbargumentet till Agent, dvs. ett medvetet handlande subjekt, och slarvar bort något gör man troligen inte med vilje.

¹⁵ Alternativet är en rörelsekonstruktion, som i *masa sig till jobbet*. Med 'rent resultativ betydelse' menas att man inte ens metaforiskt kan läsa in en rörelsebetydelse (jfr *kämpa sig till toppen* i sammanhang som inte gäller bergsklättring).

har jämfört med andra partiklar går det att ta ställning till om en högre generalitetsnivå är tillämplig.

I det senare fallet behöver inte det ena utesluta det andra, utan *bortkonstruktionen* kan behandlas som en mer specifik variant av partikel-diton. Faktum är att man tämligen ofta kan dela upp en konstruktion i flera – eller, tvärtom, gruppera flera konstruktioner i en mer generell. Inom CxG betraktar man språket som ett hierarkiskt nätverk av konstruktioner, där mer specifika konstruktioner utgör instanser av mer generella – precis som konkreta yttranden (s.k. konstrukter) utgör instanser av de konstruktioner de realiserar. Vilken abstraktionsnivå som är mest relevant är inte givet på förhand. Det är denna möjlighet till generaliseringar på olika nivåer som gör det möjligt för CxG att hantera alla typer av språkliga mönster från generella regler till specifika lexem.

Som exempel kan vi utgå från det som ibland kallas nominalfraser med adjektiviskt huvudord (se t.ex. Prytz 2000, SAG 3:40ff.). Här antar Fillmore m.fl. (u.u.) tre olika konstruktioner: ADJECTIVE-AS-NOMINAL.HUMAN (10), ADJECTIVE-AS-NOMINAL.ABSTRACT (11) och ADJECTIVE-AS-NOMINAL. ANAPHORIC (12). Motsvarande konstruktion finns även i svenskan (13a–c).¹⁶ För tydlighets skull har jag markerat de till synes ”saknade” substantiven.

- (10) a. Examine the plight of the very poor __ .
b. The elderly __ and the extremely young __ are most at risk.
- (11) a. The unimaginable __ happened.
b. [...] distinguish the abstract __ and the concrete __ .
- (12) a. Their outfits; range from the flamboyant __ to the functional __ .
b. For all you can tell, the new situation is no better than the old __ .
- (13) a. Ny svensk forskning visar att de verkligt blyga __ verkar vara födda med anlag för detta.
b. Vi har kämpat hårt för det vi har men det materiella __ är ju inte allt eller hur?
c. De inaktiva delarna klipps bort och bara de aktiva __ skickas ut i cellen för att tillverka protein.

Gemensamt för alla exemplen är de innehåller uttryck som fungerar som nominalfraser, fast med ett adjektiv i stället för ett substantiv som

¹⁶ Även om de svenska och de engelska konstruktionerna liknar varandra kan de inte förutsättas ha samma egenskaper. Tvärtom finns det sannolikt vissa skillnader mellan dem, kanske såpass stora att en annan indelning är motiverad för svenskans del. Detta är i grunden en empirisk fråga.

huvudord, liksom att de alla inleds med en bestämd artikel. Utmärkande för konstruktionen i (10), liksom i (13a), är att nominalfraserna är plurala och syftar på människor, närmare bestämt den mängd människor som har den egenskap adjektivfrasen uttrycker (poor, extremely young, verkligt blyga). Exempelen i (11) och (13b) utmärks i stället av singular, i svenskan neutrum, och syftning på något abstrakt. Slutligen kan de anaforiska strukturerna i (12) och (13c) betraktas som ellipser. Det är i dessa fall det ligger närmast till hands att tänka sig ett underförstått substantiv, som då är koreferent med något tidigare känt.¹⁷

Även om det finns skillnader mellan de tre mönstren som gör det motiverat att se dem som tre olika konstruktioner, finns det också tillräckligt stora likheter för att man också ska kunna se dem som varianter av samma konstruktion: ADJECTIVE-AS-NOMINAL / ADJEKTIV-SOM-NOMINAL. De här analyserna motsäger inte varandra, utan ligger bara på olika abstraktionsnivå. De specifika konstruktionerna är instanser av den generella och ärver vissa egenskaper av den men tillför också sina egna specifika restriktioner. Och den mer överordnade konstruktionen ärver i sin tur egenskaper från ännu mer generella konstruktioner, dels nominalfrasegenskaper, dels adjektiviska egenskaper.

Det är av detta skäl som konstruktikonets beskrivningsformat innehåller ett fält för arvsrelationer. Beskrivningar på olika abstraktionsnivå ryms inom samma system och kan ganska enkelt relateras till varandra. Notera att en konstruktion kan ärva egenskaper från flera håll (s.k. multiple inheritance).

Förekomsten av arvshierarkier väcker frågan om redundans kontra ekonomi i systemet. Ska egenskaper som nominalfrasfunktion, adjektiviskt huvudord och obligatorisk bestämd artikel redovisas för var och en av konstruktionerna i (10–12) eller räcker det att ange detta för den överordnade konstruktionen? Den här sortens avvägningar är alltid en balansgång.

Å ena sidan är ett visst mått av redundans önskvärt. Den som analyserar en viss konstruktion kan inte förutsättas ha överblick över resten av systemet, och eftersom databasen är under uppbyggnad (och förhoppningsvis ständig utveckling) bör man minimera risken att modifieringar av en analys får oönskade konsekvenser för andra. Och inte minst är det viktigt att en användare inte ska behöva gå mellan flera olika konstruktioner för att kunna tolka beskrivningen av en enda. Å andra sidan bör

¹⁷ Kay (2006) skiljer mellan rent anaforiska exempel som dem i (12), där det implicita ledet verkligen syftar tillbaka på kända referenter, och fall som avser nya referenter men av känd typ, t.ex. *Jag har flera __ hemma, Kan vi få två till __ här borta*, s.k. typanaforer (se Borthen 2004).

inte kravet på explicitthet drivas för långt. Man måste kunna klassificera en konstruktion som en nominalfras utan att därmed tvingas räkna upp allt som utmärker nominalfraser.

Vår ambition är att varje konstruktionsbeskrivning ska vara explicit nog för att vara användbar utan att andra konstruktionsbeskrivningar måste konsulteras, men i övrigt så enkel som möjligt. Idealet vore kanske en högst begränsad redundans i själva resursen, men en rikare redundans i gränssnitt och andra presentationer av materialet; i nuläget har vi dock inte möjlighet att separera redovisningarna på ett sådant sätt.

4.4. Berörda ämnesområden¹⁸

Konstruktioner kan uppfattas som perifera från såväl grammatisk som lexikal synvinkel, men de är knappast perifera i språkssystemet. ”The number of idioms and constructions that speakers know is of a comparable order of magnitude to the number of words, and the frequency of such constructions in text and conversation is very high” (Jackendoff 2007:57, jfr 1997:156).

Ett svenskt konstruktikon bidrar rimligtvis till ökad kunskap om svenska språket, både empiriskt och genom att erbjuda ett annat perspektiv. Det bör också vara av generellt teoretiskt värde för utvecklingen av CxG, särskilt då det genom sin storskalighet på sikt ger en möjlighet att studera mer omfattande konstruktionsnätverk än vad som tidigare funnits tillgängligt. Även för de didaktiska tillämpningar av CxG som börjar växa fram (Holme 2010) kan ett konstruktikon utgöra en stor tillgång.

Utöver denna ganska självklara potential till vinster för forskningen om svenska och konstruktionsgrammatik är det tre berörda områden som bör kommenteras lite mer utförligt: lexikografi, svenska som andraspråk och språkteknologi.

Lexikografi: Ett konstruktikon är ett slags lexikografisk resurs, fast med konstruktioner snarare än ord. Konstruktioner, som länge har försumrats inom lexikografin, uppmärksammas idag mer och mer i ordböcker genom valensangivelser, språkbruksexempel m.m. Täckningen är dock endast partiell, då ordböckerna tenderar att favorisera färgstarka fasta uttryck på bekostnad av anonymare

¹⁸ Det här avsnittet bygger till stor del på en projektansökan som samförfattats med Lars Borin, Julia Prentice, Rudolf Rydstedt, Emma Sköldberg och Sofia Tingsell.

konstruktioner med många luckor (variabla led), trots att de senare kan ha större relevans för inlärare än t.ex. idiom, som används ganska sällan (Farø & Lorentzen 2008). Vidare är ordböcker till sin natur begränsade av dels sitt omfång, dels sin grundstruktur baserad på uppslagsord. Särskilt det senare gör det svårt att hantera mönster som inte utgår från enskilda ord.

Här erbjuder elektroniska lexikon nya möjligheter genom alternativa sökvägar och närmast obegränsat utrymme. Inte desto mindre fordrar konstruktioner mer än rent lexikala verktyg. Hur ska man t.ex. representera grammatisk struktur, optionalitet, variabla led och ordföljdsvariation? Exempelvis har det visat sig att även idiom, som ofta betraktas som en sorts fasta uttryck, rymmer ett visst mått av variabilitet (Sköldberg 2004), vilket dock har svårt att komma fram i ett vanligt ordboksformat. Och hur visar man att det rör sig om en produktiv konstruktion? Här kan lexikografi och konstruktionsgrammatik ha ömsesidig nytta av varandra (se t.ex. Hanks 2008), och ett konstruktikon bör på sikt kunna fungera som ett komplement till ett elektroniskt lexikon.

Svenska som andraspråk: Vi vill som sagt särskilt fokusera konstruktioner som är relevanta för andra- och främmandespråksinlärning, vilka kan identifieras bl.a. genom analys av inlärarspråk, och tror också att konstruktikonet som helhet indirekt kan bli en resurs för inlärning. Modern flerspråkighetsforskning (t.ex. Gyllstad 2007, Prentice 2010, Prentice & Sköldberg 2011) visar hur idiom, kollokationer och liknande konstruktioner är problematiska för inlärare på flera sätt.

Ett grundläggande problem är att konstruktioner som sagt ofta saknas i ordböcker och grammatikor och dessutom tenderar att negligeras i undervisningen. Följaktligen är det inte särskilt förvånande att inlärare inte lär sig dem i någon större utsträckning, utan att även avancerade inlärare främst förlitar sig på ord och generella regler (till skillnad från förstaspråkstalare, som använder färdiga mönster i betydligt högre grad). Dels riskerar de därmed att uppfattas som mindre idiomatiska. Dels är det en kognitivt mer krävande uppgift att sätta samman meningar utan prefabricerat material (Ekberg 2004:272), vilket i sig leder till att det blir svårare att tala sitt andraspråk än det hade behövt vara.

Språkteknologi: Konstruktioner är notoriskt problematiska för språkteknologiska tillämpningar (Sag m.fl. 2002). Om man kan automatiskt identifiera konstruktioner i autentisk text med någorlunda korrekthet, reduceras t.ex. mycket av den ambiguitet som alltid har plågat maskinell satsanalys, där även relativt korta meningar kan få tiotals eller hundra-

tals syntaktiska analyser. Sådana metoder saknas dock än så länge och förutsätter rimligen att det finns formella konstruktionsbeskrivningar att utgå från. Ett konstruktikon kan alltså ligga till grund för språk- teknologisk metodutveckling, vilket också hör till projektets syften, där språkteknologi utgör en stor och viktig del.

Att identifiera konstruktioner maskinellt är här både mål och medel – en värdefull stödfunktion men också ett angeläget forskningsmål i sig. Vi vill utforska metoder för att med hjälp av språkverktyg och andra datorprogram leta fram konstruktionskandidater ur stora mängder text och så småningom även utveckla sätt att identifiera specifika konstruktioner. De konstruktioner som är mest realistiska ur denna synvinkel är delvis fyllda, där en del konstruktionselement är lexikalt givna och resten varierar inom vissa ramar, som t.ex. |[minuttal] *i/över* [timtal]|, |[så [adjektiv|adverb] *som möjligt*]|. Om man kan identifiera sådana halvfabrikat automatiskt, kan efterföljande texttolkningsanalyser både förenklas och förbättras avsevärt.

I detta perspektiv är det viktigt att konstruktikonets informationsmodell är sådan att den ger tillräcklig information om konstruktionerna för användning i system för automatisk textanalys. Informationen måste bl.a. vara alltigenom formell samt vara i samklang med andra typer av information som en språkteknologisk analys kan identifiera. Navet i integreringen är som sagt Saldo, ett stort lexikon för textanalys med morfologisk och semantisk information, till vilket andra resurser i Språkbanken är länkade (se Borin, Forsberg & Lönngrén 2008, Borin m.fl. 2012). Detta innebär dels att alla lexikala enheter i konstruktikonet anges i Saldo-format, dels att alla konstruktioner även registreras som Saldo-enheter.

Sammanfattningsvis: Konstruktikonet är en ny typ av resurs som kan bidra med ny kunskap på flera områden:

- Empiriskt fyller det en stor lucka, eftersom storskaliga konstruktionsbeskrivningar hittills saknas både i Sverige och internationellt, såväl generellt som med andraspråksfokus.
- Metodologiskt fordrar projektet utveckling av konstruktionsgrammatiska, lexikografiska och språkteknologiska metoder, även detta med såväl internationell som svensk relevans.
- Teoretiskt bör projektet ge nya insikter om förhållandet mellan grammatik och lexikon, i synnerhet om konstruktioners roll i detta samspel.
- Funktionellt bör konstruktikonet vara användbart för lingvistiska, språkteknologiska och i förlängningen lexikografiska och pedagogiska syften, inte minst med avseende på svenska som andraspråk.

Referenser

- FrameNet* <<https://framenet.icsi.berkeley.edu/fndrupal/>>
Konstruktikon <<http://spraakbanken.gu.se/swe/resurs/konstruktikon>>
Konstruktionsgrammatik i Göteborg
 <http://www.svenska.gu.se/forskning/forskningsprojekt/konstruktion_sgrammatik/>
Korp <<http://spraakbanken.gu.se/korp/>>
Saldo <<http://spraakbanken.gu.se/resurs/saldo>>
SweFN++ <<http://spraakbanken.gu.se/swefn/>>
- Andersson, Maria (2006). *Himalayas glaciärer hotar att smälta* – en beskrivning av konstruktionen NPpatient [V att INF]. (Kursuppsats i konstruktionsgrammatik.) Inst. f. svenska språket, Göteborgs universitet.
- Bergen, Benjamin & Nancy Chang (2005). Embodied Construction Grammar in Simulation-Based Language Understanding. I: Östman, Jan-Ola & Miriam Fried (red.), *Construction Grammars: Cognitive grounding and theoretical extensions*. Amsterdam: John Benjamins
- Blenselius, Kristian (2009). *Springa flåsande* och andra progressiva participkonstruktioner i svenskan. *Språk och stil* NF 19: 172–201.
- Boas, Hans and Ivan Sag (red.) (u.u.), *Sign-Based Construction Grammar*. Stanford: CSLI.
- Borin, Lars, Dana Dannélls, Markus Forsberg, Maria Toporowska Gronostaj, Dimitrios Kokkinakis (2010). The past meets the present in Swedish FrameNet++. *14th EURALEX International Congress*. Leeuwarden: EURALEX. S. 269–281.
- Borin, Lars, Markus Forsberg & Lennart Lönngren (2008). The hunting of the BLARK – SALDO, a freely available lexical database for Swedish language technology. *Resourceful language technology. Festschrift in honor of Anna Sågvall Hein*, red. Joakim Nivre, Mats Dahllöf & Beáta Megyesi. Acta Universitatis Upsaliensis: Studia Linguistica Upsaliensia 7. S. 21–32.
- Borin, Lars, Markus Forsberg, Leif-Jöran Olsson, Jonatan Uppström (2012). The open lexical infrastructure of Språkbanken. *Proceedings of LREC 2012*. Istanbul: ELRA.

- Borin, Lars, Markus Forsberg & Johan Roxendal (2012). Korp – the corpus infrastructure of Språkbanken. *Proceedings of LREC 2012*. Istanbul: ELRA.
- Borthen, Kaja (2004). The Norwegian type-anaphor *det*. I: A. Branco, T. McEnery & R. Mitokov (red.). *Proceedings of the 5th Discourse Anaphora and Anaphor Resolution Colloquium* (DAARC 2004).
- Börjesson, Linnéa (2012). *Kämpa på!* En undersökning av konstruktionen VERBA PÅ med fortsättningsbetydelse. (Kandidatuppsats.) Inst. f. svenska språket, Göteborgs universitet.
<<http://gupea.ub.gu.se/handle/2077/29105>>
- Croft, William (2001). *Radical Construction Grammar*. Oxford & New York: Oxford University Press.
- Ekberg, Lena (2004). Grammatik och lexikon i svenska som andraspråk på nästan infödd nivå. I: Hyltenstam & Lindberg (red.) *Svenska som andraspråk – i forskning, undervisning och samhälle*. Lund: Studentlitteratur. S. 259–276.
- Farø, Ken & Henrik Lorentzen (2009). De oversete og mishandlede ordforbindelser – hvilke, hvor og hvorfor? *LexicoNordica* 16: 75–101.
- Fillmore, Charles (1982). Frame Semantics. I: Linguistics Society of Korea (red.), *Linguistics in the morning calm*. Seoul: Hanshin Publishing Co. S. 111–138.
- Fillmore, Charles (2008). Border Conflicts: FrameNet Meets Construction Grammar. I: Bernal, Elisenda and Janet DeCesaris (red.), *Proceedings of the XIII EURALEX International Congress*. Barcelona: Universitat Pompeu Fabra. S. 49–68.
- Fillmore, Charles (2009). Frames and Constructions: Putting Them Together. Plenary talk at *Frames and Constructions. A Conference in Honor of Charles J. Fillmore*, July 31–August 2, 2009. University of California, Berkeley.
- Fillmore, Charles & Paul Kay (1996). *Construction Grammar*. Berkeley/Stanford: CSLI Lecture Notes.
- Fillmore, Charles, Russell Lee-Goldman & Russell Rhodes (u.u.). The FrameNet Constructicon I: Boas & Sag (red.).
- Fried, Mirjam (u.u.). Construction Grammar. I: A. Alexiadou & T. Kiss (red.), *Handbook of syntax* (2nd ed.) Berlin: de Gruyter.
- Fried, Mirjam & Jan-Ola Östman (2004). Construction Grammar: a thumbnail sketch. I: Mirjam Fried & Jan-Ola Östman (red.), *Construction Grammar in a cross-language perspective*. Amsterdam & Philadelphia: John Benjamins. S. 11–86.

- Goldberg, Adele (1995). *Constructions. A Construction Grammar Approach to Argument Structure*. Chicago & London: University of Chicago Press.
- Goldberg, Adele (2006). *Constructions at Work. The Nature of Generalization in Language*. Oxford & New York: Oxford University Press.
- Gyllstad, Henrik 2007. *Testing English Collocations, Developing receptive tests for Use with Advanced Swedish Learners*. Lund: Lund university.
- Hanks, Patrick (2008). The Lexicographical Legacy of John Sinclair. *International Journal of Lexicography* 21: 219–229.
- Hoffman, Thomas & Graeme Trousdale (red.) (u.u.). *Handbook of Construction Grammar*. Oxford University Press.
- Holme, Randal (2010). A construction grammar for the classroom. *IRAL* 48: 355–377.
- Jackendoff, Ray (1997). *The Architecture of the Language Faculty*. Cambridge, MA: MIT Press.
- Jackendoff, Ray (2007). *Language, Consciousness, Culture: Essays on Mental Structure*. Cambridge, MA: MIT Press.
- Jansson, Håkan (2006): Har du ölat dig odödlig? En undersökning av resultativkonstruktioner i svenskan. (Meddelanden från Institutionen för svenska språket 57) Inst. f. svenska språket, Göteborgs universitet.
- Kay, Paul & Charles Fillmore (1999). Grammatical constructions and linguistic generalizations: The *What's X doing Y?*-construction. *Language* 75/1: 1-33.
- Kay, Paul (2006). Null instantiation of nominal complements. Ms. University of California, Berkeley.
- Langacker, Ronald (1987). *Foundations of Cognitive Grammar. Volume I: Theoretical prerequisites*. Stanford, CA: Stanford University Press.
- Langacker, Ronald (1991). *Foundations of Cognitive Grammar. Volume II: Descriptive Application*. Stanford, CA: Stanford University Press.
- Lindström, Jan & Per Linell (2007). *Roli å roli. X-och-x som samtalspraktik och grammatisk konstruktion*. I: Engdahl, Elisabet & Ann-Marie Londen (utg.), *Interaktion och kontext*. Lund: Studentlitteratur. S. 19-89.
- Linell, Per & Kerstin Norén (2009): *Vågar vågar ni väl men...* – en reaktiv konstruktion i svenskan. *Språk och stil* NF 19: 72-104.
- Lyngfelt, Benjamin (2007). Mellan polerna. Reflexiv- och deponenskonstruktioner i svenskan. *Språk och stil* NF 17: 86–134.

- Lyngfelt, Benjamin (2011). Om subjekt med infinitiv, särskilt i passiv. I: *Svenskans beskrivning* 31, red. Ann-Catrine Edlund & Ing-Marie Mellenius. Umeå: Umeå universitet. S. 230–241.
- Norén, Kerstin & Per Linell (2007): "Meaning Potentials and interaction between lexis and contexts: an empirical substantiation". *Pragmatics* 17: 387-416.
- Olofsson, Joel (2010). "Bollen susade in i bortre gaveln" – En studie i rörelsekonstruktioner och produktivitet. (Magisteruppsats.) Inst. f. svenska språket, Göteborgs universitet.
- Olofsson, Joel (2011). "Det känns lite som att googla in i en frysbox" – Om rörelsekonstruktioner och produktivitet. I: *Svenskans beskrivning* 31, red. Ann-Catrine Edlund & Ing-Marie Mellenius. Umeå: Umeå universitet. S. 294–303.
- Pollard, Carl & Ivan Sag (1994). *Head-Driven Phrase Structure Grammar*. Chicago & London: University of Chicago Press.
- Prentice, Julia (2010). *På rak sak. Om ordförbindelser och konventionaliserade uttryck bland unga språkbrukare i flerspråkiga miljöer*. (Göteborgsstudier i nordisk språkvetenskap 13.) Göteborg: Göteborgs universitet.
- Prentice, Julia (2011). "Jag är född på andra november" Konventionaliserade tidsuttryck som konstruktioner – ur ett andraspråksperspektiv. (Kursuppsats i konstruktionsgrammatik.) Inst. f. svenska språket, Göteborgs universitet.
- Prentice, Julia & Emma Sköldberg (2011). Figurative word combinations in texts written by adolescents in multilingual school environments. I: Källström, Roger & Inger Lindberg (red.), *Young Urban Swedish. Variation and change in multilingual settings*. (Göteborgsstudier i nordisk språkvetenskap 14.) Göteborg: Göteborgs universitet. S. 195–217.
- Prytz, Johanna (2000). SAG:s behandling av substantiverade adjektiv. I: Elisabet Engdahl & Kerstin norén (utg.) *Att använda SAG. 29 uppsatser om Svenska Akademiens grammatik*. (MISS 33.) Inst. f. svenska språket, Göteborgs universitet.
- Ruppenhofer, Josef, Michael Ellsworth, Miriam Petruck, Christopher Johnson & Jan Scheffczyk (2010). *FrameNet II: Extended Theory and Practice*. (a.k.a. The Book)
<https://framenet.icsi.berkeley.edu/fndrupal/the_book>
- Rydstedt, Rudolf (under utg.). *En matchningsdriven semantisk modell. Mellan ordboken och den interna grammatiken*. (Göteborgsstudier i nordisk språkvetenskap 19.) Göteborg: Göteborgs universitet.
- Sag, Ivan (2010). English Filler-Gap Constructions. *Language* 86: 486–545.

- Sag, Ivan, Timothy Baldwin, Francis Bond, Ann Copestake & Dan Flickinger (2002). Multiword Expressions: A Pain in the Neck for NLP. (LinGO Working Paper No. 2001-03.) I: Alexander Gelbukh (red.), *Proceedings of CICLING-2002*.
- Sjögreen, Christian d.y. (2007): *Går det att äta bort sin huvudvärk?* - en studie av den indirekt kausativa *bort*-konstruktionen i svenskan. (D-uppsats.) Institutionen för svenska språket, Göteborgs universitet.
- Sköldberg, Emma (2004). *Korten på bordet. Innehålls- och uttrycksmässig variation hos svenska idiom*. Göteborg: Meijerbergs.
- Stefanowitsch, Anatol & Stefan Gries (2003). Collostructions: Investigating the interaction between words and constructions. *International Journal of Corpus Linguistics* 8.2: 209–43.

Bilaga 1: Illustration av preliminärt gränssnitt

Indirekt_kausativ_bort

type	Cx
category	vbm
evokes	Causation_scenario
definition	[Någon] _{Agent} utför [en aktion] _{Activity} som indirekt leder (eller antas leda) till att [en företeelse] _{Theme} [försvinner] _{Result}
structure	vb <u>bort</u> ¹ NP
cee	<u>bort</u> ¹
coll	<u>spela</u> ² <u>spela</u> ⁶
internal construction elements	<ul style="list-style-type: none"> ■ role: name=Activity cat=vb ■ role: lu=<u>bort</u>¹ name=Result ■ role: name=Theme cat=NP
external construction elements	<ul style="list-style-type: none"> ■ role: name=Agent cat=NP
examples	<ul style="list-style-type: none"> ■ Kan [man]_{Agent} [[äta]_{Activity} [bort]_{Result} [sin huvudvärk?]_{Theme}]_{Indirekt_kausativ_bort} ■ [Folkpartiet]_{Agent} vill [[undervisa]_{Activity} [bort]_{Result} [extremhögern]_{Theme}]_{Indirekt_kausativ_bort} . ■ Han har ändå ingenting att göra nu, sedan [sommarvikarien]_{Agent} [[spelat]_{Activity} [bort]_{Result} [kommunens kassa]_{Theme}]_{Indirekt_kausativ_bort} . på bingo.
comment	Strukturanalysen är förenklad. En semantiskt mer rättvisande analys är: vb [bort NP]
reference	Sjögreen, Christian d.y. (2007). Går det att äta bort sin huvudvärk? - en studie av den indirekt kausativa bort-konstruktionen i svenskan. (D-uppsats) Inst f svenska språket, Göteborgs universitet

<http://spraakbanken.gu.se/swe/resurs/konstruktikon/utvecklingsversion#Indirekt_kausativ_bort>

Bilaga 2: exempelkonstruktion i xml

```

<entry xml:id="Indirekt_kausativ_bort">
  <type>Cx</type>
  <cat>vbm</cat>
  <inheritance/>
  <evokes>
    Causation_scenario
  </evokes>
  <definition>
    <e name="Agent">Någon</e>
    utför
    <e name="Activity">en aktion</e>
    som indirekt leder (eller antas leda) till att
    <e name="Theme">en företeelse</e>
    <e name="Result">försvinner</e>
  </definition>
  <structure>
    vb bort..1 NP
  </structure>
  <cee>bort..1</cee>
  <coll>
    spela..2
    spela..6
  </coll>
  <construction_elements>
    <internal>
      <role name="Activity" cat="vb"/>
      <role name="Result" lu="bort..1"/>
      <role name="Theme" cat="NP"/>
    </internal>
    <external>
      <role name="Agent" cat="NP"/>
    </external>
  </construction_elements>
  <examples>
    <example>
      Kan
      <e name="Agent">man</e>
      <e name="Indirekt_kausativ_bort"><e name="Activity">äta</e>
      <e name="Result" cee="bort..1">bort</e>
      <e name="Theme">sin huvudvärk?</e></e>
    </example>
    <example>
      <e name="Agent">Folkpartiet</e>
      vill
      <e name="Indirekt_kausativ_bort"><e name="Activity">undervisa</e>

```

```

    <e name="Result" cee="bort..1">bort</e>
    <e name="Theme">extremhögern</e></e>.
</example>
<example>
  Han har ändå ingenting att göra nu, sedan
  <e name="Agent">sommarvikarien</e>
  vill
  <e name="Indirekt_kausativ_bort"><e name="Activity">spelat</e>
  <e name="Result" cee="bort..1">bort</e>
  <e name="Theme">kommunens kassa</e></e>
  på bingo.
</example>
</examples>
<comment>
  Strukturanalysen är förenklad. En semantiskt mer rättvisande analys är: vb
  [bort NP]
</comment>
<reference>
  Sjögreen, Christian d.y. (2007). Går det att äta bort sin huvudvärk? - en studie
  av den indirekt kausativa bort-konstruktionen i svenskan. (D-uppsats) Inst f
  svenska språket, Göteborgs universitet
</reference>
</entry>

```


GU-ISS, Forskningsrapporter från Institutionen för svenska språket, är en oregelbundet utkommande serie, som i enkel form möjliggör spridning av institutionens skriftliga produktion. Det främsta syftet med serien är att fungera som en kanal för preliminära texter som kan bearbetas vidare för en slutgiltig publicering. Varje enskild författare ansvarar för sitt bidrag.

GU-ISS, Research reports from the Department of Swedish, is an irregular report series intended as a rapid preliminary publication forum for research results which may later be published in fuller form elsewhere. The sole responsibility for the content and form of each text rests with its author.

Forskningsrapporter från institutionen för svenska språket, Göteborgs universitet
Research Reports from the Department of Swedish

ISSN 1401-5919

www.svenska.gu.se/publikationer/GU-ISS