

Institutionen för lingvistik
Stockholms Universitet
VT-2000

DATORSTÖDD SPRÅKGRANSKNING OCH ANDRASPRÅKSINLÄRARE

Lena Öhrman

I denna uppsats undersöks vilka fel några andraspråksinlärare gör när de lär sig svenska. Utifrån en korpus med texter av andraspråksinlärare skapas en typologi över inlärnarnas fel. Vidare undersöks hur och i vilken utsträckning språkgranskningsprogrammet Granska detekterar fel som inlärnarna gör. Med utgångspunkt i resultatet från körningen av Granska, diskuteras vilka problem man ställs inför när man utvecklar ett skrivstödsprogram som ska vara till nytta för andraspråksinlärare. Slutligen diskuteras språkgranskningsprogrammets roll och hur det kan utvecklas i framtiden.

D-uppsats i datorlingvistik, 10 poäng
Handledare: Ola Knutsson (IPLab, Nada, KTH),
Rickard Domeij (institutionen + IPLab, Nada, KTH)

INNEHÅLLSFÖRTECKNING

1	INLEDNING	3
1.1	SYFTE	3
2	BAKGRUND	3
2.1	SVENSKA SOM ANDRASPRÅK	4
2.2	SPRÅKGRANSKNINGSPROGRAMMET GRANSKA	6
3	MATERIAL	8
3.1	ASU-KORPUSEN	8
4	METOD	10
5	RESULTAT	10
5.1	FELTYPOLOGIN MED EXEMPEL OCH FÖRKLARINGAR	11
5.1.1	<i>Fördelning av fel i feltypologin</i>	16
5.2	RESULTAT AV KÖRNING AV GRANSKA PÅ TEXTERNA	17
5.2.1	<i>Stavfel</i>	18
5.2.2	<i>Kongruens</i>	18
5.2.3	<i>Fel tempus</i>	18
5.2.4	<i>Felaktig ordföljd</i>	19
5.2.5	<i>Felaktig preposition</i>	20
5.2.6	<i>Semantiskt fel, lexikonfel</i>	20
5.2.7	<i>Antal falska alarm</i>	20
5.3	FELAKTIG ORDFÖLJD OCH REGLER	21
5.3.1	<i>Regel 1</i>	22
5.3.2	<i>Regel 2</i>	22
5.3.3	<i>Regel 3</i>	23
6	DISKUSSION	23
6.1	INLÄRARNAS SPRÅK OCH SVÄRIGHETER	23
6.2	RÄTTNING OCH KATEGORISERING AV FEL	24
6.3	DETEKTION	24
6.4	SPRÅKGRANSKNINGSPROGRAMMETS ROLL OCH INVERKAN	25
6.5	SLUTSATSER	25
7	REFERENSER	27
	APPENDIX A: SUC-TAGGAR	28
	APPENDIX B: TEXTER UR ASU-KORPUSEN	30

1 INLEDNING

Det här arbetet görs som en del i projektet *Integrerade språkverktyg för skrivande och dokumenthantering* på Nada vid Tekniska Högskolan i Stockholm, där ett program för datorstödd språkgranskning (Granska) är under utveckling.

1.1 Syfte

Syftet med uppsatsen är att undersöka hur språkgranskningsprogrammet Granska klarar av texter skrivna av personer som lär sig svenska som andraspråk. I uppgiften ingår att analysera och beskriva de fel som andraspråksinlärare gör och att undersöka om och hur Granska klarar av att detektera dessa. Till uppgiften hör även att, om möjligt, komma med förslag på hur man kan anpassa Granska så att det även passar att användas i detta sammanhang.

2 BAKGRUND

I skrivande stund använder jag själv mig av skrivstödsprogram. För mig är det till stor hjälp till exempel vid slarvfel eller som stavningskontroll. Jag kan föreställa mig till vilken hjälp det skulle vara för en invandrare som kommer till Sverige. Att tillägna sig språket som talas i det land man har valt att bo är ett av de första och största stegen till att komma in i samhället.

Skrivstödsprogrammet Granska (se 2.2 nedan) har hittills inte testats på texter skrivna av personer som håller på att lära sig svenska som andraspråk. Frågan är hur bra Granska skulle klara av att hitta de fel som dessa personer gör. Min hypotes är att de gör andra typer av fel jämfört med oss som har svenska som modersmål och därmed borde Granska ha svårare att detektera deras fel. Om man skulle få Granska att i större utsträckning klara dessa fel skulle användningsområdet för datorstödda skrivstödsprogram vidgas. Man kan tänka sig att dessa program i framtiden skulle kunna användas i större utsträckning i inlärningssammanhang.

Många frågor och utmaningar uppstår i samband med datorstödd språkgranskning. Programmen måste anpassas till olika syften, målgrupper och så vidare. Det finns en distinktion mellan kontrollerande och instruerande program. Kontrollerande program hittar fel i en text och markerar dem. Instruerande program försöker inte bara finna fel utan även förklara för användaren vad som blivit fel. Andraspråksinlärare har inte så stor nytta av ett enbart kontrollerande program. Ett problem med instruerande program är att den som har goda språkkunskaper har liten nytta av instruktioner, medan den som har sämre kunskaper ofta har svårt att förstå vad han eller hon har gjort för fel. Språkgranskningsprogrammets dilemma är att de som kan behöver inte, de som behöver kan inte (Domeij, 1997). Programmets feldiagnoser och rättningsförslag blir svåra att förstå om de till exempel innehåller grammatisk eller lingvistisk terminologi. I dag anser många att lingvistiska uttryck gör förklaringarna svåra att förstå men samtidigt finns det de som tycker att det är bra att kalla saker vid deras rätta namn, inte minst i bildningssyfte. Det är i allmänhet svårt att hitta en lämplig nivå på granskningsprogrammets förklaringar. Språkgranskningsprogrammet måste anpassas efter användaren. Man måste ställa sig en rad frågor: vilken är målgruppen? Vad är syftet? Exempelvis verkar det som att man i Lingsofts språkkontroll för Microsoft Word 2000 har riktat in sig på en person med svenska som modersmål och som skriver mycket, snabbt och slarvigt. Eftersom användare kan bli irriterade när det kommer upp för många falska alarm har man utgått från principen att ge så få falska alarm som möjligt. Att till exempel skriva *sig* som *sej* måste man låta slinka igenom eftersom det finns en fisk som heter *sej*. Ett annat exempel är *dom* som kan betyda *dem* men också en *dom* (i domstol).

Falska alarm slår väldigt olika beroende på vem användaren är. Om jag får ett falskt alarm så märker jag oftast det på en gång och kan ignorera grammatik- och stavningskontrollen. För andraspråksinlärare däremot kan falska alarm vara ett stort problem. Det kan vara mycket svårt för dem att inse att det rör sig om ett falskt alarm. Om de inte har så stora egna kunskaper kanske de har för stor tillit till programmen och om de inte känner till begränsningarna i programmet kan de bli bedragna. Därför borde det vara extra viktigt att program som riktar sig till andraspråksinlärare inte ger falska alarm. För att kunna förstå de fel som andraspråksinlärare gör måste man först studera inlärningsprocessen av svenska som andraspråk.

2.1 Svenska som andraspråk

Alla som någon gång har lärt sig ett andraspråk vet hur svårt det kan vara och hur många fel man gör i början. Men även om det kan vara frustrerande att göra en massa fel så är det trots allt genom felen man lär sig.

I början av inlärningsprocessen är andraspråksinlärares svenska full av fel. Men antalet fel säger egentligen inte så mycket om en inlärares språknivå; ett fel kan snarare ses som ett tecken på att en regel håller på att läras in än som ett misslyckat försök till att behärska målspråket¹. Om man endast ser till antalet fel en inlärare gör kan det ge en falsk bild av hur den personen egentligen behärskar grammatiken och ordförrådet.

Termen andraspråk kan referera till alla språk som man lär sig efter förstaspråket, även om det är tredje eller fjärde språket som lärs in. Vidare kan andraspråksinläring syfta såväl på inläring i målspråksmiljö som utanför målspråksmiljön, det vill säga som främmande språk (Ellis 1994).

Man måste tänka på att det som är självklart i grammatiken för oss som har svenska som modersmål ofta kan vara främmande för andraspråksinlärare. Mycket av det som vi inte behöver tänka på, som till exempel hur ett substantiv böjs i plural, kan utgöra stora problem för andraspråksinlärare.

När man studerar sitt eget modersspråks grammatik upptäcker man att det finns fler svårigheter än man tror. Ofta kan grammatiska regler vara ologiska, som t.ex. vid bildning av sammansättningar. Hur ska en andraspråksinlärare veta om det ska vara ett fugeelement mellan de två leden eller inte? Jämför till exempel *tidlös* och *tidskrävande* där ett fuge-s ska läggas till i det senare.

I inlärningsprocessen försöker andraspråksinläraren upptäcka och tillägna sig språkets regler. Övergeneraliseringar som *springde*, *liggde*, osv. visar att inläraren har lärt sig en regelbundenhet i språket och försöker tillämpa den. Att en regel övergeneraliseras innebär att den används utanför sitt begränsade tillämpningsområde (Ekerot, 1995:17). Inläraren har kommit så långt att han eller hon börjar kunna reglerna men behärskar inte undantagen och då krävs inte bara regelkunskap utan även lexikonkunskap. För till exempel pluralbildning av substantiv måste man kunna reglerna men dessutom behövs lexikonkunskap för att kunna böja undantag. Ett annat område där lexikala faktorer har betydelse är ordbildningen, alltså bildandet av sammansättningar och avledningar som i exemplet med fuge-s ovan.

Förutom övergeneralisering finns det ett antal generella drag som karakteriserar inlärarespråket på alla nivåer såsom förenkling, undvikande, innovationer och kompensatoriska element samt

¹ Målspråk är det språk som inläraren har som mål att lära sig.

element från andra språk än målspråket (Viberg, 1987:30). Alla dessa strategier kan betraktas som en del av inlärarspråket, det vill säga det språksystem som utvecklas i riktning mot målspråket.

Inlärningsprocessen kan ses som olika på varandra följande faser som går igenom och eftersom alla inlärare i stort sett följer samma inlärningsgång blir också felen av samma typ. De fel inläraren gör beror alltså på i vilken inlärningsfas han eller hon befinner sig (Kotsinas, 1991:163).

En intressant fråga är hur stor förstaspråkets roll är vid andraspråksinläring. När man studerar andraspråksinläring jämför man möjligheterna att föra över strukturer och begrepp i förstaspråket till andraspråket, s.k. transfer (Ekerot, 1995:36). Tidigare skilde man ofta mellan positiv och negativ transfer, men numer anses transfer omfatta varje typ av inflytande från modersmålet (Viberg, 1987:51).

Det finns olika hypoteser om betydelsen av skillnaderna mellan modersmål och målspråk. Det är ett faktum att det finns en korrelation mellan hur lätt det går att lära sig målspråket och hur likt det är förstaspråket. Exempelvis innebär andraspråksinläringen av tempus, bestämdhet och numerus i svenskan trots allt ganska överkomliga svårigheter om dessa grammatiska kategorier finns i förstaspråket som fallet är i engelska, tyska, franska och spanska (Ekerot, 1995:36). Genom att jämföra språken med hjälp av s.k. kontrastiv analys skulle man alltså kunna förutse vilka svårigheter en inlärare har att vänta. Det har visat sig att förutsägelseerna inte alltid stämmer. Inlärare gör inte alltid fel där första- och andraspråk skiljer sig åt utan ibland även på områden där de båda språken är lika (Kotsinas, 1991:164).

Diskussionen kring skillnader mellan modersmål och målspråk handlar numer i stor utsträckning om ommarkerade och markerade strukturer. Ommarkerade strukturer förekommer allmänt i världens språk, medan markerade bara uppträder i ett begränsat antal språk. En allmän princip är att ommarkerade strukturer är lättare att lära in än markerade. Exempel på ommarkerade strukturer som inlärare använder i sitt interimsspråk² är ordföljden *subjekt + negation + verb* (Han inte komma) och ordföljden *adverbial + subjekt + verb* (Nu han kommer), som bägge är vanligare i världens språk (och i viss mån enklare) än svenskans konstruktioner (Ekerot, 1995:27).

Ordföljd i allmänhet, både markerade och ommarkerade strukturer, är något som inlärare av svenska som andraspråk har svårt för. Deras misstag kan hjälpa oss att få större insikt i vårt språk vad gäller ordföljd. Man har upptäckt att svenskan jämfört med andra språk har en mer grammatikaliserad ordföljd och att den fyller andra funktioner än i många andra språk. För en inlärare framstår svenskans ordföljd som ganska speciell, inlärare har också en tendens att göra liknande fel med svenskans ordföljd nästan oberoende av modersmål. Exempel på svårtillgängliga strukturer är placeringen av negationen i huvud- och bisats, liksom inversionen i påståendesatser inledda av icke-subjekt: *Igår var jag i Köpenhamn* (Ekerot, 1995:65-66).

I ett lingvistiskt perspektiv framstår ordföljden som ett grammatiskt uttrycksmedel bland andra, så att skillnader som uttrycks med ordföljden i ett språk kan uttryckas med andra grammatiska medel i ett annat och tvärtom. I stället för ordföljdsskillnader kan språken använda sig av: böjningsmorfologi (finska), formord (t.ex. frågepartikeln *czy* i polska) och konstruktion (t.ex. *do*-omskrivningar i ja/nej-frågor i engelskan) (Ekerot, 1995:74).

² Interimsspråk kallas det språk som andraspråksinläraren använder sig av på vägen mellan att inte alls kunna målspråket och att behärska det så som en infödd talare gör (Kotsinas, 1991:14).

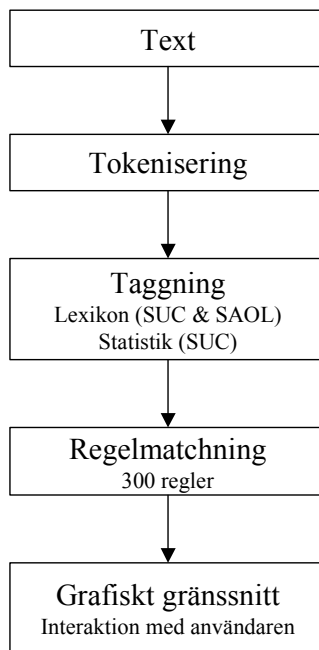
2.2 Språkgranskningsprogrammet Granska

Granska är ett program för datorstödd språkgranskning som utvecklas inom projektet *Integrerade språkverktyg för skrivande och dokumenthantering*³ på KTH i Stockholm. Projektet har som mål att utveckla datorfunktioner för språkgranskning på svenska, att införa dessa i en interaktiv miljö för datorstött skrivande och att testa dem empiriskt med användare (Domeij et al., 1998).

Granskningsfunktionen analyserar språket i en text och markerar de problem den finner i texten. Till exempel kan programmet hitta stavfel, stilavvikelse och grammatikfel. För att finna dessa fel använder sig Granska av en uppsättning regler. Reglerna beskriver olika typer av fel och kan till exempel vara baserade på Svenska språknämndens skrivregler. Felaktigheterna i en text detekteras genom att programmet undersöker ordsekvenser i texten och ser efter om det finns någon regel som matchar. Det finns till exempel regler för kongruensfel som säger att det inte får finnas ett substantiv i neutrum med en determinant i utrum framför, som i ”en hus”. Rättstavningsmodulen *Stava* som är en del av Granska sköter detektion av stavfel och kan även hantera sammansättningar (Kann et al., 1998).

Före matchning mot regler ordklasstaggas hela texten. För taggning använder sig Granska av ett lexikon som är skapat utifrån den ordklassmärkta textdatabasen SUC. Lexikonet har kompletterats med ord och ordformer från andra källor för att Granska lättare ska kunna hitta felaktigheter.

En överblick av Granskas funktionalitet visas i figuren nedan.



Figur 1 Översikt av Granska
(Domeij et al., 1999)

³ Mer information om Granskaprojektet finns på: <http://www.nada.kth.se/theory/projects/granska/>

Ett problem med taggning är att många ord kan ha flera betydelser och olika tolkningar. Granska löser detta problem genom att med hjälp av en träningstext föra statistik på sekvenser av två och tre taggar i följd. Med en s.k. Markovmodell kan sedan den mest sannolika taggningen för varje enskilt ord beräknas och på detta vis disambigueras orden.

Ett annat problem som försvårar taggning är att man på svenska kan konstruera nya ord genom sammansättningar. I en träningstext kan man bara få med en liten del av alla svenska ord. Följden blir att när taggaren ska tagga en text kommer den att stöta på många okända ord som måste tilldelas en tagg. Med hjälp av statistik om taggsekvenser, på samma sätt som ovan, och med morfologisk analys kan Granska gissa taggar. Den morfologiska analysen väljer den tagg som är mest sannolik för andra, kända ord som slutar med samma ändelse.

Granska hittar även språkliga fraser, som till exempel nominalfraser (NP). Med hjälp av så kallade hjälpreglar kan dessa språkliga fraser identifieras. Hjälpreglerna kan i sin tur användas av andra regler. På det här viset kan man skriva regler på en högre nivå, genom att man i reglerna refererar till språkliga fraser i stället för att bara ange följderna av ordklasser.

Korrektion uppnås genom att felet markeras i texten med teckenprecision, det vill säga att användaren kan se exakt var felet är. Dessutom ska ersättningsförslag kunna sättas in direkt i texten.

Hur ser då en regel ut? En regel består av två led som avskiljs med en pil. Det första ledet innehåller ett matchningsvillkor. I det andra ledet anges vad som ska göras när villkoren i det första ledet är uppfyllda. En matchningsvariabel, t.ex. X, som står i första ledet, beskriver det område som ska matchas i indata, det vill säga ett token (ord). I det första ledet står även reguljära uttryck som variablerna matchas mot.

Exempel:

```
{
X1/wordcl=jj/,
X2/wordcl=dt & gender=X1.gender & num=X1.num & spec=X1.spec/,
X3/wordcl=nn & gender=X2.gender & num=X2.num & spec=X2.spec/
-->
/X1 X2/X2 X1 X3/"info"/granskning/
}
```

I den här fiktiva regeln matchas tre tokens (X1, X2 och X3) mot ordklasserna adjektiv, determinerare och nomen (*jj*, *dt* och *nn*)⁴. Alla tre tokens ska även ha samma genus, numerus och species. I det andra ledet (efter -->) talar det första fältet (X1 X2) om att orden X1 och X2 ska markeras. Det andra fältet (X2 X1 X3) talar om vad som ska returneras (korrigeras). I vårt exempel byter regeln alltså plats på adjektiv och determinerare. Slutligen indikerar det sista fältet att detta är en granskningsregel. Regelen ovan skulle till exempel korrigera ”Vi bor i *litet ett hus*” genom att byta plats på *litet* och *ett*.

Förutom ordklasser och en mängd särdrag som till exempel genus, numerus, species osv. kan man även matcha textsträngar. För att kombinera särdrag går det också att använda logiskt eller (med tecknet '|') samt och (med tecknet '&') (Knutsson 1999).

⁴ Se Appendix A för en beskrivning av SUC-taggar.

När man kör Granska på en text tillämpas reglerna och de matchas mot varje mening i texten. Så här kan utdata för en mening till exempel se ut:

Vi bor på en hus med 5 rum och en kök.

§74 - : Vi bor på [KONGRUENSFEL-2: en/FEL hus] med 5 rum och en kök.

§74 - : Vi bor på en hus med 5 rum och [KONGRUENSFEL-2: en/FEL kök].

I meningen ovan har Granska korrekt detekterat två kongruensfel (och missat en felaktig preposition (på)). Man ser även vilken regel som har hittat dem, i det här fallet KONGRUENSFEL-2.

I skrivande stund är Granska implementerat under operativsystemet Unix. Denna lösning är tillfällig och målet är att på sikt byta till Windows. Det finns också en webbversion som fungerar direkt i användarens webbläsare⁵.

Granska är ett pågående projekt och programmet utvecklas kontinuerligt. För min undersökning har jag använt Granska i den senaste version som fanns tillgänglig i början av juni 1999.

3 MATERIAL

Som material har jag använt delar av ASU-korpusen. ASU är en förkortning för Andraspråkets strukturutveckling. Korpusen togs fram i ett projekt med syftet att undersöka olika individers utveckling vid inläring av ett andraspråk (Hammarberg 1997). ASU-korpusen beskriver ett antal individers språkliga utveckling under en längre tid, vilket är till stor fördel för min undersökning då det är möjligt att de fel som görs i början av inlärningsprocessen skiljer sig från felen som görs senare. På det här viset har jag fått med en stor del av inlärningsprocessen.

3.1 ASU-korpusen

ASU-korpusen innehåller dels transkriberade konversationer, dels skriftliga uppsatser som producerats av vuxna som håller på att lära sig svenska som andraspråk. Korpusen beskriver personernas individuella utveckling över tiden vilket ger möjlighet till jämförelse både mellan individer och mellan olika stadier hos en enskild individ.

Inlärarna bestod av tio personer i åldrarna 19-28 år som alla deltog i den förberedande kursen i svenska för utländska studenter som organiseras av *Institute for English-Speaking Students* (IES) på Stockholms Universitet. Namnet är något missvisande då kursens deltagare kommer från alla världens länder och har varierande modersmål. Inlärarna hade utbildning på gymnasienivå och tidigare erfarenheter av att lära sig främmande språk. De var motiverade eftersom de behövde svenska för fortsatta akademiska studier i Sverige och inom loppet av ett till två år gick de alla från att vara nybörjare till en relativt hög nivå. De hade mycket gemensamt som till exempel ålder, utbildning, målet att studera vidare i Sverige. Det som skilde dem åt var främst deras modersmål och kulturella bakgrund. Det fanns tre huvudgrupper bland förstaspråken: kinesiska, grekiska och portugisiska/spanska. Utöver dessa fanns det en inlärare med polska som förstaspråk och en som var tvåspråkig i tyska och

⁵ <http://www.nada.kth.se/theory/projects/granska/>

engelska. I materialet kan man se skillnader mellan européer och icke-européer vad gäller deras framsteg i inlärningsprocessen.

Den skriftliga delen av ASU-korpusen (den del som jag har använt mig av) består av 220 uppsatser som skrevs under 11 tillfällen. Totalt innehåller den ca 50 000 ord.

Uppsatserna skrevs under lektionstid och vid varje tillfälle gavs inlärnarna/studenterna två timmar för att skriva två separata texter med givna ämnen. De första 9 tillfällena sträcker sig över läsåret 1990/91. Tillfälle 10 och 11 var uppföljningar som inträffade under våren 1992 respektive våren 1993.

Man kan dela in uppgifterna som inlärnarna fick i följande kategorier:

- berättande av ett händelseförlopp utifrån en bildserie
- berättande av ett minne
- beskrivning av ett givet ämne
- diskussion kring olika ämnen t.ex. tidningsartiklar
- övrigt

Vissa uppgifter repeterades, till exempel var båda ämnena vid tillfälle 11 samma som vid tillfälle 7 (Hammarberg 1997).

I min undersökning har jag använt mig av 44 texter (12 576 ord) skrivna av fyra personer. Då jag fick tillgång till materialet insåg jag ganska snart att jag endast skulle kunna undersöka de inlärare som hade högst kunskapsnivå i svenska. Jag var tvungen att göra ett urval på grund av att vissa texter var så svårlästa att det nästan var omöjligt att kategorisera felen, därför valde jag att begränsa mig till att studera europeiska inlärare.

Tabellen nedan beskriver de försökspersoner för ASU-korpusen som jag valde att studera.

Person	Kön	Ålder vid kursens början ⁶	Uppväxt i	Modersmål	Tidigare andraspråk	Vistelse i Sverige före kursstart	Deltog i kursen	Klarade kunskaps-testet ⁷
E2	M	28	Österrike, Indien, Kenya	tyska & engelska	swahili	2 månader	HT 90 VT 91	maj 91
G3	M	19	Grekland: Thessaloniki	grekiska	engelska, franska	23 dagar	HT 90 VT 91	maj 91
P1	K	20	Polen: Gdansk	polska	engelska, ryska, tyska, italienska	6 månader	HT 90 VT 91	maj 91
Q1	M	23	Portugal: Coimbra	portugisiska	engelska, franska, spanska	10 dagar	HT 90 VT 91	aug 91

Tabell 1 - Information om några av försökspersonerna i ASU-projektet.

⁶ Kursens början var 27 augusti 1990.

⁷ National Proficiency Test (språktest som utlänningar behöver genomföra för att komma in på svenska universitet och högskolor)

De jag valde att undersöka kommer alla från Europa. Alla 4 modersmål tillhör den indoeuropeiska språkfamiljen. Tyska och engelska är germanska språk, grekiska bildar en egen gren. Polska är ett slaviskt språk och portugisiska är ett romanskt språk.

4 METOD

En del av syftet med uppsatsen var att analysera de fel som andraspråksinlärare gör. För att få fram felen började jag med att läsa igenom de utvalda texterna och rätta dem för hand. Jag har strävat efter att felmarkera allt som jag anser kan räknas som språkfel. Till min hjälp har jag använt *SAOL*, *Svenska Språknämndens skrivregler*, *Thorells Svensk grammatik*, *Svensk handordbok* och Ulla Stroh-Wollins bok *Koncentrerad nusvenska formlära och syntax*.

Allt eftersom jag rättade texterna arbetade jag fram en kategorisering av felen som resulterade i en feltypologi (se 5.1 under Resultat nedan). Felkategorierna skapades delvis med inspiration av Hultman och Westmans (1977) felmarkeringssystem.

Efter att ha felsökt texterna för hand undersökte jag i vilken utsträckning felen hittades av språkgranskningsprogrammet Granska. Jag undersökte resultatet jag fick fram när jag granskade texterna med hjälp av Granska och jämförde det med min egen rättning. Jag fick fram antal träffar, antal missar, antal falska alarm och feldiagnoser. Jag undersökte frekvensen för de olika feltyperna och hur bra Granska klarade av dem. Vidare har jag också försökt analysera varför Granska hade svårigheter att upptäcka vissa typer av fel.

Som ett exempel på hur man kan förbättra Granskas detektion av andraspråksinlärares fel, har jag gjort några förslag på nya regler i Granska som detekterar vissa ordföljdsfel.

5 RESULTAT

Resultatet består av tre delar. Första delen (5.1) innehåller en typologi över inlärares fel i ASU-korpusen med en kort beskrivning av varje feltyp. I avsnittet redovisas även fördelningen av fel på olika feltyper. Vidare följer resultatet av körningen med Granska på texterna i ASU-korpusen (5.2). De vanligast förekommande felen diskuteras mer ingående och även möjligheterna till att detektera dem med Granska. Ett vanligt förekommande fel som Granska hade problem med att detektera var felaktig ordföljd. I den sista delen av resultatet (5.3) visar jag därför några exempel på hur man kan förbättra detektionen av felaktig ordföljd i Granska.

5.1 Feltypologin med exempel och förklaringar

Tabellen nedan beskriver alla feltyper och ger autentiska exempel från texterna som jag har undersökt. Själva felet i exempelmeningen är markerat kursivt.

Ortografiska fel

Typ av fel	Kommentar	Exempel
Stavfel	ord som bryter mot svenskans stavningsregler	...tex språk, <i>literatur</i> eller historia...
Utländskt/okänt ord	användning av ett utländskt eller okänt ord	...jag är en <i>blend</i> .
Felböjt verb	verbet är felböjt vid uttryckande av tempus	vi <i>sovade</i>
Felböjt substantiv	substantivet är felböjt i numerus	Så mycket <i>träder</i> , ...
Felböjt adjektiv	adjektivet är felböjt vid komparation	...Hamlets slott från den <i>berömdaste</i> pjäs av Shakespeare...
Felaktigt sammansatt sammansättning	foge-s saknas eller överflödigt	<i>lägenhetannonser, cigarettspaket</i>
Särskrivnen sammansättning	särskrivning av två led som ska stå ihop	<i>väder rapport</i>
Versalfel	ord som ska skrivas med liten/stor initialbokstav	Hur är vintern i <i>sverige</i> ?
Felaktig förkortning	felaktiga förkortningar	...som i framtiden kan betyda <i>tom</i> inbördeskrig.
Genitiv med apostrof	genitiv som skrivs med apostrof	...av detta <i>barn's</i> stolta föräldrar.

Grammatiska fel

Kongruensfel	stämmer inte överens med omgivning i numerus, genus eller species	<i>Vilket</i> musik tycker du om?
Bestämthetsfel	användning av bestämd form av substantiv där det ska vara obestämd form och tvärtom	Jag har just beställt tid för <i>besöket</i> hos en psykoterapeut.
Adjektiv/adverbfel	användning av adjektiv där det ska vara adverb och tvärtom	På en gång fick jag intrycket att han levde <i>intensiv</i> men samtidigt <i>farlig</i> .
Fel pronomenform	fel val av pronomenform	Jag träffar en vän i Berlin och vi stannar med <i>hon</i> .
Fel tempus	verbet står i ett tempus som inte stämmer med kontexten	...men, <i>tror</i> mig, det är faktiskt inget speciellt.
Felaktig ordföljd	fel ordning bland orden i en mening	...och det intrycket som <i>gav han mig</i> ...
Verbfrasfel med ”att”	meningar där ett onödigt ”att” lagts in	... för jag behövde <i>att</i> tänka om...
Verbfrasfel utan ”att”	meningar med att-konstruktion där ”att” har utelämnats	...utan att kunna <i>tinga de betala</i> .
Saknas artikel	utebliven artikel eller determinerare före nominalfras	... och <i>att sådan</i> händelse kan lämna starka spår på hans hälsa.
Saknas genitiv-s	substantiv som ska stå i genitiv men saknar genitiv-s	Den <i>sorten</i> annonser är polska tidningar intresserade av.

Semantiska fel

Semantiskt fel *	brott mot betydelsesammanhanget (Hultman et al.)	<i>De är ett enda sätt att informera om de, som har avlidit, om begravningsplats och ibland om livet av de som inte kan säga det själva.</i>
Överflödigt konstituent	ett ord i en mening som inte fyller någon funktion	Här står <i>det</i> allt som kallas för det viktigaste...

Lexikonfel

Lexikonfel *	fel ordval (Hultman et al.)	Sen <i>hittade</i> de en idé i deras järnar.
Fel preposition	fel preposition, överflödigt preposition eller preposition saknas	Vad läser du <i>i</i> din fritid?

Tabell 2 – Feltypologin med kommentarer och exempel

Feltyperna i tabellen ovan är de jag upptäckte när jag rättade texterna. Indelningen i feltyper blev till efterhand jag rättade och feltypernas benämningar är delvis hämtade från Hultman och Westman (1977). I deras undersökning av gymnasisters svenska har de använt sig av ett felmarkeringssystem där de presenterar sju huvudtyper av fel som gymnasterna har gjort. För att kunna fånga skillnaden mellan de fel personer med svenska som förstaspråk gör och de fel som andraspråksinlärare gör, har jag valt att anpassa deras indelning till mitt material.

Semantiskt fel och lexikonfel (som är markerade med * i tabell 2) skiljer sig från resten av min feltypologi i det avseendet att de är svåra att analysera statistiskt. Ett exempel på lexikonfel har vi i meningen ”Sen *hittade* de en idé i deras järnar”. Om man känner till kontexten förstår man att inläraren med *järnar* menar *hjärnor* (stavfel) och att de *kom på* en idé eller *fick* en idé. Ordet *hittade* blir alltså ett lexikonfel. Däremot, när betydelsesammanhanget är så svagt att jag kanske inte ens förstår vad inläraren menar, har jag kategoriserat felen som semantiska fel. Till skillnad från lexikonfel som endast består av ett ord, kan ett semantiskt fel bestå av en hel mening (som exemplet i tabellen ovan) eller en del av en mening. Svårbegripliga eller obegripliga passager som har någorlunda korrekt syntax har jag alltså kategoriserat som semantiska fel.

Nedan följer en kort beskrivning av alla feltyper förutom lexikonfel och semantiska fel.

Stavfel: När jag rättade stavfelen kunde jag se att inlärare generellt gör en annan typ av stavfel än dem man är van att se. Inlärare kan göra sådana fel som att skriva *underbad* i stället för *underbart*, *studiera* i stället för *studera*. Det kan också vara så att de byter ut en vokal som i t.ex. *liva* (leva) och att det felstavade ordet blir ett annat, existerande ord än det som åsyftas. Av vissa felstavade ord kan man gissa att inläraren har stavat så som han eller hon uttalar dem, eller uppfattar att de uttalas. En av inlärarna skrev under en period *varför* i stället för *varför*.

Man kan framför allt se en viss skillnad mellan de stavfel som görs i början av inlärningsprocessen och de som görs när inläraren har större kunskap om målspråket. Längre fram i inlärningsprocessen, då inläraren har större ordförråd och större kunskap om stavning utgörs stavfelen av sådana fel som även någon med svenska som modersmål skulle kunna göra, t.ex. *kontroversiel*, *kulturelt*, *obeffintlig* osv.

Utländskt/okänt ord: När en inlärare inte hittar det svenska ordet händer det att han eller hon blandar olika språk. Ofta kunde jag under rättningen se att det handlar om inlån från modersmålet eller engelska. Det här är en typ av fel som jag endast upptäckte i början av inlärningsprocessen och det är inte alls särskilt vanligt hos de inlärare jag har undersökt.

Felböjt verb/substantiv/adjektiv: kan vara att verb är felböjda vid uttryckande av tempus, att substantiv är felböjda i numerus eller att adjektiv är felböjda vid komparation. Jag har i min feltypologi valt att skilja på ord som är felstavade och ord som är böjda på fel sätt. Då jag rättade texterna upptäckte jag att det är vanligt bland inlärare i ett visst stadium av inlärningsprocessen att använda sig av övergeneraliseringar (dvs. böja ett ord efter mönster som de behärskar och tillämpa det mönstret i överkant). Exempel på felböjt substantiv är ”vänner och *bekanter*” och ett exempel på felböjt verb har vi i meningen: ”Sen jag *vetade* att vill leva i Sverige”.

Felaktigt sammansatt sammansättning: foge-s saknas eller överflödigt foge-s. I svenskan är möjligheterna att bilda nya ord genom sammansättningar i princip obegränsade. Hur de bildas är däremot svårt att veta för den som inte har svenska som modersmål. Många av felen i min undersökning gäller sådana sammansättningar där fogeelementet –s är med på fel ställen som i till exempel *lägenhetannons* eller *cigarettspaket*.

Särskrivnen sammansättning: särskrivning av två led som ska stå ihop. Enligt Svenska Språknämndens skrivregler ska alla sammansättningar skrivas som ett ord. Om man har svenska som modersmål kan man oftast höra på betoning och intonation om det ska vara ett ord eller flera, jämför t.ex. *fem ton – femton*. För en inlärare blir det förstås svårare. Ändå verkar det inte vara någon större skillnad mellan de särskrivningar som inlärarna har gjort och de särskrivna sammansättningar man kan se dagligen, gjorda av någon med svenska som modersmål (Öhrman, 1998). Särskrivningar som *favorit ämnesområde* och *jätte fina* existerar överallt, oavsett författarens modersmål.

Versalfel:

Stor bokstav: ord som felaktigt har skrivits med stor initialbokstav. Detta fel har i huvudsak gjorts av inläraren med tyska och engelska som förstaspråk, möjligen på grund av transfer (se 2.1 ovan).

Liten bokstav: ord som saknar stor initialbokstav. Hit hör namn som skrivits med liten initialbokstav som till exempel *sverige*, *stockholm*. Detta fel har precis som *Stor bokstav* gjorts av inläraren med tyska och engelska som modersmål, vilket förmodligen beror på en övergeneralisering av huvudregeln att substantiv skrivs med inledande gemen på svenska.

Felaktig förkortning: punkt eller mellanrum saknas vid förkortning. Svenska Språknämndens skrivregler rekommenderar att punkt sätts ut vid förkortningar som till exempel *bl.a.* eller *fr.o.m.* eftersom ”*bla*” och ”*from*” kan förvirra eller irritera läsaren. Felet gör texten mer svårläst. I början av inlärningsprocessen borde förkortningar som *tom*, *etc*, *tex* och *sk* undvikas eftersom texterna kanske redan är svårlästa nog.

Genitiv med apostrof: genitiv skrivs med apostrof som i exemplet ”...av detta *barn's* stolta föräldrar”. Inte helt oväntat har den inlärare som har gjort det här felet engelska (och tyska) som modersmål.

Kongruensfel: Kongruens definieras i Thorells Svensk grammatik som att ett substantiv kräver ett visst numerus, genus eller species hos det styrda leDET vilket i sin tur kan bestå av artiklar, adjektiv, particip och pronomen. Även predikativ kongruerar med subjekt eller objekt i fråga om numerus och genus. De kongruensfel jag har hittat utgörs till stor del av substantiv som inte stämmer överens med omgivningen när det gäller numerus och genus, men alla ovan nämnda kongruensvarianter finns med i mitt material.

Bestämthetsfel: användning av bestämd form av substantiv där det ska vara obestämd form och tvärtom. Vid bestämthetsfel har jag sett till kontexten och om ordet i fråga har förekommit tidigare i berättelsen eller inte. De bestämdhetsfel jag har hittat är antingen substantiv i obestämd form när de borde stå i bestämd form som t.ex. ”Snart kom *kväll*, så jag måste resa till min hotel”, eller tvärtom: ”Nästa *dagen* åkte jag till...”. Ibland gränsar den här feltypen till en annan feltyp, nämligen ”saknas artikel”. I exemplet i tabellen ovan ”Jag har just beställt tid för *besöket* hos en psykoterapeut” kan man konstatera att *besöket* ska stå i obestämd form, men ska det inte i så fall dit en obestämd artikel också?

Adjektiv/adverbfel: användning av adjektiv där det ska vara adverb och tvärtom. Ofta kunde jag se detta fel i situationer där endast ett *t* skiljde adverb från adjektiv, som i meningen ”*Plötslig* tog han ut en liten låda”.

Fel pronomenform: *De*, *dem* eller *dom* är något som ställer till problem även för oss med svenska som modersmål. Även om *de* ofta används i objektsform föredras ändå formen *dem* i Svenska Akademiens ordlista. I min felsamling hade oftast pronomenet subjektform när det borde ha objektsform som i exemplet ”...vi stannar med *hon*”. Valet mellan possessiv konstruktion *hennes* och reflexiv konstruktion *sin* verkar också ställa till problem som i exemplet: ”Hon lämnade *hennes* hat (hatt) på golvet, tog *sin* handväska med sig och försökte att krypa in genom biljettluckan”.

Fel tempus: verbet står i ett tempus som inte stämmer överens med kontexten. Jag fann en mängd tempusfel i min undersökning. Att få verbet i rätt tempusform verkar utgöra ett problem för inlärarna. Problemet uppstår i första hand vid ett tidigt stadium av inlärningsprocessen då de flesta verb skrivs i infinitiv eller presensform, kanske för att det är de enda verbformer inläraren känner till än så länge. Att uttrycka dåtid verkar inlärarna behärska först senare när de lär sig hur verb böjs och vilket tempus som krävs av kontexten. I min undersökning har jag markerat som fel när ett verb inte följer tempusharmonin i resten av meningen, som i exemplet ”Allting *går* som de *ville* och de *betala* ingenting, en gång till”. En annan vanlig typ av tempusfel jag upptäckt är att inlärarna använder fel verbform i samband med hjälpverb, t.ex. ”Men, man bör aldrig *glömmar* och titta på bilder...”. Jag har också märkt att inlärarna ofta gör fler än ett tempusfel i en given mening.

Felaktig ordföljd: Majoriteten av ordföljdsfelen i min undersökning var fel som hade att göra med ordföljd i bisatser. Framför allt verkade placering av adverb utgöra ett problem för alla fyra inlärare. För att kunna bilda underordnade satser måste man veta vilka ordföljdsförändringar som sker i en bisats, något som kräver ganska goda kunskaper i svenska språket.

I huvudsatser däremot var det vanligaste felet inlärarna gjorde att använda rak ordföljd i meningar som borde ha omvänd ordföljd som t.ex. meningen ”Emellertid, *man kunde* följa någon riktlinjer”. När något annat led än subjektet inleder en påståendesats har satsen omvänd ordföljd, dvs. subjektet följer efter det finita verbet (Stroh-Wollin, 1998:44). Denna ordföljd är en markerad struktur i svenskan och utgör ett problem för de flesta inlärare oavsett modersmål.

Verbfrasfel med ”att”: verbfrasfel där ett onödigt ”att” har lagts till. De flesta fel av den här typen innehöll kombinationen *försökte att* + ett verb i infinitiv som till exempel i meningen: ”Jag försökte *att* prata med henne”. Enligt Thorells Grammatik ska hjälpverb som *behöva*, *börja* och i det här fallet *försöka* + infinitiv i regel konstrueras utan ”att”.

Verbfrasfel utan ”att”: det saknas ett ”att” i en verbfras där det normalt borde finnas ett. Ofta handlar det om infinitivkonstruktioner som till exempel i meningen: ”*Du kommer faktiskt kämpa många gånger*”, där hjälpverbet *kommer* (+ det *att* som saknas) har en futural betydelse. Men i vissa fall har inlärarna missat vissa att-konstruktioner som *jag tror/tycker + att* eller som i meningen: ”*När jag såg ett av hans skosnören var löst*”.

Saknas artikel: utebliven artikel eller determinerare före nominalfras. Meningar som ”*Hon var duktig dansös*” (Hon var *en* duktig dansös) där det saknas en obestämd artikel är ett vanligt exempel på den här typen av fel.

Saknas genitiv-s: substantiv som ska stå i genitiv men saknar genitiv-s. Jag upptäckte några meningar där ett substantiv saknar genitiv-s som till exempel: ”de döda flugorna är *framtiden* melodi!”.

Överflödig konstituent: Felet innebär att ett ord i en mening inte passar in eller inte behövs i sammanhanget som till exempel i meningen: ”Jag måste också jobba *min* själv på språklabb”.

Fel preposition: fel preposition, överflödig preposition eller preposition saknas. Ett vanligt fel som inlärarna gjorde var att använda fel preposition som i exemplet: ”Hon var så imponerad *med* mig”. Det var inte heller ovanligt att det i texten fanns överflödiga prepositioner som i meningen: ”Jag kommer att studera ekonomi *i* nästa år”. Under rättningen har jag i tveksamma fall använt mig av *Svensk Handordbok* där konstruktioner och fraseologi beskrivs.

En typ av fel som jag inte har med i min feltypologi är interpunktionsfel. Jag valde att utesluta dessa fel dels för att Granska inte detekterar interpunktionsfel och dels därför att t.ex. kommatering varierar väldigt mycket från individ till individ.

5.1.1 Fördelning av fel i feltypologin

Tabellen nedan visar hur inlärarnas fel fördelar sig på de olika feltyperna i min feltypologi.

<i>Felkategori</i>	<i>Typ av fel</i>	<i>Antal</i>	<i>Andel av totala fel</i>
Ortografiska fel	stavfel	289	26,1%
	utländskt/okänt ord	35	3,2%
	felböjt verb	14	1,3%
	felböjt substantiv	19	1,7%
	felböjt adjektiv	1	0,1%
	felaktigt sammansatt sammansättning	5	0,5%
	särskrivnen sammansättning	26	2,3%
	versalfel	11	1,0%
	felaktig förkortning	13	1,2%
	genitiv med apostrof	1	0,1%
Grammatiska fel	kongruensfel	156	14,1%
	bestämthetsfel	42	3,8%
	adjektiv/adverbfel	22	2,0%
	fel pronomenform	9	0,8%
	fel tempus	71	6,4%
	felaktig ordföljd	76	6,9%
	verbfrasfel med "att"	11	1,0%
	verbfrasfel utan "att"	12	1,1%
	saknas artikel	15	1,4%
	saknas genitiv-s	4	0,4%
Semantiska fel	semantiskt fel	69	6,2%
	överflödig konstituent	16	1,4%
Lexikonfel	lexikonfel	92	8,3%
	fel preposition	100	9,0%
	Totalt	1109	

Tabell 3 – Frekvenser för olika feltyper

Totalt har jag i min undersökning funnit 1109 fel i de 44 texterna. De vanligaste typerna av fel är markerade i tabellen. Vi kan konstatera att stavfel med 26,1 % är den i särklass mest förekommande typen av fel. Efter stavfel följer kongruensfel, felaktig preposition, lexikonfel, felaktig ordföljd, fel tempus och semantiskt fel. Resterande feltyper har vardera en frekvens som är lägre än 4 %. De två vanligaste felen, stavfel och kongruensfel, utgör tillsammans 40 % av alla fel.

5.2 Resultat av körning av Granska på texterna

Resultatet av körning av Granska på texterna visas i tabellen nedan. I kolumnen *ej detekterade* visas andelen fel som Granska inte har funnit. I kolumnen *detekterade* återges andelen fel som Granska har funnit. Slutligen i kolumnen *detekterade med fel diagnos* finner man fel som Granska hittat men som har tilldelats fel diagnos.

Typ av fel	Ej detekterade		Detekterade		Detekterade med fel diagnos	
	Antal	Andel	Antal	Andel	Antal	Andel
stavfel	93	32%	192	66%	4	1%
utländskt/okänt ord *	17	49%	17	49%	1	3%
felböjt verb *		0%	13	93%	1	7%
felböjt substantiv *	8	42%	11	58%		0%
felböjt adjektiv *	1	100%		0%		0%
felaktigt sammansatt sammansättning *		0%	5	100%		0%
särskriven sammansättning	14	54%	10	38%	2	8%
versalfel	10	91%		0%	1	9%
felaktig förkortning	9	69%		0%	4	31%
genitiv med apostrof	1	100%		0%		0%
kongruensfel	68	44%	87	56%	1	1%
bestämthetsfel	40	95%		0%	2	5%
adjektiv/adverbfel	22	100%		0%		0%
fel pronomenform	7	78%	2	22%		0%
fel tempus	68	96%		0%	3	4%
felaktig ordföljd	72	95%	4	5%		0%
verbfrasfel med "att"	2	18%	9	82%		0%
verbfrasfel utan "att"	9	75%	3	25%		0%
saknas artikel	15	100%		0%		0%
saknas genitiv-s	3	75%		0%	1	25%
semantiskt fel	62	90%		0%	7	10%
överflödig konstituent	16	100%		0%		0%
lexikonfel	86	93%		0%	6	7%
fel preposition	100	100%		0%		0%
Totalt	723	65%	353	32%	33	3%

Tabell 4 – Antal och andel fel i feltypologin som detekterats respektive ej detekterats samt de fel som detekterats men med fel diagnos.

(*) För dessa fel har jag betraktat diagnosen *stavfel* som riktig.

Av de 1109 felen i min undersökning detekterades nästan en tredjedel (32 %) med rätt diagnos av Granska. Ytterligare 3 % detekterades men med felaktig diagnos. Övriga 65 % detekterades inte alls.

Av alla detekterade fel utgör de två största grupperna, stavfel och kongruensfel, hela 79 %. Då dessa två typer utgör 40 % av totala antalet fel, kan man dra slutsatsen att Granska detekterar dessa feltyper betydligt bättre än andra feltyper.

I avsnitten nedan diskuteras de markerade feltyperna i tabellerna ovan mer ingående och även möjligheterna till detektion. I slutet av detta avsnitt finns även en översikt över de falska alarm som Granska genererat.

5.2.1 Stavfel

Inte helt oväntat utgjorde stavfel den största andelen fel i min undersökning (26,1 %). Av de 289 stavfelen hittade Granska 192 eller 66 %, medan 32 % inte detekterades och 1 % detekterades men fick fel diagnos.

Felen som i feltypologin följer efter stavfel, dvs. utländskt/okänt ord, felböjt verb/substantiv/adjektiv och felaktig sammansättning, detekteras av Granska som stavfel. Trots detta har jag valt att i min feltypologi göra den här finare indelningen för att mer i detalj se vilken sorts stavfel inlärarna gör. Om man slår ihop de ovan nämnda kategorierna får man samma fördelning mellan detekterade, ej detekterade och detekterade med fel diagnos, som man får för enbart stavfel.

En anledning till att Granska inte har lyckats hitta vissa stavfel är att det felstavade ordet sammanfaller med ett annat ord som rent syntaktiskt skulle kunna fungera i satsen. Detta ser vi till exempel i meningarna ”Sverige är ett *kalt* land” (ska vara *kallt*), ”Hon lämnade hennes *hat* på golvet” (ska vara *hatt*) och ”Vi kan gå till stranden och *svimmar* i havet”. Detta problem förvärras av att det i svenskan finns i princip oändligt många sammansatta ord. Förutom att felstavade ord kan bli ett annat ord så kan de även råka bli en kombination av flera ord som utgör en möjlig eller omöjlig sammansättning. Till exempel har ordet *underbart* felstavats som *underbad*, men Granska hittade inte felet eftersom ordet tolkades som en sammansättning av *under* och *bad*. Om ett ord inte känns igen delas det nämligen upp för att se om det finns ett möjligt förled och ett möjligt efterled.

5.2.2 Kongruens

Näst efter stavfel är kongruensfel den största felgruppen. Den representerar 14,2 % av andelen totala fel. Även dessa fel tillhör en typ av fel som Granska under normala omständigheter detekterar (enligt det senaste resultatet med 60 % täckning och 97 % precision). I min undersökning detekterade Granska 56 % av kongruensfelen; 44 % detekterades inte vilket har flera förklaringar. I första hand beror det på att ett ord är felstavat och därmed också feltaggat. Eftersom Granskas regler måste matcha en viss ordsekvens krävs det att orden är rätt stavade och rätt taggade för att ett fel ska kunna detekteras. Som exempel kan vi ta meningen ”Det är en fantastisk ö, inte så mycket utländskt folk, *ingen hotel*”. Här har *hotel*, som är felstavat, taggats som <pm.nom> och därför har ingen kontroll gjorts om artikeln kongruerar med substantivet. Ett annat exempel där kongruensfelet blir omöjligt att hitta för Granska är meningen ”...och därför har jag *ingen annan val*”. Med stor sannolikhet menar författaren till den här meningen *val* i betydelsen *alternativ* och därför är detta ett kongruensfel, men *val* är också ett ord som existerar i en annan betydelse med utrum som genus.

5.2.3 Fel tempus

Användning av fel tempus är en feltyp som är vanligt förekommande hos de andraspråksinlärare som jag har studerat. Granska har inte hittat något tempusfel med rätt diagnos. Möjligheterna att detektera tempusfel är små eftersom det kräver en god förståelse av kontexten, vilken kan sträcka sig betydligt längre än en mening. I meningarna med fel tempus fann jag få genomgående mönster med undantag för att inlärarna använde mycket presens- och infinitivformer i början av inlärningsprocessen. Ibland växlade de felaktigt mellan olika

tempus inom en mening, ibland stämde inte tempus med kringliggande meningar osv. Möjligen skulle man kunna förbättra detektionen av tempusfel i samband med hjälpverb, till exempel kontrollera att infinitiv används efter modala hjälpverb.

5.2.4 Felaktig ordföljd

Felaktig ordföljd är en annan stor felgrupp (6,9 % av alla fel) som Granska har problem med att detektera. Av 76 ordföljdsfel har Granska hittat 4. Med andra ord har 95 % av dessa meningar alltså inte detekterats. I min undersökning såg jag snart att vissa fel följde vissa bestämda mönster, t.ex. verkar de flesta inlärare ha svårt att placera adverbial på rätt plats i satsen. Detta gäller framför allt i bisatser, som till exempel i följande mening ”Nu min enda mål är att sluta *framgångsrikt* mina studier här”. Ett annat mönster som jag kunde urskilja var att inlärarna hade en tendens att använda rak ordföljd i nästan samtliga meningar, det vill säga med subjektet före det finita verbet. Detta gäller framför allt i påståendesatser där något annat än subjektet inleder meningen som t.ex. ”Kanske *det är* också en grund varför Amerika är så angagerat”. Nästan hälften (29) av meningarna med felaktig ordföljd (69) var bisatser där det oftast var adverbet som hamnat fel, och ungefär lika många var påståendesatser där subjektet felaktigt hamnat före det finita verbet.

Detektion av felen som görs i bisatser hänger mycket på hur Granska klarat av att tagga bisatsinledaren. Om man klarar av att identifiera subjunktionen, som oftast inleder en bisats, och tilldela den rätt tagg så är chanserna stora att finna felet. Ett problem är att inte alla bisatser inleds med en subjunktion, men för det mesta gör de det. I min undersökning fann jag åtta bisatsinledare som visas i tabellen nedan. Dessutom visas hur många av varje enskild bisatsinledare som förekom i meningarna och vilken tagg de tilldelats.

Bisatsinledare	Tagg	Antal
där	<ab> (adverb)	1
som	<hp> (frågande/relativt pronomen)	8
att	<ie> (infinitivmärke)	2
att	<sn> (subjunktion)	12
när	<ha> (frågande/relativt adverb)	2
hur	<ha> (frågande/relativt adverb)	1
om	<sn> (subjunktion)	1
varför	<ha> (frågande/relativt adverb)	1
därför	<ab> (adverb)	1

Tabell 5 – Fördelning och taggning av bisatsinledare.

Möjligheterna till detektion skulle vara mycket större om fler av bisatsinledarna taggades som subjunktioner. Om bisatsinledaren är taggad som en subjunktion är det lättare att detektera bisatsen eftersom alla subjunktioner följs av en bisats (Ejerhed et al., 1992). Vet man att man har hittat en bisats kan man enklare detektera ett eventuellt ordföljdsfel. Som vi kan se i tabellen taggas många bisatsinledare med taggarna <hp> och <ha> som täcker in massor med frågande/relativa pronomen respektive adverb. Vissa av dessa bisatsinledare är i själva verket subjunktioner.

Vi ser också i tabellen att ordet *att* felaktigt taggats som infinitivmärke när det egentligen varit en subjunktion. Exempelen visar vilken betydelse en korrekt taggning har för möjligheterna att detektera fel i allmänhet.

För diskussion kring möjliga regler som kan hjälpa Granska att finna meningar med felaktig ordföljd, se 5.3 nedan.

5.2.5 Felaktig preposition

Sett till antalet fel kommer felaktig preposition efter kongruensfel. Möjligheterna till detektion av dessa fel är små för att inte säga näst intill obefintliga. Enligt Thorells Svensk grammatik kan prepositionernas olika betydelser inte sammanfattas i enkla och klara regler utan måste studeras i ett lexikon för varje preposition. Till exempel kan prepositionen ”i” i vissa uttryck ha en konkret betydelse, vanligen rum: *i Stockholm*, tid: *i kväll*, sätt eller medel: *hand i hand*, i andra åter en mera abstrakt betydelse, ofta kausal eller final (orsak eller syfte): *i smyg* (Thorell, 1987:172). Med alla dessa alternativ blir det förstås svårt att hitta regler för hur Granska ska kunna detektera prepositionsfel. Man skulle kunna utnyttja att vissa verb tillsammans med en preposition bildar ett så kallat ”prepositionsverb”, som t.ex. verbet *njuta* som följs av prepositionen *av* (Hon njöt av sin ledighet). Ett problem om man skriver språkgranskningsregler utifrån prepositionsverb är att ibland så kan verbet följas av en annan preposition som är en del av t.ex. ett adverbial (Hon njöt på morgnarna).

5.2.6 Semantiskt fel, lexikonfel

Dessa feltyper är på grund av sin natur svåra att analysera (se vidare diskussion under 5.1 ovan). Av samma skäl är de nästan omöjliga att detektera för Granska. Programmet har inte detekterat något semantiskt fel eller lexikonfel med rätt diagnos.

Ett exempel på en mening som jag felmarkerat som semantiskt fel är: ”Vi åt på en fin kinesisk restaurang men *vi villste inte så långa väntade*”. Andra halvan av meningen är omöjlig att tyda. Det Granska har hittat är ett stavfel, ordet *villste*. Det är av naturliga skäl omöjligt för Granska att upptäcka att betydelsesammanhanget i meningen är obefintligt. Det finns situationer där en mening kan vara syntaktiskt och ortografiskt korrekt, men betydelsesammanhanget svagt. Sådana fel skulle endast kunna upptäckas om språkgranskningsprogrammet hade samma djupgående kunskap om vår omvärld som vi människor har. Resonemanget gäller i synnerhet semantiska fel, men även lexikonfel. I det senare fallet kan jag ibland utifrån kontexten förstå att inläraren egentligen menar ett annat ord än det han eller hon skriver, något Granska självklart inte klarar av och därmed upptäcks inte felet.

5.2.7 Antal falska alarm

Totalt har Granska gett 61 falska alarm på texterna i min samling. Med tanke på att det finns över 12 000 ord i min samling är detta ett gott resultat. Antalet falska alarm bör också sättas i relation till de 353 fel som Granska har hittat. Delar man antalet korrekt detekterade fel med det totala antalet rapporterade fel, inklusive falska alarm, får man fram att precisionen är 85 %.

Jag har delat in de falska alarmen i grupper beroende på anledningen till att de har uppkommit.

Egennamn: Egennamn som inte finns i Granskas lexikon detekteras som stavfel, t.ex. ”Det är 20 kilometer från *Zafra* till Portugal”. Ordet *Zafra* har detekterats som ett stavfel.

Frågor: Frågor är ofta ofullständiga satser vilket Granska har problem med. Exempelvis rapporterar Granska att ett finit verb saknas i frågan: ”*Vart?*”. I och för sig är det så att ett finit verb saknas, men likväl är det vanligt även i skriftspråk med frågor som inte är fullständiga satser. I detta avseende kan man betrakta dessa alarm som falska.

Utrop och tilltal: Liksom med frågor har granska problem med utrop och tilltal som t.ex. ”*Vad hemskt!*”, ”*Lycka till!*” eller ”*Kära Anna!*” . Även här hittar Granska felet att finit verb saknas och precis som med frågor så är detta visserligen sant, men ändå att betrakta som falskt alarm.

Övrigt: Till denna grupp hör övriga falska alarm. Ett exempel är meningen: ”Men svenska tidningarna säljer i *genomsnitt flera exemplar*”. Här hittar Granska ett särskrivningsfel och föreslår att man ändrar *genomsnitt flera* till *genomsnittflera*.

I tabellen nedan redovisas hur många falska alarm som går att hänföra till de olika grupperna.

Grupper	Antal falska alarm
egennamn	7
frågor	5
utrop	11
övrigt	38
totalt	61

Tabell 6 – Fördelning av falska alarm på olika grupper.

5.3 Felaktig ordföljd och regler

Som exempel på förbättringar till Granskas regler som kan underlätta för andraspråksinlärare, har jag tagit fram några regler som detekterar felaktig ordföljd.

Som vi sett tidigare förekom många av ordföljdsfelen i bisatser. I de två första förslagen till regler nedan detekteras bisatser genom att leta efter ord taggade som subjunktioner eller ordet *som* och efterföljande subjekt och predikat. I det tredje förslaget ges en regel som försöker hitta felaktig ordföljd i huvudsatser som inte inleds med subjekt.

NP i reglerna nedan syftar på en hjälpregel som definierar en nominalfras. I exempelmeningarna har jag kursiverat de ord som utlöser regeln.

5.3.1 Regel 1

I många av de bisatser där inlärnarna haft felaktig ordföljd ska adverbet stå före verbet. Regeln hittar situationer när adverbet står efter verbet i bisatser. För att regeln inte ska ge falska alarm vid situationer där adverbet fungerar som bestämning till efterföljande adverb eller adjektiv lägger vi till att adverbet inte får följas av ett adjektiv eller ytterligare ett adverb.

Exempelmening: Det är tydligt *att han vill bara kontrollera* oljen som tidigare.

```
{
X1/(wordcl=sn | text="som") & sed!=sen/,
(NP,X2)//,
X3/wordcl=vb/,
X4/wordcl=ab/,
X5/wordcl!=jj & wordcl!=ab/
-->
/*X1 X2 X4 X3 X5///granskning/
}
```

Första variabeln (X1) matchar en subjunktion eller ordet "som", dock inte i början av en mening. Nästa variabel matchar en nominalfras. Tredje och fjärde variabeln matchar verbet och adverbet, som har hamnat i fel ordning. Slutligen får den femte variabeln inte matcha adjektiv och adverb för att regeln inte ska orsaka falska alarm enligt ovan.

5.3.2 Regel 2

Den här regeln bygger på att ordet *inte* ska stå före verbet i bisatser. Regeln överlappas delvis av regel 1 eftersom *inte* är ett adverb. Eftersom regeln bygger på förekomst av ord snarare än ordklasser borde den ha lägre täckning och högre precision jämfört med regel 1.

Exempelmening: Jag hoppas *att jag träffar inte* honom igen.

```
{
X1/wordcl=sn | text="som"/,
(NP,X2)//,
X3/wordcl=vb/,
X4/text="inte"/
-->
/*X1 X2 X4 X3///granskning/
}
```

Första variabeln (X1) matchar en subjunktion eller ordet "som". Nästa variabel matchar en nominalfras. Tredje och fjärde variabeln matchar verbet respektive ordet "inte".

5.3.3 Regel 3

När en huvudsats inleds med något annat än subjektet blir det omvänd ordföljd mellan subjekt och predikat. Ofta inleds sådana satser av ett adverbial som består av en preposition följt av en nominalfras.

Exempelmening: *På studentshem det finns folk från Tyskland, Finland, Kina och Spanien.*

```
{
  X1/wordcl=pp & sed=sen/,
  (NP,X2)//,
  X3/wordcl=pn/,
  X4/wordcl=vb & (vbf=prs | vbf=prt)/
  -->
  /*X1 X2 X4 X3///granskning/
}
```

I regeln ovan matchas den första variabeln mot en preposition som påbörjar en sats. Den andra variabeln matchas mot en nominalfras, med hjälp av en hjälpregel. Tredje variabeln matchas av ett pronomen och den fjärde av ett verb som står i presens eller preteritum.

De tre reglerna ovan kördes på de meningar som innehöll ordföljdsfel. Ungefär en tredjedel av 76 ordföljdsfel hittades, vilket ska jämföras med de 4 felen som Granska klarade tidigare. Flertalet fel hittades av regel 1.

Idéerna till reglerna ovan kan utvecklas och kombineras med varandra. Noggranna empiriska undersökningar av reglerna på ett större textmaterial krävs för att förhindra falska alarm som kan utgöra ett stort problem för andraspråksinlärare.

6 DISKUSSION

I avsnitten nedan diskuteras först inlärarnas svenska och vilka fel de gör. Därefter följer en diskussion kring hur inlärarnas fel kan rättas och kategoriseras samt problemställningar kring detta. Sedan diskuteras Granskas detektion av andraspråksinlärares fel och vilka svårigheter som uppstår i samband med detta. Därefter diskuteras språkgranskningsprogrammets roll och eventuella inverkan. Slutligen redovisas de slutsatser jag har kunnat dra.

6.1 Inlärarnas språk och svårigheter

För mitt arbete hade jag stor glädje av att få tillgång till ASU-korpusen. Den har många fördelar för mitt syfte, t.ex. att den spänner över en längre tidsperiod så att man kan fånga alla fel som görs, från början till slutet av inlärningsprocessen. Jag har kunnat se att inlärarna när de blivit bättre på målspråket gör andra typer av fel, mer lika de fel som infödda gör, jämfört med i början av inlärningsprocessen.

En nackdel med min metod är att jag har relativt få individer i min undersökning. Jag var tvungen att begränsa mig till europeiska inlärare eftersom de andra inlärarnas texter, särskilt i början av inläringen, var svårbegripliga och på gränsen till ogenomträngliga. För att kunna analysera vad som är fel i en mening och rätta det, måste jag förstå vad författaren menar. En fördel med att studera en relativt sett mer homogen grupp är att undersökningen av felen blir mer konsekvent.

Olika inlärare gör olika fel, en inlärare med grekiska som modersmål har till exempel en tendens att använda felaktig ordföljd mer än andra. Skulle det ge ett bättre resultat om man kunde ställa in programmet så att det vet att här kommer text skriven av en person med grekiska som modersmål? För att kunna dra en sådan slutsats skulle man behöva jämföra de enskilda inlärarnas resultat med varandra. Generellt har jag inte sett några större skillnader att tala om men så är ju också alla inlärarna i min undersökning från Europa. Men även om det fanns större skillnader mellan inlärare med olika förstaspråk, tror jag att det inte skulle tillföra så mycket att anpassa programmet efter användarens språk. Har man skapat ett system som detekterar ordföljdsfel väldigt bra och som någon med grekiska som modersmål har stor glädje av, finns det ingen anledning att inte tillämpa den funktionen när t.ex. en person med engelska som modersmål använder programmet.

I och med att jag upptäckte att det inte var så stora skillnader mellan vilka fel de enskilda inlärarna gjorde, valde jag att presentera mitt resultat i form av en sammanslagning av de fyra inlärarna. Likheterna mellan de olika inlärarnas fel stödjer teorin om s.k. naturliga inlärningsgångar, det vill säga att språkinläring sker enligt vissa mönster.

6.2 Rättning och kategorisering av fel

Att avgöra vad som är ett fel är inte alltid lätt. Något som verkar rätt i en mening kan vara helt fel när man känner till kontexten. I min undersökning av inlärarnas fel var det min tolkning av vad författaren ville få fram som avgjorde hur felen kategoriserades. Rättningen av texterna blev i viss uträkning subjektiv eftersom gränsdragningen mellan rätt och fel samt mellan olika typer av fel är svår, i synnerhet när inlärarna befinner sig på en lägre språknivå.

En svårighet vid rättning och räkning av felen var att felen många gånger överlappade varandra, t.ex. i meningen ”Kan du visa mig *bästa platser* i Stockholm?”. Saknas det en artikel (*de*)? Är det ett bestämdhetsfel (*platserna*)? Eller ett lexikonfel (*bra* i stället för *bästa*)? Eller är det en kombination?

Det är svårare att rätta texter som andraspråksinlärare har skrivit än texter som är skrivna av personer med svenska som modersmål. Inlärarna gör fler fel och ofta överlappar de varandra. Dessutom är deras fel av en annan karaktär. Deras språk präglas av den språkbild, de grammatiska strukturer och syntax som de har med sig från sitt förstaspråk och ibland även från andra andraspråk. För den som ska rätta felen och som har målspråket som modersmål, kan det vara svårt att sätta sig in i inlärarens sätt att tänka och hur han eller hon skapar meningar. Alla dessa faktorer försvårar rättningen.

6.3 Detektion

Ett stort problem vid detektion är att många fel inte detekteras på grund av att orden är felstavade. Även om stavfelet i sig detekteras, missar språkgranskningsprogrammet på grund av stavfelet andra kringliggande eller relaterade fel. Dessutom kan det felstavade ordet ha råkat bli ett nytt ord så att inte ens ett stavfel detekteras.

För att taggning och regler ska fungera krävs det att orden går att identifiera. Eftersom andraspråksinlärare gör många och annorlunda stavfel, är det väsentligt att stavningskontrollen lyckas. Det skulle kunna vara en fördel för andraspråksinlärare om Granska först rättade stavning och sedan kördes en gång till för att bättre kunna detektera grammatikfelen och korrigera dem.

Stavfel är inte det enda som orsakar problem med detektion i Granska. Det kan också vara så att fel ordval förstör chanserna till detektion.

Problem med detektion och falska alarm kommer kanske aldrig att kunna lösas fullt ut. Eftersom språket är levande och till sin natur komplext, är det svårt att göra ett system med vattentäta regler som hittar alla fel och aldrig ger falska alarm. För att göra detta skulle systemet behöva kunna göra en fullständig analys av indata, såväl grammatisk som semantisk.

På kort sikt skulle Granska kunna förbättras genom att man förutom taggning även skulle göra en mer omfattande syntaktisk analys (parsning) och finna satsdelar, satser, med mera. Då skulle man kunna skriva bättre regler genom att använda satsdelar i reglerna. Regler som till exempel BIFF-regeln (i Bisats står Inte Före Finit verb) skulle kunna läggas in enklare om man kunde matcha bisats, subjekt, etc.

På längre sikt skulle man kunna förbättra Granskas detektion genom att föra in en semantisk dimension. Detta skulle innebära att programmet på något plan förstod betydelsen av texten och kunde använda detta för att detektera olika typer av fel.

6.4 Språkgranskningsprogrammets roll och inverkan

Språkgranskningsprogrammets funktion kan vara tveeggad för en andraspråksinlärare. Fördelarna är många men det finns flera fallgröpar: falska alarm, feldiagnoser, dåliga rättningsförslag och svårförstådda förklaringar.

I första hand ska programmet kunna detektera så många fel som möjligt. Sedan måste felen också ha rätt diagnos. Dessutom är det viktigt att det är så få falska alarm som möjligt. För någon med ganska goda språkkunskaper är falska alarm och feldiagnoser irriterande, för en andraspråksinlärare kan det däremot utgöra ett stort problem, förutom möjligtvis i undervisningssituationer med lärarstöd.

När man skapar regler i språkgranskningsprogram finns det ett motsatsförhållande mellan att de ska hitta så många fel som möjligt, det vill säga ha en hög täckning och att reglerna inte ger falska alarm, det vill säga att de har hög precision. Eftersom andraspråksinlärare i större utsträckning får problem med falska alarm och feldiagnoser, så bör man i program som riktar sig till dem i högre grad satsa på precisa regler som ger få falska alarm, även om det minskar antalet detekterade fel. Det blir en avvägning mellan högre detektion med fler falska alarm eller lägre detektion med färre falska alarm. Ett exempel på detta är Regel 3 i avsnitt 5.3.3 ovan. Om man släpper kravet att regeln bara matchar början på satser, ökar antalet detektioner av ordföljdsfelen från 2 till 6. Samtidigt ökar antalet falska alarm för regeln från 0 till 3.

Ett stort problem för språkgranskningsprogram utöver att detektera fel är hur man ska kunna förklara felen och kunna ge förslag till rättning. Om språkgranskningen ska vara till hjälp för andraspråksinlärare är det speciellt viktigt att feldiagnoser och rättningsförslag är enkla och pedagogiska.

6.5 Slutsatser

Jag har funnit att de typer av fel som inlärarna gör motsvarar ungefär vad jag förväntade mig. De mest frekventa feltyperna för inlärarna är (i fallande storleksordning): stavfel, kongruensfel, fel preposition, lexikonfel och felaktig ordföljd. Av dessa är stavfel, fel preposition och lexikonfel sådana fel som borde utgöra ett problem för alla inlärare som lär sig ett nytt (indoeuropeiskt) språk. Kongruensfel och felaktig ordföljd är fel som är mer språkspecifika.

Av ovan nämnda fel klarar Granska att detektera de två största grupperna, stavfel och kongruensfel, med en relativt hög frekvens. För de övriga vanligt förekommande feltyperna är frekvensen låg. Ska man förbättra detektion i Granska bör man koncentrera sig på att förbättra

detektion av fel som dels är vanligt förekommande, dels är möjliga att detektera med dagens teknik. Felaktig ordföljd är ett exempel på en sådan typ av fel.

Inlärarna i min undersökning gör ungefär samma typer av fel oavsett vilket deras förstaspåk är. Detta ger en indikation på att system som skräddarsys efter inlärnas förstaspåk endast har ett begränsat värde. Vad man borde satsa på är att finna de markerade strukturerna i svenskan, det vill säga sådant som inlärare generellt har svårt för.

Eftersom andraspråksinlärare får problem med falska alarm och feldiagnoser, så bör man i språkgranskningsprogram som riktar sig till dem i hög grad satsa på regler som ger få falska alarm, även om det blir på bekostnad av antalet detekterade fel. När man skapar regler i Granska finns det ett motsatsförhållande mellan att hitta så många fel som möjligt, det vill säga ha en hög täckning och att reglerna inte ger falska alarm, det vill säga att de har hög precision. Det behövs mer forskning för att undersöka hur falska alarm och feldiagnoser påverkar andraspråksinlärare.

Stavfelen utgör en stor anledning till att många andra fel inte går att detektera med hjälp av skrivstödsprogrammet Granska. För att hjälpa inlärare på en låg nivå är det lämpligt att först rätta stavfelen så att de övriga språkgranskningsreglerna sedan kan bearbeta materialet på ett riktigt sätt.

På kort sikt kan detektion av andraspråksinlärares fel i Granska förbättras genom att införa en mer omfattande syntaktisk analys av texten. På så sätt kan fler språkregler göras om till Granska-regler. Ett sådant system skulle i första hand vara till hjälp för den något mer avancerade inläraren, vars språk börjar få så pass mycket struktur att reglerna har något att ta fäste vid. Riktigt bra språkgranskning av andraspråksinlärares texter kommer att kräva att man konstruerar program med semantisk förståelse. Ett sådant program skulle precis som en mänsklig rättare kunna förstå vad det är som inläraren *egentligen* menar.

Det är en utmaning att göra ett språkgranskningsprogram som kan vara till hjälp för andraspråksinlärare som befinner sig på en mindre avancerad nivå. Förutom en stavningskontroll i två faser enligt ovan, krävs att man lägger en stor ansträngning på att utforma feldiagnoser och förslag till korrektion så att de blir så lättförståeliga och pedagogiska som möjligt.

7 REFERENSER

- Domeij, Rickard, 1997. *Datorn och språkriktigheten. Svenskan i IT-samhället. Ord och stil 28*. Hallgren och Fallgren studieförlag, Uppsala.
- Domeij, R., Knutsson, O., Carlberger, J. & Kann, V. 1999. *Granska – ett effektivt hybridsystem för kontroll av svensk grammatik*, NoDaLiDa, december 1999.
- Domeij, R., Knutsson, O., Larsson, S., Severinson Eklund, K. & Rex, Å., 1998. *Granskaprojektet 1996-1997*. IPLab, Nada, KTH, Stockholm.
- Ejerhed, E., Källgren, K., Wennstedt, O., Åström, M., 1992. *The linguistic annotation system of the Stockholm-Umeå corpus project – Description and guidelines*. Språkvetenskapliga institutionen, Umeå universitet.
- Ekerot, Lars-Johan, 1995. *Ordföljd, tempus, bestämdhet. Föreläsningar om svenska som andra språk*. Gleerups förlag, Malmö.
- Ellis, Rod, 1994. *The Study of Second Language Acquisition*. Oxford University Press.
- Hammarberg, Björn, 1997. *Manual of the ASU Corpus, a longitudinal text corpus of adult learner Swedish*. Institutionen för lingvistik, Stockholms Universitet.
- Hultman, T., Westman, M., 1977. *Gymnasistsvenska*. Bröderna Ekstrands Tryckeri AB, Lund.
- Kann, V., Domeij, R., Hollman, J. & Tillenius, M. 1998. *Implementation aspects and applications of a spelling correction algorithm*, Journal of Quantitative Linguistics.
- Knutsson, Ola, 1999. *Granskas nya regelspråk*. IPLab, Nada, KTH, Stockholm.
- Kotsinas, Ulla-Britt, 1991. *Invandrare talar svenska*. Akademitryck, Edsbruk.
- Stroh-Wollin, Ulla, 1998. *Koncentrerad nusvensk formlära och syntax*. Studentlitteratur, Lund.
- Svensk Handordbok. Konstruktioner och fraseologi*, 1999. Svenska språknämnden och Norstedts Ordbok AB, Almqvist & Wiksell, Uppsala.
- Svenska Akademiens ordlista över svenska språket*, (SAOL), 1986. Norstedts förlag, Stockholm.
- Svenska skrivregler utgivna av Svenska språknämnden*, 1991. Almqvist & Wiksell, Stockholm.
- Thorell, Olof, 1987. *Svensk grammatik*. Norstedts Tryckeri, Stockholm.
- Viberg, Åke, 1987. *Vägen till ett nytt språk 1, andraspråksinläring i ett utvecklingsperspektiv*. Bokförlaget Natur & Kultur, Stockholm.
- Öhrman, Lena, 1998. *Felaktigt särskrivna sammansättningar*. C-uppsats i datorlingvistik, Institutionen för lingvistik, Stockholms Universitet.

APPENDIX A: SUC-TAGGAR

SUC-tag	Betydelse	Exempel på ord som har taggen
AB	adverb	inte
AKT	aktiv form	spelar
AN	förkortning	t.ex.
DEF	definit	bilen
DL	skiljetecken	.
DT	determinerare (artikel)	den
GEN	genitiv	bilens
HA	frågande/relativt adverb	när
HD	frågande/relativ determinerare	vilken
HP	frågande/relativt pronomen	vem
HS	frågande/relativ possessiv	vems
IE	infinitivmärke	att
IMP	imperativ	spring
IN	interjektion	ja
IND	indefinit	bil
IND/DEF	indefinit/Definit	gula
INF	infinitiv	spela
JJ	adjektiv	gula
KN	konjunktion	och
KOM	komparativ	gulare
KON	konjunktiv form	vare
MAD	skiljetecken i slutet av en mening	.
MAS	maskulinum	gule
MID	skiljetecken inom en mening	,
NEU	neutrum	huset
NN	nomen (Substantiv)	bilen
NOM	nominativ	bilen
OBJ	objektform	mig
PC	particip	kastad
PL	partikel	om
PLU	plural	bilar
PM	egennamn	Svensson
PN	pronomen	hon
POS	positiv	gul
PP	preposition	till
PRF	perfekt	kastad
PRS	presens	spelar
PRT	preteritum	spelade
PS	possessiv	mina
RG	räkneord grundtal	två
RO	räkneord ordningstal	andra
SFO	s-form, passiv eller deponens	behövdes

SIN	singular	bil
SIN/PLU	singular/plural	boxande
SMS	sammansättning	pojck- (och flickrum)
SN	subjunktion	om
SUB	subjekt	jag
SUB/OBJ	subjekt/objekt	den
SUP	supinum	spelat
SUV	superlativ	gulast
UO	utländskt ord	action
UTR	utrum	bil
UTR/NEU	utrum/neutrum	gula
VB	verb	spela

APPENDIX B: TEXTER UR ASU-KORPUSEN

Nedan följer två exempel på texter ur ASU-korpusen. Båda texterna är skrivna av samma person. Den första texten är från början av inlärningsprocessen medan den andra är från slutet.

Jag vill alldrig glömma min första dag i Sverige. Det var i höst men det var väldigt fantastiskt höst dag. Himmelen var stark blå och luften var kallt och frisk. Tredet varat gult och orange. De set ut lika statluetor på jorden. Jag trodde att sverige var så olika från alla andra länder. En fantastiska nordika atmosfären. Vi gick på en lång promenad för tre timmen i en stor skoggen. Vi hadde så många till tala. Vi pratade om vår liva, om vår framtiden och också om Afrika.

Sedan gick vi hemma och drack en hella flask rödvin. Vi skålade tillander och vi tittade på vårt ögon. Då vi gick ut, ut till frisk luften.

Vi åt på en finn kinesiska restaurang men vi villste inte så långa väntade. Jag tycker det va en viktigt moment för mig. Vi långsam springade hem och jag tänkte, et är roligt men jag ha enorm positive vibrationer. Sen jag vetade att vill leva i Sverige. Sen jag vetade Sverige hadde en hand på mig. Ibland jag tror min väg hadde vara så. Men nu jag tänker olika om Sverige. Jag lärarde många om Sverige människor och deran liver. Från alla saker enting är inte olika. Jag älskarde Sverige på min första besök, och jag älskar Sverige nu också. Det är viktigt på mig.

Nu ska jag berätta en väldigt otrolig händelse som jag upplevde några veckor sedan när jag åkte ned till Frankrike. Det var nämligen så att jag har en gammal klass kamarat som jag inte har sett sedan vi avslutade gymnasiet. En dag fick jag ett brev från honom att jag skulle hälsa på honom och just det bestämde jag mig för. Det var en fredag kväll när jag flög in till Lyon. Han hämtade mig från flyplatsen och vi åkte direkt in i stan. Det var så mycket att prata om och jag var väldigt nyfiken om hur han har haft det de sista åren och hur han har blivit. "Fransk mat ska det bli" bestämde han och vi gick in i en vanlig men gemutlig restaurang.

Vi bäge två åt fisk och drack ett excellent vin. Han berättade om hans liv och det var fylld med kvinnor - och sprit. På en gång fick jag intrycket att han levde intensiv men samtidigt farlig. Han chansade på allting och han var så fantastisk omedvetande om faktum. Plötslig tog han ut en lite låda som han hade i fickan. På den tiden hade vi nästan ätit upp allting och med en otrolig lugnhet la han en död flygga på tallriken. Han ropade till kyparen och blev plötslig väldigt upprörd. Han frågade kyparen varför man måste dela sin mat med en flygga och överhuvudtaget vilka tillstånd de har i restaurangen. Kyparen blev hellt ställt och innan jag fattade hur otrolig skicklig min vän var, gick vi ut utan att har betalt en öre. Jag sa ingenting för att jag tyckte det var hellt fantastiskt och överhuvudtaget tycke jag att han skulle ha blivit en skådspeare. Jag skrattade bara och hade en känsla att jag började att lära känna min gammal vän. Vi pratade över gamla tider och promenerade genom halva stan.

"Här ska vi gå in" sa han och då hamnade vi i en väl etablerat Café. Vi tog en röd vin av bästa kvalitet och njutade. Innan vi drack upp allting slängde han en flygga i glaset. Det var ett reprise och kyparen stod bara där och kunde inte säga ett ord. Vi gick ut och då begrepp jag att min vän hade hittat ett livsstil där man får ett väldigt speciellt förhållande till flyggor.