



LUNDS
UNIVERSITET

Bedömning av korrekthet och komplexitet i ett muntligt språkfärdighetstest

Anders Agebjörn

Handledare: Anna Flyman Mattsson

Språk- och litteraturcentrum, Nordiska språk
Masteruppsats i Svenska som andraspråk, 30 hp
27 januari 2015

Abstract

In this thesis, I investigate the relation between assessed speaking proficiency and measured complexity and accuracy in twenty test-takers' performance in the oral part of the Swedish certificate test Swedex B2. Accuracy and complexity have been measured in seven different ways; both global and local measures from the CAF (*Complexity–Accuracy–Fluency*) triad have been used. Also, each test-taker's level of processability capacity, according to the Processability theory (PT), has been estimated. Hence, I have tested correlations between in total eight measurements using Pearson's correlation coefficient (r). Most measures co-vary with each other with some degree of significance. Furthermore, correlations between measurements and grades have been tested. The effect of accuracy and complexity on both a specific grade for 'Formal accuracy and complexity' and two holistic grades – one for the oral part of the test and one for the whole test, consisting of four parts – have been investigated. Results indicate that most measurements co-vary with the holistic grade for the oral exam. However, the specific grade co-varies with foremost global accuracy measures, indicating a halo-effect. In a final qualitative investigation, four test-takers are selected and examined closer. From this study, all seven CAF measures can be criticised; only PT level seems to capture the level of proficiency. Quantitative as well as qualitative results are considered in a discussion of both language testing and language proficiency measures.

Tack

Jag vill tacka min handledare Anna för hennes stora engagemang och hennes uppmuntran, men framför allt för hennes mycket proffsiga handledning. Jag vill också tacka Andreas Eraybar på Swedex, Folkuniversitetet, som inte bara hjälpt mig så mycket med att samla in mitt material utan också visat ett så stort och uppriktigt intresse och för mitt projekt. Dessutom är jag mycket tacksam mot de tjugo testtagare som låtit mig ta del av inspelningar och betygsprotokoll från deras muntliga språktest. Även om det inte framgår av studien – den handlar mest om AS-enheter, underordningskvot, V2, kongruens med mera – tycker jag mig ha lärt känna dessa människor en aning, och jag önskar dem alla lycka till med det svenska språket och med livet i Sverige.

Innehållsförteckning

Tack	i
Innehållsförteckning	ii
Figurförteckning	iv
Tabellförteckning	v
Diagramförteckning	vi
1 Inledning	1
1.1 Syfte och frågeställning	1
1.2 Uppsatsens upplägg	2
2 Bakgrund	3
2.1 Att testa (muntlig) språkfärdighet	3
2.1.1 Reliabilitet och validitet	4
2.1.2 Den historiska utvecklingen av (muntliga) språktest	5
2.1.3 Kommunikativ språkförmåga	6
2.2 Gemensam europeisk referensram för språk (CEFR)	7
2.2.1 Språkliga aktiviteter – VAD talaren kan göra	8
2.2.2 Språkliga kompetenser – HUR talaren kan göra det	9
2.2.3 CEFR om bedömning	11
2.3 Komplexitet, korrekthet och språkfärdighet	11
2.3.1 CAF-triaden	12
2.3.2 Komplexitet – definitioner och operationaliseringar	13
2.3.3 Korrekthet – definitioner och operationaliseringar	15
2.3.4 Studier om komplexitet, korrekthet och språkfärdighet	16
2.4 Grammatiska utvecklingsstadier och språkfärdighet	18
2.4.1 Processbarhetsteorin	19
2.4.2 Studier om Processbarhetsteorin och allmän språkfärdighet	20
2.5 Sammanfattning av bakgrunden	22
3 Metod och material	22
3.1 Swedex muntliga B2-test	22
3.2 Material	25
3.3 Etiskt övervägande	26
3.4 Transkribering och kodning	27
3.4.1 Kodning av komplexitet	29
3.4.2 Kodning av korrekthet	33
3.4.3 Inter-bedömarreliabilitet	35
3.4.4 Bestämning av PT-nivå	35
3.5 Statistisk analys	37

4 Resultat	38
4.1 <i>Korrelationer mellan mått</i>	38
4.1.1 Korrelationer mellan komplexitetsmått	39
4.1.2 Korrelation mellan korrekthetsmått	39
4.1.3 Korrelationer mellan korrekthetsmått och komplexitetsmått	40
4.1.4 Korrelationer mellan PT-nivå och korrekthets- och komplexitetsmått	40
4.2 <i>Korrelationer mellan bedömningar</i>	41
4.3 <i>Korrelationer mellan mått och bedömningar</i>	41
4.4 <i>Sammanfattning av den kvantitativa studien</i>	43
5 Närstudie av fyra testtagare	44
5.1 <i>Urval av testtagare för närstudien</i>	45
5.2 <i>Testtagarnas pratighet och delaktighet</i>	49
5.3 <i>Komplexitet och korrekthet hos fyra testtagare</i>	50
5.3.1 PT-nivå	50
5.3.2 Nominalfraser	51
5.3.3 Underordning	52
5.3.4 Infinitivfraser	55
5.3.5 Verbfraser	56
5.3.6 Finithet – obligatoriska verb och subjekt	58
5.4 <i>Sammanfattning av närstudien</i>	59
6 Diskussion och slutsatser	60
6.1 <i>Språktestning – bedömningar som beroende variabler</i>	61
6.1.1 Diffusa gränser mellan nivåer	62
6.1.2 Diffusa gränser mellan kriterier	63
6.2 <i>Språkfärdighet – mätvärden som beroende variabler</i>	64
6.3 <i>Framåtblick</i>	65
7 Avslutning	66
Referenser	69
Bilaga 1: Informerat samtycke	73
Bilaga 2: Transkriptionsnyckel	75
Bilaga 3: Tabeller	76
Bilaga 4: Diagram	79
Bilaga 4.1 <i>Korrekthetsmått</i>	79
Bilaga 4.2 <i>Komplexitetsmått</i>	79
Bilaga 4.3: <i>Korrekthetsmått och komplexitetsmått</i>	81
Bilaga 4.4: <i>Komplexitets-/korrekthetsmått och PT-nivå</i>	82
Bilaga 4.5: <i>Komplexitets-/ korrekthetsmått och kriterium 3</i>	83
Bilaga 4.6: <i>Komplexitets-/korrekthetsmått och slutbetyg</i>	84

Figurförteckning

Figur 1.	Samband mellan validitet och reliabilitet	4
Figur 2.	Kommunikativ språkförmåga efter Bachman och Palmer	7
Figur 3.	Tre aspekter av språklig performans efter Skehan	13
Figur 4.	Olika typer av komplexitet efter Bulté och Housen	14
Figur 5.	Uppsatsens fokus: Swedex B2-test, den muntliga delen, kriterium 3	24
Figur 6.	Mätning av komplexitet – kodschema	30
Figur 7.	Mätning av korrekthet – kodschema	34
Figur 8.	Fastställande av PT-nivå – kodschema	36

Tabellförteckning

Tabell 1.	CEFR:s färdighetsnivåer	8
Tabell 2.	CEFR:s skala Allmän muntlig produktion, nivå B1–C1	9
Tabell 3.	CEFR:s skala Allmän muntlig interaktion, nivå B1–C1	9
Tabell 4.	CEFR:s skala Allmän språklig repertoar, nivå B1–C1	10
Tabell 5.	CEFR:s skala Grammatisk korrekthet, nivå B1–C1	10
Tabell 6.	PT-nivåer	19
Tabell 7.	Deskriptorer för kriterium 3	24
Tabell 8.	Bedömningar av testtagarna muntliga performans samt slutbetyg	25
Tabell 9.	Samtliga mått med beskrivningar	29
Tabell 10.	Korrelationer mellan mätresultat	39
Tabell 11.	Korrelationer mellan bedömningar	41
Tabell 12.	Korrelationer mellan bedömningar och mätresultat	42
Tabell 13.	Bisatsinledare som används av NAG, FAG, MCU och JCU	53
Tabell 14.	Hjälperb som används av NAG, FAG, MCU och JCU	57
Tabell 15.	Alla fall av (pluskvam)perfekt hos NAG, FAG, MCU och JCU	57
Tabell 16.	Partikelverb och reflexiva verb producerade av NAG, FAG, MCU och JCU	57
Tabell 17.	Passiviserade verb hos NAG, FAG, MCU och JCU	58
Tabell 18.	Kvantitativ sammanfattning av den kvalitativa närstudien	60
Tabell 19.	Absoluta tal som kodningen av materialet resulterat i	76
Tabell 20.	Mätvärden för alla testtagare	77
Tabell 21.	Bedömningar och PT-nivå	78

Diagramförteckning

Diagram 1.	Sammanfattning av den kvantitativa undersökningen	44
Diagram 2.	Samband mellan satskorrekthet och AS-längd ($r=,51^*$)	46
Diagram 3.	Samband mellan satskorrekthet och andelen relativsatser ($r=,32$)	46
Diagram 4.	Samband mellan INV-korrekthet och satslängd ($r=,37$)	47
Diagram 5.	Samband mellan INV-korrekthet och andelen bisatser ($r=,69^{**}$)	47
Diagram 6.	Samband mellan INV-korrekthet och andelen INV-kontexter ($r=,30$)	48
Diagram 7.	Token	49
Diagram 8.	Type	49
Diagram 9.	Type/token	49
Diagram 10.	INV-korrekthet~Satskorrekthet ($r=,69^{**}$)	79
Diagram 11.	AS-längd~Satslängd ($r=,52^{**}$)	79
Diagram 12.	AS-längd~BS-andel ($r=,90^{***}$)	79
Diagram 13.	AS-längd~RS-andel ($r=,75^{***}$)	80
Diagram 14.	RS-andel~Satslängd: ($r=,54^*$)	80
Diagram 15.	RS-andel~BS-andel ($r=,64^{**}$)	80
Diagram 16.	BS-andel~INV-kontextandel ($r=,64^{**}$)	80
Diagram 17.	RS-andel~INV-kontextandel ($r=,42.$)	80
Diagram 18.	Satskorr.~AS-längd ($r=,51^*$)	81
Diagram 19.	INV-korr.~AS-längd ($r=,72^{***}$)	81
Diagram 20.	Satskorr.~BS-andel ($r=,60^{**}$)	81
Diagram 21.	INV-korr.~BS-andel ($r=,69^{**}$)	81
Diagram 22.	INV-korr.~RS-andel ($r=,54^*$)	81
Diagram 23.	AS-längd~PT-niåv ($r=,53^*$)	82
Diagram 24.	BS-andel~PT-niåv ($r=,58^{**}$)	82
Diagram 25.	RS-längd~PT-nivå ($r=,42.$)	82
Diagram 26.	INV-kontextandel~PT-nivå ($r=,49^*$)	82
Diagram 27.	Satskorr.~PT-nivå ($r=,54^*$)	82
Diagram 28.	INV-korr.~PT-nivå ($r=,61^{**}$)	82
Diagram 29.	AS-längd~Kriterium 3 ($p=.49^*$)	83
Diagram 30.	BS-andel~Kriterium 3 ($p=.43.$)	83
Diagram 31.	Satskorrekthet~Kriterium 3 ($p=.59^{**}$)	83
Diagram 32.	INV-korrekthet~Kriterium 3 ($p=.49^*$)	83
Diagram 33.	AS-längd~Slutbetyg ($r=,46^*$)	84
Diagram 34.	Satslängd~Slutbetyg ($r=,40.$)	84
Diagram 35.	Satskorr.~Slutbetyg ($r=,48^*$)	84
Diagram 36.	INV-korr.~Slutbetyg ($r=,49^*$)	84

1 Inledning

En fotbollsspelare befinner sig offside när hon eller han är närmare motståndarlagets mål än både bollen och motståndarlagets näst bakersta spelare (Fédération Internationale de Football Association, FIFA, 2014, s. 35). I en analys av alla 337 offside-vinkningar under slutspelen i fotbolls-VM 2002 visar Helsen, Gilis and Weston (2006) att mer än en fjärdedel av dessa domslut var felaktiga. Inom FIFA (2014) är man medvetna om att offside-beslut är subjektiva: i *The Law of the Game* står uttryckligen att dessa beslut beror på "the opinion of the referee" (s. 35).

En annan subjektiv bedömningssituation är det muntliga språktestet. Precis som linjedomaren i fotboll ska bedömare av muntliga språktest, under tidspress och utifrån ett antal samverkande faktorer, fatta mycket viktiga beslut (Sandlund & Sundqvist 2014, s. 125; Eklund Heinonen 2009, s. 11; Europarådet 2001, s. 19–20). I dagens kommunikativa språktest ingår nästan alltid en muntlig del, och i föreliggande uppsats undersöks den muntliga delen av det svenska certifikatstestet Swedex, nivå B2. Två av de faktorer som bedömare av Swedex B2-test har att vara uppmärksamma på är grammatisk korrekthet och komplexitet; dessa kvaliteter ska bedömas under kriterium 3. Detta kriterium står i fokus i uppsatsen, dels därför att det finns ett behov av att utreda relationen mellan rent lingvistisk kompetens, som grammatisk behärskning, och en mer allmän muntlig språkfärdighet (se t.ex. Hulstijn 2007), dels därför att relationen mellan just korrekthet och komplexitet också är problematisk (se t.ex. Spoelman & Verspoor 2010; Baten & Håkansson 2014). Med denna uppsats vill jag bidra till forskningen om vad språkfärdighet egentligen är och hur denna färdighet kan testas och bedömas.

1.1 Syfte och frågeställning

Uppsatsen har två syften. För det första utgör den en kritisk granskning av fenomenet muntliga språktest: är bedömare förmögna att under ett tjugo minuter långt samtal mellan flera testtagare, på ett rättssäkert sätt, särskilja och bedöma en rad olika komponenter som exempelvis grammatisk korrekthet och komplexitet? För det andra undersöks relationen mellan grammatisk kompetens och en mer allmän muntlig språkfärdighet. Genom att på olika sätt mäta komplexitet och korrekthet i testtagares produktion och sedan testa hur dessa mått korrelerar med bedömningar undersöker jag alltså både vad som faktiskt bedöms – korrekthet eller komplexitet? – och vilken roll olika typer av korrekthet och komplexitet spelar för den allmänna språkfärdigheten. För att uppfylla det första syftet använder jag mina mätningar av testtagarnas produktion som oberoende variabler; med dem försöker jag "avslöja" vad som faktiskt bedöms i ett muntligt språktest.

För att uppfylla det andra syftet använder jag tvärtom bedömningarna som oberoende variabler, det vill säga jag antar att bedömningarna i testet faktiskt säger vilken färdighetsnivå testtagarna är på. I det senare fallet utgör alltså mina mätningar själva forskningsobjektet.

Detta dubbla syfte återfinns hos bland andra Iwashita, Brown, McNamara och O'Hagen (2008, s. 25) som uttrycker det så här: "In general, then, the relationship of different aspects of performance to overall judgments of proficiency is an issue both for theories of the nature of language proficiency and for the interpretability of test scores." Att kombinera det tillämpade forskningsfältet språktestning och det mer teoretiska fältet andraspråksinlärning kan sägas utgöra ett syfte i sig (jmf. Eklund Heinonen 2009, s. 61). Bachman och Cohen (1998, s. 3–4) påpekar att dessa forskningsfält, även om de har olika fokus, delar flera intressen. Forskning om såväl språktestning som andraspråksinlärning handlar till stor del om individuell variation, både sådan variation som beror på språkanvändarens språkliga kompetens och sådan som beror på andra förmågor, exempelvis strategisk kompetens.

Uppsatsens två syften uppfylls med hjälp två forskningsfrågor. Den första frågan besvaras i uppsatsens huvudsakliga del, som utgörs av en kvantitativ undersökning av tjugo testtagares performans under Swedex B2-test. Den andra frågan besvaras i en kvalitativ närstudie av fyra av dessa testtagare.

1. (Hur) korrelerar komplexitet, korrekthet och bedömning i ett muntligt språktest?
2. (Hur) kan till synes avvikande bedömningar förklaras?

1.2 Uppsatsens upplägg

Uppsatsen är disponerad enligt följande: I avsnitt 2 ger jag en bakgrund till ämnet språktestning och språkfärdighet; här diskuteras både vad språkfärdighet är och hur den kan mätas. Fokus ligger på relationen mellan komplexitet, korrekthet och allmän muntlig språkfärdighet. I avsnitt 3 presenteras den kvantitativa studiens material och metod, och i avsnitt 4 presenteras resultaten av denna studie, det vill säga svaret på den första forskningsfrågan. Avsnitt 5 utgörs av en kvalitativ närstudie av fyra testtagare som utifrån de kvantitativa resultaten kan betraktas som typiska eller otypiska; i denna del besvaras den andra forskningsfrågan. Därefter diskuteras, i avsnitt 6, såväl kvantitativa som kvalitativa resultat utifrån uppsatsen två syften; här drar jag också några slutsatser samt ger också förslag på hur språktestning kan utvecklas och utforskas vidare. Avsnitt 7 utgörs av en epilög.

2 Bakgrund

I avsnitt 2.1 presenteras några grundläggande begrepp inom testforskningen, framför allt de två centrala testkvaliteterna reliabilitet och validitet. Den historiska utvecklingen av språktest har beskrivits i termer av en spänning mellan dessa kvaliteter; med dagens kommunikativa språktest ska vi ha lämnat reliabiliteten till förmån för validiteten. I avsnitt 2.2 diskuteras denna historieskrivning med hjälp av Fulcher (2000, 2003). Utvecklingen av kommunikativa språktest har bland annat resulterat i det ambitiösa standardiseringsprojektet *Gemensam europeisk referensram för språk* (CEFR) (Europarådet 2001). CEFR och dess nivåskalor, som Swedextestet utgår från, beskrivs i avsnitt 2.3. CEFR-nivåerna har dessvärre kritiserats för att inte svara mot inlärningsgångar som andraspråksforskningen kunnat belägga (Hulstijn 2007) och för att vara vaga och svåransända (Kuiken & Vedder 2014, s. 283). De senaste åren har forskare därför, bland annat inom SLATE-nätverket (Second Language Acquisition and Testing in Europe, 2014), studerat relationer mellan CEFR-nivåerna och mätbara lingvistiska kompetenser (Bartning, Martin & Vedder 2010). Mycket av denna forskning har bedrivits inom CAF-traditionen, där CAF står för *Complexity–Accuracy–Fluency* (t.ex. Gyllstad, Granfeldt, Bernardini & Källkvist 2014; Iwashita m.fl. 2008; jmf. även Martin, Mustonen, Reiman & Seilonen 2010); denna forskning refereras i avsnitt 2.4. Metoderna som används inom CAF-forskningen har dock kritiserats för att inte heller de svara mot de inlärningsgångar som andraspråksforskningen kunna belägga. Baten och Håkansson (2014) menar att Processbarhetsteorin (PT) är ett bättre verktyg för att fastställa andraspråksinlärares språkbehärskning. Några forskare har också undersökt relationen mellan PT och CEFR (eller allmän kommunikativ språkförmåga) (se Granfeldt & Ågren 2013; Eklund Heinonen 2009; se även Ellis 2008); denna forskning refereras i avsnitt 2.5.

2.1 Att testa (muntlig) språkfärdighet

Inom testteori skiljer man på bakåtblickande, summerande, så kallad summativ bedömning som syftar till att fastställa testtagares kunskaper eller färdigheter, och framåtblickande formativ bedömning som syftar till att exempelvis bestämma vilka vidare pedagogiska insatser som behövs – ”en bedömning inte av, utan för lärande” (Tornberg 2009, s. 205–206; se även Olofsson & Sjöqvist 2013, s. 686; Europarådet 2001, s. 181). Vidare skiljer man på *achievement test* som testar ett avgränsat material, till exempel ett kapitel i en kursbok eller en bestämd uppsättning glosor, och *proficiency test*, som syftar till att fastställa generella färdigheter (Tornberg 2009, s. 207; Bachman 2004, s. 9; Europarådet 2007, s. 181; Luoma 2004, s. 3). Swedextestet är ett summativt färdighetstest, ett så kallat

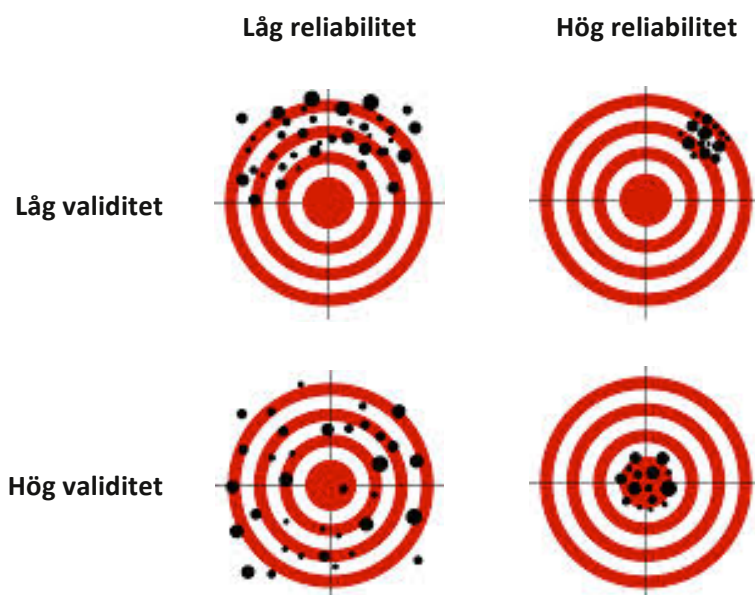
certifikatstest som testtagare betalar för att ta för att få papper på sin kompetens (Folkuniversitetet 2014). Testet beskrivs i avsnitt 3.1.

I sin klassiska *Fundamental Considerations in Language Testing* skiljer Bachman (1990, s. 23–24) på begreppen testning (*testing*), mätning (*measurement*) och bedömning (*assessment*). Testning innebär någon form av systematisk observation; mätning innebär att observationer kvantifieras, det vill säga tillskrivs ett visst värde; och bedömning innebär någon form av beslut, till exempel om vilket betyg en student ska få eller vilka pedagogiska åtgärder som ska vidtas. I denna uppsats refererar jag med *test* till den konkreta test-situationen, med *bedömning* till de beslut bedömarna fattar under testet, och med *mätning* och *mått* till de kvantitativa analyser jag gör av det inspelade materialet.

Det är möjligt att bedöma något utan att testa det eller mäta det – det gör vi hela tiden. Det är också i princip möjligt att tillskriva något ett visst värde utan att ha utfört ett test, men testet är avgörande för att det tillskrivna värdet ska äga någon vetenskaplig eller rättslig giltighet. Testets kvalitet avgörs i sin tur av dess reliabilitet och validitet (Bachman 1990, s. 24–25; se dock Fulcher 2000 som menar att validitet inte är en egenskap hos själva testet utan ligger i tolkningen av resultatet).

2.1.1 Reliabilitet och validitet

I testsammanhang innebär hög reliabilitet att testresultat inte beror på irrelevanta omständigheter utan blir desamma oavsett av exempelvis vem som bedömer. Hög validitet innebär att testresultatet verkligen säger något om det som ska testas, i detta fall muntlig språkfärdighet. Relationen mellan dessa två kvaliteter illustreras i Figur 1.



Figur 1. Samband mellan validitet och reliabilitet

Låg validitet och låg reliabilitet innebär att testresultaten varken är förutsägbara eller säger något om det som ska testas. Ett sådant test har givetvis inget värde. Hög validitet i kombination med låg reliabilitet innebär att resultaten är oförutsägbara – det finns alltså en rättsosäkerhet – men som framgår av bilden är ”träffarna” ändå centrerade runt ”målet”, det vill säga testet tenderar att testa det som ska testas. Hög reliabilitet i kombination med låg validitet innebär tvärtom att resultaten är förutsägbara – gör du samma test i dag som i går kan du räkna med samma resultat – men inte testas det som ska testas. Vilket scenario av de två senare som är minst dåligt är inte självklart; som vi ska se i nästa avsnitt kan validitet och reliabilitet betraktas som två konkurrerande kvaliteter.

Önskvärt är givetvis ett test med hög reliabilitet och validitet, men Bachman (1990, s. 30–32) konstaterar att språktest aldrig kan vara riktigt valida och reliabla. För det första vet vi inte exakt vad det är vi vill testa – en exakt definition av språkförmåga saknas – och för det andra vet vi inte hur det vi ändå tror oss vilja testa kan observeras och kvantifieras. Detta, skriver Bachman (1990, s. 32–39), beror bland annat på att språkförmågan inte är ett fysiskt fenomen, utan ett mentalt, som alltså måste mätas indirekt; språktestaren får anta att testtagares performans i någon mån återspeglar en underliggande kompetens (se även Europarådet 2001, s. 185). Testkonstruktörens uppgift består följaktligen i att både definiera och operationalisera konstrukten, alltså det som ska testas (Bachman 1990, s. 40–45; se även Luoma 2004, kapitel 6, och Fulcher 2003, kapitel 2). En konstrukt är, skriver Fulcher (2003, med referens till Kerlinger & Lee 2000) ”a concept that is deliberately defined for a ‘special scientific purpose’ [...] in a way that can be observed and measured” (s. 18).

2.1.2 Den historiska utvecklingen av (muntliga) språktest

I Fulcher (2003) berättas det muntliga språktestets historia, en historia som enligt författaren måste förstås utifrån sin politiska kontext; det är samhälls- och militära behov som har styrt utvecklingen av nya språktest (s. 1). Spolsky (refererad hos Fulcher 2000, s. 438) delar upp utvecklingen av språktest i tre perioder (se även Eklund Heinonen 2009, s. 44). Under den för-vetenskapliga perioden, före andra världskriget, litade man på bedömares och lärares erfarenhet och expertis och ställde inga frågor om reliabilitet och validitet. Under andra världskriget växte, framför allt i USA, det militära behovet av språktestning och med det också ett vetenskapligt intresse av språktestning; denna psyko-metriska period (1930–1960) förknippas med behaviorismen i allmänhet och med Robert Lado i synnerhet. Under den tredje perioden, den psykolingvistiska–sociolingvistiska (1960–1970), menade man dock att Lado och hans samtida, i sin vetenskapliga strävan

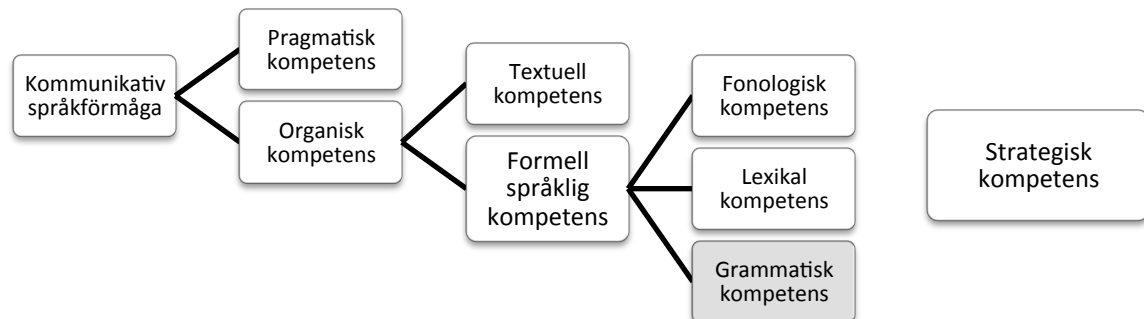
efter objektivitet, lade allt för stort fokus på reliabiliteten och glömde validiteten, vilket bland annat ska ha tagit sig uttryck i det föraktade *multiple choice*-testet. Enligt företrädarna för *The Communicative Movement*, som Fulcher (2000) kallar den rörelse som lett fram till dagens kommunikativa språkundervisning och -testning, finns en inneboende motsättning mellan reliabilitet och validitet, men Fulcher själv hävdar att denna motsättning är falsk, liksom hela historieskrivningen. Alla de frågor som diskuteras inom språktestningen i dag, skriver han, diskuterades redan under 1940-talet: för- och nackdelar med holistisk och analytisk bedömning, beskrivningar av färdighetsnivåer, partiskhet, relationen mellan lingvistiska och kommunikativa kompetenser med mera (Fulcher 2003, s. 10). Inte desto mindre har *The Communicative Movement*, med sin antipati mot statistik och förkärlek för "common sense" (Fulcher 2000, s. 485), vunnit stora framgångar: "It would be difficult to market a new large-scale test that did not claim to be 'communicative' – whatever the term may mean to different users", skriver Fulcher (2000, s. 493). Vad kommunikativ språkförmåga är och hur den ska testas vet vi fortfarande inte, enligt honom.

2.1.3 Kommunikativ språkförmåga

Det var som en reaktion mot den tidens fokus på grammatisk kompetens i såväl forskning som undervisning och testning som Hymes i början av 1970-talet introducerade begreppet *kommunikativ kompetens* (Eklund Heinonen 2009, s. 45–48; Green 2014, s. 202–203). Begreppet utvecklades under 1980-talet av Canale och Swain, som fokuserade på andraspråkstalare. Enligt dem bestod kommunikativ kompetens av tre huvudsakliga komponenter: *grammatisk*, *sociolingvistisk* och *strategisk kompetens*. Bachman (1990) och Bachman och Palmer (1996, refererad i Eklund Heinonen 2009 och i Green 2014) utgår från dessa kompetenser i sin modell *Communicative Language Ability*, som är ett försök att definiera den konstrukt som ett kommunikativt språktest ska testa. Enligt modellen består kommunikativ språkförmåga av organisk och pragmatisk kompetens. Till den organiska kompetensen hör det vi kan kalla lingvistisk kompetens, med ordförråd, uttal och grammatik. Utanför modellen men ändå central är också den strategiska kompetensen. Distinktionen mellan språklig och strategisk kompetens är grundläggande för Bachman och Palmer (2010). De skriver: "It is the combination of language knowledge and strategic competence that provides language users with the ability, or capacity, to create and interpret discourse [...]" (s. 44).

Bachman och Palmer (2010) skriver inte vad ett språktest ska testa, utan betonar att varje testkonstruktör måste vara medveten om vad ett visst test ska testa. Oavsett vad som ska testas måste testaren vara medveten om alla de andra faktorerna som alltid på-

verkar mänsklig kommunikation; därför finns den strategiska kompetensen med som en del av modellen över språklig kommunikativ kompetens. I Figur 2 återges delar av modellen.



Figur 2. Kommunikativ språkförmåga efter Bachman och Palmer (2010, s. 45)

I dag är modellen dominerande inom språktestning (Eklund Heinonen 2009, s. 46). I många språktest ingår också strategisk kompetens i den konstrukt som bedöms; det gäller exempelvis de svenska SFI-testen, där stort fokus läggs på strategisk kompetens (Skolverket 2012). Modellen ligger också till grund för *Gemensam europeisk referensram för språk* (CEFR) (Europarådet 2001, s. 106), som Swedextestet utgår från (Folkuniversitetet 2014), och även här utgör den strategiska kompetensen en hörnsten.

2.2 Gemensam europeisk referensram för språk (CEFR)

Sedan 1970-talet har ett europeiskt standardiseringsarbete pågått för att skapa ett gemensamt synsätt på språkundervisning och språktestning i Europa. År 2001 publicerade Europarådet resultatet av detta projekt: *Gemensam europeisk referensram för språk* (*Common European Frame of Reference for Language*, fortsättningsvis CEFR). CEFR syftar till att ge språkutbildare och -testare, samt deras kunder, ett gemensamt metaspråk. Ett mer övergripande syfte är också att stärka den europeiska demokratin genom att göra det lättare för européer att kommunicera med varandra (Europarådet 2001, s. xi-xii). CEFR har fått stor spridning i Europa och måste betraktas som ett mycket framgångsrikt projekt (Hulstijn 2007, s. 663).

Språksynen i CEFR är handlingsorienterad; språkanvändaren ses som en "social aktör" som måste klara av vissa uppgifter (Europarådet 2001, s. 9). Vilka uppgifter språkbrukare på olika färdighetsnivåer klarar av beskrivs i CEFR:s fjärde kapitel. För att klara av dessa uppgifter behöver språkanvändare olika kompetenser, inte bara rent lingvistiska utan även andra, exempelvis sociala och kognitiva, kompetenser (jmf. avsnitt 2.1.3). Vilka kompetenser som krävs för att klara av uppgifter på olika färdighetsnivåer beskrivs i

CEFR:s femte kapitel. En grundläggande fråga för forskarna inom SLATE-nätverket och andra är huruvida den kommunikativa kompetensen – den som styr VAD inläraren kan göra – utvecklas parallellt med den språkliga kompetensen – den som styr HUR inläraren kan göra det (SLATE 2014; Hulstijn 2007).

Vad språkanvändaren kan göra och hur hon eller han gör det beskrivs med så kallade deskriptorer, korta beskrivningar, i de nivåskalor som CEFR är mest känt för, med sex färdighetsnivåer under tre övergripande band, se tabell 1. En nivåskala måste ha en tillräcklig mängd nivåer för att vara användbar men inte så många att det blir svårt att bestämma nivå objektivt. Ungefär sex nivåer brukar anses lagom (Europarådet 2001, s. 20–23; Luoma 2004, s. 80).

Tabell 1. CEFR:s färdighetsnivåer

A Nybörjare		B Självständig användare		C Avancerad användare	
A1 <i>Breakthrough</i>	A2 <i>Waystage</i>	B1 <i>Threshold</i>	B2 <i>Vantage</i>	C1 <i>Effective operational proficiency</i>	C2 <i>Mastery</i>

Viktigt att veta när skalorna läses är att varje nivå inbegriper nivåerna under (Europarådet 2001, s. 37). Vidare är skalorna inte ekvidistanta och ska alltså inte användas som linjaler; exempelvis brukar det ta dubbelt så lång tid att nå B1 från A2 som det tar att nå A2 från A1 (Europarådet 2001, s. 18). B2-nivån ska dock ligga lika långt över B1 som B1 ligger över A2 (Europarådet 2001, s. 33). Att nå B2-nivån, som står i fokus i denna uppsats, innebär, enligt CEFR, "att inläraren, efter att sakta men säkert ha tagit sig fram över 'mellanplattan', märker att han/hon har kommit till ett stadium där saker och ting ser annorlunda ut och nya perspektiv öppnas där inläraren kan betrakta sin omgivning på ett nytt sätt" (s. 35). Konkreta saker som händer på B2-nivån är, skriver man, att inläraren blir bättre på att bland annat ta och behålla ordet i den sociala diskursen, samt att inläraren, som en följd av en större språklig medvetenhet, rättar sig själv mer.

2.2.1 Språkliga aktiviteter – VAD talaren kan göra

Exempel på språkliga aktiviteter som beskrivs i CEFR:s fjärde kapitel är *skriva, läsa, lyssna* och *tala*. Muntlig produktion och muntlig interaktion ses som olika aktiviteter. Under muntlig produktion finns fem skalor, bland annat skalorna "Längre monolog" och "Offentliga meddelanden". Under muntlig interaktion finns nio skalor, bland

annat "Samtal" och "Informell diskussion (med vänner)". För att visa hur B2-nivån, som står i fokus i denna uppsats, skiljer sig från nivåerna över och under återges i Tabell 2 och 3 deskriptorerna för nivå B1–C1 ur skalorna "Allmän muntlig produktion" och "Allmän muntlig interaktion". Viktigt att notera är att det i dessa skalor framgår inte bara vad talaren kan göra utan också hur hon eller han kan göra det: "flytande, korrekt", "relativt flytande", "med god grammatisk behärskning". Hur talaren löser sina uppgifter – den fråga som också står i fokus i denna uppsats – ska ju egentligen presenteras i CEFR:s femte kapitel. Uppenbarligen är det svårt att hålla VAD och HUR isär.

Tabell 2. CEFR:s skala Allmän muntlig produktion, nivå B1–C1 (Europarådet 2001, s. 60)

Nivå	Allmän muntlig produktion
C1	Kan ge tydliga, detaljerade beskrivningar och presentationer av komplicerade ämnen, där han/hon integrerar underteman, utvecklar särskilda punkter och avrundar med en lämplig slutsats.
B2	Kan ge tydliga, systematiskt utformade beskrivningar och presentationer, där viktiga punkter lyfts fram på lämpligt sätt, och ge relevanta understödjande fakta. Kan ge tydliga, detaljerade beskrivningar och presentationer av en rad olika ämnen med anknytning till sitt intresseområde och utveckla och ge stöd för tankar med hjälp av understödjande detaljer och belysande exempel.
B1	Kan relativt flytande ge en enkel beskrivning av ett av flera ämnen inom sitt intresseområde och presentera den i logisk följd.

Tabell 3. CEFR:s skala Allmän muntlig interaktion, nivå B1–C1 (Europarådet 2001, s. 74)

Nivå	Allmän muntlig interaktion
C1	Kan uttrycka sig ledigt och spontant, praktiskt taget utan ansträngning. Har ett brett ordförråd och kan då han/hon saknar ord kringgå detta genom att uttrycka sig på ett annat sätt. Söker inte alltför tydligt efter ord och uttryck eller använder sig av undvikandestrategier; endast begreppsmässigt komplicerade ämnesområden kan förhindra ett naturligt och flytande språk.
B2	Kan använda språket flytande, korrekt och effektivt inom en mängd olika allmänna, studie-, arbets- och fritidsrelaterade ämnen och förmedlar effektivt tankegången. Kan kommunicera spontant med god grammatisk behärskning utan att visa särskilt många tecken på att han/hon behöver begränsa vad han/hon vill säga, och väljer en formalitetsnivå som lämpar sig för situationen.
B1	Kan med viss tillförsikt uttala sig om fakta som rör kända rutinmässiga och icke-rutinmässiga ämnen inom sitt intresse- och yrkesområde. Kan utbyta, kontrollera och bekräfta information, hantera mindre rutinmässiga situationer och förklara varför någonting utgör ett problem. Kan uttrycka tankar om mer abstrakta kulturella ämnen som filmer, böcker, musik etc.

2.2.2 Språkliga kompetenser – HUR talaren kan göra det

Språkanvändarens kompetenser, som CEFR:s femte kapitel handlar om, är uppdelade i två huvudgrupper: generella kompetenser som används vid alla handlingar, även icke-språkliga, och specifika språkliga kompetenser (jmf. avsnitt 2.1.3). Till de språkliga

kompetenserna räknas fonologisk, semantisk, ortografisk, sociolingvistisk, pragmatisk, funktionell samt grammatisk kompetens (Europarådet 2001, s. 109–125). I denna uppsats fokuseras den grammatiska kompetensen. I Tabell 4 återges deskriptorerna från skalan "Allmän språklig repertoar" och i Tabell 5 deskriptorerna från skalan "Grammatisk korrekthet". På B2-nivån i den förra av dessa skalor står att talaren kan använda "vissa komplexa satsstrukturer". I övrigt står här egentligen bara att talaren har det språk som krävs för att göra det som talaren, enligt tabellerna i kapitel 4, kan göra. Skalan "Grammatisk korrekthet" handlar om i vilken utsträckning talaren begår fel, om felen märks, om talaren rättar sina fel, om felen är systematiska och om de leder till missförstånd.

Tabell 4. CEFR:s skala Allmän språklig repertoar, nivå B1–C1 (Europarådet 2001, s. 107)

Nivå	Allmän språklig repertoar
C1	Kan välja en lämplig formulering från en bred språklig repertoar för att uttrycka sig tydligt utan att behöva begränsa det han/ hon vill säga.
B2	Kan uttrycka sig tydligt och utan att ge nämnvärt intryck av att behöva begränsa det han/hon vill säga. Har tillräckligt bred språklig repertoar för att kunna ge tydliga beskrivningar, uttrycka åsikter och utveckla argument utan att alltför uppenbart behöva leta efter ord. Kan använda vissa komplexa satskonstruktioner för att göra detta.
B1	Har tillräckligt stor språklig repertoar för att kunna beskriva oväntade situationer, förklara huvudpunkterna i en idé eller ett problem förhållandevis exakt och uttrycka tankar om abstrakta eller kulturella ämnen som musik och filmer.

Tabell 5. CEFR:s skala Grammatisk korrekthet, nivå B1–C1 (Europarådet 2001, s. 111)

Nivå	Grammatisk korrekthet
C1	Visar genomgående en hög grad av grammatisk korrekthet; gör sällan fel och eventuella fel är svåra att upptäcka.
B2	Har god grammatisk behärskning, även om det fortfarande förekommer enstaka felsägningar och icke-systematiska fel och små misstag i meningsstrukturen, vilka dock är sällsynta och ofta kan korrigeras i efterhand. Visar en relativt hög grad av grammatisk behärskning. Gör inga fel som leder till missförstånd.
B1	Kommunicerar någorlunda korrekt i välbekanta sammanhang och har i regel god kontroll, även om modersmålet har märkbar inverkan. Fel förekommer, men det framgår tydligt vad han/hon försöker säga.

Som en del av sin studie om relationen mellan CEFR-nivåerna och komplexitet i inlärares skriftliga produktion undersökte Gyllstad m. fl. (2014) användningen av termerna *complex* och *simple/basic* i CEFR:s skalor för skriftlig produktion. Resultatet visade att språket beskrivs som enkelt på nivå A1 till A2 och som mer eller mindre komplext från nivå B2 och uppåt. I deskriptorerna för nivå B1 används begreppen växlande. Författarna påpekar att det ingenstans i CEFR finns en definition av begreppet komplexitet (se också Ellis 2008).

2.2.3 CEFR om bedömning

Kapitel 8 i CEFR handlar om bedömning. Här diskuteras reliabilitet och validitet, men också genomförbarhet (Europarådet 2001, s. 185). I viss mån, skriver man, konkurrerar dessa tre krav med varandra. CEFR har försökt åstadkomma reliabilitet genom att testa vilka deskriptorer som fungerar; erfarenheten visar "att lärare och studerande kan tolka deskriptorer på ett mer enhetligt sätt om deskriptorerna inte bara beskriver VAD studerande kan göra utan också HUR BRA han/hon gör det" (Europarådet 2001, s. 178). I CEFR påpekas dock att betyg på såväl muntlig som skriftlig performans till viss del alltid är subjektiva (Europarådet 2001, s. 40).

Nivåskalorna med deskriptorer kan användas för bedömning på tre olika sätt: som checklista, skala eller tablå (Europarådet 2001, s. 179). Checklista innebär att några aspekter för en viss nivå fokuseras och bockas av under testet. Skala används vid holistisk nivåbestämning. Båda dessa varianter är endimensionella. Bedömning utifrån tablå är däremot tvådimensionell och innebär att bedömaren både ska hålla koll på olika aspekter av performansen och för varje aspekt bestämma nivå; man talar om analytisk bedömning, men också om diagnostisk eftersom bedömningen kan användas för att ställa "diagnos" och svara på frågan vad inläraren behöver utveckla mer. Denna variant används av Swedex.

Fördelen med analytisk bedömning är att bedömaren tvingas göra noggranna iakttagelser och inte lika lätt kan låta en enskildhet bli utslagsgivande, vilket är risken vid holistisk bedömning. Nackdelen är att det är mycket svårt att särskilja olika komponenter i performansen; risken för så kallade halo-effekter – *cross-contamination* eller "the tendency to give the same grade across categories" (Fulcher 2003, s. 90; se även Eckes 2009, s. 5; Tonkyn 2012) – är stor. Mer än fyra–fem, maximalt sju, kategorier kan därför inte bedömas samtidigt utan att bedömaren blir kognitivt överbelastad (Europarådet 2001, s. 188; se även Suoma 2003, s. 80).

2.3 Komplexitet, korrekthet och språkfärdighet

Sedan 1980-talet har forskare studerat muntliga språktest ur en rad olika aspekter. Man har undersökt hur olika egenskaper hos såväl testtagare som testledare, samt relationen och interaktionen mellan dessa båda, påverkar testresultatet (se Eklund Heinonen 2009, s. 51–52). Till exempel använder sig Sandlund och Sundqvist (2014) av CA-verktyg för att undersöka ett muntligt test i engelska i en svensk skola; enligt dem tycks elevens villighet "att spela med i spelet" (s. 135) avgörande för vilket betyg som delas ut.

Andra studier använder muntliga språktest för att kartlägga den muntliga språkfärdighetens multikompositionella struktur, det vill säga: ta reda på vari själva

språkfärdigheten egentligen består. De Jong, Steinel, Florijn, Schoonen och Hulstijn (2012) delar upp denna typ av studier i två grupper. Den första typen utgörs av jämförelser mellan holistiska subjektiva bedömningar och analytiska subjektiva bedömningar. I en av de första studierna av denna typ jämförde Adams (1980, refererad i Iwashita m. fl. 2008 och i De Jong m. fl. 2012) holistiska och analytiska betyg utdelade till 834 testtagare i ett amerikanskt språktest. Adams visade att komponenterna grammatik och ordförråd hade störst betydelse för det holistiska betyget medan flyt och accent bidrog minst. I en senare studie av samma typ visade Higgs och Clifford (1982, refererade i Iwashita m. fl. 2008 och i De Jong m. fl. 2012) att olika komponenter har olika stor betydelse beroende på vilken färdighetsnivå talaren befinner sig på: uttal och ordförråd är, enligt dem, de viktigaste komponenterna på låga färdighetsnivåer medan betydelsen av grammatik och flyt ökar på högre färdighetsnivåer.

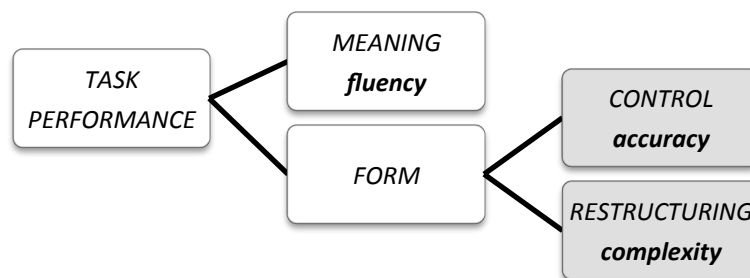
I den andra typen av studier testas hur subjektiva bedömningar av språkfärdighet korrelerar med objektiva mätningar av testtagares performans (De Jong m. fl. 2012). Det är framför allt denna typ av forskning som bedrivs inom SLATE-nätverket (Second Language Acquisition and Testing in Europe 2014), där man under de senaste tio åren undersökt relationer mellan kommunikativ språkförmåga, så som den beskrivs i CEFR, och specifik språklig kompetens, så som den mäts av andraspråksforskare (Bartning, Martin & Vedder 2010, s. 5–6). Dessa undersökningar är givetvis avhängiga de så kallade objektiva mätmetoder som forskarna har att tillgå. Inom CAF-forskningen, där CAF står för *complexity*, *accuracy* och *fluency*, har en rad mätverktyg utvecklats, och forskarna i SLATE-nätverket använder ofta dessa metoder. Andra, till exempel Martin m. fl. (2010), försöker mäta lingvistisk kompetens också på andra sätt. Eftersom korrekthet och komplexitet står i fokus i denna uppsats har jag valt att hämta mycket av min metod från CAF-forskningen, vars grundtanke beskrivs i avsnitt 2.3.1. I avsnitt 2.3.2 beskrivs hur komplexitet kan operationaliseras och i 2.3.3 hur korrekthet kan operationaliseras. I 2.3.4 refereras några studier där dessa metoder använts.

2.3.1 CAF-triaden

Grundtanken inom CAF-forskningen är att komplexitet (*complexity*), korrekthet (*accuracy*) och flyt (*fluency*) utgör grundläggande komponenter i det multikompositionella system som språkfärdigheten utgör (Housen, Kuiken & Vedder 2012, s. 1–3). Komplexitet, korrekthet och flyt har mätts på en rad olika sätt och mätresultaten har länge fungerat som beroende variabler inom tillämpad andraspråksforskning, där man intresserar sig för effekten av exempelvis ålder eller typ av uppgift på inlärares performans. Med den kognitiva vändningen inom andraspråksforskningen har CAF också börjat användas som

en oberoende variabel; CAF antas säga något om talarens språkfärdighet (Bulté & Housen 2012, s. 21; se också Baten & Håkansson 2014, s. 2).

Under 1990-talet skapade Skehan (1998, refererad hos Ellis & Barkhuizen 2005, s. 141–144) en sammanhållen, psykologisk modell över relationen mellan dessa tre komponenter. Enligt teorin måste andraspråkstalare, på grund av arbetsminnets begränsade kapacitet, välja mellan att, i sin språkliga produktion, fokusera på antingen innehåll eller form. Fokus på innehåll resulterar i flyt medan den som fokuserar på form i nästa steg måste välja mellan att fokusera på antingen korrekthet eller komplexitet, se Figur 3.



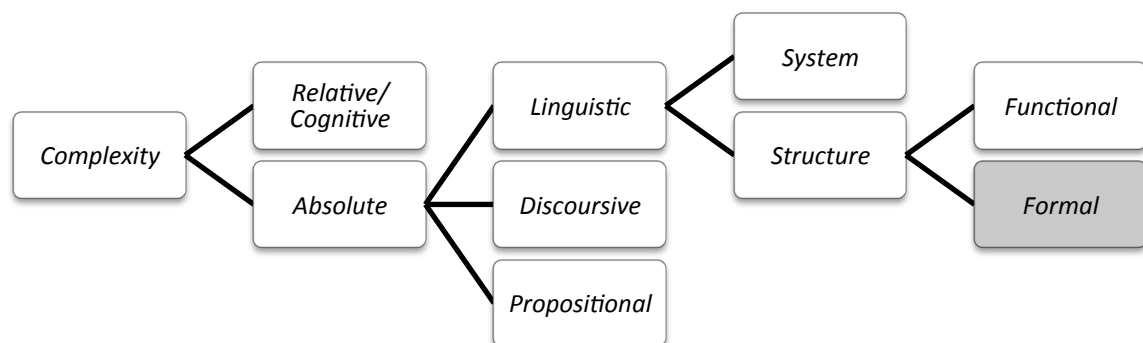
Figur 3. Tre aspekter av språklig performans efter Skehan (Ellis & Barkhuizen 2005, s. 143)

Modellens attraktionskraft ligger bland annat i att den kan förklara den osystematiska variation som teorier om generella inlärningsgångar inte kan förklara (Ellis & Barkhuizen 2005, s. 142). Grundtanken är som sagt att de tre komponenterna konkurrerar om arbetsminnets kapacitet, men numera talar man inte bara om *trade off*-effekter utan har visat att de tre aspekterna av språkfärdighet också interagerar på andra sätt. Housen, Kuiken och Vedder (2012, s. 7) skriver: "Accumulative evidence indicates that complexity, accuracy and fluency do not develop collinearly in SLA, that they interact in intricate ways and that this interaction is sometimes mutually supportive and sometimes competitive." Vissa menar att valet mellan fokus på form och innehåll är universellt medan konkurrensen mellan korrekthet och komplexitet bara gäller andraspråkstalare (Ellis & Barkhuizen 2005, s. 144). Andraspråkstalaras val mellan komplexitet och korrekthet ska enligt många styras av talarens personlighetstyp; så kallade risktagare försöker gärna producera ett mer komplext språk än vad de egentligen klarar av.

2.3.2 Komplexitet – definitioner och operationaliseringar

Komplexitet brukar definieras som bredden och djupet hos andraspråkstalaras ordförråd och grammatik och är den mest svårdefinierade av de tre konstrukterna i CAF-triaden (Housen, Kuiken & Vedder 2012, s. 2–4; Gyllstad m. fl. 2014, s. 6). I en metastudie av fyrtio undersökningar där olika komplexitetsmått använts konstaterar Bulté och Housen (2012, s. 22) att definitionerna av komplexitet ofta är vaga och ibland rent cirkulära. "In general,

we can say that empirical CAF research has taken a rather narrow, reductionist, perhaps even simplistic view on and approach to what constitutes L2 complexity.” (s. 34) För att definiera komplexitet måste man, skriver de, först och främst skilja på relativ eller kognitiv komplexitet – alltså det som är svårt, något som kan skifta från person till person – och absolut komplexitet. Den senare typen beräknas objektivt utifrån antalet komponenter någonting består av och antalet kopplingar mellan dessa komponenter (Bulté & Housen 2012, s. 22–24). Komponenterna kan vara innehållsmässiga, diskursiva eller rent lingvistiska. Absolut, lingvistisk komplexitet delar de upp i systemkomplexitet och strukturkomplexitet, där den förra termen refererar till det språkliga systemets bredd, det vill säga de variationsmöjligheter språk användaren har, medan den senare termen refererar till enskilda strukturers komplexitet. Strukturkomplexitet kan slutligen delas upp i funktionell och formell komplexitet, där den förra termen refererar till antalet funktioner en viss struktur har eller antalet komponenter som styr strukturens funktion, medan den senare refererar till antalet komponenter som ingår i själva strukturen. Dessa olika typer av komplexitet illustreras i Figur 4. I den här uppsatsen fokuseras lingvistisk komplexitet, framför allt formell strukturkomplexitet.



Figur 4. Olika typer av komplexitet efter Bulté och Housen (2012)

Inom såväl första- som andraspråksforskning har man använt sig av en rad olika komplexitetsmått – längdmått, koordineringsmått, variationsmått – men det populäraste inom andraspråksforskningen har länge varit underordningskvoten, det vill säga andelen bisatser av alla satser (Norris & Ortega 2009, s. 558; se även Ellis & Barkhuizen 2005, s. 156). Norris och Ortega (2009, s. 563) menar dock att forskare som är intresserade av språkutveckling måste mäta komplexitet på olika sätt för att fånga denna utveckling, detta eftersom språket utvecklas först genom paratax (samordning), sedan genom hypotax (underordning) och slutligen genom olika typer av grammatiska metaforer, som nominaliseringar och satsförkortningar, vilka leder till lägre grad av underordning. ”At worst,” skriver de, ”on the basis of subordination measures alone we may completely misinterpret whether an increase or decrease is indicative of a positive or negative change in perfor-

mance, because a decrease in subordination at the highest levels of proficiency may be related to an increase in the overall complexity of the language performance.” (s. 566) Norris och Ortega (2009, s. 574) menar för det första att komplexitet måste mätas både globalt, med generella metoder, och lokalt, med specifika metoder, där man tittar på användningen av vissa utvalda strukturer, exempelvis hjälpverb, passivkonstruktioner och infinitivfraser. Enligt Fulcher (2003, s. 28) hör till exempel relativsatser till de strukturer som lärs in sent. För det andra, skriver Norris och Ortega, måste både underordningskomplexitet och fraskomplexitet mätas (samt samordningskomplexitet vid studier av talare på mycket låga färdighetsnivåer). Det enklaste sättet att mäta fraskomplexitet är, enligt författarna, att beräkna satsernas genomsnittliga längd (underordnade satser räknas för sig); detta mått bör säga något om hur utbyggda fraserna är. Tonkyn (2012) är dock mer noggrann i sina mätningar av fraskomplexiteten; han tittar på användningen av framförställda och efterställda attribut inom nominalfrasen, olika typer av hjälpverb och olika typer av adverbial.

2.3.3 Korrekthet – definitioner och operationaliseringar

Korrekthet refererar till i vilken utsträckning andraspråkstales performans och underliggande interimspråk överensstämmer med en norm som vanligen utgörs av förstaspråkstales interna grammatik (Housen, Kuiken och Vedder 2012, s. 4). Även om korrekthet är den minst problematiska av de tre konstrukterna i CAF-triaden är den inte oproblematisk: dels är det svårt att säga i vilken utsträckning en form avviker från normen – något kan ju vara mer eller mindre fel –; dels är det inte alltid självklart vad normen är – förstaspråkstalarare utgör ju inte en homogen grupp. Housen, Kuiken och Vedder (2012, s. 4) menar därför att det kan vara lämpligt att tala inte bara om korrekthet (*accuracy*) utan även om *appropriateness* eller *acceptability*.

Precis som komplexitetsmått delas korrekthetsmått upp i globala och lokala mått, eller generella och specifika. Exempel på generella mått är andelen felfria yttranden av alla yttranden, antalet fel per hundra ord eller andelen rättade fel av alla fel (Ellis & Barkhuizen 2005, s. 150). Logiken bakom denna typ av mått är att lyssnares irritation skulle vara ”directly predictable from the number of errors [...] regardless of error type” (Albrechtsen, Henrikson & Færch 1980, citerade i Tonkyn 2012, s. 224). Exempel på specifika mått är andelen producerade plural-s och tredjeperson-s (i engelskan) av alla obligatoriska kontexter eller andelen korrekta verbfraser av alla verbfraser. Ellis och Barkhuizen (2005, s. 151–150) påpekar att risken med specifika mått är att vissa andraspråkstalarare kan ha problem med just den struktur som forskaren undersöker eller

kanske inte alls producerar denna struktur. Det bästa är således, skriver de, en kombination av olika typer av mått.

Det finns två olika syften med att titta på specifika mått. För det första kan vissa typer av fel anses allvarigare än andra, antingen därför att de leder till missförstånd eller därför att de har ett signalvärde och skapar irritation. Fulcher (2003, s. 27) menar till exempel att prepositionsfel och tempusfel, i språk som engelskan, sällan leder till allvarliga missförstånd medan ordföljdsfel och utelämnningar av obligatoriska led nästan alltid kan anses grava. Att fastställa en gravhierarchy för språkfel har dock visat sig svårt, skriver Tonkyn (2012, s. 225); dessutom, skriver han, tvärtemot Fulcher, råder det enighet om att problem inom verbfrasen är allvarigare än problem inom nominalfrasen. För det andra kan syftet med att studera korrektheten i specifika strukturer vara att behärsksningen av vissa strukturer berättar något om var i språkutvecklingen talaren befinner sig, se avsnitt 2.4.

2.3.4 Studier om komplexitet, korrekthet och språkfärdighet

I en av de första studierna där subjektiva, holistiska bedömningar jämfördes med objektiva, lingvistiska mätningar (se avsnitt 2.3) analyserade Magnan (1988, refererad i Iwashita m. fl. 2008 och i De Jong m. fl. 2012) fyrtio testtagares performans i ett muntligt test i franska. Där fanns en signifikant korrelation mellan testtagarnas korrekthet och betyg, men relationen var inte alltid linjär: "[...] at higher levels, learners attempt more complex grammatical notions, and consequently make more errors" (Iwashita 2008, s. 26). Detta fenomen, *non-linearity*, är välkänt inom andraspråksforskningen. Fulcher (2003, s. 28) skriver: "In teaching and assessing, we should be aware of the fact that sometimes quite serious errors are made simply because the learner is making progress in learning the language."

Iwashita m. fl. (2008) testade korrelationer mellan testtagares performans och ett holistiskt femgradigt betyg i ett muntligt språktest i engelska. Genom att mäta tre olika konstrukter – lingvistiska resurser (ordförråd och grammatik), fonologisk korrekthet samt flyt – på mer än 20 olika sätt hos 200 testtagare visade de att mått från vart och ett av de tre områdena kunde användas för att särskilja de fem färdighetsnivåerna, även om överlappningen mellan nivåerna var stor. Globala mått visade sig korrelera mer med bedömningar än vad specifika mått gjorde. Ordförråd och flyt tycktes vara de viktigaste komponenterna av språkfärdigheten. Även om grammatisk korrekthet, precis som i tidigare studier, också var viktig ville författarna nedtona korrekthetens betydelse.

Jin och Mak (2013) undersökte relationen mellan 66 testtagares muntliga performans och betyg i ett kinesiskt språktest. I studien användes, förutom mått för exempelvis flyt

och ordförråd, ett korrekthetsmått – andelen korrekta meningar (*sentences*) – och ett komplexitetsmått – meningarnas genomsnittliga längd. Resultatet visade att både komplexitet och korrekthet korrelerade med betyget – korrekthet starkt ($p < 0,01$) och komplexitet svagt ($p < 0,05$) – men att dessa båda komponenter inte korrelerade med varandra. Båda komponenterna befanns alltså vara viktiga, men de utvecklas inte nödvändigtvis parallellt. Precis som Iwashita m. fl. (2008) konstaterar Jin och Mak (2013) att färdighetsnivåerna överlappar varandra; gränserna är inte distinkta.

Även Tonkyn (2012) undersökte relationen mellan subjektiva bedömningar och objektiva mätningar av engelska andraspråkstalares muntliga performans, men i hans studie fanns en tredje dimension: genom att mäta och låta bedöma informanternas språk före och efter en intensivkurs i engelska kunde han också se vilka mått som bäst åter speglade en faktisk utveckling. Grammatisk komplexitet mättes med tio mått och korrekthet med sju. Både generella och specifika mått användes. Resultaten visade att de generella komplexitetsmåten – det totala antalet ord, *Syntactic complexity measure* (en poäng för strukturers utbyggnad) och underordningskvoten – mätte progressionen bättre än de mått som ska säga något om hur utbyggda fraserna är, även om också verbfras- och adverb-komplexiteten visade att en inläring skett. Av korrekthetsmåten var det också de globala måten som gav tydligast utslag, men även verbfras-korrektheten visade att inläring skett. Holistiska bedömningar var inte lika bra som objektiva mätningar för att upptäcka den inläring som skett under den korta intensivkursen. De holistiska bedömningarna påverkades också av halo-effekter (se avsnitt 2.2.3); särskilt komplexiteten tycktes vara svår att särskilja från andra konstrukter (Tonkyn 2012, s. 238).

Gyllstad m. fl. (2014) hörsammade uppmaningen från Norris och Ortega (2009) i sin studie av 240 texter skrivna på engelska, franska eller italienska av 120 svenska grundskole- och gymnasieelever. Texterna nivåplacerades av erfarna CEFR-bedömare och för varje text beräknades T-enheternas genomsnittliga längd (T-enhet=huvudsats med alla tillhörande bisatser; makrosyntagm), underordningskvoten samt satsernas genomsnittliga längd (där underordnade satser är borträknade; fraskomplexitet, se avsnitt 2.3.2). Alla tre måten korrelerade signifikant med CEFR-nivå men korrelationerna för satslängd var svagare än för de två andra komplexitetsmåten. Det förklarar författarna med att de flesta texterna bedömts vara på färdighetsnivå A1–B2. Enligt Norris och Ortega ska ju fraskomplexitet fånga utvecklingen av komplexitet först hos avancerade inlärare (se avsnitt 2.3.2).

Martin m. fl. (2010) diskuterar vad komplexitet egentligen är; de menar att många av de globala mått som används inom CAF-forskningen är problematiska, speciellt för ett språk som finska, där långa utbyggda meningar anses tillhöra dålig stil. I deras projekt Cefling har 669 finska texter skrivna av andraspråksinlärare nivåbestämts av CEFR-

bedömare, och man har skapat en modell för att undersöka språkutveckling i dessa texter genom att titta på specifika strukturer. I det delprojekt som redovisas i Martin m. fl. (2010) undersöktes hur användningen av lokativt kasus, transitiva konstruktioner och passivkonstruktioner korrelerar med CEFR-nivå. Resultaten visade att användningen av passiv ökar stadigt från nivå A1 till C2. Användningen av lokativt kasus minskar samtidigt som den metaforiska användningen av samma kasus ökar från nivå A1 till C1; språket blir mer abstrakt. De transitiva konstruktionerna, slutligen, ökar i användning framför allt mellan nivå A2 och B1 och blir sedan korrekta på nivå B2.

I projektet *What is Speaking Proficiency* undersöktes relationen mellan muntlig språkfärdighet, så som den bedöms utifrån CEFR, och mätbar språklig kompetens – både lexikal och grammatisk kunskap och processkunskap – hos 181 inlärare av nederländska som andraspråk (Hulstijn, Schoonen, de Jong, Steinel & Florijn 2011; De Jong m. fl. 2011). Den delstudie som rapporteras i Hulstijn m. fl. (2011) skiljer sig från studierna som refererats ovan i det att den lingvistiska kompetensen inte mättes utifrån testperformansen; i stället mättes den med experimentella metoder under totalt 3,5 timmar per deltagare. Erfarna CEFR-bedömare bestämde sedan CEFR-nivå oberoende av dessa experiment. Med statistiska metoder testades vilka lingvistiska kompetenser som krävs för att uppnå en viss CEFR-nivå. Nivåerna B1 och B2 stod i fokus i undersökningen och alla lingvistiska kompetenser utom talhastighet kunde åtminstone delvis användas för att skilja B1-or från B2-or. Den bästa prediktorn visade sig vara ordförrådets storlek, men även grammatisk kompetens och uttal var viktiga komponenter.

Slutligen ska nämnas den forskning som bland andra Spoelman och Verspoor (2010) bedriver inom Dynamisk systemteori. I sin longitudinella studie av kasusfel och nominalfrasers komplexitet i 54 texter skrivna av en inlärare av finska som andraspråk visade de att utvecklingen av korrekthet och komplexitet inte är linjär. Ordens utbyggnad och nominalfrasernas utbyggnad utvecklades parallellt medan nominalfraskomplexitet och satskomplexitet snarare konkurrerade med varandra, även om denna konkurrens upphörde på högre färdighetsnivåer.

2.4 Grammatiska utvecklingsstadier och språkfärdighet

Att komplexitet och korrekthet inte utvecklas linjärt, vilket flera av studierna ovan visar, gör dessa konstrukter olämpliga att använda för att mäta språkutveckling, menar Baten och Håkansson (2014). I sin longitudinella studie av tyska och svenska andraspråkstalar visar de, som många andra, att andelen bisatser – underordningskvoten – inte ökar linjärt över tid. De skriver: "[...] CAF measures in general appear to be valid measures to reveal variation (which is caused by a variety of factors, such as different tasks and different

learning contexts). However, the CAF measures cannot be used as indicators of L2 development, because they are not operationalized as such.” (s. 23) För att fastställa var i språkutvecklingen en talare befinner sig är det i stället bättre, skriver Baten och Håkansson, att använda mätinstrument utvecklade inom Processbarhetsteorin (PT) (se även Håkansson & Norrby 2007). I 2.4.1 beskrivs PT och i 2.4.2 refereras två studier om relationen mellan PT och allmän språkfärdighet.

2.4.1 Processbarhetsteorin

Processbarhetsteorin (PT) är en psykolingvistisk teori som förklarar den morfosyntaktiska utvecklingen hos en andraspråkstalare (se Ellis 2008). Teorin utvecklades under 1990-talet av Pienemann (1998) utifrån Levelts modell för talproduktion. I denna modell, som utgörs av tre huvudsakliga beståndsdelar – *Conceptualizer*, *Grammatical encoder* och *Articulator* –, antas en rad autonoma processorer arbeta parallellt i en inkrementell process: "the next processor can start working on the still incomplete output of the current processor" (Levelt 1989, citerad hos Pienemann 2011, s. 29). Detta innebär att en talare som påbörjar ett yttrande – exempelvis "De där blommorna ..." – vid samma tidpunkt inte har hela yttrandet färdigplanerat. För att yttrande ska kunna fullbordas krävs att grammatisk information lagras i vad Levelt kallar *Grammatical memory store*, ett omedvetet, specialiserat korttidsminne. Utan detta grammatiska minne skulle inte yttrandet kunna avslutas med pluralform på adjektivet: "... är gräslika."

Enligt PT finns det en implikationell hierarki, en femgradig skala, för hur andraspråksinlärare bygger upp dessa automatiserade processer. Att hierarkin är implikationell innebär att en nivå inte kan uppnås utan att nivåerna under också uppnåtts. I Tabell 6 sammanfattas och exemplifieras denna hierarki.

Tabell 6. PT-nivåer

Nivå	Processbar domän	Morfologi & syntax	Exempel
5	meningar	distinktion mellan huvudsats- och bisatsordföljd	<i>de röda blommorna som inte är dyra vill jag ha</i>
4	satser	predikativ kongruens; V2-inversion	<i>de röda blommorna som är inte dyra vill jag ha</i>
3	fraser	attributiv kongruens; infinitiv efter hjälpverb	<i>de röda blommorna som är billig jag vill ha</i>
2	ordklasser	plural; bestämd form; presens; preteritum	<i>jag vill ha röd blomma</i>
1	ord/lemma	invarianta former; helfraser	<i>jag ha blomma; skulle vilja</i>

På nivå 1 lagras ingen grammatisk information; talaren arbetar med enskilda ord – lemmen – och med fraser som är inlärda i sin helhet. På nivå 2 processas grammatisk

information på ordnivå; orden får ordklassstillhörighet och kan till exempel böjas i plural. På nivå 3 kan grammatisk information processas inom fraser, vilket bland annat innebär att inläraren klarar av kongruenser inom nominal- och verbfraser. På nivå 4 kan inläraren processa information över frasgränser; här möjliggörs predikativ kongruens och inversion efter spetsställt icke-subjekt. På nivå 5 kan grammatisk information slutligen processas över satsgränser, vilket möjliggör systematisk distinktion mellan huvudsats- och bisatsordföljd.

Utifrån PT är det möjligt att förutsäga i vilken ordning grammatiska strukturer i olika språk lärs in. I en metastudie testade Pienemann och Håkansson (1999) PT mot resultaten från fjorton undersökningar av svenska som andraspråk, och inte i ett enda fall falsifierade resultaten från dessa studier teorins prediktioner. I en tvärspråklig, skandinavisk studie undersökte Glahn, Håkansson, Hammarberg, Holmen, Hveneklide och Lund (2001) utvecklingen av attributiv kongruens, predikativ kongruens och bisatsordföljd hos inlärare av svenska, norska och danska. Också de fann tydligt stöd för PT: attributiv kongruens lärs in före predikativ kongruens som lärs in före bisatsordföljd.

PT är inte en teori om komplexitet och korrekthet, utan en teori om grammatiska utvecklingsstadier. I studier av PT är man inte intresserad av huruvida inlärare gör rätt eller fel eller av hur komplexa strukturer inlärarna producerar; man intresserar sig i stället för vilken processningskapacitet inläraren har. Större processningskapacitet ökar dock möjligheten för inlärare att producera mer komplexa strukturer på ett korrekt sätt och jag menar därför att det är rimligt att använda PT i ett arbete om komplexitet och korrekthet.

2.4.2 Studier om Processbarhetsteorin och allmän språkfärdighet

Eftersom PT kan användas för att mäta var i språkutvecklingen en inlärare befinner sig, eller åtminstone hur väl utvecklade de morfosyntaktiska processerna är, har forskare använt teorin i undersökningar av språktestning. Grandfeldt och Ågren (2013) jämförde språkfärdighet, så som den bedöms utifrån CEFR, och språkutveckling, så som den mäts med PT-verktyg. Materialet utgjordes av 76 texter skrivna på franska av 38 elever i en svensk skola. Texterna nivåbestämdes av två erfarna CEFR-bedömare. I 61 av texterna fanns tillräckligt med material för att PT-nivån skulle kunna fastställas. De två variablerna visade sig korrelera starkt med varandra ($p < 0,001$), även om överlappningen också visade sig vara stor på framför allt de högre färdighetsnivåerna. Samtidigt som det finns ett samband mellan PT-nivå och generell kommunikativ förmåga är det alltså möjligt för en inlärare, åtminstone efter att en viss lingvistisk nivå uppnåtts, att utvecklas kommunikativt utan att utvecklas lingvistiskt. Grandfeldt och Ågren (2013, s. 29) diskuterar dessa

resultat utifrån Hulstijns distinktion mellan *Basic Language Cognition* och *Higher Language Cognition*. För att uppnå CEFR:s högre nivåer krävs, enligt Hulstijn, mer än lingvistisk kompetens; denna utveckling beror bland annat på språkoberoende faktorer som intelligens och utbildning.

I experiment visade Ellis (2008) att PT endast är giltig vid test av implicit kunskap i så kallade *online*-uppgifter. Det är principiellt möjligt för en inlärare som skriver en text att med hjälp av sin explicita kunskap producera bisatsordföljd utan att ha uppnått nivå 5, varför teorin bör lämpa sig bäst för analys av talspråk. I sin avhandling använde Eklund Heinonen (2009) PT för att undersöka den muntliga delen av det svenska Tisustestet, som ger samma behörighet som gymnasiesvenska, det vill säga behörighet till universitetsstudier. Tisus utgår inte från CEFR men testar samma konstrukt som beskrivs i CEFR, nämligen kommunikativ språkförmåga. Tisusnivån motsvarar CEFR:s C1-nivå.

Eklund Heinonen fastställde PT-nivå för 33 godkända och 33 underkända testtagare utifrån tre syntaktiska och två morfologiska kriterier. Inom PT säger man att en nivå är uppnådd när det finns belägg på systematisk användning, med kontrasterande exempel mellan till exempel predikativ singular- och pluralkongruens; man talar om *The Emergence Criterion* (Eklund Heinonen 2009, s. 75). Detta kriterium säger ingenting om hur automatiserad en process är, utan endast om inläraren kan utföra processen. Eftersom det är svårt att veta om ett fall av till exempel predikativ kongruens verkligen bevisar att grammatisk information processas över frasgränser, eller om det snarare är ett fall av helfrasinlärning eller ren slump, använde Eklund Heinonen inte bara *The Emergence Criterion* utan också, precis som Glahn m. fl. (2001), ett 50-procents- och ett 80-procentskriterium (dvs. för att inversion ska anses inlärda ska där vara inversion i 50/80 procent av alla obligatoriska kontexter). Resultaten visade att de flesta godkända testtagare utifrån 80-procentskriteriet nått upp till nivå 4, flera även till nivå 5, medan endast några få av de underkända testtagarna nått upp till nivå 4.

Precis som i andra studier fanns i Eklund Heinonens studie en viss överlappning mellan de bedömda språkfärdighetsnivåerna. Denna överlappning diskuterade hon i en närstudie av en godkänd testtagare som med knapp marginal bedömts ha uppnått PT-nivå 3 och en underkänd testtagare som med god marginal uppnått PT-nivå 4. Genom samtalsanalys visade hon hur kommunikativ förmåga inte måste komma hand i hand med morfologisk utveckling. Enligt Eklund Heinonen (2009) tyder närstudien på "att den pragmatiska kompetensen, det vill säga förmågan att hantera de kommunikativa funktionerna, liksom den strategiska kompetensen och kontrollen, varit avgörande för det resultat testtagarna fått" (s. 165).

2.5 Sammanfattning av bakgrunden

I detta avsnitt har jag beskrivit hur utvecklingen av kommunikativa språktest, med CEFR (*Gemensam europeisk referensram för språk*) som främsta exempel, har resulterat i en beskrivning av språkfärdighet som gör språkfärdigheten svårtestad. Språkets främsta syfte är kommunikation och i ett kommunikativt språktest ska förmåga att kommunicera testas, men att skilja den kommunikativa förmågan från den rent lingvistiska tycks svårt. En rad studier visar att lingvistisk kompetens – framför allt ordförråd och grammatik – är viktiga komponenter i språkfärdigheten, framför allt på låga färdighetsnivåer, men föga förvånande finns det inga absoluta samband. En grundläggande fråga är i vilken utsträckning den kommunikativa kompetensen förutsätter lingvistisk kompetens. En del resultat tyder på att den som kommit över en viss "tröskel" kan utveckla sin kommunikativa förmåga oberoende av den lingvistiska kompetensen. Den av de refererade studierna som framför allt angriper denna frågeställning är Eklund Heinonen (2009), som också varit en viktig utgångspunkt för min egen undersökning, som jag berättar om i nästa avsnitt.

3 Metod och material

För att besvara uppsatsens första fråga – (Hur) korrelerar komplexitet, korrekthet och bedömning i ett muntligt språktest? – har jag spelat in, transkriberat, kodat och med statistiska verktyg analyserat tjugo testtagares performans under ett autentiskt muntligt språktest: den muntliga delen av Swedex B2-test. Detta test beskrivs i avsnitt 3.1. I avsnitt 3.2 beskriver jag mitt primära material, bestående av tio ljudfiler. De etiska överväganden som berör insamlingen och hanteringen av detta material diskuteras i avsnitt 3.3. Omvandlingen av det inspelade materialet till data genom transkribering och kodning beskrivs i avsnitt 3.4, och den statistiska analysen av dessa data beskrivs slutligen i avsnitt 3.5.

3.1 Swedex muntliga B2-test

Swedex (Swedish Examinations) är ett svenskt certifikatstest som sedan 2005 ges av studieförbundet Folkuniversitetet (2014) och som utvecklades av Folkuniversitetet tillsammans med bland andra myndigheten Svenska Institutet. Testet utgår från CEFR och ges i dag för tre av CEFR:s sex färdighetsnivåer: A2, B1 och B2. I denna uppsats fokuseras B2-testet, som bland annat tas av många så kallade EU-immigranter som söker jobb inom svensk sjukvård. Personer som har till exempel läkar- eller sjuksköterskelegitimation från länder utanför EU/ESS och Schweiz måste uppvisa dokumenterade kunskaper i svenska språket motsvarande CEFR:s C1-nivå – till exempel ett godkänt Tisustest – för att

Socialstyrelsen (2014) ska utfärda motsvarande svenska legitimation. Medborgare i EU/ESS-länder och Schweiz får dock sin svenska legitimation utfärdad automatiskt om de har motsvarande utbildning från sitt hemland; det är sedan upp till landstingen själva att bestämma vilka språkkunskaper som krävs för anställning. Praxis i dessa fall är att kräva dokumenterad B2-nivå (Stockholms läns landsting 2012, s. 3; Region Skåne 2011, s. 1), vilket den har som gått ut svensk grundskola med godkänd svenska, avslutat SFI eller tagit Swedex B2-test. Swedex B2-test är alltså ett så kallat *high stake*-test (se t.ex. Bachman 2004, ss. 12–13); ett felaktigt godkänt eller underkänt betyg kan medföra allvarliga konsekvenser för såväl testtagaren som samhället.

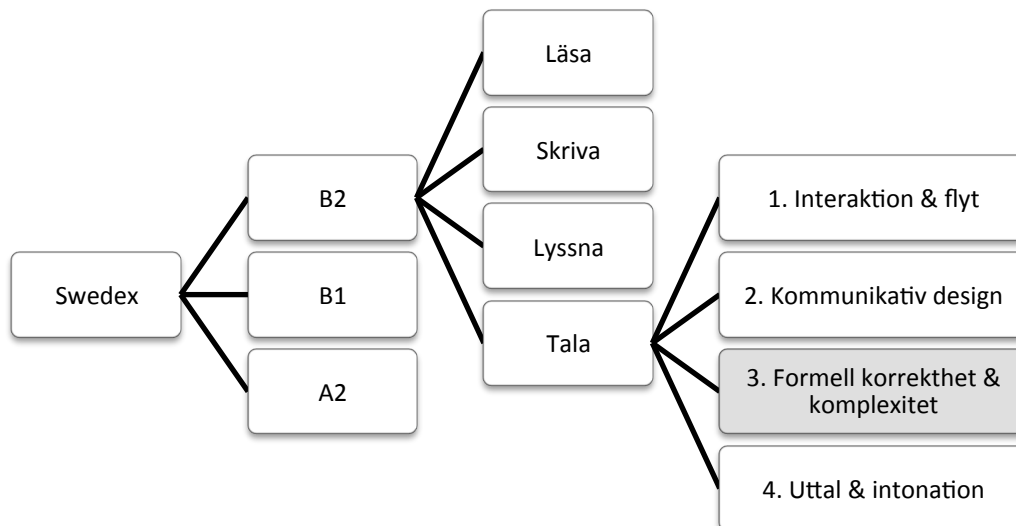
Swedex testar "alla språkliga färdigheter" och består av fyra deltest: läsa, skriva, lyssna och tala (Folkuniversitetet 2011). Det muntliga testet görs av två, ibland tre testtagare tillsammans; dessa har normalt aldrig träffat varandra före testet. De får välja ett av några ämnen att förbereda femton minuter före testet. Efter att ha presenterat sig för varandra ska de dels hålla varsin monolog på 3–4 minuter om sina ämnen, dels diskutera dessa ämnen med varandra. Totalt tar testet cirka 20 minuter.

För varje deltest finns tre betygsnivåer: en godkänd nivå, en miniminivå och en underkänd nivå. För att bli godkänd på hela Swedex måste testtagaren uppnå minst godkänd-nivån på tre delar och miniminivån på maximalt en del. "Minimigränsen", skriver testkonstruktörerna, "bör representera en nivå där testtagaren kan tänkas ligga på nivå B2 men av någon anledning (stress, black out eller mildare störning i miljön) inte har kunnat prestera sitt bästa på hela testdelen [...]" (Folkuniversitetet 2011). Den som blir underkänd på ett deltest blir underkänd på hela testet och måste, vid ett eventuellt nytt försök, göra om hela testet.

Den muntliga delen bedöms analytiskt (se avsnitt 2.2.3) enligt fyra kriterier på en tregradig skala, där A betyder att testtagaren uppnått B2-nivå med marginal eller befinner sig över B2-nivå, B betyder att testtagaren befinner sig på B2-nivå och C att hon eller han inte uppnått B2-nivå. För att bli godkänd på den muntliga delen krävs minst B på alla fyra kriterier. För att uppnå miniminivån krävs minst B på tre av kriterierna. Det är alltså möjligt att bli godkänd på hela Swedex trots ett C på ett av kriterierna (Folkuniversitetet 2011). Distinktionen mellan nivåerna A och B är inte relevant för det slutgiltiga testresultatet men behövs för att "göra det lättare att urskilja vilka som faktiskt har uppnådd B2-nivå från dem som har en god B2-nivå eller kanske t o m ligger över B2. Risken blir då mindre att man lägger en lägstanivå för uppnådd B2 för högt", skriver testkonstruktörerna (Folkuniversitetet 2010).

De fyra kriterierna som bedöms är 1) Interaktion och flyt, 2) Kommunikativ utformning, 3) Formell korrekthet och komplexitet, och 4) Uttal och intonation (Folkuniversitetet

2010). Bedömningsunderlaget utgörs av tolv deskriptorer, en för varje nivå och kriterium, som beskriver vad testtagaren kan göra och hur hon eller han gör det på respektive nivå. För C-nivån är deskriptorerna framför allt negativt formulerade, det vill säga de beskriver vad testtagaren inte klarar av. Deskriptorerna för kriterium 3, som här står i fokus, beskriver testtagarens "grammatiska behärskning" och användning av "komplexa satskonstruktioner". Fokus ligger på i vilken utsträckning testtagaren gör grammatiska fel och i vilken utsträckning dessa fel stör eller hindrar kommunikationen. I testkonstruktörernas "Kommentarer till bedömningskriterierna för muntlig uppgift" påpekas att "kravet på grammatisk korrekthet/behärskning inte kan vara lika högt ställt i muntlig interaktion som i skriftlig" (Folkuniversitetet 2010). Testtagarna bedöms av två bedömare gemensamt. Uppsatsens avgränsade fokusområde illustreras i Figur 5, och bedömningsunderlaget för kriterium 3 återges i Tabell 7.



Figur 5. Uppsatsens fokus: Swedex B2-test, den muntliga delen, kriterium 3

Tabell 7. Deskriptorer för kriterium 3 (Folkuniversitetet 2010)

3. Formell korrekthet och komplexitet	
Har god grammatisk behärskning. Fel förekommer, men sällan genomgående. Felen stör dock inte kommunikationen. Använder vissa komplexa satskonstruktioner.	A
Har relativt god grammatisk behärskning. En hel del fel förekommer, vissa t o m genomgående. Felen hindrar dock inte kommunikationen. Använder vissa komplexa satskonstruktioner.	B
Upprepade grammatiska fel hindrar kommunikationen. Använder genomgående enkla satskonstruktioner.	C

3.2 Material

Vi tre testtillfällen har jag, vid ett testcenter i Sverige, spelat in tio muntliga B2-test med totalt 22 testtagare. Två testtagare ville inte delta i studien (se avsnitt 3.3), varför de är exkluderade ur materialet. Materialet utgörs således av ljudinspelningar av 20 testtagare som var och en talar i knappt tio minuter.

Eftersom jag endast studerar relationen mellan språklig performans och bedömning har jag inte samlat in information om testtagarnas förstaspråk, vistelsetid i Sverige, utbildning med mera (jmf. Eklund Heinonen 2009 s. 71). Vidare refererar jag till testtagarna utan att avslöja deras kön. Inte heller redovisas någon information om de fyra bedömare som i olika kombinationer bedömde de 20 testtagarna. Dessa beslut har jag fattat dels därför att informationen inte är relevant för min frågeställning, dels därför att materialet ändå är för litet för att några slutsatser skulle kunna dras utifrån sådan information. I Tabell 8 redovisas hur de 20 testtagarna bedömts i det muntliga testet.

Tabell 8. Bedömningar av testtagarnas muntliga performans samt slutbetyg

Testtagare	1. Interaktion & flyt	2. Kommunikativ design	3. Formell korrekthet & komplexitet	4. Uttal & intonation	Sammanvägning*	Slutbetyg
AAG	A	A	A	A	8	Godkänd
BBU	B	C	B	B	3	Underkänd
CBG	B	B	B	B	4	Godkänd
DBU	A	A	B	A	7	Underkänd
EAG	A	A	A	A	8	Godkänd
FAG	A	A	A	B	7	Godkänd
GBG	B	A	B	B	6	Godkänd
HBG	C	B	B	B	3	Godkänd
IBU	B	C	B	B	3	Underkänd
JCU	B	C	C	B	2	Underkänd
KBG	A	B	B	B	5	Godkänd
LBU	A	B	B	B	5	Underkänd
MCU	A	B	C	B	4	Underkänd
NAG	A	A	A	A	8	Godkänd
OBG	A	A	B	A	7	Godkänd
PBU	A	B	B	B	5	Underkänd
QBG	B	C	B	B	3	Godkänd
RBG	B	B	B	B	4	Godkänd
SBU	B	C	B	B	3	Underkänd
TBU	B	C	B	B	3	Underkänd

*I statistiken räknas C som 0, B som 1 och A som 2. Under rubriken "sammanvägning" har bedömningarna på de fyra muntliga kriterierna vägts samman.

I mina statistiska beräkningar ger jag C värdet 0, B värdet 1 och A värdet 2 (se avsnitt 3.5). I näst sista kolumnen har jag vägt samman bedömningarna för de fyra kriterierna: 8 poäng

betyder A på alla fyra kriterier och 0 poäng betyder C på alla fyra kriterier. Denna sammanvägning motiveras av att det inte är säkert att bedömarna kan särskilja distinkta konstrukter utan låter ett visst särdrag i testtagarens performans påverka bedömningen under andra kriterier; detta är känt som halo-effekter (se avsnitt 2.3.3). Det är alltså möjligt att komplexitets- och korrekthetsmått påverkar andra bedömningar än den under kriterium 3. Sammanvägningen kan sägas motsvara en holistisk bedömning av testtagarnas muntliga performans. I sista kolumnen redovisas slutbetyg på hela Swedex B2-test: godkänd eller underkänd. Testtagarnas koder, i första kolumnen, består av en identifierande bokstav, bedömningen under kriterium 3 (A, B eller C) samt slutbetyget på hela testet (G=godkänd; U=underkänd).

Sju av de 20 testtagarna har fått C på ett av de fyra kriterierna och en testtagare – JCU – har fått C på två av kriterierna. Det är alltså bara JCU som blivit underkänd på den muntliga delen; de sju testtagare som bara fått ett C har ju uppnått miniminivån och har alltså ändå haft möjlighet att bli godkända på hela testet. Vi ser dock att bara en dessa sju – QBG – klarade det. Ingen testtagare har fått C på kriterium 4, en testtagare har fått C på kriterium 1 och två testtagare har fått C på kriterium 3. Det kriterium som flest testtagare fått C på är kriterium 2, Kommunikativ design, där ordförråd, koherens och förmågan att tala om det valda ämnet bedöms. Här är betygsfördelningen jämn: sju testtagare har fått A, sju har fått B och sex har fått C. På kriterium 3, som står i fokus i denna studie, ser vi fyra A, fjorton B och, som sagt, endast två C. Förutom JCU är det MCU som inte anses ha uppnått B2-nivån när det gäller Formell korrekthet och komplexitet.

3.3 Etiskt övervägande

Under samlingen en stund före testet blev testtagarna informerade av testledaren om min närvaro. Jag berättade kort – på svenska – om min undersökning och delade ut ett papper med samma information på svenska och engelska; där fanns också ett medgivande för underteckning (se Bilaga 1). Jag bad dem att ge medgivandet till mig efter testet om de ville ge mig tillstånd att använda deras inspelade tal och deras betyg i min undersökning. På så sätt gav jag dem tid att tänka efter om de ville delta, och gjorde det också lättare för dem att avstå från att delta, vilket, som nämndes ovan, två av 22 testtagare valde att göra.

Min forskning medför ingen uppenbar fysisk eller psykisk risk för testdeltagarna, och metoden kräver inte heller att känsliga personuppgifter om exempelvis etniskt ursprung eller politisk eller religiös tillhörighet lagras, varför jag inte ansåg att något tillstånd från en etisk nämnd behövdes (se Vetenskapsrådet 2011, s. 49). Eventuella privata uppgifter som testtagarna berättar om sig själva under testet, liksom all information som kan avslöja deras identitet, transkriberas inte och utgör inte en del av mina data.

Trots detta kan det finnas etiska problem med insamlingen av materialet. Eckert (2013, s. 21) skriver om den press att delta som personer kan känna när forskare samlar in material i institutionella sammanhang, som i skolan, akademien och militären. För att undvika sådan press bör, skriver Eckert, insamlingen ske så att en överordnad eller – i det här fallet – betygsättande testledare inte vet om vilka som väljer att delta eller inte delta. Kanske borde jag ha informerat testtagarna om forskningen först efter testet för att inte riskera att stressa dem i en redan pressad situation. Sedan borde jag bett dem att skicka medgivandet direkt till mig efter att de fått sina resultat. Här finns anledning att vara självkritisk.

Men etiken kan också diskuteras ur en annan synvinkel. Enligt Eckert (2013, s. 25) är det viktigt att fråga sig vilken nytta forskningen har för dem som deltar och för samhället i stort. För testtagarna är det givetvis viktigt att det test de betalar för att ta och som är så betydelsefullt för dem granskas; min närvaro under testet kan ha ingett en trygghets-känsla hos testtagarna, en känsla av att bli rättvist bedömda. För samhället är det också viktigt att certifikatstest som Swedex granskas, eftersom testet bland annat är avgörande för vilka som får eller inte får arbeta i den allmänna sjukvården (se avsnitt 3.1). I den mån det funnits risk att enskilda testtagare blivit stressade av min närvaro – vilket jag inte tror att de blev – måste denna risk ställas i relation till undersökningens syfte.

3.4 Transkribering och kodning

Transkribering innebär i sig en analys- och tolkningsprocess (Eklund Heinonen 2009, s. 84). "Valet av transkriptionsmetod", skriver Eklund Heinonen (2009, s. 73), "måste styras av undersökningens syfte." Kendall (2008, citerad hos Nagy & Sharma 2013, s. 235) skriver: "Each decision that is made while transcribing influences and constrains the resulting possible readings and analyses." I detta avsnitt beskrivs de beslut jag fattat under transkriberingen och kodningen av mitt material.

Vid transkribering och kodning av materialet har jag använt mig av programmet CLAN och följt de grundläggande CHAT-konventionerna, båda utvecklade inom CHILDES-projektet (MacWhinney 2000). En fördel med att använda ett program som CLAN har varit att jag kunnat länka transkriptionerna till ljudfilerna, vilket gjort det möjligt för mig att under kodningen mycket enkelt kunna lyssna på varje yttrande många gånger. Det primära materialet har således hela tiden varit ljudfilerna; transkriptionerna är endast ett analysverktyg.

De inspelade testen har transkriberats i sin helhet med undantag för de inledande presentationerna, som mest syftar till att skapa en god stämning och få testtagarna att slappna av. Testledarnas tal har inte transkriberats och inte heller, som nämndes ovan,

material som kan hota testtagarnas anonymitet. Totalt består materialet av cirka 20.000 ord när upprepningar och självrättningar är borträknade, eller – om man så vill – 2.970 satser.

Jag har följt skriftspråksnormer med undantag för några få ord som har etablerade uttalsformer i skrift, till exempel *nån*, *nåt*, *nåra*, *dom*, *sen*, *dar*. Eftersom testtagarnas grammatiska kompetens står i fokus för undersökningen vore det inte bara slöseri med tid att använda någon form av fonetiskt notationssystem, utan den skriftspråkliga ortografin har också hjälpt mig att under analysprocessen fokusera på det som är relevant. Vad beträffar morfologiska ändelser är det ofta svårt att skilja mellan grammatiska och uttalsmässiga problem. Verb i konjugation 1 som i preteritum saknar ändelse (*jag mejla henne nyss*) transkriberas utan ändelse eftersom denna form är så vanlig i talspråk. Framför allt en testtagare – PBU – saknar *r*-ljud, men denna testtagare har en i övrigt så korrekt morfosyntax att jag antar att uteblivna *r*-ändelser är ett uttalsproblem snarare än ett grammatiskt problem. Därför har jag, för PBU, skrivit *r* när jag misstänker att *r*-et ”finns” där.

Tvekljud som i exempel (1) har transkriberats med konsekvent stavning så att det är enkelt att exkludera dessa ur ordräkningar, och långa eller markanta pauser har markerats med fyrkant (#) som i exempel (2); dock har varken tvekljud eller pauser transkriberats med stor exakthet. Däremot har självrättningar som i exempel (1) och (2) och upprepningar som i exempel (3) transkriberats och kodats konsekvent. Se transkriptionsnyckeln i Bilaga 2.

- (1) KBG: men kanske < om framtiden > [//] i framtiden om femtio år till exempel äh vatten kan bli en problem [...].
- (2) PBU: men kanske dom kommer inte till dig när du har problem när du # < blir > [//] mår inte så bra .
- (3) IBU: < bara > [/] < bara > [/] bara beror < om > [/] om din nationalitet < eller > [/] eller din religion eller din bakgrund .

I transkriptionen markeras inte samtidigt tal. Trots att förmågan att interagera ska testas är materialet nämligen mycket monologiskt. I de fall då samtidigt tal ändå förekommer skrivs yttrandena var för sig. I exemplet (4) yttrar sig BBU tre gånger under det att AAG talar. I exempel som detta utgör transkriptionen ingen särskilt sann eller begriplig återspeglning av det verkliga samtalet, men valet av transkriptionsmetod ska som sagt styras av undersökningens syfte. Vid stöttning, det vill säga när den ena talaren griper in och ger den andra ett ord eller fras, styckar jag upp yttrandet, som i exempel (5).

- (4) AAG: jag menar inte < att > [/] att ha utbildning som till exempel äh äh teoretisk < bi > [/] utbildning och sånt .
 BBU: ja .
 ta ett prov .
 ja ja .
- (5) CBG: och nu har jag den möjligheten .
 DBU: att utveckla dig ?
 CBG: < att utvecklas > [/] att utvecklas och att bli bättre .

Utifrån transkriptionerna har jag mätt testtagarnas komplexitet på fem sätt och deras korrekthet på två sätt. Dessutom har jag bestämt testtagarnas PT-nivå. Dessa totalt åtta mått sammanfattas i Tabell 9. I avsnitt 3.4.1 beskrivs hur komplexiteten mätts och i 3.4.2 hur korrektheten mätts. I 3.4.3 redogörs för ett inter-bedömarreliabilitetstest. I 3.4.4 beskriver jag slutligen hur PT-nivån har fastställts.

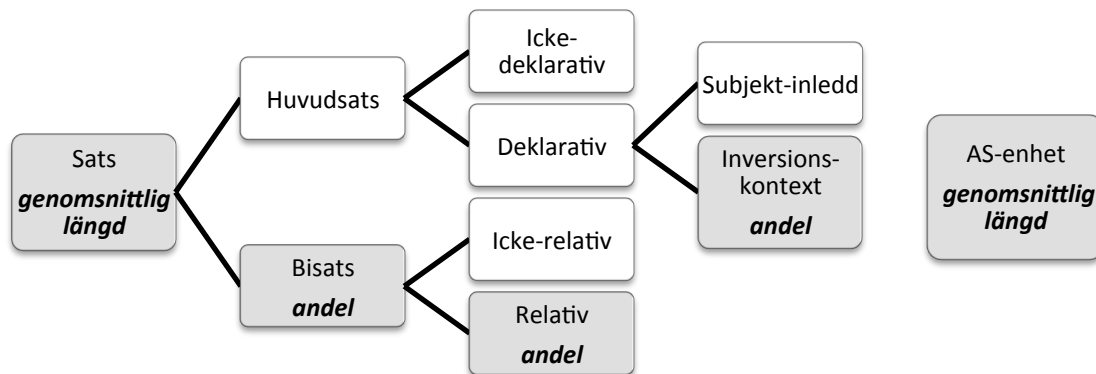
Tabell 9. Samtliga mått med beskrivningar

Typ av mått	Mått	Beskrivning
Komplexitets- mått	1. AS-längd	Det genomsnittliga antalet ord per AS-enhet.
	2. Satslängd	Det genomsnittliga antalet ord per sats. Varje sats räknas för sig, dvs. ord som ingår i en underordnad sats räknas inte till den överordnade satsen. Måttet antas mäta utbyggnaden av fraser.
	3. Andelen bisatser (BS-andel)	Andelen bisatser av alla satser.
	4. Andelen relativsatser (RS-andel)	Andelen relativsatser av alla satser.
	5. Andelen obligatoriska inversionskontexter (INV-kontextandel)	Andelen obligatoriska inversionskontexter av alla deklarativa huvudsatser.
Korrektthets- mått	6. Satskorrekthet (Satskorr.)	Andelen korrekta satser av alla satser. Varje sats räknas för sig, dvs. en sats räknas som korrekt även om en underordnad sats inom satsen är felaktig.
	7. Inversionskorrekthet (INV-korr.)	Andelen fall av inversion av alla obligatoriska inversionskontexter.
Utvecklings- mått	8. PT-nivå	En två-gradig skala som anger om testtagaren kan processa grammatisk information över frasgränser eller inte.

3.4.1 Kodning av komplexitet

Som framgår av tabellen i föregående avsnitt har komplexitet mätts på fem sätt. Ordflödet har segmenterats i syntaktiskt självständiga enheter (AS-enheter, se nedan) och i satser

(jmf Norris & Ortega 2009, avsnitt 2.3.2); för båda typerna av enheter har medellängd i ord beräknats. Satserna har kodats som huvudsats/bisats för att underordningskvoten skulle kunna beräknas. Bisatserna har kodats för om de utgör relativsats eller inte eftersom relativsats brukar läras in sent (Flyman Mattsson & Håkansson 2010, s. 59; Fulcher 2003, s. 28). Huvudsatserna har slutligen, precis som hos Eklund Heinonen (2009, s. 84), kodats för om de utgör obligatorisk inversionskontext eller inte eftersom vi vet att spetsställning av icke-subjekt är en mer komplex struktur än subjektet först (Pienemann & Håkansson 1999). Se kodschemat i Figur 6.



Figur 6. Mätning av komplexitet – kodschemat

I en inflytelserik metastudie skriver Foster, Tonkyn och Wigglesworth (2000) att mätningar av fenomen som komplexitet och korrekthet måste relateras till någon språklig enhet för att bli meningsfulla. I fler än hälften av de 87 artiklarna författarna går igenom definieras aldrig denna enhet, och i de fall då definitioner förekommer är de ofta vaga. Hur enheter som används för att segmentera ordflödet definieras och operationaliseras, skriver de, måste styras av forskarens syn på vad språk och språkfärdighet är. I studier av förstaspråksinlärning kan enheten definieras som ett eller flera ord som produceras i följd utan paus, men i undersökningar av andraspråksinlärning fungerar inte denna definition eftersom andraspråkstalare ofta kan producera mycket tal trots låg språkfärdighet (Foster m. fl. 2000, s. 355–356). Om man ser muntlig språkfärdighet som talarens förmåga att utifrån syntaktiska system skapa yttranden, snarare än som en förmåga att plocka fram memorerade ord och fraser ut minnet, ligger utmaningen i att identifiera en enhet som både äger psykologisk validitet – det vill säga en enhet som motsvarar den segmentering som talaren gör när hon eller han planerar sina yttranden – men som samtidigt kan identifieras med en tillräckligt hög grad av reliabilitet.

Foster m. fl. (2000) går igenom en lång rad enheter som forskare använt – alla definierade utgår från antingen semantiska, intonala eller syntaktiska kriterier – och

introducerar den så kallade AS-enheten (*Analysis of Speech unit*). AS-enheten är en syntaktisk enhet som operationaliseras med stöd av intonation. Den liknar den ofta använda T-enheten men är anpassad för talspråk, särskilt för andraspråkstales interaktion. "An AS-unit is a single speaker's utterance consisting of an independent clause, or a sub-clausal unit, together with any subordinate clause(s) associated with either." (s. 365; originalets kursiv) En *sub-clausal*-enhet utgörs av en eller flera fraser som kan ses som en elliptisk huvudsats. Psykolingvistisk forskning ger, enligt författarna, stöd för att den syntaktiska huvudsatsen äger psykolingvistisk validitet; åtminstone förstaspråkstalare tycks planera sitt tal huvudsats för huvudsats. Denna definition har jag utgått ifrån när jag transkriberat och segmenterat ordflödet i mitt inspelade material. Eftersom identifieringen av enheten kräver att jag skiljer på huvudsatser och bisatser var det enkelt att utifrån segmenteringen i AS-enheter också ta fram måtten andelen bisatser av alla satser och genomsnittlig satslängd; det senare måttet antas som sagt säga något om hur utbyggda fraserna inom satserna är (se avsnitt 2.3.2).

Foster m. fl. (2000, s. 363–369) identifierar en rad problem som uppstår när andraspråkstales interaktion ska segmenteras syntaktiskt. Till exempel är det svårt att veta om vad som ser ut som en final icke-nödvändig bisats verkligen är inkorporerad i föregående sats. Särskilt gäller det, skriver de, satser inledda med *because*, vilka ofta tycks fungera som självständiga yttranden. Ordet *because* fungerar i dessa fall som diskursmarkör med betydelsen 'jag säger detta eftersom' snarare än som underordnare. Lösningen för Foster m. fl. är tonal och temporal. Finala, icke-nödvändiga satser ska räknas som en del av föregående sats om 1) de faller under samma intonationskurva och 2) pausen mellan satserna inte är mer än en halv sekund.

Jag har bara delvis följt dessa kriterier när jag bestämt om en sats är huvudsats eller bisats. Min främsta kritik mot 0,5-sekundersregeln är att en strikt tillämpning av den skulle innebära att talare med dåligt flyt automatiskt skulle tillskrivas lägre komplexitet, vilket är problematiskt om man ser flyt och komplexitet som två distinkta konstrukter. Liknande problem får man om initiala bisatser, som till formen ser ut som huvudsatser, inte ses som en del av den större syntaktiska strukturen; om man, som jag i denna uppsats, vill hålla komplexitet och korrekthet isär är det viktigt att inte kalla bisatser för huvudsatser bara för att talaren inte har de morfosyntaktiska redskap som krävs för att på ett korrekt sätt inkorporera satsen i en överordnad sats.

När jag segmenterat ordflödet utifrån syntaktiska principer och bestämt vad som är huvudsats och vad som är bisats har jag därför i första hand utgått från om en sats ser ut att ha en funktion i en annan sats; jag kallar detta "funktionskriteriet". I exempel (6) är det tydligt att den icke-nödvändiga satsen, inledd med *eftersom*, ingår i en större syntaktisk

struktur trots den långa pausen (#). Med 0,5-sekundersreglen skulle den räknas som en egen AS-enhet. I exempel (7) är det tvärtom tydligt att satsen inledd med *eftersom* inte kan utgöra en kausal bisats i föregående fråga, och i (8) är det lika tydligt att satsen inledd med *så att* inte kan utgöra en final bisats till föregående påstående; *eftersom* tycks snarast betyda 'anledningen till att jag ställer denna fråga är att' medan *så att* bör tolkas som 'det är alltså så att'. Dessa satser utgör alltså egna AS-enheter, varför de börjar på ny rad. Varje AS-enhet avslutas med stort skiljetecken i enlighet med CHAT-konventionen.

- (6) CBG: många bestämmer sig att följa ah sina < föräldrarsna > [/] föräldrarnas yrke #
eftersom kanske det finns redan ett företag # att dom kan arbeta där .
- (7) CBG: men är det viktigt att varje land ah behåller sin egen språk ?
eftersom jag tror att språk visar personalitet av landet .
- (8) OBG: äh alltså dom vill äh ha en framgång utan äh att arbeta hårt för detta # och ta
risker .
så att äh < dom vill > [/] dom vill liksom ha den enkla vägen mot en succé .

Många gånger är det, även utifrån funktionskriteriet, svårt att säga om en sats är självständig eller utgör en final, icke-nödvändig bisats till föregående sats. För konsekvensens skull utgår jag, om funktionskriteriet inte tvingar mig till något annat, från etablerade beskrivningar av vilka ord som fungerar som subjunktioner respektive konjunktioner. Framför allt innebär det att satser inledda med *för* eller *så* kodas som egna AS-enheter medan satser inledda med *för att* eller *så att* kodas som bisatser – se exempel (9) –, detta trots att alla dessa ord i tal kan inleda såväl huvudsatser som bisatser, som mina påhittade exempel i (10) visar. I exempel (11) och (12) är det dock tydligt, utifrån funktionskriteriet, att *därför* respektive *så* inleder bisats – här får talaren alltså "komplexitetspoäng" men inte "korrekthetspoäng".

- (9) DBU: så vi # som USA dom pratar engelska alla vill lära dig engelska **för att** kanske jag
vill gå och ha jobb där eller dit .
- (10) ... så/för (att) jag inte kommer
... så/för (att) jag kommer inte
- (11) DBU: jag ska prata om språk därför vi är här i dag .
- (12) DBU: ... du sa att < det > [/] det är viktigt att man < har > [/] har jobb eller arbete så det
är fjärdedel av livet .

Ibland kan AS-enheterna bli mycket långa därför att en talare radar upp många bisatser, som i exempel (13), och satser kan bli långa därför att talaren radar upp många fraser,

som i exempel (14). Dessa uppradade bisatser och fraser skulle kunna betraktas som egna elliptiska AS-enheter (*sub-clausal units*). Hur Foster m. fl. (2000) skulle behandla dem vet jag inte, men jag har valt att konsekvent låta dem ingå i den större AS-enheten. Även om jag inte tror att talaren "planerar" yttranden som dessa i sin helhet vore det omöjligt för mig att bestämma vilka samordnade led som ska räknas som elliptiska, självständiga AS-enheter och vilka som ska anses inkorporerade i en överordnad struktur. Nackdelen med detta förfaringssätt är att måttet satslängd dessvärre kommer att säga lika mycket om samordning inom satserna som om syntaktisk utbyggnad av fraserna inom satserna.

(13) BBU: till exempel om du gå till en anställningsintervju om du < har > [/] har klär på sig då ah smutsigt kläder och du har inte tvättat håret < ja dom > [/] dom ska titta på dig .

(14) AAG: utan det är bättre < att > [/] att fokusera på förhållanden till andra människor ha en äh sällskap skaffa en familj ha barn och sånt .

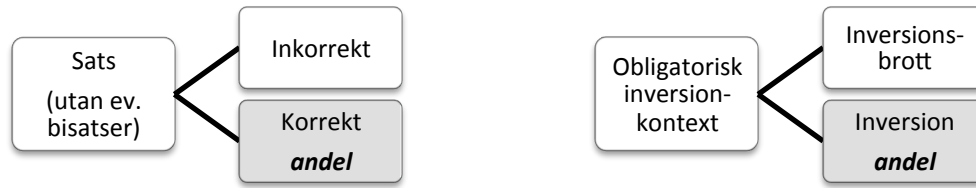
I kodningen av relativsatser har jag inte stött på så många problem, men jag ska ändå nämna att vad som ser ut som en relativsats i utbrytningskonstruktioner inte kodats som relativsats. Beträffande kodning av obligatoriska inversionskontexter ska nämnas att huvudsatser inledda av *till exempel*, *alltså*, *i alla fall* och liknande ord inte kodats som obligatoriska inversionskontexter, detta eftersom denna typ av ord också inleder huvudsatser med rak ordföljd i talspråklig svenska; man brukar tala om en annexposition. Exempel (15) har följaktligen också kodats som korrekt (se nästa avsnitt). Tydligast är det att initialt *kanske*, som i (16), inte kräver inversion.

(15) CBG: i alla fall arbetet spelar stor roll .

(16) FAG: och kanske jag kan jämföra olika produkter .

3.4.2 Kodning av korrekthet

Som framgick av Tabell 9 har jag mätt korrektheten på två sätt, ett specifikt och ett generellt. Det generella måttet, andelen korrekta satser av alla satser, har jag fått fram genom att koda varje sats (där underordnade satser räknas för sig) som korrekt eller inkorrekt. Detta mått ska ge en uppfattning om hur stor mängd fel lyssnaren utsätts för, men det säger ingenting om typen av fel. Det specifika måttet har jag fått fram genom att koda varje obligatorisk inversionskontext (se föregående avsnitt) för inversion eller icke-inversion. V2-brott kan ha ett signalvärde, men framför allt räknas behärsksningen av inversion till PT-nivå 4 eftersom den kräver att grammatisk information processas över frasgränser; strukturen säger alltså något om inlärares grammatiska utvecklingsnivå (se avsnitt 2.4.1). Se kodschemat i Figur 7.



Figur 7. Mätning av korrekthet – kodschema

Det är morfosyntaktisk korrekthet som kodats och inte lexikal, varför exempel (17) och (18) kodats som korrekta (jmf. eng. *I mean/it means*). Det är emellertid inte alltid självklart vad som är ett lexikalt fel och vad som är ett morfosyntaktiskt fel. I strikt mening är genusfel som *en land* lexikala, men de har ändå kodats som felaktiga eftersom de antagligen uppfattas som grammatiska fel. Det är dock den sista versionen av yttrandet som räknats (se Foster m. fl. 2000), varför (19) ändå kodats som korrekt. Felaktig användning av formord som prepositioner, konjunktioner och hjälpverb har också kodats som inkorrekt.

(17) BBU: < jag > [/] jag betyder att ...

(18) PBU: ... vad framgång menar .

(19) AAG äh om man bor i < en > [//] ett land där ...

Jag bedömer korrektheten snarare än lämpligheten (jmf. Housen, Kuiken & Vedder 2012, s. 4). I svenskan använder vi ofta nakna substantiv, så bara om ett substantiv är räknebart och har en uppenbar specifik referent har jag kodat frånvaro av artikel som inkorrekt; exempel (20) har följaktligen kodats som korrekt trots att bestämd form – *hälsan* – hade varit det idiomatiskt riktiga. Huvudsatsordföljd i hävdade bisatser och indirekta frågebisatser, som i (21), har kodats som korrekt eftersom denna ordföljd förekommer i naturligt svenskt talspråk. Olika typer av dubblingar, som det dubblade subjektet i (22), är vanliga i materialet och har kodats som korrekta. Överanvändning av hjälpverb i bisatser, som i (23), kodas också som korrekt eftersom användningen inte är omöjlig. Bruket av reflexiva pronomen är så pass vacklande i svenskan att jag inte heller kodat satser som (24) som inkorrekta. Neutrumböjning på adjektiv i predikativ ställning kodas också oftast som korrekt eftersom dessa så kallade pannkaksmeningar (*pannkakor är gott*), som i (25), är helt naturliga i svenskan. Oavslutade eller ofullständiga satser inkluderas och kodas om där finns tillräckligt med information för att jag ska kunna säga

om det är en huvudsats eller en bisats; i exempel (26) är ordet *barn* utelämnat, men satsen är ändå kodad som en korrekt huvudsats.

- (20) JCU: och det är bättre < för > [/] < för häl >
 [/] < för > [/] < för hälsa > [/] för hälsa .
- (21) EAG: att det beror på
 hur äh och vem skapar dom .
- (22) FAG: och dina föräldrar dom är rika människor .
- (23) IBU: när hon ska bli äh gammal .
- (24) OBG: att dom kan definiera framgång för dom själva också .
- (25) OBG: men är vänner mer viktigt för dig än familjerelationer ?
- (26) JCU: < så > [/] så < du > [/] du bestämde för ditt äh +//.

3.4.3 Inter-bedömarreliabilitet

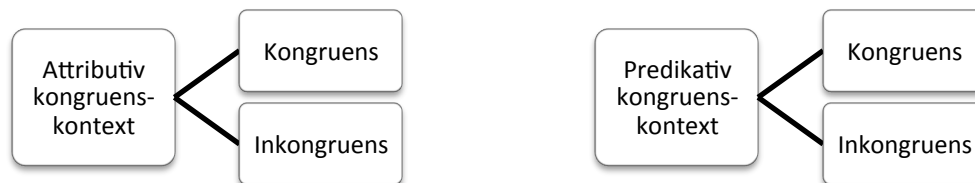
För att undersökningen ska kunna replikerat måste inter-bedömarreliabiliteten vara hög (Bachman 2004, s. 169). Jag testade inter-bedömarreliabiliteten genom att låta en kollega – en erfaren svensklärare och bedömare av andraspråkstalare – läsa detta metodavsnitt och sedan koda hundra satser, de tio första satserna från de tio första talarna, för satstyp och korrekthet. Med hjälp av Watkins och Pacheco (2001) beräknade jag Cohans kapp (k). Värdet för k kan variera från +1, som innebär total överensstämmelse till -1, som innebär totalt motsatt kodning. Resultatet av detta test blev för satstyp $k=0,91$ och för korrekthet $k=0,80$. Alla värden över 0,75 indikerar "excellent agreement" (Watkins och Pacheco 2001, s. 209).

3.4.4 Bestämning av PT-nivå

Vid PT-analysen har jag använt *The Emergence Criterion* (se avsnitt 2.4.2), framför allt därför att de relevanta kontexterna i materialet varit så få att ett 50-procentskriterium inte skulle säga någonting. *The Emergence Criterion* innebär dock inte att det räcker med endast ett fall av vad som ser ut som inversion eller predikativ kongruens för att testtagaren ska bedömas ha uppnått PT-nivå 4; testtagaren måste i stället ha uppvisat ett fall av systematisk användning med kontrasterande exempel.

Jag har inte excerperat obligatoriska BIFF-kontexter (icke-hävdade bisatser med satsadverbial) eftersom dessa – det märkte jag snabbt – var så få; således kan jag inte säga om några av testtagarna nått upp till nivå 5, men eftersom inte många av Tisustesttagarna i Eklund Heinonens (2009) studie nådde upp till nivå 5 såg jag det inte heller som troligt

att särskilt många Swedextesttagare skulle nå upp till denna nivå (Tisus testar C1-nivån; Swedex B2-nivån). För att avgöra om nivå 3 var uppnådd excerperade jag alla obligatoriska kontexter för attributiv kongruens. Hos några testtagare är dessa kontexter så få att jag inte kunnat fastställa huruvida nivå 3 uppnåtts. I dessa fall har jag också tittat på verbfraserna för att se om där finns några fall av systematisk växling mellan finita och infinita verbformer. På detta sätt har jag kommit fram till att alla tjugo testtagare uppnått nivå 3, även om de grammatiska processerna i många fall inte automatiserats. Frågan är alltså om är testtagarna börjat processa grammatisk information över frasgränserna, det vill säga, om de uppnått PT-nivå 4. För att en testtagare ska anses ha uppnått nivå 4 krävs tydliga fall av antingen predikativ kongruens eller inversion. I Figur 8 återges kodschemat för fastställandet av PT-nivå 3 och 4 (inversionskontexterna hade redan excerperats i komplexitets- och korrekthetsmätningarna).



Figur 8. Fastställande av PT-nivå – kodschema

Fastställandet av PT-nivå har gjorts genom en kvalitativ bedömning. Jag har, för varje testtagare, samlat alla "bevis" för och emot nivå 3 och 4 och vägt dessa mot varandra. I denna process har jag konsulterat en erfaren PT-forskare. Talaren LBU producerar till exempel flera meningar med korrekt pluralböjning på adjektivet *framgångsrik*, som i exempel (27), men när samma talare sedan felaktigt använder samma form, som i (28), förlorar de tidigare korrekta exemplen sitt värde; LBU tycks alltid använda ordet i dess pluralform. Med ett kvantitativt tillvägagångssätt, där jag räknar andelen kongruens av alla obligatoriska kontexter, hade jag inte gjort denna iakttagelse.

(27) LBU: ... dom är framgångsrika ...

(28) LBU: ... hon är inte framgångsrika ...

Vissa vanliga adjektiv, som *sann*, *svår*, *viktig* och *rolig*, står mycket ofta – hos alla testtagarna – i neutrum, predikativt efter *det är*, som i (29). Dessa satser kan mycket väl vara inlärd i sin helhet och utgör således inte bevis för att grammatisk information processas över frasgränser. Detsamma gäller vanliga adjektiv som används i sin grundform i exempel som (30). Den systematik som FAG uppvisar i exempel (31) är dessvärre

ovanlig i materialet; mindre tydliga exempel har fått användas som bevisning för att information processas på nivå 4.

- (29) HBG: ... det är mycket svårt för mig ...
- (30) BBU: ... man kan bli glad ...
- (31) FAG: ... utbildningen är inte så viktig ...
 ... det är viktigt ...
 ... dom är viktiga ...

Hos nio av de tjugo testtagarna finns endast ett eller ett par exempel på inversion, men dessa tas ofta inte som bevis på att talaren uppnått nivå 4, antingen därför att talaren uppvisar tvekan som i (32) och (33), eller därför att inversionen bara förekommer med vanliga verb-pronomen-kombinationer som *kan man*, som i (33) och (34); här kan hel-frasinlärningen inte heller uteslutas.

- (32) KBG: men nu så äh är jag här .
- (33) SBU: så för mig inom fem år äh kan man hjälpa ingenting .
- (34) HBG: också inom arbete kan man äh presentera sina personliga kunskaper .

3.5 Statistisk analys

Det statistiska verktyg som används är Pearsons korrelationskoefficient r (se Ezra Johnson 2013, s. 306; Bachman 2004, s. 85). Det är ett värde som säger i vilken utsträckning två variabler är associerade med varandra (korrelerar eller samvarierar), alltså kan förutsägas utifrån varandra. Koefficienten 1 står för en positiv och fullständig korrelation, vilket innebär att den ena variabeln ökar helt förutsägbart med den andra. Koefficienten -1 står för en negativ och fullständig korrelation, vilket också innebär att de två variablerna kan förutsägas utifrån varandra: om den ena variabeln ökar vet man att den andra minskar. Koefficienten 0 betyder att det inte finns någon korrelation mellan variablerna, det vill säga, värdet på den ena variabeln kan bara gissas utifrån värdet på den andra.

Pearsons r är ett mycket simpelt statistiskt verktyg som bara kan upptäcka linjära relationer. Därför använder jag också diagram där sambanden mellan variablerna visualiseras (se Ezra Johnson 2013, s. 304). Viktigt att nämna är att korrelationsstudier inte säger något om orsakssamband. Fulcher (2003, s. 203) skriver till exempel att det faktum att skrivfärdighet hos barn så tydligt samvarierar med längd givetvis inte innebär att skrivförmågan har något med längden att göra; de båda variablerna är i stället relaterade till en tredje variabel, i det här fallet ålder.

Huruvida en korrelation är statistiskt signifikant eller inte beror till stor del på hur många observationspunkterna är. I det här fallet är de ganska få: tjugo testtagare. Graden av signifikans (p) anges enligt följande: tre stjärnor (***) betyder att sannolikheten att den observerade korrelationen beror på slump är mindre än 0,001; två stjärnor (**) betyder att sannolikheten inte är större än 0,01; och en stjärna (*) betyder att sannolikheten inte är större än 0,05. I dessa fall talar man om signifikanta korrelationer ($p < 0,5$ =svag korrelation; $p < 0,01$ =medelstark korrelation; $p < 0,001$ = stark korrelation). Om sannolikheten att korrelationen beror på slump inte är större än en på tio ($p < 0,1$) säger man att resultatet är marginellt signifikant, och det markeras med en punkt (.).

De statistiska analyserna av har jag gjort i programmet R Studio (se Fox & Weisberg 2011). Som nämnts är bedömningarna C, B och A i statistiken representerade av värdena 0, 1 och 2. Betygen underkänd och godkänd representeras av värdena 0 och 1.

4 Resultat

I detta avsnitt presenteras resultaten av undersökningens kvantitativa del. I avsnitt 4.1 testas korrelationerna mellan samtliga åtta mått som mina mätningar resulterat i, och i avsnitt 4.2 testas korrelationerna mellan bedömningar och slutbetyg. Slutligen testas, i avsnitt 4.3, i vilken utsträckning måtten kan användas för att förutsäga bedömningar. De kvantitativa resultaten sammanfattas i avsnitt 4.4. I Tabell 19 i Bilaga 3 redovisas de absoluta tal för varje testtagare som måtten är beräknade utifrån, och i Tabell 20 i samma bilaga redovisas mätvärdena för varje testtagare.

4.1 Korrelationer mellan mått

I detta avsnitt testas för det första hur de fem komplexitetsmåtten korrelerar med varandra och för det andra hur de två korrekthetsmåtten korrelerar med varandra. För att fastställa enskilda måtts validitet är det viktigt att på detta sätt testa dem mot andra mått för samma konstrukt (Ellis & Barkhuizen 2005, s. 145). För det tredje testas hur dessa två typer av mått korrelerar med varandra, och för det fjärde hur korrekthetsmått och komplexitetsmått korrelerar med PT-nivå. Samtliga korrelationskoefficienter återfinns i Tabell 10. Diagram som visar signifikanta och marginellt signifikanta korrelationer mellan samtliga mått återfinns i Bilaga 4.1–4.4.

Tabell 10. Korrelationer mellan mätresultat

MÅTT*	1.	2.	3.	4.	5.	6.	7.	8.
1. AS-längd	1	,52*	,90***	,75***	,29	,51*	,72***	,53**
2. Satslängd	,52*	1	,22	,54*	-,25	,13	,32	-,01
3. BS-andel	,90***	,22	1	,64**	,42.	,60**	,69**	,58**
4. RS-andel	,75***	,54*	,64**	1	,12	,37	,54*	,42.
5. INV-kontextandel	,29	-,25	,42.	,12	1	,10	,30	,49*
6. Satskorrekthet	,51*	,13	,60**	,37	,10	1	,69**	,54*
7. INV-korrekthet	,72***	,32	,69**	,54*	,30	,69**	1	,61**
8. PT-nivå	,53*	-,01	,58**	,42.	,49*	,54*	,61**	1

*I den övre raden står, av utrymmesskäl, endast måttens nummer.

4.1.1 Korrelationer mellan komplexitetsmått

För det första ser vi hur komplexitetsmåten korrelerar med varandra. Den starka positiva korrelationerna mellan AS-längd och andelen bisatser ($r=,90$; $p<0,001$) och den svaga korrelationen mellan AS-längd och andelen relativsatser ($r=,75$; $p<0,05$) är båda väntade eftersom måtten delvis mäter samma sak; AS-enheten är ju definierad som en huvudsats med tillhörande bisatser, varför en talare med fler bisatser (inkl. relativsatser) också producerar längre AS-enheter. Samma sak gäller den medelstarka korrelationen mellan andelen bisatser och andelen relativsatser ($r=,64$; $p<0,01$); talare med många bisatser producerar uppenbarligen också generellt fler relativsatser. Mer intressant är då korrelationen mellan AS-längd och satslängd ($r=,52$; $p<0,05$) i kombinationen med den obefintliga korrelationen mellan satslängd och andelen bisatser ($r=,22$); talare kan uppenbarligen producera långa AS-enheter antingen genom att producera bisatser eller genom att bygga ut sina fraser på andra sätt, men det är inte självklart att den som gör det ena också gör det andra. Korrelationen mellan andelen relativsatser och satslängd ($r=,54$; $p<0,05$) tyder däremot på att den som bygger ut sina nominalfraser med relativsatser också tenderar att bygga ut sina fraser på andra sätt. Slutligen ser vi att andelen obligatoriska inversionskontexter inte korrelerar med andra komplexitetsmått; endast en marginellt signifikant korrelation mellan detta mått och andelen bisatser kan konstateras ($r=,42$; $p<0,1$).

4.1.2 Korrelation mellan korrekthetsmått

För det andra ser vi att korrelationen mellan det generella och det specifika korrekthetsmåttet är positiv och mycket signifikant ($r=,69$; $p<0,001$), vilket är väntat eftersom inversionskorrektheten utgör en delmängd av den generella korrektheten, det vill säga: varje sats med V2-brott är kodad som inkorrekt, varför en talare med många V2-brott inte bara får låg inversionskorrekthet utan antagligen också låg satskorrekthet.

4.1.3 Korrelationer mellan korrekthetsmått och komplexitetsmått

För det tredje ser vi hur komplexitetsmåten och korrekthetsmåten korrelerar med varandra. Ovan såg vi överlag positiva, signifikanta korrelationer, vilket var väntat eftersom det var olika mätningar av i grunden samma konstrukt som testades mot varandra. Här förväntar vi oss inga korrelationer eftersom det är relationen mellan två olika konstrukter som testas; i materialet kan det ju finnas talare med komplexitet-korrekthets-kombinationerna hög-hög, låg-låg, hög-låg och låg-hög. Så är dock inte fallet. Till att börja med kan vi konstatera att de flesta korrelationerna är positiva; några är neutrala; inga av dem är negativa. Den som har ett mer komplext språk tenderar alltså att ha ett mer korrekt språk och vice versa.

Det av korrekthetsmåten som korrelerar mest med komplexitetsmåten är det specifika utvecklingsmättet inversionskorrekthet; här ser vi korrelationer – av olika grad av signifikans – med AS-längd ($r=,72$; $p<0,001$), andelen bisatser ($r=,69$; $p<0,01$) och andelen relativsatser ($r=,54$; $p<0,05$). Det generella satskorrekthetsmättet korrelerar svagt med AS-längd ($r=,51$; $p<0,05$) och starkare med andelen bisatser ($r=,60$; $p<0,01$).

De komplexitetsmått som korrelerar mest med korrekthetsmåten är AS-längd och andelen bisatser. De talare som begår färre morfosyntaktiska fel producerar alltså generellt fler bisatser och därmed längre AS-enheter. Däremot är korrelationerna mellan satslängd och korrekthetsmåten inte signifikanta, och andelen relativsatser korrelerar signifikant endast med inversionskorrektheten ($r=,54$; $p<0,05$). Det komplexitetsmått som korrelerar minst med korrekthetsmåten är andelen obligatoriska inversionskontexter; som vi såg ovan korrelerar andelen inversionskontexter inte heller med andra komplexitetsmått.

4.1.4 Korrelationer mellan PT-nivå och korrekthets- och komplexitetsmått

För det fjärde ser vi att PT-nivå korrelerar positivt och signifikant med alla mått utom med satslängd, där korrelationskoefficienten är i det närmaste noll ($r=-0,01$), och andelen relativsatser, där korrelationen är marginellt signifikant ($r=,42$; $p<0,1$); dessa två mått har ovan visat sig korrelera ganska lite även med andra mått. Tydligast korrelerar PT-nivån med inversionskorrektheten ($r=,61$; $p<0,01$), vilket är väntat eftersom inversionskorrektheten använts – och i tre fall varit utslagsgivande – vid fastställandet av PT-nivå (se Tabell 21 i Bilaga 3). Mer intressant är den medelstarka korrelationen mellan PT-nivå och andelen bisatser ($r=,58$; $p<0,01$). Dessutom kan vi konstatera att PT-nivå är det enda mått som andelen inversionskontexter korrelerar signifikant, om än svagt, med ($r=,49$; $p<0,05$).

4.2 Korrelationer mellan bedömningar

I Tabell 11 ser vi hur bedömningarna för kriterium 1–4, sammanvägningen av dessa bedömningar samt slutbetyget för hela Swedex B2-test korrelerar med varandra. Generellt korrelerar bedömningarna för de fyra kriterierna tydligt med varandra; varje specifik bedömning korrelerar starkt ($p < 0,001$) med sammanvägningen. Den som bedöms ha uppnått en viss nivå under ett kriterium tenderar alltså att bedömas likadant under de andra kriterierna. Kriterium 3, som står i fokus här, korrelerar dock något svagare med sammanvägningen än de andra kriterierna. Däremot är det bara kriterium 2 och 3 som korrelerar signifikant med slutbetyget. De kompetenser som bedöms under kriterium 3 tycks alltså vara viktiga för att klara hela Swedex B2-test. Vad som bedöms under kriterium 3 ska vi se i nästa avsnitt.

Tabell 11. Korrelationer mellan bedömningar

BETYG	1. Flyt & interaktion	2. Kommunikativ design	3. Formell korr. & komplexitet	4. Uttal & intonation	Sammanvägning	Slutbetyg
1. Flyt & interaktion	1	,59**	,33	,54*	,76***	,01
2. Kom. design	,59**	1	,57**	,68***	,91***	,59*
3. For. korr. & komp.	,33	,57**	1	,54*	,74***	,54*
4. Uttal & intonation	,54*	,68***	,54*	1	,82***	,29
Sammanvägning	,76***	,91***	,74***	,82***	1	,45*
Slutbetyg	,01	,59*	,54*	,29	,45*	1

4.3 Korrelationer mellan mått och bedömningar

I detta avsnitt testas i vilken utsträckning de olika måtten – korrekthetsmått, komplexitetsmått och PT-nivå – korrelerar med bedömningarna av den muntliga performansen och med slutbetyget för hela Swedex B2-test. Korrelationskoefficienterna redovisas i Tabell 12. Diagram som visar signifikanta och marginellt signifikanta korrelationer mellan komplexitets- och korrekthetsmåtten å ena sidan och bedömningen under kriterium 3 och slutbetyget å andra sidan återfinns i Bilaga 4.5–4.6. Relationen mellan PT-nivå och bedömningar/slutbetyg visualiseras inte i diagram, utan i Tabell 21 i Bilaga 3.

Vi fokuserar på kriterium 3. Av komplexitetsmåtten är det bara ett – AS-längd – som korrelerar signifikant med kriterium 3 ($r = ,49$; $p < 0,05$). Korrelationen med andelen bisatser är marginell ($r = ,43$; $p < 0,1$). Båda korrekthetsmåtten korrelerar positivt och signifikant med bedömningen under kriterium 3, men det korrekthetsmått som korrelerar tydligast med komplexitetsmåtten – inversionskorrekthet – korrelerar mindre med bedömningen ($r = ,49$; $p < 0,05$) än vad den generella korrektheten gör ($r = ,59$; $p < 0,01$). Det tyckas alltså vara framför allt generell korrekthet som bedöms under kriterium 3.

Som framgår av låddiagrammen i Bilaga 4.5 kan de tre mått som korrelerar signifikant med kriterium 3 – AS-längd, satskorrekthet och inversionskorrekthet – användas för att skilja de två testtagare som fått C från de fyra som fått A; här finns ingen överlappning. Inom den stora grupp som fått B – fjorton testtagare – är spridningen däremot stor; gränserna mellan C och B och mellan B och A är i samtliga fall mycket diffusa, om de alls finns.

Tabell 12. Korrelationer mellan bedömningar och mätresultat

	1. Interaktion & flyt	2. Kommunikativ design	3. Formell korrekthet & komplexitet	4. Uttal & intonation	Sammanvägning	Slutbetyg
1. AS-längd	,27	,40.	,49*	,57**	,51*	,46*
2. S-längd	,45*	,39.	,18	,50*	,46*	,40.
3. BS-andel	,23	,37	,43.	,45*	,44*	,33
4. RS-andel	,46*	,26	,35	,48*	,45*	,31
5. INV-andel	,07	,04	-,08	-,04	0	,06
6. Satskorr.	,23	,61**	,59**	,35	,57**	,48*
7. INV-korr.	,17	,43.	,49*	,43.	,46*	,49*
8. PT-nivå	,18	,18	,54*	,29	,35	,19

Det generella korrekthetsmåttet är också det mått som korrelerar starkast med den sammanvägda bedömningen för det muntliga testet ($r=,57$; $p<0,01$), men även komplexitetsmåten – alla utom andelen inversionskontexter – korrelerar signifikant med denna sammanvägning, trots att de flesta komplexitetsmåten alltså inte korrelerar signifikant med kriterium 3. Förklaringen till detta är att komplexitetsmåten korrelerar med andra kriterier än kriterium 3, framför allt med kriterium 4, Uttal och intonation. Detta kan bero antingen på att det som mina komplexitetsmått mäter kommer hand i hand med intonation och uttal (precis som läsförmåga kommer hand i hand med ökad längd; se avsnitt 3.5) eller på att bedömarna uppfattar komplexiteten som positiv men inte kan sätta fingret på vad det är som är positivt. I det senare fallet skulle vi kunna tala om en halo-effekt (se avsnitt 2.2.3).

Kriterium 2 och 3 korrelerar starkt med varandra och med generell korrekthet. Dessa två kriterier korrelerar också med slutbetyget. Detta talar för att ordkunskap (som ju ska bedömas under kriterium 2) och grammatisk korrekthet, som är vad som verkar bedömas under kriterium 3, dels hänger ihop, dels är viktiga för att klara hela Swedex B2-test.

Slutbetyget för hela Swedex B2-test kan förutsägas utifrån de båda korrekthetsmåten – satskorrekthet ($r=,48$; $p<0,05$) och inversionskorrekthet ($r=,49$; $p<0,05$) – och utifrån AS-längd ($r=,46$; $p<0,05$), samt marginellt utifrån satslängd ($r=,40$; $p<0,1$). Att satslängden,

som annars korrelerar relativt lite med övriga mått och inte heller med bedömningen under kriterium 3, uppvisar en svag korrelation med slutbetyget kan tyda på att den kompetens som krävs för att bygga ut fraserna kommer mer till nytta i de skriftliga delarna av provet än i den muntliga delen.

PT-nivå, som ju ska säga något allmänt om testtagarnas morfosyntaktiska utveckling, kan användas endast för att förutsäga kriterium 3 ($r=.54$; $p<0,05$), vilket talar för att det som bedöms under detta kriterium är just morfosyntaktiska utvecklingen. PT-nivån är ett av endast två mått som inte uppvisar någon signifikant korrelation med den sammanvägda bedömningen ($r=.35$), och PT-nivån samvarierar inte heller med slutbetyget ($r=.19$).

Som framgår av Tabell 21 i Bilaga 3 har de två testtagare som fått C på kriterium 3 inte uppnått PT-nivå 4 medan de fyra som fått A alla gjort det. Hos FAG finns visserligen inget säkert fall av inversion, men den predikativa kongruensen är tydlig, som framgick av exempel (31) ovan. Av de 14 testtagare som fått B på kriterium 3 har precis hälften nått upp till nivå 4. PT-nivå kan alltså – precis som AS-längd, satskorrekthet och inversionskorrekthet, som vi såg ovan – användas för att särskilja dem som fått C från dem som fått A. Bedömningen B framstår däremot som ett passivt, slumpmässigt icke-beslut.

4.4 Sammanfattning av den kvantitativa studien

Alla statistiskt signifikanta eller marginellt signifikanta korrelationer mellan mått och mått, bedömningar och bedömningar och mått och bedömningar är sammanfattade i Diagram 1. En bra bild av resultaten ges också i Diagram 2–6 i avsnitt 5.1. Sambanden är många och komplexa och vi ska komma ihåg att en "statistiskt signifikant" korrelation mellan två variabler mycket väl kan bero på slump; det rör sig trots allt bara om tjugo testtagare.

Vi har sett att framför allt generell korrekthet bedöms under kriterium 3, men komplexitetsmått har trots detta en positiv effekt på sammanvägningsbetyget. Vi kan möjligen tala om en halo-effekt. Samtidigt korrelerar PT-nivå med kriterium 3 men inte med sammanvägningsbetyget, vilket kan tolkas som att det som bedöms under detta kriterium är just morfosyntaktisk utveckling.

Det som bedöms under kriterium 3 – oavsett om det nu är generell korrekthet eller morfosyntaktisk utveckling – tycks viktigt för att klara hela testet. Genomsnittlig satslängd, som korrelerar ganska lite med andra mått, är ett av två komplexitetsmått som också samvarierar med slutbetyget, vilket jag tolkar som att den kunskap som krävs för att bygga ut språket inom fraserna kommer mer till nytta i Swedextestets skriftliga delar än i den muntliga delen.

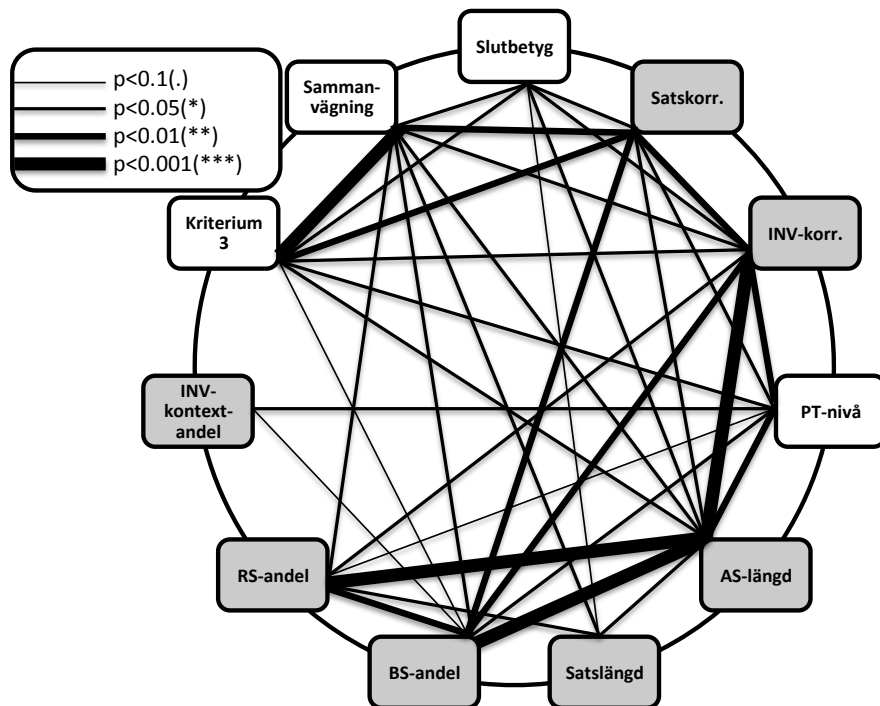


Diagram 1. Sammanfattning av den kvantitativa undersökningen

Korrelationen mellan satslängd och andelen relativsatser tyder på att den som bygger ut sina fraser med relativsatser också bygger ut dem på andra sätt. AS-enheternas längd varierar med andelen bisatser och med satslängd, men satslängd och andelen bisatser korrelerar inte med varandra. Den som bygger ut sitt språk inom fraserna behöver alltså inte nödvändigtvis bygga ut sitt språk med bisatser. I Diagram 4 och 5 i avsnitt 5.1 ser vi till exempel att testtagaren AAG har en låg andel bisatser men en hög genomsnittlig satslängd; kanske har denna avancerade inlärare kommit till det tredje stadiet som Norris och Ortega (2009) skriver om, där språket utvecklas genom grammatiska metaforer i stället för med hypotax (se avsnitt 2.3.2).

Andelen inversionskontexter, som brukar räknas till nivå 3 i PT-litteraturen, är det mått som korrelerar minst med andra mått och bedömningar. Det korrelerar signifikant endast med PT-nivå, och marginellt signifikant med andelen bisatser.

5 Närstudie av fyra testtagare

I föregående avsnitt presenterades uppsatsens kvantitativa resultat. Där såg vi vilka generella relationer som tycks finnas mellan komplexitet och korrekthet, PT-nivå samt bedömningar i ett muntligt språktest. I detta avsnitt studeras fyra av de tjugo testtagarna närmare, två som bedömts uppnå nivå A och två som bedömts uppnå nivå C under kriterium 3. Den ena av dem som fått A och den ena av dem som C borde dock, enligt de

flesta av mina mätningar, snarare fått B, och syftet med denna närstudie är att svara på uppsatsens andra fråga: (Hur) kan till synes avvikande bedömningar förklaras? I avsnitt 5.1 redogör jag för hur de fyra testtagarna valts ut för denna närstudie och i 5.2 beskrivs dessa testtagares deltagande i testet i termer av pratighet (*talkativness*) och interaktion. I avsnitt 5.3 studeras korrektheten och komplexiteten hos de fyra testtagarna närmare på en rad olika sätt. Resultaten av dessa undersökningar sammanfattas i 5.4.

5.1 Urval av testtagare för närstudien

I syfte att välja ut testtagare till denna närstudie skapade jag fem så kallade sambandsdiagram där testtagarna är utplacerade med sina kodnamn i koordinatsystem med komplexitetsmått på x-axeln och korrekthetsmått på y-axeln. Satskorrektheten kombineras med AS-längd i Diagram 2 och med andelen relativsatser i Diagram 3; inversionskorrektheten kombineras med satslängd i Diagram 4, med andelen bisatser i Diagram 5 och med andelen obligatoriska inversionskontexter i Diagram 6. På så sätt använder jag samtliga komplexitets- och korrekthetsmått och visar både signifikanta och icke-signifikanta korrelationer.

Jag vill här påminna om att testtagarnas kod berättar hur kriterium 3 bedömdes (den andra bokstaven) och om hela testet blev godkänt eller underkänt (tredje bokstaven). I dessa diagram har jag dessutom lagt till PT-nivån testtagaren uppnått i denna kod. På så vis kan diagrammen sägas vara fyrdimensionella: för varje testtagare framgår komplexitet, korrekthet, PT-nivå och betyg. De fyra testtagare som valts ut för denna närstudie är markerade.

Vi ser att spridningen bland de 14 testtagare som fått B på kriterium 3 är stor, både beträffande korrekthet, komplexitet och PT-nivå. Några av dem, till exempel OBG och QBG, befinner sig ofta högt upp till höger, det vill säga deras värden för såväl korrekthet som komplexitet är höga. De har också uppnått PT-nivå 4. Andra – framför allt BBU, DBU, SBU och RBG – är ofta placerade långt ner till vänster, det vill säga deras värden för såväl korrekthet som komplexitet är låga. BBU har uppnått PT-nivå 4, men inte de tre andra. Självklart skulle det vara intressant att titta närmare på dessa testtagare, men jag fokuserar här på dem som fått C eller A på kriterium 3; dessa är mer intressanta eftersom bedömningarna A och C verkar utgöra mer aktiva beslut från bedömarnas sida.

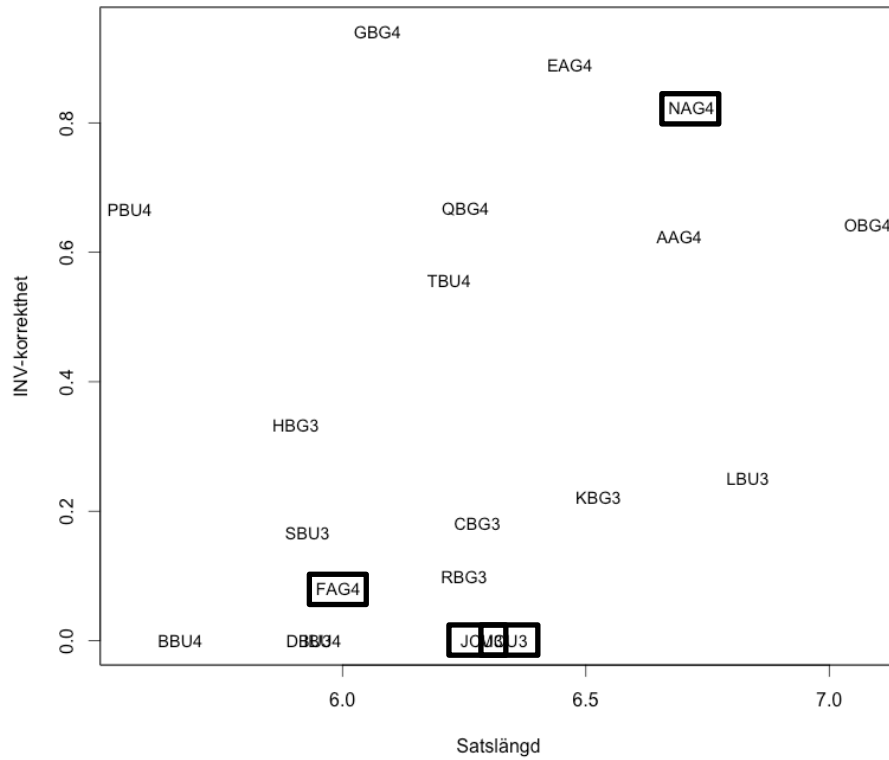


Diagram 4. Samband mellan inversionskorrekthet och satslängd ($r=,37$)

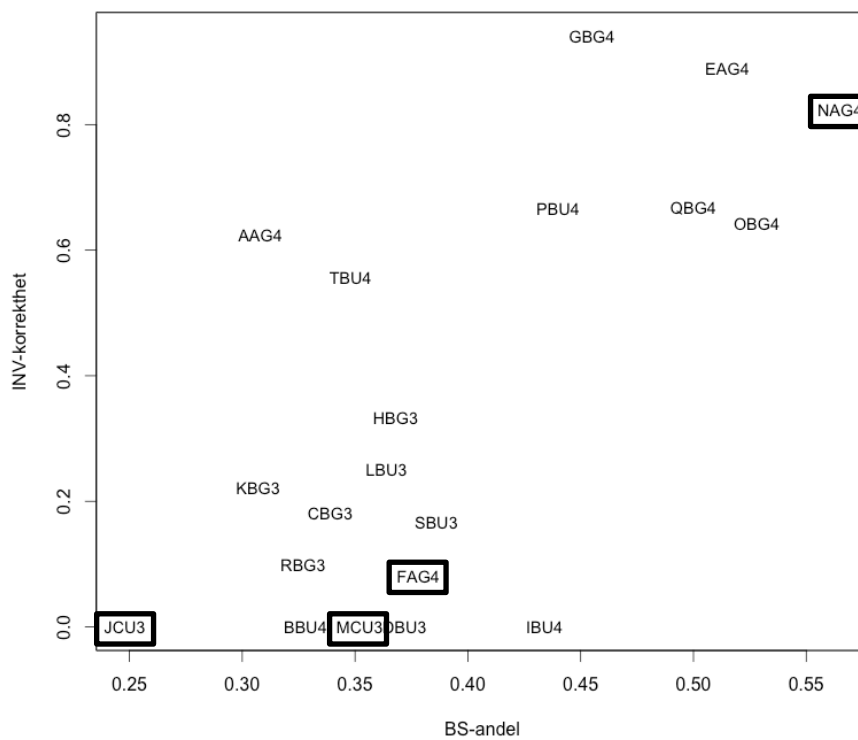


Diagram 5. Samband mellan inversionskorrekthet och andelen bisatser ($r=,69^{**}$)

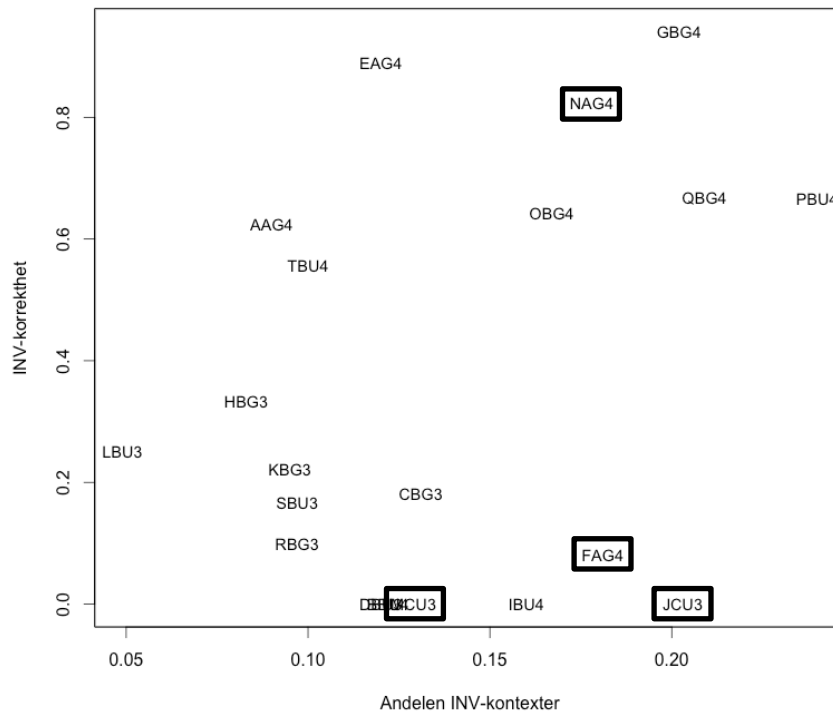


Diagram 6. Samband mellan inversionskorrekthet och andelen inversionskontexter ($r=,30$)

Av diagrammen framgår att av de fyra testtagarna som fått A på kriterium 3 har EAG och NAG (nästan) alltid höga korrekthetsvärden och komplexitetsvärden (NAG har inte så hög andel inversionskontexter, se diagram 6). EAG har för det mesta något högre korrekthet och något lägre komplexitet än NAG, även om EAG producerar fler relativsatser. NAG får ingå i närstudien som en testtagare med väntat bra betyg. De två andra som fått A, AAG och FAG, placerar sig närmare mitten av diagrammen, bland dem som fått B. AAG har visserligen höga korrekthetsvärden medan FAG har låga. Båda har låga komplexitetsvärden även om FAG har en hög andel obligatoriska inversionskontexter och AAG har en hög genomsnittlig satslängd. FAG får ingå i närstudien som en testtagare med oväntat bra betyg.

Av de två som fått C på kriterium 3 är det JCU som urskiljer sig som testtagaren med lägst korrekthets- och komplexitetsvärden, även om andel obligatoriska inversionskontexter är mycket hög. JCU får därmed ingå i närstudien som en testtagare med väntat lågt betyg. MCU har ingen inversion, men den generella satskorrekthet, det vill säga det mått som korrelerade tydligast med såväl bedömning som med slutbetyg, är genomsnittlig. MCU är också den testtagare med kortast AS-längd och med näst lägst andel relativsatser, men andelen bisatser – också det ett mått som visat sig viktigt – är genomsnittlig. MCU

placeras sig ofta bland dem som fått B, varför MCU får ingå i närstudien som en testtagare med oväntat lågt betyg.

MCU och FAG är placerade mer eller mindre nära varandra i alla fem diagrammen. I denna närstudie försöker jag förstå varför MCU och FAG bedömts ha så olika grad av korrekthet och komplexitet trots att deras performans, enligt mina mätningar, inte skiljer sig mycket åt. På vilket sätt liknar FAG NAG och MCU JCU?

5.2 Testtagarnas pratighet och delaktighet

Utän att ha undersökt interaktionen kan jag försäkra att de fyra testtagarna deltar i "testspelet" (se Sandlund & Sundqvist 2014); de är alla engagerade i såväl monologer som diskussioner. De gör också alla ett öppet och sympatiskt intryck. MCU och – framför allt – JCU talar dock långsammare än NAG och FAG varför de, som framgår av Diagram 7, får mindre sagt. Sett till antalet token är FAG den som talar mest, men NAG producerar något fler unika ord (type), se Diagram 8. Eftersom de talar olika mycket säger inte type/token-proportionen så mycket, ett mått som annars används mycket inom CAF-forskningen.

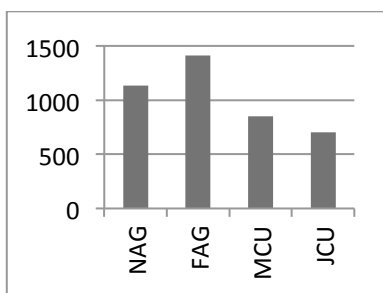


Diagram 7. Token

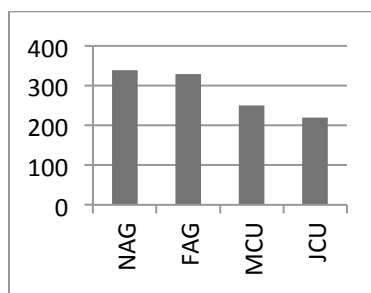


Diagram 8. Type

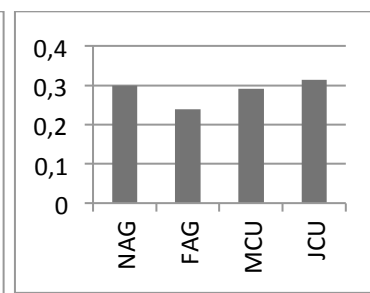


Diagram 9. Type/token

Inom CAF-forskningen ses ordförrådets bredd och djup som en del av konstrukten komplexitet (Bulté & Housen 2012), och även om ordförrådet inte ska bedömas under kriterium 3, utan under kriterium 2, kan det inte uteslutas att det påverkat bedömarens uppfattning av MCU:s och FAG:s grammatiska behärskning; så kallade halo-effekter är, som nämnts, väldokumenterade inom testforskningen (Fulcher 2003, s. 90; Eckes 2009, s. 5; Tonkyn 2012, s. 238). Ordförrådets storlek tenderar också att korrelera med den övriga språkutvecklingen och kan således säga något om den grammatiska förmågan som mina mätningar inte fångat; ett rikt ordförråd tycks vara en av de bästa indikatorerna på generell hög språkfärdighet (se t.ex. Hulstijn m. fl. 2012; Iwashita m.fl. 2008; Tonkyn 2012). Resten av denna närstudie handlar dock fortfarande om det som ska bedömas under kriterium 3: Formell korrekthet och komplexitet.

5.3 Komplexitet och korrekthet hos fyra testtagare

I detta avsnitt studeras de fyra testtagarnas morfosyntaktiska komplexitet och korrekthet närmare. I avsnitt 5.3.1 försöker jag förklara varför MCU och FAG bedömts så olika utifrån deras PT-nivå. Alla fyra har i någon mån tillägnat sig attributiv kongruens och alltså uppnått PT-nivå 3, men i avsnitt 5.3.2 tittar vi närmare på deras kunskaper om nominalfrasens interna struktur och deras förmåga att bygga ut nominalfraserna. I avsnitt 5.3.3 studeras bisatserna närmare, i 5.3.4 infinitivfraserna och i 5.3.5 verbfraserna, både i termer av komplexitet och korrekthet. I 5.3.6 studeras slutligen de fyra testtagarnas behärskning av två fenomen som är relaterade till finithet: obligatoriska subjekt och verb.

5.3.1 PT-nivå

Som framgår redan i sambandsdiagrammen i avsnitt 5.1 utgör PT-nivån en skillnad mellan FAG och MCU. FAG har uppnått PT-nivå 4, precis som de tre andra som fått A, medan MCU, precis som JCU, är på nivå 3. Detta kan dock inte i sig förklara skillnaden i bedömning mellan FAG och MCU; bland de testtagare som fått B finns ju både de som uppnått nivå 4 och de som inte gjort det. Att bestämma PT-nivå var heller inte någon enkel sak, som framgick i avsnitt 3.4.4. Jag har därför gått tillbaka till de strukturer jag excerperade för att fastställa PT-nivå och tittat närmare på hur de fyra testtagarnas PT-nivå fastställdes.

Minst ett av två kriterier skulle vara uppfyllt för att en testtagare skulle anses ha uppnått nivå 4: ett fall av antingen systematisk inversion eller systematisk predikativ kongruens. NAG uppfyller med marginalt kravet på inversion och med tveksamhet kravet på predikativ kongruens. FAG uppfyller inte kravet på inversion: här finns 23 fall av V2-brott och bara två fall av inversion; i båda dessa fall sker inversionen mellan vanliga hjälpverb och subjektpronomen, varför helfrasinlärning inte kan uteslutas. Däremot uppfylls kravet på kongruens kanske tydligast av alla tjugo testtagare, med exempel som (31), här upprepat i (35).

- (35) FAG: ... utbildningen är inte så viktig ...
 ... det är viktigt ...
 ... dom är viktiga ...

Varken MCU eller JCU uppfyller något av kraven för nivå 4. Hos JCU finns inget fall av inversion och bara tre kontexter för predikativ kongruens; trots att dessa kongruerar har det inte räckt för att jag ska betrakta det som ett fall av systematisk användning. Hos MCU finns ett fall av inversion, också mellan ett hjälpverb och ett subjektspronomen. Här finns fem adjektiv som står i predikativ ställning: *gammal*, *hungrig* och *törstig* kongruerar med *jag* eller *du* medan *jätteviktigt* och *tråkigt* kongruerar med *det*. Detta har inte heller räckt

för att tala om systematisk variation; det är troligt att dessa ord aldrig förekommer i andra former hos denna inlärare.

Sammanfattningsvis avslöjar de strukturer som används för att fastställa PT-nivå 4 – predikativ kongruens och inversion – skillnader mellan FAG och MCU som övriga mått i den kvantitativa undersökningen inte avslöjade. Vi tittar mer på testtagarnas kunskap om nominalfrasens interna struktur i nästa avsnitt.

5.3.2 Nominalfraser

När jag fastställde PT-nivå utifrån alla kontexter för attributiv kongruens var jag osäker på om MCU och JCU ens nådde upp till nivå 3; det var genom att studera verbfraserna som nivå 3 ändå kunde fastställas. Här tittar vi närmare på nominalfraserna, som enligt Axelson (1994, s. 99) utvecklas i tre steg, med komplexa nominalfraser med framförställda attribut och dubbel bestämdhet i tredje och sista steget. Alla lexikala nominalfraser som de fyra testtagarna producerade har excerperats.

Vi såg i den kvantitativa studien att FAG, precis som MCU och JCU, producerar ganska få relativsatser, men här ser vi ändå att de är såväl fler och längre som mer korrekta än de relativsatser som MCU och JCU producerar. Överlag är det tydligt att FAG bygger ut sina nominalfraser mer än MCU, både med framförställda och efterställda attribut. Exempel på några av testtagarnas mest komplexa nominalfraser ges i (36)–(39).

(36) NAG: ... den människa alltid som James Bond kämpa emot som hade en stort vattenreservoar xxx där han höll vatten tillbaka ...

(37) FAG ... den här radion som har inte för mycket reklam att [som] till exempel P1 .

(38) MCU: ... ett år här i Sverige ...

(39) JCU: ... lite problem med < blod > [//] höj blodtryck och hjärtinfarkt ...

Vi tittar närmare på testtagarnas behärskning av nominalfrasens interna struktur. Exempel som (40), med både pluralböjning och dubbel bestämdhet, visar att behärsningen hos NAG är god; här finns bara ett fall av attributiv inkongruens, se exempel (41). Även FAG uppvisar, med exempel som i (42), stora kunskaper, men FAG begår också misstag i betydligt högre utsträckning än NAG, se (43); processerna tycks inte lika automatiserade. De flesta fel FAG begår har att göra med framförställda bestämda artiklar eller demonstrativa pronomen. MCU producerar mest korta nominalfraser med betydligt färre kontexter för attributiv kongruens än både NAG och FAG, och begår därmed inte lika många fel som FAG. Här finns några få enkla korrekta exempel som i (44), men inga exempel på plural-kongruens eller bestämdhetskongruens. Genusfel och kongruensfel som i (45) är vanliga.

Även JCU producerar få kontexter för attributiv kongruens. Här finns ett antal genusfel, se (46), men exemplen i (47) tyder på kunskaper om pluralkongruens.

- (40) NAG: ... såna länder som ligger i den äh arabiska delen av världen ...
- (41) NAG: ... denna svenska samhället ...
- (42) FAG: ... kvinnliga kläder ...
... fackliga ämnen ...
... en annan punkt ...
... den andra punkten ...
- (43) FAG: ... den här kompisarna ...
... ett olyckliga < barn > [/] barn ...
... min föräldrar ...
... den här reklam ...
- (44) MCU: ... en het dag ...
- (45) MCU: ... några riktigt vän ...
... första dag här i Sverige ...
... en annat land ...
- (46) JCU: ... en stor äh äh företag ...
... mycket gott äh mat ...
- (47) JCU: ... rika människor ...
... gamla < roman > [/] romaner ... [försöker säga *romare*]

Även i detta avsnitt har vi sett skillnader mellan FAG och MCU som den kvantitativa undersökningen inte visade. FAG producerar mer komplexa nominalfraser och uppvisar större kunskaper om nominalfrasens interna struktur än MCU. FAG begår också många fel, men dessa kan relateras till de mer komplexa strukturerna.

5.3.3 Underordning

I det här avsnittet undersöks bisatsernas komplexitet och korrekthet hos de fyra testtagarna. Andraspråksinlärare brukar lära sig adverbiala bisatser före nominala och nominala före relativa (Flyman Mattsson & Håkansson 2010, s. 59), varför jag börjar med att undersöka vilka typer av bisatser som används. I Tabell 13 anges vilka bisatsinledare varje talare använder minst en gång. De fetmarkerade används inkorrekt.

Tabell 13. Bisatsinledare som används av NAG, FAG, MCU och JCU

BISATS- INLEDARE*	NAG	FAG	MCU	JCU
Adverbiella	<i>eftersom</i> <i>så att</i> <i>för att</i> <i>när</i> <i>efter</i> <i>om</i> <i>ju mer som</i> <i>även om</i> <i>som (än)</i>	<i>(där)för att</i> <i>eftersom</i> <i>(efter)som</i> <i>när</i> <i>om</i>	<i>(där)för att</i> <i>eftersom att</i> <i>(efter)som</i> <i>när</i> <i>efter</i> <i>innan</i> <i>om</i>	<i>därför (att)</i> <i>om</i> <i>än (om)</i> <i>varför</i> <i>när</i>
Nominala	<i>att</i> <i>(att)</i> <i>om</i> <i>hur</i> <i>vad</i>	<i>att</i> <i>(att)</i> <i>om</i> <i>det (att)</i> <i>vad</i>	<i>att</i> <i>(att)</i> <i>om</i> <i>vad</i> <i>var</i>	<i>(att)</i> <i>vad</i> <i>om (att)</i> <i>av (vad)</i>
Relativa	<i>som</i> <i>(som)</i> <i>där</i>	<i>som</i> <i>att (som)</i>	<i>(som)</i> <i>vem</i>	<i>att (som)</i>

* Material som inte uttalas står inom parentes. Fetmarkering betyder att bisatsinledaren använts inkorrekt, eller, om den står inom parentes, att utelämnningen är inkorrekt.

Vi ser att NAG inleder sina adverbiella bisatser med fler och mer raffinerade underordnare än de andra. Vi ser också att JCU aldrig använder *att* som inledare av nominal bisats; JCU använder *att* som infinitivmärke (men utelämnar oftast infinitivmärket) och en gång som relativpronomen (se nedan). I stället för *att* används *om*, som i (48), och i stället för *om* används *än*, se (49). Oftast utelämnas bisatsinledaren *att*, även när det är ogrammatiskt som i (50). FAG använder vid ett tillfälle *det* i stället för *att* och *att* i stället för *som*, men uppvisar inga systematiska problem som JCU. MCU uppvisar inga sådana problem.

(48) JCU: gamla romaner [romare] sade om mat dödar mer än svärd [upprepningar och självrättningar utelämnade]

(49) JCU: ... varför sjukdoms ähm kommer # < senare > senare än [om] äh du äter äh < en > [/] < en bra > [/] en bra mat .

(50) JCU: så ähm vi måste respektera ähm människor kommer < från > [//] ifrån < länder > [//] äh äh olika länder .

Tittar vi i stället på vilka underordnare som inleder relativbisatser ser vi en likhet mellan MCU och JCU, som skiljer dem från FAG och NAG: varken MCU eller JCU producerar relativsatser inledda av *som*. Vi har redan tidigare sett att FAG och MCU ligger nära varandra när det gäller andelen relativsatser; NAG producerar många relativsatser – 16 stycken – medan FAG och MCU producerar få – sju respektive tre stycken. JCU producerar bara en

relativsats. Andelen relativsatsers tycks alltså inte kunna förklara varför FAG och MCU bedömts så olika. Men – både NAG och FAG behärskar relativsatsens struktur (även om FAG alltså inte använder den så mycket) medan varken MCU eller FAG producerar en enda korrekt relativsats. NAG uppvisar större kunskaper än de andra med korrekt utelämnning av *som* och ett relativt *där*. FAG:s sju relativsatsers är alla utom en i grunden korrekta, som i exempel (51). MCU:s tre relativsatsers är alla felaktiga; *som* utelämnas felaktigt i ett fall, se (52), och *vem* används i stället för *den som* i två fall, se (53). JCU använder *att* i stället för *som* för att inleda sin enda relativsats, se exempel (54). Alla fyra behärskar alltså adverbiala bisatser, men JCU behärskar inte de nominala och varken JCU eller MCU behärskar de relativa.

(51) FAG: ... dom männen som vanligen går till gym .

(52) MCU: den är min vän ska vara min vän också efter . [den som är min vän ...]

(53) MCU: vem är i våran ålder ... [den som är i våran ålder ...]

(54) JCU: nummer att styr dig [nummer som styr dig=*bäst-före-datum*]

Jag har även undersökt bisatsernas "grad" eller "djup". En talare som inleder många yttranden med *jag tror att* får ett ganska högt värde för andelen bisatser trots att språket knappast upplevs som komplext. Talare kan också samordna många bisatser, vilket vi sett exempel på tidigare (se avsnitt 3.4.1), något som vid en enkel bisatsräkning ger samma komplexitetsvärde som bisatser inkorporerade i bisatser. Jag har därför excerperat alla andragsbissatser – det vill säga bisats i bisats –, tredjegradsbissatser (bisats i bisats i bisats) och så vidare.

NAG producerar en stor mängd andra- och tredjegradsbissatser – 20 respektive sex stycken – samt två fjärdegradsbissatser och utklassar därmed de övriga. Ofta är språket raffinerat, som i (55). FAG producerar en del långa men inte så djupa strukturer, som i (56); här finns 14 andragsbissatser men inga tredje- eller fjärdegradsbissatser. MCU producerar med viss möda tre andragsbissatser, som i (57), och tyckas därmed ha ett mindre komplext språk än både FAG och NAG. JCU producerar fyra andragsbissatser och en tredjegradsbissats, men det är uppenbarligen över den egna förmågan, se exempel (58). Som nämnts ovan har både FAG och MCU genomsnittliga värden för andelen bisatser, men denna undersökning visar alltså att FAG bäddar in sina bisatser mer än MCU.

(55) NAG: men jag tror därför det är också viktigt att man lär sig språket så att man förstår lite mer vad dom tycker om att prata om .

(56) FAG: om jag har kompisar och den här kompisarna dom är rika människor och till exempel < dom > [/] dom fattar beslut att vi ska åka till en fest och nä men < du > [/] du kan inte du är fattig person du kan inte följa med < du känner sig > [//] du känner dig < besvär > [/] besvär .

(57) MCU: om jag bestämmer [att] jag ger till dig [...] det är helt okej .

(58) JCU: därför äh forskarna tänker < en god mat äh > [//] om en man äh äter < en god > [/] en god mat kan leva < mycket > [/] ähm mycket bra varför sjukdoms ähm kommer < # senare > [/] senare än äh du äter äh < en > [/] < en bra > [/] en bra mat .

Avslutningsvis ska vi se hur svenskans bisatsordföljd behärskas. Både NAG och FAG visar att de behärskar bisatsordföljd i indirekta frågor, som i (59) och (60), och hos NAG finns dessutom ett fall av pre-verbal negation i en icke-hävdad bisats, se exempel (61); det kan tala för att NAG uppnått PT-nivå 5. Både MCU och JCU producerar några indirekta frågor, men medan JCU använder bisatsordföljd, som i (62), använder MCU alltid huvudsatsordföljd, som i (63). Här framstår alltså MCU som den med minst utvecklad grammatik medan NAG återigen uppvisar större språklig kompetens.

(59) FAG: ... det beror på ja vad dom har också .

(60) NAG: ... man förstår lite mer vad dom tycker om att prata om .

(61) NAG: ... eftersom dom inte får så mycket vatten < som > [/] som det behövs ...

(62) JCU: eller är bättre barn # känner själv vad äh han < vill få > [//] vill göra ?

(63) MCU: du vill inte berätta vad gör ont för dig .

5.3.4 Infinitivfraser

Som exempel på en specifik struktur att studera nämner Norris och Ortega (2009, s. 560) infinitivfrasen; denna satsförkortning kan ju ibland ersätta bisatsen hos inlärare som kommit långt i språkutvecklingen. Alla fyra testtagare producerar infinitivfraser: NAG 11, FAG 20, MCU 8 och JCU 15. Alla utom JCU behärskar strukturen, även om MCU inte placerar satsadverbialet rätt i (64). Det är ibland svårt att avgöra om JCU producerar infinitivfraser med presensform på verbet eller bisatser med utelämnat subjekt, se exempel (65), och dessa fraser/satser är ofta korta, även om just exempel (65) är utbyggt med en infinitivfras i infinitivfrasen. MCU producerar också korta infinitivfraser medan FAG är den som producerar de längsta och mest komplexa infinitivfraserna, se (66). Även NAG producerar ett par mer utbyggda infinitivfraser, se (67). Återigen ser vi alltså att MCU har ett enkelt språk, men ganska korrekt, medan JCU har ett komplext språk men ganska

felaktigt. Även här ser vi skillnader mellan MCU och FAG som den kvantitativa undersökningen inte fångade. MCU framstår mer som en risktagare.

(64) MCU: ... att va alltid mellan folk ...

(65) JCU: ... att studera mycket för att öh i framtiden jobbar och tjänar mycket pengar .

(66) FAG: ... att uppmuntra dom att anställa mig .
... att kolla om kvaliteten av den här produkten är bra .
... att välja den här radion som har inte för mycket reklam att [som] till exempel P1

(67) NAG: ... att ge vatten tillbaka till dom som behövs .

5.3.5 Verbfraser

Enligt Tonkyn (2012) är verbfrasernas komplexitet och korrekthet en bra indikator på språkutveckling, även om Fulcher (2003, s. 28) menar att fel i verbfrasen inte är att beteckna som allvarliga. För verbfraserna finns en inlärningsordning: enkla verbfraser (presens/preteritum) kommer före komplexa verbfraser (hjälpverb+infinitiv) som kommer före perfekt och pluskvamperfekt (Philipsson 2007, s. 93; Flyman Mattsson & Håkansson 2010, s. 50). Tonkyn (2012) tittar även på användningen av *catenative verbs*, alltså verb som kan följas av ett annat verb trots att de inte är hjälpverb, till exempel *begin*. Norris och Ortega (2009, s. 560), Martin m. fl. (2010) och andra skriver att passivkonstruktion är en komplex struktur som kan säga något om talares grammatiska behärskning. Dessutom är partikelverb och reflexiva verb strukturer som även avancerade inlärare kan ha problem med (Ekberg 2013, s. 271). Allt detta har jag studerat genom att excerpera alla verbfraser som de fyra testtagarna producerar.

I Tabell 14 ser vi vilka hjälpverb som används. Alla använder vanliga hjälpverb som följs av infinitiv, men FAG och NAG har en något bredare repertoar med verben *borde*, *får* och *kommer*. Alla utom JCU producerar verbfraser med hjälpverbet *har* (perfekt) och FAG och MCU producerar *catenative verbs* som *börja*. I Tabell 15 ser vi också att MCU producerar betydligt fler verbfraser i perfekt än FAG och NAG. I Tabell 16 ser vi att NAG producerar en större mängd partikelverb och reflexiva verb än de andra medan FAG och MCU är på samma nivå. JCU producerar ett partikelverb: *tycker om*. I Tabell 17 ser vi slutligen alla fall av passiv. Även här utklassar NAG de övriga, med tre verb i passiv varav ett används fem gånger, bara en gång inkorrekt. FAG och MCU producerar ett passivverb vardera (korrekta) och JCU inte ett enda. Sammantaget kan undersökningen av verbfraserna inte förklara skillnaden i bedömning mellan FAG och MCU; på detta område befinner de båda sig någonstans mellan NAG och JCU.

Tabell 14. Hjälpverb som används av NAG, FAG, MCU och JCU

Typ av hjälpverb	NAG	FAG	MCU	JCU
Hjälpverb som följs av infinitiv	<i>ska</i> <i>kan</i> <i>vill</i> <i>måste</i> <i>kommer</i>	<i>ska</i> <i>kan</i> <i>vill</i> <i>borde</i> <i>får</i>	<i>ska</i> <i>kan</i> <i>vill</i> <i>måste</i>	<i>ska</i> <i>kan</i> <i>vill</i> <i>måste</i>
Perfektiva	<i>har</i>	<i>har</i>	<i>har</i>	
Catenative verbs		<i>behöver</i> <i>börjar</i> <i>bestämmer att</i>	<i>behöver</i> <i>börjar</i>	

Tabell 15. Alla fall av (pluskvam)perfekt hos NAG, FAG, MCU och JCU

NAG	FAG	MCU	JCU
<i>har sagt</i> <i>har valt</i> <i>har samlats</i> <i>har förbrukat</i> <i>∅ använt</i> <i>∅ tagit ut</i>	<i>har sagt (4)</i> <i>har pratat</i> <i>har berättat (2)</i> <i>har diskuterat</i> <i>har flyttat</i> <i>har märkt</i> <i>har kämpat</i> <i>∅ fått</i>	<i>har pratat</i> <i>har valt</i> <i>har hänt</i> <i>har saknats</i> <i>∅ haft</i> <i>∅ flyttat</i> <i>∅ bott</i> <i>∅ förstått</i> <i>∅ glömt</i> <i>∅ försökt</i> <i>∅ tränat</i> <i>∅ druckit</i> <i>∅ glömt dricka</i>	

Tabell 16. Partikelverb och reflexiva verb producerade av NAG, FAG, MCU och JCU

NAG	FAG	MCU	JCU
<i>känner mig</i> <i>lär sig</i> <i>utvecklar sig</i> <i>tycker om</i> <i>kommer ihåg</i> <i>tagit ut</i> <i>måste fylla på</i> <i>kan blanda sig mer i</i> <i>kan få ut</i> <i>kan ta upp</i> <i>går bara förlorat</i> <i>hålla kontakt</i> <i>komma in i</i>	<i>håller med</i> <i>fattar beslut</i> <i>känner dig</i> <i>kan inte följa med</i> <i>bestämmer att skilja sig</i>	<i>måste slappna av</i> <i>kan vi ta bort</i> <i>ta kontakt</i>	<i>tycker om</i> <i>att lära mig</i> <i>att lära dig</i>

Tabell 17. Passiviserade verb hos NAG, FAG, MCU och JCU

NAG	FAG	MCU	JCU
<i>visas behövs (5 gånger) har samlats</i>	<i>påverkas</i>	<i>har saknats</i>	

5.3.6 Finithet – obligatoriska verb och subjekt

Platzack (2001) skriver att såväl första- som andraspråksinlärare, liksom patienter med afasi, har större problem med syntaktiska regler som är relaterade till finithet – V2, obligatoriskt subjekt och obligatoriskt verb – än med andra syntaktiska regler, till exempel för objektens och verbpartikelns placering. Fulcher (2003, s. 28) skriver att utelämnningar av obligatoriska led alltid kan betraktas som allvarliga fel. Även hos modersmålstalare utelämnas ibland obligatoriska led, och Platzack (2001) väljer att dra en 2-procentsgräns: den som utelämnar mindre än 2 procent av obligatoriska led sägs ha uppnått förstaspråkstalarens nivå. Jag har excerperat alla utelämnningar av obligatoriska verb eller subjekt.

Vid något tillfälle sväljer möjligen NAG ett platshållar-*det*, men det finns inget säkert belägg på utelämnning av obligatoriskt subjekt. Obligatoriska verb utelämnas aldrig. Inte heller hos FAG finns några säkra belägg på utelämnning av obligatoriskt subjekt, men däremot utelämnas obligatoriska verb vid två tillfällen, se (68). FAG producerar dock 217 verbfraser, så Platzacks 2-procentsregel bryts inte. MCU utelämnar totalt nio obligatoriska subjekt och sex obligatoriska verb – i båda fallen är det mer än två procent. Vi såg ovan att MCU utelämnar ett obligatoriskt subjekt-*som*. I exempel (69) utelämnas platshållar-*det*, och i (70) ser vi flera exempel på hur språket förenklas genom att *det är* utelämnas. Som vi sett tidigare producerar MCU många verbfraser i perfekt, och här tycks också finnas en systematisk variation mellan huvudsats och bisats: *har* utelämnas alltid i bisats, men dessvärre några gånger även i huvudsats, se (71). JCU utelämnar också nio obligatoriska subjekt; det är mer än två procent. Platshållar-*det* saknas nästan helt, se (72). Däremot finns alltid verben på plats, även om de ofta står i infinitiv. Här ser vi alltså återigen en tydlig skillnad mellan MCU och FAG, och likheter mellan FAG och NAG å ena sidan och mellan MCU och JCU å andra sidan.

(68) FAG: den här reklam siktat till vem ? är den siktat till kvinnor eller till män ?
men i den här fallet också den nackdelen för dom att det kan ...

(69) MCU: och finns många kompisar .

- (70) MCU: när man måste inte ge tillbaka [är det] okej .
och första gång [är det] okej .
MCU: [det är] inga problem för att det har hänt med mig
så för mig [är det] jättefarligt .
- (71) MCU: jag glömt inte .
jag försökt också från första dag här i Sverige att ...
- (72) JCU: för dig är fel ja .
< för > [/] < för > [/] för dig är bra ?
är mycket .

5.4 Sammanfattning av närstudien

I denna närstudie har jag försökt besvara uppsatsens andra fråga genom att undersöka varför testtagarna FAG och MCU bedömdes så olika under kriterium 3, Formell korrekthet och komplexitet. Det har jag gjort genom att jämföra dem med NAG och JCU på en rad olika sätt. Av diagrammen i avsnitt 5.1 får vi uppfattningen att de fyra testtagarna placerar sig på en fallande skala – NAG > FAG > MCU > JCU – där NAG har högst grammatisk förmåga och JCU lägst. Denna bild bekräftas i denna närstudie: MCU har definitivt ett mer utvecklat språk än JCU, som bland annat har stora problem med underordning. NAG har också ett betydligt mer utvecklat språk än FAG; det har vi sett nästan hela tiden. I diagrammen i avsnitt 5.1 tycks dock skillnaden mellan FAG och MCU ganska liten, medan NAG och JCU befinner sig i ytterkanterna. Denna bild har nyanserats här.

FAG och MCU producerar ungefär samma mängd bisatser och relativsatser, och har ungefär samma generella och specifika korrekthet. Men jämfört med MCU bäddar FAG in sina bisatser mer och bygger ut sina infinitivfraser mer; till skillnad från MCU behärskar FAG relativsatsernas struktur, predikativ kongruens samt bisatsordföljd i indirekta frågor. FAG har dessutom större kunskaper om nominalfrasens interna struktur och utelämnar varken obligatoriska verb eller subjekt i samma utsträckning som MCU. Det enda av mina åtta mått som i den kvantitativa studien fångade något av detta var PT-nivå.

I denna kvalitativa närstudie har jag försökt beskriva de fyra testtagarnas språk genom att ge många exempel på strukturer de använder och behärskar (eller inte behärskar). I Tabell 18 försöker jag att sammanfatta resultaten kvantitativt. Tittar man på denna tabell framstår skillnaden i bedömning mellan MCU och FAG som mer rimlig än när vi tittade på Diagram 2–6 ovan. Inte desto mindre skulle jag behöva undersöka fler testtagare lika noggrant för att kunna säga huruvida skillnaden i bedömning verkligen var rättvis. Att definiera några exakta gränser mellan C och B och mellan B och A är givetvis mycket svårt.

Tabell 18. Kvantitativ sammanfattning av den kvalitativa närstudien

Grammatik som testtagaren har/producerar	NAG	FAG	MCU	JCU
PT-nivå 4	+	+	-	-
Ett fall av systematisk inversion	+	-	-	-
Ett fall av systematisk predikativ kongruens	+	+	-	-
Korrekta adverbiala bisatser	+	+	+	+
Korrekta nominala bisatser inledda av att/om	+	+	+	-
Korrekta relativsatser	+	+	-	-
Bisatsordföljd i indirekta frågor	+	+	-	+
BIFF	+	-	-	-
Tredjegradsbisats	+	-	-	+
Infinitivfras utbyggd med sats eller infinitivfras	+	+	-	+
Dubbel bestämdhet	+	+	-	-
Attributiv kongruens, plural (adjektiv, inte pronomen)	+	+	-	+
Attributiv kongruens, bestämd form	+	+	-	-
Perfektiv	+	+	+	-
<i>Catentative verbs</i>	-	+	+	-
Verb i passiv	+	+	+	-
Partikelverb	+	+	+	+
Reflexivt verb	+	+	-	+
Utelämnar mindre än 2 % av alla obligatoriska verb	+	+	-	+
Utelämnar mindre än 2 % alla obligatoriska subjekt	+	+	-	-

6 Diskussion och slutsatser

Genom en kvantitativ undersökning av tjugo testtagares performans under ett muntligt språktest har jag försökt besvara uppsatsens första fråga, och genom en kvalitativ närstudie av fyra testtagare har jag sökt svar på uppsatsen andra fråga. De två frågorna upprepas här:

1. (Hur) korrelerar komplexitet, korrekthet och bedömning i ett muntligt språktest?
2. (Hur) kan till synes avvikande bedömningar förklaras?

Med reservation för att resultaten inte är generaliserbara ska jag här svara på de två frågorna. Svaret på fråga 1 är, kortfattat, att korrekthet och komplexitet generellt korrelerar med varandra, ibland starkt. Det är framför allt den specifika inversionskorrektheten – ett utvecklingsmått som säger något om testtagarens förmåga att processa information över frasergränser – som korrelerar med komplexitetsmått. Samtidigt är det den generella korrektheten, som ju inte säger något om vilka fel som begås, som framför allt korrelerar med bedömningen under kriterium 3, Formell korrekthet och komplexitet. Komplexitet uppfattas som positivt, men bedömarna lyckas inte "fånga" denna konstrukt under "rätt"

kriterium. Jag tolkar det som en halo-effekt (se avsnitt 2.2.3). Generell korrekthet är också det mått som bäst förutsäger vilka som klarar hela Swedex B2-test.

Svaret på fråga 2 modifierar svaret på fråga 1 något. Trots de generella korrelationerna mellan komplexitet, korrekthet och bedömning blir det i den kvalitativa undersökningen tydligt att de mått jag använt inte alltid säger så mycket om testtagarnas språkliga kompetens. Två testtagare – MCU och FAG – bedömdes olika trots att de enligt mina mätningar producerade ungefär samma slags språk. Genom att närstudera dem fann jag viktiga skillnader i deras språkutveckling som bara ett av måtten i den kvantitativa undersökningen – utvecklingsmättet PT-nivå – lyckades fånga.

Uppsatsen har två syften: För det första utgör den en kritisk granskning av fenomenet muntliga språktest; för det andra utgör den en undersökning av relationen mellan morfosyntaktisk korrekthet, komplexitet och allmän muntlig språkfärdighet. För att uppfylla dessa båda syften ska jag diskutera resultaten utifrån två olika utgångspunkter. I avsnitt 6.1 betraktar jag mina mått som oberoende variabler och bedömningarna som beroende variabler; här är det med andra ord bedömningarna som är forskningsobjektet. I avsnitt 6.2 betraktar jag tvärtom bedömningarna som oberoende variabler – jag antar att de faktiskt säger något om testtagarnas språkfärdighet – medan mina mått får fungera som beroende variabler; här är det alltså mina mätningar som är forskningsobjektet.

6.1 Språktestning – bedömningar som beroende variabler

”Objektiva kriterier för att beskriva språkfärdighet kommer att underlätta det ömsesidiga erkännandet av examina från olika platser”, kan vi läsa i CEFR (Europarådet 2001, s. 1). Men hur objektiva är beskrivningarna i CEFR:s nivåskalor? De många deskriptorerna har inte tagits fram med lingvistiska mätningar av språkfärdighet och andraspråksinlärning utan genom statistiska metoder: de deskriptorer som lett till högst inter-bedömarreliabilitet har ansetts vara de bästa (Europarådet 2001, s. 25). Detta förfaringsätt kan ifrågasättas. Eklund Heinonen (2009, med referens till Shohamy m. fl. 1992) skriver: ”Även om bedömarna är överens om ett resultat, det vill säga interbedömarreliabiliteten är hög, betyder inte det automatiskt att validiteten i bedömningen är hög. De kan ju vara överens på felaktiga grunder.” (s. 53) Detta problem är CEFR-författarna också medvetna om. De skriver: ”Att fastställa skiljelinjer mellan nivåer är alltid ett subjektivt förfarande.” (Europarådet 2001, s. 32). Det är också vad FIFA (2014) skriver om bedömningar av offside-beslut; man vet att besluten är subjektiva. Definierade nivåer behövs dock, argumenterar man i CEFR, ”av praktiska skäl” (s. 17), precis som offside-regeln behövs av praktiska skäl (utan den skulle det bli tråkigt att titta på fotboll eftersom båda lagen skulle ha spelare stående högt uppe vid motståndarlagets mål; bollen skulle

bara sparkas fram och tillbaka över planen). Språktest kan aldrig vara helt valida (se avsnitt 2.1.1), varför valideringsarbetet är ”en ständigt pågående och teoretiskt oändlig process” (Europarådet 2001, s. 21). Att denna process hålls levande är också det enda som egentligen krävs av testinstitutioner som Folkuniversitetet, enligt de riktlinjer som European Association for Language Testing and Assessment (EALTA) (2014) ger ut. Föreliggande uppsats kan förhoppningsvis användas i detta arbete. Swedex muntliga B2-test bedöms analytiskt, tvådimensionellt, med fyra kriterier längs ena dimensionen och tre nivåer längs den andra (jmf. avsnitt 2.2.3). Längs båda dimensionerna finns problem. De diffusa gränserna mellan nivåer diskuteras i 6.1.1 och de diffusa gränserna mellan kriterier diskuteras i 6.1.2.

6.1.1 Diffusa gränser mellan nivåer

Resultaten från den kvantitativa studien tyder på att det inte finns några distinkta gränser mellan de bedömda nivåerna C och B eller mellan B och A; den gräns som går att finna är i stället den mellan C och A. Detta gäller samtliga mått som kan användas för att förutsäga bedömning, även det mått som klarade sig bäst genom granskningen i den kvalitativa studien, PT-nivån: precis hälften av dem som fick B på kriterium 3 hade uppnått PT-nivå 4 medan den andra hälften inte gjort det. Precis hälften av dem som fick B klarade också hela Swedex. Bedömningen B framstår alltså som ett neutralt, slumpmässigt icke-beslut som fattas när bedömaren inte lagt märke till något särskilt – varken positivt eller negativt – och jag menar att B:s status som icke-beslut borde erkännas. I stället för att, som nu, hävda att B betyder att B2-nivån är uppnådd, medan A betyder att nivån är uppnådd med marginal och C betyder att nivån inte är uppnådd, vore det ärligare att säga att B betyder att bedömaren inte kan säga säkert om nivån är uppnådd eller inte och att man därför friar hellre än faller. I praktiken skulle detta inte innebära någon skillnad, men förutom att det vore ärligare tror jag att det vore bra om pressen på bedömarna att fatta aktiva beslut för varje kriterium minskade; man skulle erkänna det som alla vet, nämligen att det är svårt att bedöma muntlig språkfärdighet.

Några som på ett mer raffinerat sätt tagit fasta på bedömarnas (o)säkerhet är Jin, Mak och Zhou (2012), som introducerat begreppet *confidence scoring*. De visar att problemet med otydliga gränser mellan nivåer kan lösas – rent statistiskt – om bedömarna också bedömer med vilken säkerhet de gjort en viss bedömning. Att implementera detta system i Swedex skulle kräva betydligt större förändringar, men om de kommunikativa testen ska behålla (och öka) sin trovärdighet är det nog hitåt vi måste gå.

6.1.2 Diffusa gränser mellan kriterier

Trots de generella positiva korrelationerna mellan korrekthetsmått och komplexitetsmått finns det anledning att diskutera huruvida korrekthet och komplexitet bör bedömas under samma kriterium. Utifrån den kunskap vi har om *non-linearity* (Fulcher 2003, s. 28) kan detta förfarande ifrågasättas. Studier som Magnan (1988, refererad i Iwashita m. fl. 2008), Håkansson och Norrby (2007), Spoelman och Verspoor (2010), Jin och Mak (2013), Baten och Håkansson (2014) med flera har visat att komplexitet och korrekthet inte utvecklas parallellt, varken med varandra eller med övrig språkutveckling, och ibland rent av konkurrerar med varandra. Att jag i min studie finner positiva korrelationer mellan komplexitet, korrekthet och PT-nivå – en del svaga, några starka – är trots detta inte förvånande: andelen bisatser i exempelvis Baten och Håkanssons (2014) studie ökar också generellt över tid, men författarnas poäng är att andelen bisatser inte ökar linjärt (se även Håkansson & Norrby 2007). I min studie har vi sett att AAG, som fått A under kriterium 3, har hög genomsnittlig satslängd men kort genomsnittlig AS-längd och låg andel bisatser. Vi har också sett att MCU och FAG har ungefär samma korrekthet och komplexitet trots att de kommit olika långt i språkutvecklingen. Trots tendenserna finns det alltså gott om "undantag".

Att jag utifrån mina resultat ifrågasätter huruvida det är rimligt att bedöma korrekthet och komplexitet under samma kriterium beror framför allt på att det verkar vara korrektheten som faktiskt bedöms under kriterium 3. En bedömare kan inte bedöma fler än fyra-fem (maximalt sju) kriterier samtidigt utan att bli kognitivt överbelastad (Europarådet 2001, s. 188). Enligt Luoma (2004, s. 80) är det viktigt att kriterierna är "conceptually independent". Med det följer, antar jag, att det som samlas under ett och samma kriterium bör vara konceptuellt relaterat. I Swedex B2-tests muntliga del bedöms flyt och interaktion under kriterium 1, ordförråd och koherens under kriterium 2, och intonation och uttal under kriterium 4. Och så har vi korrekthet och komplexitet under kriterium 3. Hur konceptuellt sammanhållna är dessa fyra kriterier? Det ser onekligen ut som att ganska många konstrukter ska bedömas samtidigt – åtta stycken (?), fler än vad som anses möjligt. Det är ett rimligt antagande att bedömaren förenklar denna uppgift genom att reducera den. När det gäller kriterium 3 fokuseras uppenbarligen korrekthet.

Detta leder i sin tur till risk för halo-effekter (se Fulcher 2003, s 90; Eckes 2005, s. 5). Om komplexitet uppfattas som positivt utan att bedömarna kan sätta fingret på vari det positiva består blir bedömningen holistisk trots att den ska vara analytisk (se avsnitt 2.2.3). Tonkyn (2012, s. 238) noterar att halo-effekter lätt drabbar just komplexiteten: "Grammatically complex language may not be perceived as such if it involved considerable repetition of structures or occurred in relatively short turns. Complex

language might also be discounted if it felt to be imprecise or irrelevant. Finally, complex language appeared to be 'hidden' in some cases if it was produced laboriously in relatively non-fluent ways."

Så om vi slår fast att det är svårt att definiera och bedöma komplexitet, och att komplexitet inte behöver utvecklas linjärt med korrekthet eller allmän språkutveckling, bör kriterium 3 omformuleras. Testkonstruktörerna bör acceptera att det som fokuseras här är korrekthet, samtidigt som man bör påpeka att korrektheten måste bedömas med hänsyn till hur komplext språket är. Huruvida felen leder till missförstånd bör däremot inte bedömas under detta kriterium; förmågan att få samtalspartnern att förstå hör snarast till kategorin interaktion. Mig veterligen vet vi inte om fel som begås på senare inlärningsstadier leder till färre missförstånd än fel som begås på tidiga inlärningsstadier (jmf. Tonkyn 2012, s. 225, som skriver att det visat sig svårt att fastställa en gravhethierarki för språkfel).

I närstudien såg vi att inga av mina mått utom PT-nivå – varken generell eller specifik korrekthet, eller andelen bisatser, relativsatser eller inversionskontexter, eller AS-längd eller satslängd – avslöjar de skillnader som finns mellan MCU och FAG. Detta gör ju, som sagt, att alla dessa mått kan ifrågasättas. Kanske måste vi helt enkelt lita på bedömarnas erfarenhet och intuition, på samma sätt som fotbollsspelare måste lita på linjedomarens. I ett radioprogram hörde jag en gång en fotbollsdomare säga att offside-regeln är mycket enkel: det är offside när domaren säger att det är offside. Samma sak skrev Vollmer (1983, citerad hos Fulcher 2003, s. 180): "proficiency' is what proficiency tests measure". I nästa avsnitt utgår jag från att bedömarnas beslut faktiskt säger vilka av talarna som har uppnått B2-nivån.

6.2 Språkfärdighet – mätvärden som beroende variabler

Om vi utgår från att Swedextestet är valitt kan vi använda bedömningarna för att bestämma värdet av mina lingvistiska mått. Generellt tycks de som har uppnått B2-nivån på den muntliga delen producera längre AS-enheter, längre satser (det vill säga, mer utbyggda fraser) samt fler bisatser och relativsatser. Framför allt begår dessa talare också färre grammatiska fel. De som klarar hela Swedex B2-test producerar också längre AS-enheter och satser samt begår färre fel. Inte desto mindre ser vi att exempelvis AAG får ett högt betyg trots korta AS-enheter och få bisatser – denna talare demonstrerar kanske sin språkliga kompetens genom att bygga ut sina fraser, vilket måttet satslängd här tycks visa. Men satslängd är inte ett säkert mått. Som nämndes i avsnitt 3.2 kan det säga lika mycket om samordning inom fraserna som om frasernas djup; exempel (13) och (14) (s. 33) får ju höga komplexitetspoäng eftersom så många fraser respektive bisatser radas på varandra.

Måttet satslängd lyckas exempelvis inte avslöja någon skillnad mellan FAG och MCU, trots att jag i min närstudie visar att FAG bygger ut sina nominalfraser mer än MCU. Detta problem behöver lösas metodologiskt.

Att PT-nivån var det enda mått som fångade de skillnader mellan MCU och FAG som sedan blev så tydliga i närstudien ger onekligen stöd åt PT och till Baten och Håkansson (2014) som menar att PT, till skillnad från CAF-måtten, kan mäta språkutveckling (se även Håkansson & Norrby 2007). Men korrelationerna mellan PT och övriga mått, och mellan PT och kriterium 3, är svaga, och till skillnad från i Eklund Heinonens (2009) studie kan PT inte användas för att förutsäga betyget på hela testet utan bara bedömningen under kriterium 3. Antagligen är PT bäst lämpad för studier om just muntlig färdighet (se Ellis 2008), även om Granfeldt och Ågren (2014) fann starka korrelationer mellan PT och bedömningar av skrivna texter.

PT kan inte förklara all språkutveckling – det har ingen heller hävdad – och utifrån min närstudie kan jag dra några slutsatser om vad som också kan bidra till definitionen av (grammatisk) språkfärdighet. Som Platzack (2001) skriver tycks fenomen kopplade till finithet signifikanta, men också fenomen kopplade till definithet. Även om både MCU och FAG i någon mån behärskar attributiv kongruens är det tydligt att FAG har större kunskaper om nominalfrasens interna struktur; FAG behärskar bland annat, till skillnad från MCU, dubbel bestämdhet. Enligt Axelsson (1994, s. 99), som undersökt tillägandet av nominalfraser hos inlärare av svenska som andraspråk, hör attributiv kongruens och dubbel bestämdhet till samma steg i inlärningsprocessen, men jag undrar om inte den dubbla bestämdheten är en mer komplex struktur än själva kongruensen. Dessutom förmår FAG, till skillnad från MCU, bygga ut nominalfraserna med relativsatser. Philipsson (2007) har studerat inlärningsgångar för verbmorfologin, men i fallet MCU/FAG kan verbmorfologin inte avslöja att de befinner sig på olika färdighetsnivåer. Uppenbarligen finns det inga enkla svar på frågan vad som, rent lingvistiskt, kännetecknar en viss allmän färdighetsnivå.

6.3 Framåtblick

Ett av de viktigaste argumenten för kommunikativa språktest är den så kallade washback-/backwash-effekten (Ellis 2008, s. 18; Bachman & Palmer 2010, s. 109). Underhill (1982, citerad hos Fulcher 2000, s. 485) skrev: "As ye teach, so shall ye test." Idén är att testen måste överensstämma med undervisningen. Om vi vill att undervisningen ska vara kommunikativ måste vi också testa kommunikativt, men frågan är vad vi menar med kommunikativ undervisning. Tomita och Spada (2013) visar att fokus på form i undervisningen kan öka studenternas kommunikativa insats i klassrummet (se också Flyman Mattsson

2003). Vi ska givetvis inte tillbaka 1950-talets audiolingvala undervisning, som byggde på behavioristisk teori och metod – ”drillning” (Tornberg 2009, s. 44–45) – men om vi erkänner behovet av att jobba aktivt med specifika strukturer i undervisningen så blir det ingen negativ backwash-effekt av att också testa specifika strukturer i språktest (jmf. Van Moere 2012).

CEFR ska enligt författarna vara ”odogmatisk” och oberoende av specifika teorier om språk och pedagogik (Europarådet 2001, s. 18), men det är tydligt att CEFR är en produkt av vad Fulcher (2000) kallar *The Communicative Movement*. Carlsen (2003, refererad i Eklund Heinonen 2009, s. 15) menar att muntliga språktest har lägre reliabilitet om de är kommunikativt inriktade, så kanske är det dags att, som Ellis (2008), ifrågasätta den dogm som säger att testning måste vara ”kommunikativ” (se avsnitt 2.1.2). Ellis (2008, s. 17–18) skriver att ”pragmatic and conversational competence are realized primarily by means of linguistic resources [...]” och han fortsätter: ”What language learners can do with language is to a very considerable extent dependent on what language they know.” Precis som Ellis undersöker Van Moere (2012) hur muntlig språkfärdighet kan testas med psykolingvistiska, experimentella metoder. Han menar att olika typer av eliciteringstekniker, som testar förmågan att producera specifika strukturer, på ett mer reliabelt sätt mäter konstrukter som korrekthet och komplexitet. Men han erkänner svagheterna med denna typ av testning – bland annat bristen på validitet. Så då är vi tillbaka där vi började (se avsnitt 2.1.2): vad ska vi välja – reliabilitet eller validitet?

Van Moere menar dock, precis som Ellis, att det finns utrymme för olika typer av test: ”a complementary approach to communicative and psycholinguistic testing will undoubtedly lead to stronger and fairer assessments” (s. 340). Här finns ett stort behov av – och stora potential för – vidare forskning. Denna forskning kommer oundvikligen, precis som denna uppsats, att handla både om språktestning och om språkfärdighet.

7 Avslutning

Avslutningsvis vill jag, i strid mot konventionen, presentera nya fakta i målet. I närstudien i avsnitt 5 synade vi MCU:s lingvistiska kompetens i sömmarna. Vad MCU talade om nämndes dessvärre inte, nämligen *vänskap*. Som framgår av exempel (73) tyckte MCU att det är svårt att skaffa nya vänner i sitt nya land.

(73) MCU: men riktigt vänner är svårt att göra igen här i en annat land .

Trots genomsnittliga värden för generell korrekthet och andelen bisatser fick MCU betyget C på kriterium 3. Kanske var betyget rättvist – jag hade behövt närstudera fler testtagare för att kunna säga det – men av betygsprotokollet framgår att MCU inte hade klarat Swedex B2-test även med ett B på kriterium 3. MCU är dessbättre inte den som ger upp; i exempel (74) ser vi hur MCU kämpar för att skaffa vänner här i Sverige, bland annat genom att alltid tala svenska. Optimismen är det inget fel på.

- (74) MCU: jag försökt också från första dag här i Sverige att va alltid mellan folk och alltid ta kontakt med människor alltid bara försöka prata svenska hela tiden äh ha nytt kontakt .
så efter vi ta fika tillsammans ?
eller vi kan gå till bio tillsammans ?
< och > [/] och det är inte alltid funkar för att dom har sin kompisar också .
men efter en tid jag tror att det kan vara lättare .

Några månader efter det underkända testet gjorde MCU ett nytt försök. Även detta test spelade jag in. Det tog ett bra tag innan jag, under forskningsprocessen, upptäckte att MCU och SBU är samma fysiska person. För undersökningens trovärdighet spelar detta ingen roll – det är två olika performanser som bedömts vid två olika tillfällen – men jag vill ändå betona att övriga arton testtagare är unika fysiska individer.

Om vi jämför SBU med MCU i diagrammen 2–6 i avsnitt 5.1 ser vi att den generella korrektheten vid det andra testtillfället var något lägre medan inversionskorrektheten var något högre; den genomsnittliga AS-enheten var längre medan satslängden var kortare; andelen bisatser och relativsatser var högre medan andelen inversionskontexter var lägre. Testtagaren bedömdes fortfarande vara på PT-nivå 3. Här är svårt att tala om någon linjär utveckling, men bedömningen under kriterium 3 höjdes i alla fall till B: SBU ansågs ha uppnått B2-nivå när det gäller formell korrekthet och komplexitet. Men som framgår av Tabell 8 (s. 25) i avsnitt 3.2 utdelades i stället ett C på kriterium 2, Kommunikativ design, där bland annat ordförråd och koherens bedöms. Dessutom ändrades bedömningen under kriterium 1, Interaktion och flyt, från A till B. Inte heller denna gång blev testtagaren godkänd på Swedex B2-test.

Jag har varken undersökt ordförråd eller textuell kompetens (kommunikativ design) eller interaktionell kompetens eller flyt och kan därför inte säga något om dessa bedömningar, men det är givetvis rimligt att fråga sig hur troligt det är att en motiverad inlärare utvecklas negativt på flera områden under några månader i Sverige. Har MCU/SBU försämrat sin svenska, eller är det snarare något diffust i testtagarens språk som gör att bedömarna anser miniminivån uppnådd men inte mer, och därför vill utdela

åtminstone ett C, mer eller mindre godtyckligt? Är bedömningen analytisk, eller är den snarare holistisk?

Frågan om hur rättssäkra kommunikativa språktest är kvarstår. Mycket talar för att testtagare – precis som fotbollsspelare – är utlämnade åt bedömarnas intuition. I fotbolls-VM blev det fel i ett fall av fyra (se avsnitt 1). Av exempel (75) drar jag åtminstone slutsatsen att MCU/SBU inte kommer att ge upp; hen kommer att försöka igen.

(75) SBU: jag upplevt hur svårt att börja när jag komma här .
men jag har valt att komma här .
jag försöka lära < som > [/] < som > [/] + ...
jag är inte svensk men jag försöker lära som en svensk .
så jag försöka göra alla regler äh betala allt .
sen försöka framme framme framme med allt .
men det tar lång tid .
ingen fara .
jag ska göra .
[...]
men jag försöka .
jag är en sportmänniska .
så jag vill vinna .
och finns ingenting omöjlighet .
så om man vill ha någonting man kan göra tror jag det .

Referenser

- Axelsson, Monica (1994). *Noun phrases development in Swedish as a second language. A study of adult learners acquiring definiteness and the semantica and morphology of adjectives*. Diss. Stockholm: Stockholms univeristet.
- Bachman, Lyle F. (2004). *Statistical analyses for language assessment*. Cambridge: Cambridge University Press.
- Bachman, Lyle F. (1990). *Fundamental considerations in language testing*. Oxford: Oxford University Press.
- Bachman, Lyle F. & Andrew D. Cohen (1998). Language testing – SLA interfaces: An update. I: Bachman, Lyle F. (red.). *Interfaces between second language aquisition and language testing research*. Cambridge: Cambridge University Press.
- Bachman, Lyle F. & Adrian Palmer (2010). *Language assessment in practice*. Oxford: Oxford University Press.
- Bartning, Inge, Maisa Martin & Ineke Vedder (2010). Foreword. I: Inge Baating, Maisa Martin & Ineke Vedder (red.). *Communicative proficiency and linguistic development: Intersections between SLA and language testing research*. European Second Language Association, EUROSLA.
- Baten, Kristof & Gisela Håkansson (2014). The development of subordinate clauses in German and Swedish as L2s. A theoretical and methodological comparison. *Studies in second language acquisition*. Publicerad på tidskriftens webbplats.
- Bulté, Bram & Alex Housen (2012). Defining and operationalising L2 complexity. I: Housen, Alex, Folkert Kuiken & Ineke Vedder (red.) *Dimensions of L2 performance and proficiency. Complexity, accuracy and fluency in SLA*. Amsterdam: John Benjamins Publishing Co.
- De Jong, Nivja. H., Margarita P. Steinel, Arjen Florijn, Rob Schoonen & Jan H. Hulstijn (2012). Facets of speaking proficiency. *Studies in second language acquisition* 34, s. 5–34.
- Eckert, Penelope (2013). Ethics in linguistic research. I: Robert J. Podesva & Devyani Sharma (red.). *Research methods in linguistics*. Cambridge: Cambridge University Press.
- Eckes, Thomas (2009). Many-facets Rasch measurement. I: *Reference supplement to the Manual for relating language examinations to the Common European framework of reference for language: learning, teaching, assessment. Section H: Many-facet Rasch measurement*. Strasbourg: Europarådet.
- Ekberg, Lena (2013). Grammatik och lexikon i svenska som andraspråk på nästan infödd nivå. I: Kenneth Hyltenstam & Inger Lindberg (red.). *Svenska som andraspråk: i forskning, undervisning och samhälle*. 2. uppl. Lund: Studentlitteratur.
- Eklund Heinonen, Maria (2009). *Processbarhet på prov. Bedömning av muntlig språkfärdighet hos vuxna andraspråksinlärare*. Diss. Uppsala: Uppsala universitet.
- Ellis, Rod (2008). Investigating grammatical difficulty in second language learning: Implications for second language acquisition research and language testing. *International journal of applied linguistics* 18(1), s. 4–22.
- Ellis, Rod & Gary Barkhuizen (2005). *Analysing learner language*. Oxford: Oxford University Press.
- Europarådet (2001). *Gemensam europeisk referensram för språk: lärande, undervisning och bedömning* (CEFR). Översatt till svenska av skolverket 2007. Hämtad på skolverkets hemsida: www.skolverket.se
- European Association for Language Testing and Assessment, EALTA (2014). *EALTAs riktlinjer för god praxis vid prov och bedömning i språk*. Dokument hämtat 15 juni 2014 från: <http://www.ealta.eu.org/documents/archive/guidelines/Swedish.pdf>

- Ezra Johnson, Daniel (2013). Descriptive statistics. I: Robert J. Podesva & Devyani Sharma (red.). *Research methods in linguistics*. Cambridge: Cambridge University Press.
- Fédération Internationale de Football Association, FIFA (2014). *Laws of the game*. Zurich.
- Flyman Mattsson, Anna (2003). *Teaching, learning, and student output. A study of French in the classroom*. Diss. Lund: Lunds universitet.
- Flyman Mattsson, Anna & Gisela Håkansson (2010). *Bedömning av svenska som andraspråk: en analysmodell baserad på grammatiska utvecklingsstadier*. Lund: Studentlitteratur.
- Folkuniversitetet (2014). Swedex: Swedish Examinations. Swedex' webbplats 9 mars 2014: <http://www.folkuniversitetet.se/Las-mer-om-sprak/Sprakexamina/Swedex/Om-Swedex/Testets-innehall-och-struktur/>
- Folkuniversitetet (2010). Swedex B2 – Kriterier – Tala 2010-12-20. Kommentarer till bedömningskriterierna för muntlig uppgift. Dokument hämtat den 5 januari 2014 på Swedex' hemsida: <http://www.folkuniversitetet.se/Las-mer-om-sprak/Sprakexamina/Swedex/Modelltest/Modelltest-B2/>
- Folkuniversitetet (2011). Swedex B2 – Regler för godkänt 2011-04-05. Regler för godkänt på hela Swedex B2. Dokument hämtat den 5 januari 2014 på Swedex' hemsida: <http://www.folkuniversitetet.se/Las-mer-om-sprak/Sprakexamina/Swedex/Modelltest/Modelltest-B2/>
- Foster, Pauline, Alan Tonkyn & Gillian Wigglesworth (2000). Measuring spoken language: A unit for all reason. *Applied Linguistics* 21(3), s. 354–375.
- Fox, John & Weisberg, Sanford (2011). *An R companion to applied regression*. 2. uppl. Thousand Oaks, Calif.: SAGE Publications.
- Fulcher, Glenn (2003). *Testing second language speaking*. London: Longman.
- Fulcher, Glenn (2000). The 'communicative' legacy in language testing. *System* 28, s. 483–497.
- Glahn, Esther, Gisela Håkansson, Björn Hammarberg, Anne Holmen, Anne Hvenekilde & Karen Lund (2001). Processability in Scandinavian second language acquisition. *Studies in second language acquisition* 23, s. 389–416.
- Granfeldt, Jonas & Malin Ågren (2013). Stages of processability and levels of proficiency in the Common European framework of reference for languages. The case of L3 French. I: Anna Flyman Mattsson & Catrin Norrby (red.). *Language acquisition and use in multilingual contexts. Theory and practice*. Lund: Lund University Press.
- Green, Anthony (2014). *Exploring language assessment and testing. Language in action*. Oxon: Routledge.
- Gyllstad, Henrik, Jonas Granfeldt, Petra Bernardini & Marie Källkvist (2014). Linguistic correlates to communicative proficiency levels of the CEFR. I: L. Roberts, I. Vedder & Jan H. Hulstijn (red.). *EUROSLA Yearbook* 14, ss. 1–30. Amsterdam: John Benjamins Publishing Co.
- Helsen, Werner, Bart Gilis & Matthew Weston (2006). Errors in judging "offside" in association football: Test of the optical error versus the perceptual flash-lag hypothesis. *Journal of sports sciences* 24(5), s. 521–528.
- Housen, Alex, Folkert Kuiken & Ineke Vedder (2012). Complexity, accuracy and fluency. Definitions, measurements and research. I: Housen, Alex, Folkert Kuiken & Ineke Vedder (red.) *Dimensions of L2 performance and proficiency. Complexity, accuracy and fluency in SLA*. Amsterdam: John Benjamins Publishing Co.
- Hulstijn, Jan H. (2007). The shaky ground beneath the CEFR: Quantitative and qualitative dimensions of language proficiency. *The Modern Language Journal* 91(4), s. 663–667.
- Hulstijn, Jan H., Rob Schoonen, Nirva H. De Jong, Margarita P. Steinel & Arjen Florijn (2011). Linguistic competences of Dutch as a second language at the B1 and B2 levels of speaking proficiency of the Common European framework of reference for language (CEFR). *Language testing* 29(2), s. 203–221.

- Iwashita, Noriko, Annie Brown, Tim McNamara & Sally O'Hagan (2008). Assessed levels of second language speaking proficiency: how distinct? *Applied Linguistics* 29(1), s. 24–49.
- Jin, Tan & Barley Mak (2013). Distinguishing features in scoring L2 Chinese speaking performance: How do they work? *Language Testing* 30(1), s. 23–47.
- Jin, Tan, Barley Mak & Pei Zhou (2012). Confidence scoring of speaking performance: How does fuzziness become exact? *Language testing* 29(1), s. 43–65.
- Kuiken, Folkert & Ineke Vedder (2014). Raters' decisions, rating procedures, and rating scales. *Language testing* 31(3), s. 279–284.
- Luoma, Sari (2004). *Assessing speaking*. Cambridge: Cambridge University Press.
- MacWhinney, Brian (2000). *The CHILDES Project: Tools for analyzing talk*. 3 uppl. Mahwah, NJ: Lawrence Erlbaum Associates.
- Martin, Maisa, Sanna Mustonen, Nina Reiman & Marja Seilonen (2010). On becoming an independent user. I: Inge Baating, Maisa Martin & Ineke Vedder (red.). *Communicative proficiency and linguistic development: Intersections between SLA and language testing research*. EUROSLA.
- Nagy, Naomi & Devyani Sharma (2013). Transcription. I: Robert J. Podesva & Devyani Sharma (red.). *Research methods in linguistics*. Cambridge: Cambridge University Press.
- Norrby, Catrin & Gisela Håkansson (2007). The interaction of complexity and grammatical processability – the case of Swedish as a foreign language. *International review of applied linguistics* 45, s. 45–68.
- Norris, John & Lourdes Ortega (2009). Towards an organic approach to investigating CAF in instructed SLA: The case of complexity. *Applied Linguistics* 30(4), s. 555–578.
- Olofsson, Mikael & Lena Sjöqvist (2013). Bedömning i svenska som andraspråk. I: Kenneth Hyltenstam & Inger Lindberg (red.). *Svenska som andraspråk: i forskning, undervisning och samhälle*. 2. uppl. Lund: Studentlitteratur.
- Philipsson, Anders (2007). *Interrogative clauses and verb morphology in L2 Swedish. Theoretical interpretations of grammatical development and effects of different elicitation techniques*. Diss. Stockholm: Stockholms universitet.
- Pienemann, Manfred (2011). The psycholinguistic basis of PT. I: Manfred Pienemann & Jörg-U. Keßler (red.). *Studying Processability theory*. Amsterdam: John Benjamins Publishing Company.
- Pienemann, Manfred (1998). Developmental dynamics in L1 and L2: Processability theory and generative entrenchment. *Bilingualism: Language and cognition* 1, s. 1–20.
- Pienemann, Manfred & Gisela Håkansson (1999). A unified approach toward the development of Swedish as L2. A processability account. *Studies in second language acquisition* 21, s. 383–420.
- Platzack, Christer (2001). The vulnerable C-domain. *Brain and language* 77, s. 364–377.
- Region Skåne (2011). Kvalifikationskrav för anställning av legitimerad hälso- och sjukvårdspersonal från EU/ESS-området. Beslut 2011-02-09. Hämtad från Region Skånes hemsida: <https://www.skane.se/Public/Jobb%20Framtid/Utländska%20läkare/Kvalifikationskrav%20leg%20yrken%20EU.pdf>
- Sandlund, Erica & Pia Sundqvist (2014). Provuuppgiftshantering som social praktik. En jämförelse mellan interaktionsanalys och bedömandata för muntligt prov i engelska. I: Christina Rosén, Per Simfors & Ann-Kari Sundberg (red.). *Språk i undervisning. Rapport från ASLA:s vårsymposium Linköping, 11–12 maj 2012*. Reviderad upplaga 2014. ASLA:s skriftserie 24. Linköping: Linköpings universitet.
- Second Language Acquisition and Testing in Europe (SLATE) (2014). Webbplats besökt den 24 november 2014: <http://www.slate.eu.org/>
- Skolverket (2012). *Utbildning i svenska för invandrare. Kursplan och kommentarer*. Förordning 2012:13. Hämtad från Skolverkets hemsida: www.skolverket.se

- Spoelman, Marianne & Marjolijn Verspoor (2010). Dynamic patterns in development of accuracy and complexity: a longitudinal case study in the acquisition of Finnish. *Applied linguistics* 31(4), s. 532–553.
- Socialstyrelsen (2014). Ansökan om legitimation för utbildade i annat land. Socialstyrelsens hemsida 9 mars 2014:
<http://www.socialstyrelsen.se/ansokaomlegitimationochintyg/legitimation/utbildadianatland>
- Stockholms läns landsting (2012). *Riktlinjer vid rekrytering och anställning av utländsk vårdpersonal, inom EU/EES och för svenskar som studerar till läkare inom EU*. Rapport hämtad på Stockholms läns landstings hemsida:
<http://www.sll.se/upload/Policies/Riktlinjer%20för%20rekrytering%20av%20utländsk%20vårdpersonal.pdf>
- Tomita, Yasuyo & Nina Spada (2013). Form-focused instruction and learner investment in L2 communication. *The modern language journal* 97(3), s. 591–610.
- Tonkyn, Allan (2012). Measuring and perceiving change in oral complexity, accuracy and fluency. Examining instructed learners' short-term gains. I: Housen, Alex, Folkert Kuiken & Ineke Vedder (red.) *Dimensions of L2 performance and proficiency. Complexity, accuracy and fluency in SLA*. Amsterdam: John Benjamins Publishing Co.
- Tornberg, Ulrika (2009). *Språkdidaktik*. 4. uppl. Malmö: Gleerup.
- Van Moere, Alistair (2012). A psycholinguistic approach to oral language assessment. *Language testing* 29(3), s. 325–344.
- Vetenskapsrådet (2011). *God forskningssed*. Vetenskapsrådets rapportserie 1:2011. Vetenskapsrådet: Stockholm.
- Watkins, Marly W. & Miriam Pacheco (2001). Interobserver agreement in behavioral research: Importance and calculation. *Journal of behavioral education* 10(4), s. 205–212.

Bilaga 1: Informerat samtycke

MASTERSPROJEKT OM BEDÖMNING AV MUNTLLIG SPRÅKFÄRDIGHET

Hej!

Jag är svensklärare och masterstudent vid Lunds universitet. Nu skriver jag min masteruppsats om bedömning av muntlig språkfärdighet. Jag spelar in och analyserar muntliga språktest för att sedan jämföra dessa analyser med testtagarnas betyg. På så vis hoppas jag kunna bidra till utvecklingen av muntliga språktest.

Din medverkan är helt frivillig och din anonymitet garanteras. Kontakta gärna mig eller min handledare på Lunds universitet om du har några frågor.

Lycka till med ditt SWEDEX-test!

Anders Agebjörn

076-248 73 59

anders.agebjorn@nordlund.lu.se

Handledare: Anna Flyman Mattsson

046-222 87 01

anna.flyman_mattsson@nordlund.lu.se

MEDGIVANDE

Jag godkänner att inspelningen från mitt SWEDEX-test och mitt betyg används enligt villkoren ovan.

Datum

Namnsteckning

Namnförtydligande

MASTER'S PROJECT IN SPEAKING PROFICIENCY ASSESSMENT

Dear SWEDEX test taker!

I am a Swedish teacher and a master student at Lunds University. My master's project is on speaking proficiency assessment. Analyses of recordings from oral proficiency tests will be compared with the test takers grades. In that way I wish to contribute to the development of speaking proficiency tests.

Your participation is voluntary and your anonymity will be guaranteed. Please, contact me or my supervisor at Lunds University if you have any questions.

Good luck with your SWEDEX test!

Anders Agebjörn

076-248 73 59

anders.agebjorn@nordlund.lu.se

Supervisor: Anna Flyman Mattsson

046-222 87 01

anna.flyman_mattsson@nordlund.lu.se

PERMISSION

I accept that the recording from my SWEDEX test together with my grade can be used according to the conditions mentioned above.

Date

Signature

Clarification of signature

Bilaga 2: Transkriptionsnyckel

Fenomen	Transkriberas
Lång eller markant paus.	#
Tvekljud (olika varianter)	äh, ähm, eh, ehm, ah, ahm, öh, öhm, mh, mhm
Otydligt tal	xxx
Talaren upprepar sig.	< tal som upprepas > [/]
Talaren omformulerar eller rättar sig.	< tal som rättas/omformuleras > [//]
Talaren avbryter sig.	+//...
Talaren blir avbruten.	+//...

Bilaga 3: Tabeller

Tabell 19. Absoluta tal som kodningen av materialet resulterat i

Testtagare	Token*	AS-enheter	Satser	Huvud-satser	Bisatser	Relativ-satser	Inversions-kontexter
AAG	879	92	130	40	89	6	8
BBU	797	87	132	44	90	10	11
CBG	870	110	130	43	84	6	11
DBU	1070	123	174	64	108	7	13
EAG	1125	84	167	87	82	28	9
FAG	1414	150	219	84	138	7	25
GBG	1047	89	153	70	84	6	17
HBG	733	81	115	42	72	0	6
IBU	1054	104	170	72	94	7	15
JCU	701	93	105	26	79	2	16
KBG	953	99	135	42	95	14	9
LBU	885	74	120	47	82	14	4
MCU	852	99	121	45	83	1	9
NAG	1134	63	159	95	73	18	17
OBG	1292	99	178	94	84	34	14
PBU	955	104	158	70	89	14	21
QBG	548	41	83	43	43	11	9
RBG	988	98	144	50	103	8	10
SBU	1258	118	195	78	124	6	12
TBU	856	89	137	48	90	7	9

*Under "Token" anges hur många token (ord) testtagarna producerade under testet. Dessa data används inte i några mått men anges här eftersom det kan vara intressant att se hur "pratig" varje testtagare var.

Tabell 20. Mätvärden för alla testtagare

Test-tagare	1. AS-längd	2. Sats-längd	3. BS-andel	4. RS-andel	5. INV-kontext-andel	6. Sats-korr.	7. INV-korr.	8. PT-nivå
AAG	9,484	6,692	0,308	0,046	0,090	0,685	0,625	4
BBU	8,536	5,667	0,328	0,075	0,122	0,590	0,000	4
CBG	9,054	6,277	0,339	0,047	0,131	0,693	0,182	3
DBU	9,036	5,931	0,372	0,041	0,120	0,587	0,000	4
EAG	13,697	6,467	0,515	0,166	0,120	0,787	0,889	4
FAG	9,809	5,991	0,378	0,032	0,181	0,649	0,080	4
GBG	11,568	6,072	0,455	0,039	0,202	0,818	0,941	4
HBG	9,464	5,904	0,368	0,000	0,083	0,693	0,333	3
IBU	10,924	5,959	0,434	0,042	0,160	0,645	0,000	4
JCU	8,769	6,286	0,248	0,019	0,203	0,333	0,000	3
KBG	9,653	6,526	0,307	0,102	0,095	0,613	0,222	3
LBU	10,609	6,833	0,364	0,109	0,049	0,636	0,250	3
MCU	8,189	6,331	0,352	0,008	0,129	0,563	0,012	3
NAG	15,148	6,717	0,565	0,107	0,178	0,720	0,824	4
OBG	14,443	7,079	0,528	0,191	0,167	0,691	0,643	4
PBU	9,564	5,563	0,440	0,088	0,240	0,704	0,667	4
QBG	12,390	6,253	0,500	0,128	0,209	0,593	0,667	4
RBG	9,010	6,250	0,327	0,052	0,097	0,484	0,100	3
SBU	9,903	5,928	0,386	0,030	0,097	0,470	0,167	3
TBU	9,562	6,219	0,348	0,051	0,100	0,609	0,556	4

Tabell 21. Bedömningar och PT-nivå

3. Formell korrekthet och komplexitet	Slutbetyg	Testtagare	PT-nivå*	Predikativ kongruens	Inversion	
A	Godkänd	AAG	4	+	+	
		EAG	4	+	+	
		NAG	4	+	+	
		FAG	4	+	-	
B		QBG	4	+	+	
		OBG	4	+	+	
		GBG	4	-	+	
		RBG	3	-	-	
		CBG	3	-	-	
		KBG	3	-	-	
		HBG	3	-	-	
		Underkänd	PBU	4	+	+
			IBU	4	+	-
			BBU	4	+	-
TBU	4		-	+		
LBU	3		-	-		
DBU	3		-	-		
SBU	3		-	-		
C	MCU	3	-	-		
	JCU	3	-	-		

*För att nivå 4 ska anses uppnått måste testtagaren ha uppvisat ett fall av systematisk inversion och/eller predikativ kongruens.

Bilaga 4: Diagram

I denna bilaga återfinns diagram som signifikanta och marginellt signifikanta korrelationer mellan mått och mellan mått och bedömningar: korrekthetsmått i Bilaga 4.1; komplexitetsmått i Bilaga 4.2; korrekthets- och komplexitetsmått i Bilaga 4.3; komplexitets-/korrekthetsmått och PT-nivå i Bilaga 4.4; komplexitets-/korrekthetsmått och kriterium 3 i Bilaga 4.5; samt komplexitets-/korrekthetsmått och slutbetyg i Bilaga 4.6. Relationen mellan bedömning och PT-nivå redovisas inte i diagramform, utan i Tabell 21 ovan.

Bilaga 4.1 Korrekthetsmått

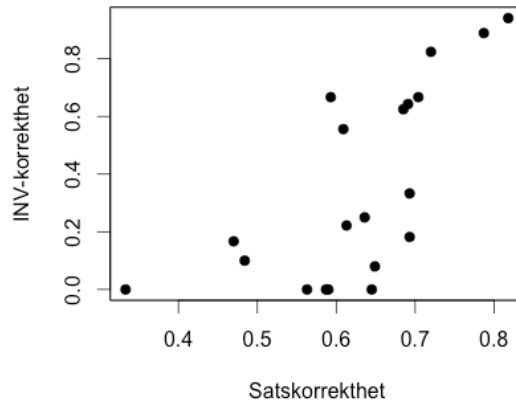


Diagram 10. INV-korrekthet~Satskorrekthet ($r=,69^{**}$)

Bilaga 4.2 Komplexitetsmått

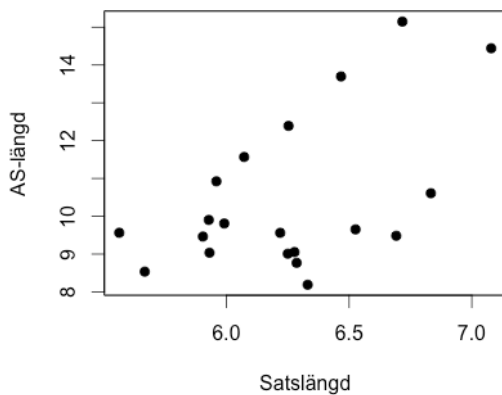


Diagram 11. AS-längd~Satslängd ($r=,52^*$)

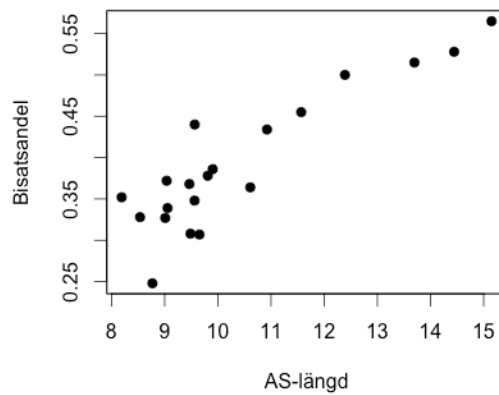


Diagram 12. AS-längd~BS-andel ($r=,90^{***}$)

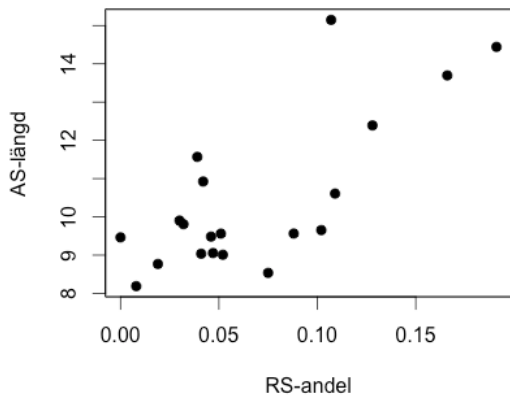


Diagram 13. AS-längd~RS-andel ($r=,75^{***}$)

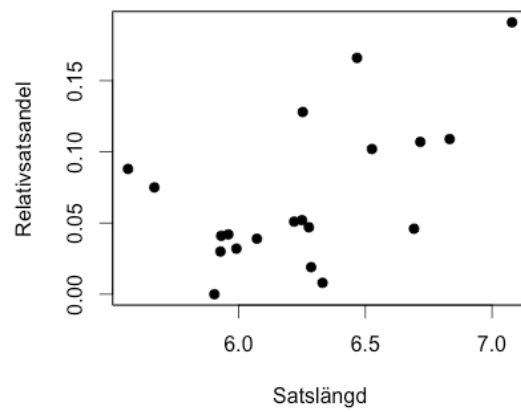


Diagram 14. RS-andel~Satslängd: ($r=,54^*$)

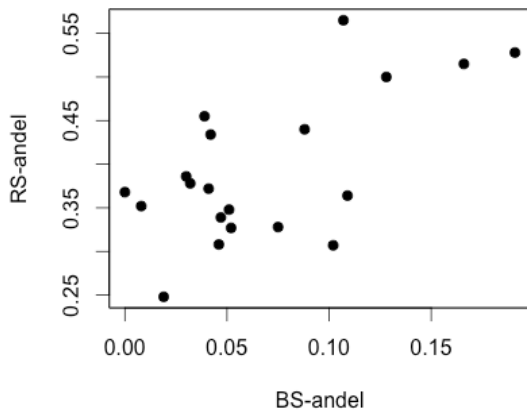


Diagram 15. RS-andel~BS-andel ($r=,64^{**}$)

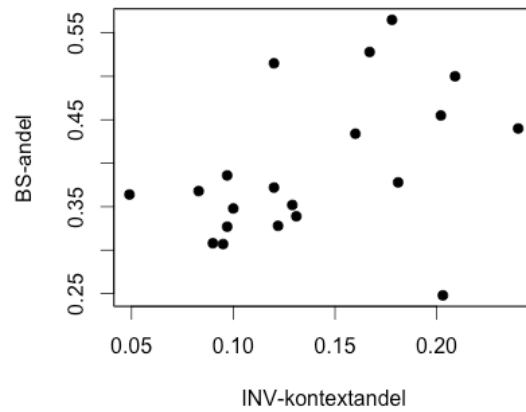


Diagram 16. BS-andel~INV-kontextandel ($r=,64^{**}$)

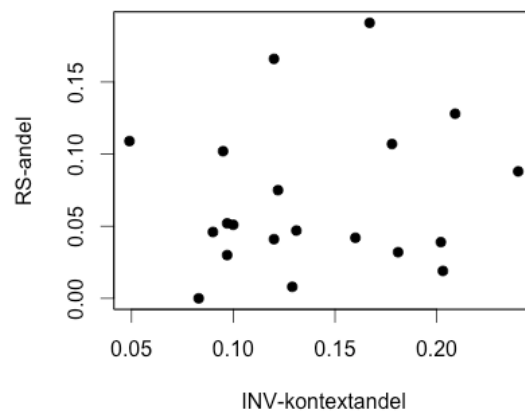


Diagram 17. RS-andel~INV-kontextandel ($r=,42.$)

Bilaga 4.3: Korrekthetsmått och komplexitetsmått

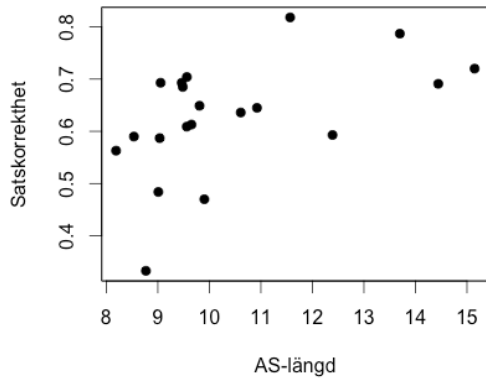


Diagram 18. Satskorr.~AS-längd ($r=,51^*$)

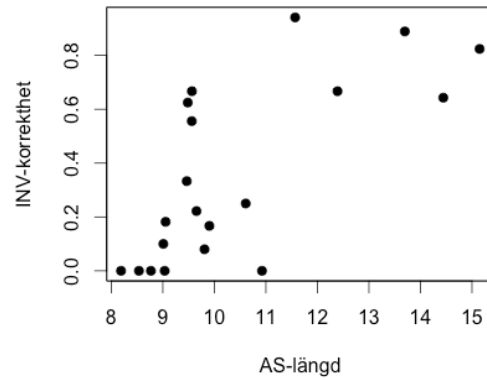


Diagram 19. INV-korr.~AS-längd ($r=,72^{***}$)

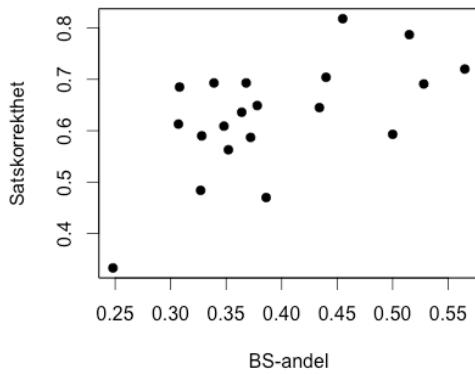


Diagram 20. Satskorr.~BS-andel ($r=,60^{**}$)

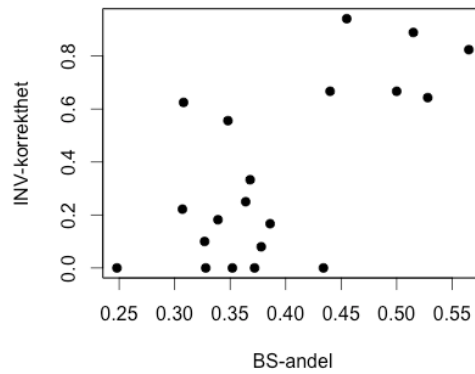


Diagram 21. INV-korr.~BS-andel ($r=,69^{**}$)

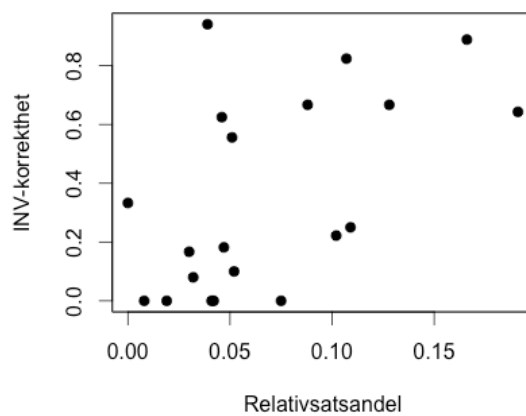


Diagram 22. INV-korr.~RS-andel ($r=,54^*$)

Bilaga 4.4: Komplexitets-/korrekthetsmått och PT-nivå

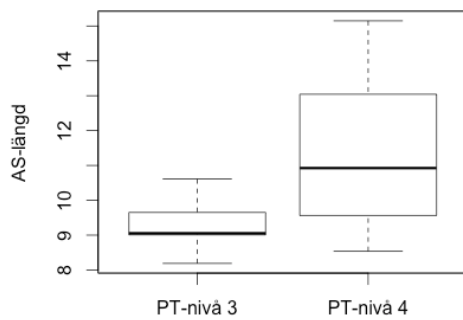


Diagram 23. AS-längd~PT-nivå ($r=,53^*$)

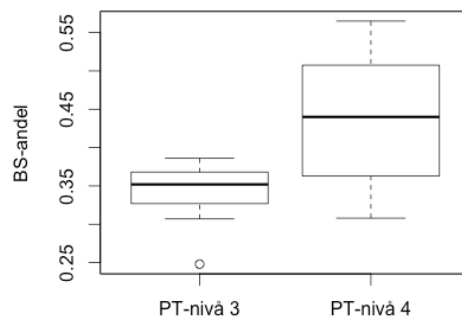


Diagram 24. BS-andel~PT-nivå ($r=,58^{**}$)

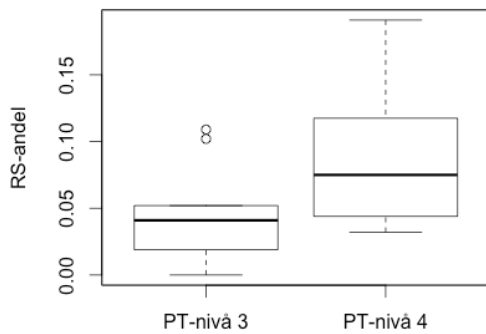


Diagram 25. RS-längd~PT-nivå ($r=,42.$)

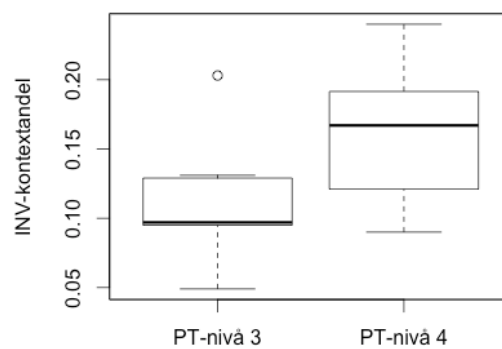


Diagram 26. INV-kontextandel~PT-nivå ($r=,49^*$)

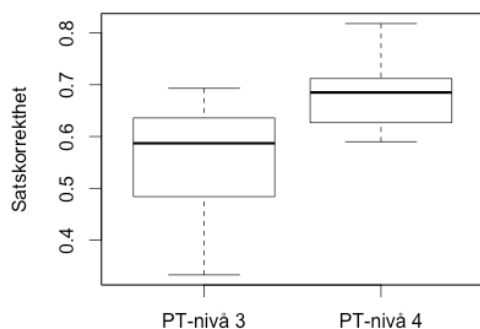


Diagram 27. Satskorr.~PT-nivå ($r=,54^*$)

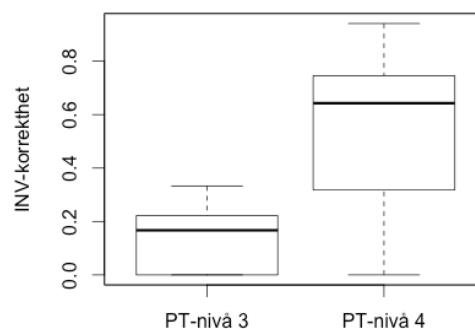


Diagram 28. INV-korr.~PT-nivå ($r=,61^{**}$)

Bilaga 4.5: Komplexitets-/ korrekthetsmått och kriterium 3

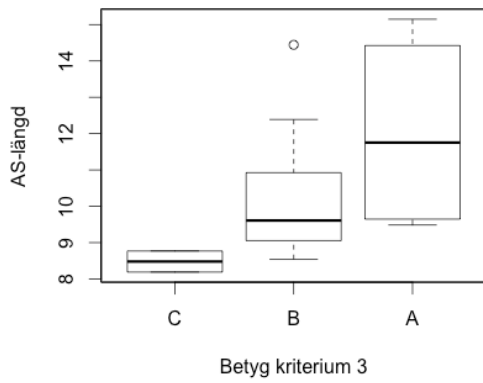


Diagram 29. AS-längd~Kriterium 3 ($p=,49^*$)

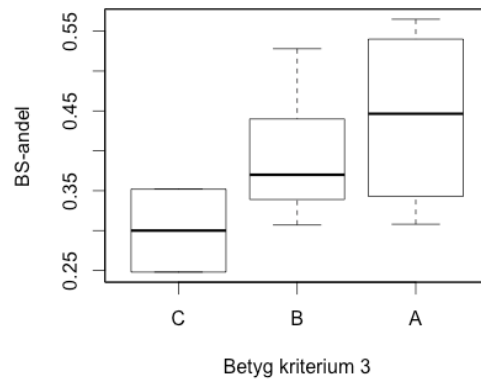


Diagram 30. BS-andel~Kriterium 3 ($p=,43.$)

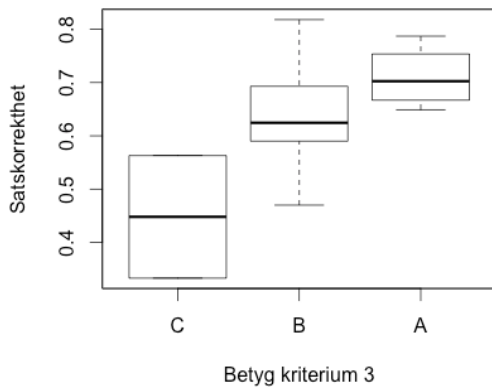


Diagram 31. Satskorrekthet~Kriterium 3
($p=,59^{**}$)

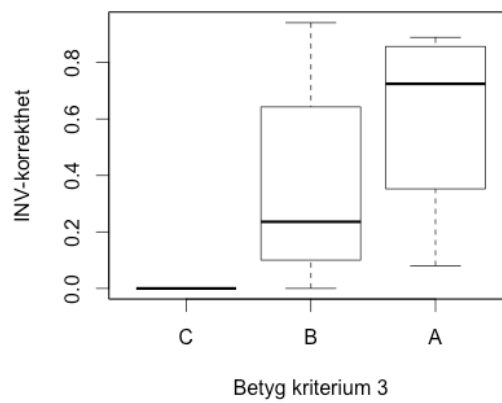


Diagram 32. INV-korrektthet~Kriterium 3
($p=,49^*$)

Bilaga 4.6: Komplexitets-/korrekthetsmått och slutbetyg

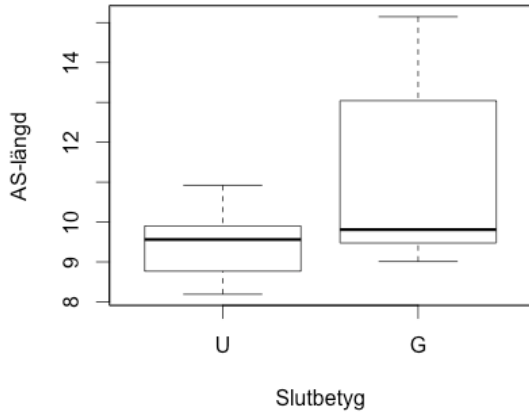


Diagram 33. AS-längd~Slutbetyg ($r=,46^*$)

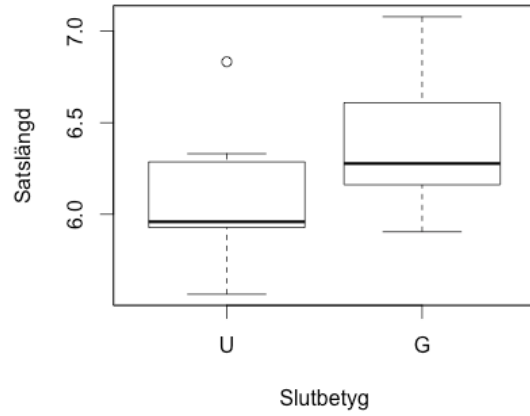


Diagram 34. Satslängd~Slutbetyg ($r=,40.$)

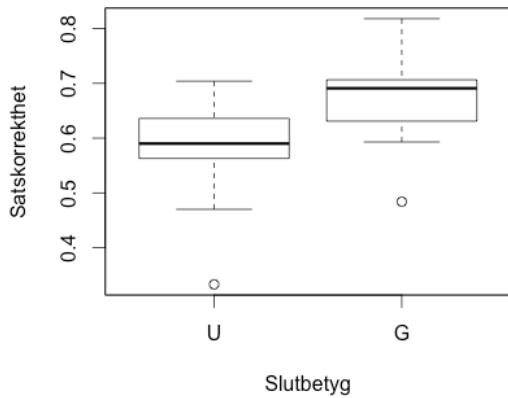


Diagram 35. Satskorr.~Slutbetyg ($r=,48^*$)

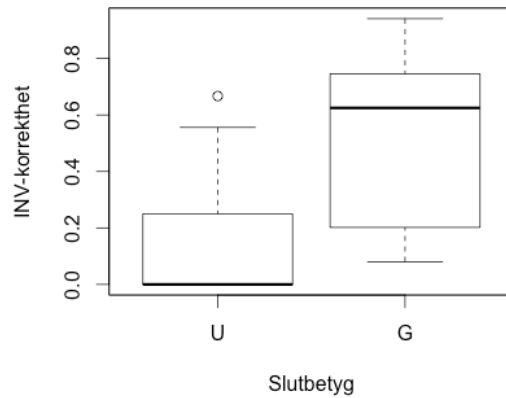


Diagram 36. INV-korr.~Slutbetyg ($r=,49^*$)