

# Käppen i hjulen



**Behärskning av svenska  
konventionaliserade uttryck bland gymnasieelever  
med varierande språklig bakgrund**

**Julia Prentice**

© ROSA Institutionen för svenska språket och författaren, 2010.

**Institutet för svenska som andraspråk,  
Institutionen för svenska språket, Göteborgs universitet**

# Käppen i hjulen

– Behärskning av svenska  
konventionaliserade uttryck bland gymnasieelever  
med varierande språklig bakgrund

Julia Prentice



ROSA (Rapporter Om Svenska som Andraspråk) ges ut av Institutet för svenska som andraspråk, Institutionen för svenska språket, Göteborgs universitet. Serien omfattar vetenskapliga arbeten av olika slag inom fältet svenska som andraspråk. Frågor och synpunkter är välkomna, och kan riktas direkt till författaren eller till redaktörerna: ROSA, Institutet för svenska som andraspråk, Institutionen för svenska språket, Göteborgs universitet, Box 200, 405 30 Göteborg eller på e-post: [rosa@svenska.gu.se](mailto:rosa@svenska.gu.se). Hit kan också bidrag för publicering skickas.

TIDIGARE UTGIVNA RAPPORTER:

- Anna-Britta Wallerstedt (1997) *Den receptiva ordförståelsen hos invandrarelever och deras inlärningsituation.*
- Monica Reichenberg Carlström (1998) *Koherens, röst och läsning på ett andraspråk.*
- Roger Källström (1999) *Svenska som andraspråk – lärarkompetens och lärarutbildningsbehov.*
- Suzanne Nordin-Eriksson (under medverkan av Anna Kumlin) (2000) *Inlärarautonomi speciellt ifråga om lågutbildade andraspråksinlärare.*
- Uno Källtén (2001) *Analys av Skolverkets rapporter och trycksaker under åren 1994 och 1999.*
- Ulla Sundemo och Monica Nilsson (2004) *Barnboksfiguren – en tillgång på olika plan.*
- Inger Lindberg och Karin Sandwall (red) (2006) *Språket och kunskapen – att lära på sitt andraspråk i skola och högskola.*
- Inger Lindberg och Sofie Johansson Kokkinakis (red) (2007) *OrdiL – en kartläggning av ordförrådet i läromedel för grundskolans senare år.*
- Qarin Franker (2007) *Bildval i alfabetiseringsundervisning – en fråga om synsätt.*
- Inga-Lena Rydén (2007) *Litteracitet och sociala nätverk ur ett andraspråksperspektiv.*
- Marie Carlson (2007) *Språk och gränser – Om språk och identitetsskapande i några skönlitterära verk*

ROSA nr 6 och senare nummer finnas tillgängliga via Göteborgs universitetsbibliotek genom länken: <http://gupea.ub.gu.se/dspace/handle/2077/19158>

© Författaren och Institutionen för svenska språket.

Institutet för svenska som andraspråk  
Institutionen för svenska språket  
Göteborgs universitet  
Box 200, 405 30 GÖTEBORG

OMSLAGSFOTO: Chaplin archives © Roy Export SAS

SÄTTNING: Janne Saaristo

TRYCK: Reprocentralen, Humanisten, Göteborgs universitet

# Abstract

This study focuses on the mastery of Swedish conventionalized expressions among adolescents with various linguistic backgrounds. The data have been collected by means of a survey that contains four tests focusing on various aspects of the mastery of conventionalized expressions. The quantitative results show that L1 students reach significantly higher test results than L2 students. Furthermore, L2 students with an early age of onset reach significantly higher results than the later L2 learners. Thus, age of onset appears to be an important factor for the participants' mastery of conventionalized expressions in Swedish. This conclusion is also confirmed by the qualitative analysis of part of the data, which shows that L2 students refrain from solving the task to a higher degree than L1 students. Moreover, the participants seem to associate between individual components of two or several different expressions in a way that is in line with certain theoretical assumptions concerning the mental representation of lexical units.

# Sammandrag

Denna studie fokuserar behärskning av svenska konventionaliserade uttryck bland ungdomar med varierade språklig bakgrund. Materialet har samlats in med hjälp av en enkät som innehåller fyra test med fokus på olika aspekter av behärsningen av konventionaliserade uttryck. De kvantitativa resultaten visar att förstaspråkseleverna når signifikant högre resultat än andraspråkseleverna. Andraspråkseleverna med tidig startålder når dessutom signifikant högre resultat än de senare andraspråksinlärarna. Startålder visar sig alltså vara en viktig faktor för informanternas behärskning av konventionaliserade uttryck i svenskan. Denna slutsats bekräftas även av den kvalitativa analysen av en del av materialet, som visar att andraspråkseleverna i högre grad avstår från att besvara uppgifterna än förstaspråkseleverna. Den kvantitativa analysen ger vissa indikationer på att informanterna verkar associera mellan enskilda komponenter av två eller flera olika uttryck på ett sätt som ligger i linje med vissa teoretiska antaganden om mental representation av lexikala enheter.

# Förord

Arbetet med själva undersökningen och denna rapport hade inte kunnat genomföras utan en rad olika människor som varit väldigt generösa med sin tid och energi.

Först och främst vill jag tacka de elever som har deltagit i enkätundersökningen och de lärare som delat med sig av sin lektionstid! Utan er hade undersökningen helt enkelt inte blivit av.

Jag vill också särskilt tacka Margit Nelson Wareborn och Carina Carlund som har öppnat viktiga dörrar för mig.

Ett stort tack till min handledare Inger Lindberg och mina bihandledare Kennet Hyltenstam och Emma Sköldberg som har läst mycket text och kommit med kloka och värdefulla synpunkter.

Roger Källström har också läst och kommenterat hela texten i slutskedet, ett hjärtligt tack för det!

Tommy Johnsson har hjälpt mig med de statistiska analyserna och Janne Saaristo har skött layouten på denna skrift, ett stort tack även för det!

Slutligen vill jag tacka redaktörerna Sofie Johansson Kokkinakis och Sofia Tingsell för allt jobb de har lagt ned på denna publikation!

Göteborg, juni 2010

Julia Prentice

# Innehåll

1. Inledning.....	1
1.1. Syfte och frågeställningar .....	2
2. Idiomaticitet, språklig konvention och konventionaliserade uttryck..	3
2.1. Idiomaticitet .....	3
2.2. Konventionalisering, kompositionalitet och genomskinlighet..	5
2.3. Idiom.....	8
2.4. Kollokationer .....	9
2.4.1. Kontextualistisk kollokationsforskning.....	10
2.4.2. Systemorienterad kollokationsforskning.....	12
3. Lexikal och fraseologisk kunskap	
– vad kan man när man ”kan” ett uttryck?.....	15
3.1. Vad kan man när man kan ett ord? .....	16
3.2. Att mäta lexikonkunskap .....	21
4. Konventionaliserade uttryck i språkutveckling .....	25
4.1. Förstaspråkutveckling .....	25
4.2. Andraspråkutveckling.....	27
4.3. Mental representation av konventionaliserade uttryck .....	29
4.4. Konventionaliserade uttryck i andra- och främmandespråksinläring .....	32
4.4.1. Studier inom andraspråksforskning .....	33
4.4.2. Korpuslingvistiska studier inom främmandespråksinläring.....	38
4.5. Konventionaliserade uttryck i utbildningsrelaterat språk.....	39

5. Material och metod .....	41
5.1. Etnisk och social bakgrund som påverkande faktorer i samband med elevers skolresultat .....	41
5.2. Skolor och elevsammansättning i deltagarklasserna.....	43
5.3. Informantgrupperna .....	46
5.4. Enkäten .....	48
5.5. Analysmetod.....	52
6. Resultat .....	55
6.1. Resultaten för de deltagande gymnasieklasserna.....	55
6.2. Resultaten i förhållande till språklig bakgrund .....	58
6.2.1. Resultatspridning.....	60
6.3. De olika deltesten .....	64
6.3.1. Skillnader mellan delresultaten inom varje grupp .....	67
6.4. De enskilda uttrycken .....	68
6.4.1. Förståelse .....	68
6.4.2. Lucktest i kontext .....	71
6.4.3. Lucktest utanför kontext.....	75
6.4.4. Kollokationer.....	79
6.5. Validitet .....	83
6.6. Kvalitativ analys.....	84
6.6.1. De svåraste uttrycken .....	86
6.6.1.1. hålla någon i schack .....	87
6.6.1.2. Sätta käppar i hjulen .....	89
6.6.1.3. få upp ögonen för något .....	91
6.6.1.4. på rak arm .....	93
6.6.1.5. livet leker för någon .....	95
6.6.1.6. med något i högsta hugg .....	97
6.6.1.7. klättra på väggarna .....	98
6.6.1.8. få någon/något på halsen .....	100
6.6.1.9. lyckligt ovetande .....	101
6.6.1.10. inte för fem öre .....	103
6.6.2. De vanligaste avvikelserna .....	105
6.6.3. Enkäterna med de högsta resultaten .....	107



7. Sammanfattning och slutsatser.....	111
7.1. Kvantitativ analys .....	111
7.1.1. Skillnader mellan deltesten.....	112
7.2. De enskilda uttrycken .....	113
7.3. Den kvalitativa analysen.....	115
7.3.1. Lexikala komponenter, kontext och mental representation .....	116
8. Avslutande diskussion.....	121
Referenser.....	125
Övriga källor .....	134
Appendix.....	135

## Tabeller och figurer

TABELL 1:	Olika aspekter av lexikonkunskap .....	18
TABELL 2:	Bakgrundsdata för deltagarklasserna.....	44
TABELL 3:	Studieresultat för relevanta utbildningar i årskurs 3 för hela landet i procent .....	45
TABELL 4:	Antal elever från varje årskurs i grupperna .....	47
TABELL 5:	Testresultat per gymnasieklass .....	57
TABELL 6:	Resultat för hela enkäten) per startåldersgrupp .....	59
TABELL 6.I:	Detaljerad redovisning av resultat för informanterna med svenska som L2 .....	60
TABELL 7:	Resultat för enkätens olika deltester (% av maximalt resultat) .....	65
TABELL 7.I:	Delresultaten i fallande ordning (% av maximalt resultat) per grupp. ....	67
TABELL 8:	Resultat på förståelsetest i del A per uttryck (i fallande ordning) .....	72
TABELL 8.I:	Resultat på lucktestet i del B per uttryck (i fallande ordning) .....	74
TABELL 8.2:	Resultat på lucktestet i del C per uttryck (i fallande ordning) .....	77

TABELL 8.3:	Resultat på matchningstestet i del D per uttryck (i fallande ordning) .....	80
TABELL 9:	Antal olika typer av avvikelser i samband med uttrycket <i>hålla någon i schack</i> per informantgrupp .....	87
TABELL 9.1:	Antal olika typer av avvikelser i samband med uttrycket <i>sätta käppar i hjulen</i> per informantgrupp .....	89
TABELL 9.2:	Antal olika typer av avvikelser i samband med uttrycket <i>få upp ögonen för något</i> per informantgrupp .....	91
TABELL 9.3:	Antal olika typer av avvikelser i samband med uttrycket <i>på rak arm</i> per informantgrupp .....	93
TABELL 9.4:	Antal olika typer av avvikelser i samband med uttrycket <i>livet leker för någon</i> per informantgrupp .....	95
TABELL 9.5:	Antal olika typer av avvikelser i samband med uttrycket <i>med något i högsta hugg</i> per informantgrupp .....	97
TABELL 9.6:	Antal olika typer av avvikelser i samband med uttrycket <i>klättra på väggarna</i> per informantgrupp.....	98
TABELL 9.7:	Antal olika typer av avvikelser i samband med uttrycket <i>få någon/något på halsen</i> per informantgrupp .....	100
TABELL 9.8:	Antal olika typer av avvikelser i samband med uttrycket <i>lyckligt ovetande</i> per informantgrupp .....	101
TABELL 9.9:	Antal olika typer av avvikelser i samband med uttrycket <i>inte för fem öre</i> per informantgrupp .....	103
TABELL 10:	Antal olika typer av avvikelser för samtliga tio uttryck per informantgrupp .....	105
TABELL 10.1:	Antal olika typer av avvikelser för samtliga tio uttryck per informantgrupp – utökad material.....	106
TABELL II:	Typer av avvikelser i de 20 enkäterna med högst resultat.....	109

FIGUR 1:	Exempel på relationer mellan ett idiom och dess lexikala komponenter.....	30
FIGUR 2:	Jämförelse av resultatspridningen mellan informantgrupperna .....	61
FIGUR 3.1:	Antal informanter som nått olika resultat i grupp L1 .....	63
FIGUR 3.2:	Antal informanter som nått olika resultat i grupp L2/tidig .....	63
FIGUR 3.3:	Antal informanter som nått olika resultat i grupp L1 .....	64
FIGUR 4:	Nätverk mellan de kontaminerade uttrycken och deras lexikala komponenter .....	117
FIGUR 4.1:	Nätverk mellan olika konventionaliserade uttryck och deras lexikala och semantiska komponenter .....	118

# 1. Inledning

Användningen av konventionaliserade uttryck utgör ofta ett av de stora hindren på vägen mot en inföddlik språkbehärskning för andraspråksinlärare och behärskning av olika typer av fasta fraser och konventionaliserade uttryck har alltmer lyfts fram som central i samband med både första- och andraspråksinlärning (Wray 2002, Abrahamsson & Hyltenstam 2009, Prentice & Sköldberg und.utg.).

Rimligtvis är det viktigt för inlärare att behärska de mest frekventa uttrycken i vardagligt språk såväl som i ett mer domänspecifikt språkbruk – t.ex. utbildningsrelaterat språk. Vilka konventionaliserade uttryck som är intressanta och relevanta ur språkbeskrivnings- och språkinlärningsperspektiv är emellertid inte bara en fråga om frekvens. Även mer lågfrekventa uttryck, som t.ex. idiom, fyller viktiga funktioner för språkanvändaren och förtjänar större uppmärksamhet än att endast betraktas och läras ut som språkspecifika kuriositeter (McCarthy 1998, O’Keefe, McCarthy & Carter 2007:80ff). O’Keefe, McCarthy & Carter påpekar att idiomatiska uttryck inte i första hand är ”colourful alternatives to their literal counterparts” (2007:82) utan att de innehåller viktig kulturell information samtidigt som de fyller diskursiva funktioner, som t.ex. att uttrycka författarens värderingar (McCarthy 1998: 131f).

Detta arbete fokuserar olika typer av det som här kallas för *konventionaliserade uttryck* (se avsnitt 2.2). En relevant typ av uttryck är den som bl.a. Ellis, Simpson-Vlach & Maynard (2008) kallar *formulaic expression* och som kan vara relativt högfrekvent i allmänspråket såväl som

i det språkbruk som elever möter i skolan. Som exempel kan uttryck som *på egen hand* eller *dra en slutsats* nämnas. I undersökningen tas dock också generellt mer lågfrekventa språkliga fenomen upp – t.ex. idiom som *prata i nattmössan* eller *sätta käppar i hjulen*.

## 1.1. Syfte och frågeställningar

Den undersökning som redovisas i denna rapport är en enkätstudie som genomförts bland gymnasieelever i flerspråkiga skolmiljöer. Studiens övergripande syfte är att undersöka i vilken mån elever med varierande språklig bakgrund behärskar, dvs. känner igen, förstår och kan producera, några olika typer av svenska konventionaliserade uttryck som de kan förväntas möta och använda. De frågeställningar som ligger till grund för undersökningen är:

- Finns det skillnader i behärskning av konventionaliserade uttryck mellan olika elevgrupper
  - som kan relateras till elevernas språkliga bakgrund, dvs. modersmål och startålder för inlärning e.d.?
  - vad gäller olika aspekter av lexikal kompetens, t.ex. förståelse och produktion?
- Finns det skillnader i behärskning av enskilda konventionaliserade uttryck mellan olika elevgrupper?
- På vilket sätt avviker eleverna från konventionaliserade mönster i sin produktion av konventionaliserade uttryck i samband med denna undersökning?

Frågorna under första punkten kommer här att undersökas med hjälp av en kvantitativ analys. Även den andra punkten utforskas med hjälp av en kvantitativ analys, som dock samtidigt ligger till grund för en kvalitativ analys av ett antal enskilda uttryck. Denna kvalitativa analys svarar mot den tredje frågeställningen.

## 2. Idiomaticitet, språklig konvention och konventionaliserade uttryck

I detta avsnitt introduceras några fraseologiska begrepp som även har haft viss relevans för forskning om andraspråksinlärning och är grundläggande för de resonemang som förs i detta arbete. Inledningsvis tas begreppet *idiomaticitet* upp, som är en viktig utgångspunkt i analysen av konventionaliserade språkliga strukturer. *Konventionalisering*, *kompositionalitet* och *genomskinlighet* kan sägas utgöra grundläggande begrepp för beskrivningen av konventionaliserade uttryck. Dessutom introduceras även två huvudtyper av konventionaliserade uttryck, nämligen *idiom* och *kollokationer*.

### 2.1. Idiomaticitet

Den kanske mest grundläggande aspekten i samband med behärskning och användning av konventionaliserade uttryck är det som särskilt i engelskspråkig litteratur ofta kallas *idiomaticitet* (eng. *idiomaticity*). Vissa forskare använder även termer som *inföddlik* (eng. *nativelike*) *språkanvändning* med liknande innebörd (se Pawley & Syder 1983). Idiomaticitet är ett tämligen svårdefinierat begrepp, vilket framgår av följande citat:

Idiomaticity is a non-phonological “accent”, not always attributable to surface language errors, but to a certain undefined quality which many frustrated L2 composition teachers define as “I don’t know what’s wrong with this, but we just don’t say that in English” (Yorio 1989:64).

Detta definitionsförsök gör tydligt att det är svårt att komma åt idiomaticitetens kärna och vad exakt det är som avgör om ett språkligt yttrande betraktas som idiomatiskt. Yorio framhäver här dessutom att grammatisk korrekthet inte är en lämplig grund för bedömning av idiomaticitet i språkbruk. Även Wiktorsson (2003:3) betonar vikten av att skilja idiomaticitet från grammatisk korrekthet i samband med sin korpuslingvistiska studie av idiomatiska uttryck i texter skrivna av svenska inlärare av engelska. En diskussion kring denna distinktion förs även av Pawley & Syder (1983:196), som anför en rad grammatiskt korrekta men icke idiomatiska sätt att uttrycka samma innehåll som i den konventionaliserade frasen *I want to marry you*, t.ex. *I wish to be wedded to you*, *I want marriage with you* och *my becoming your spouse is what I want*. Pawley & Syder skriver vidare att endast en liten andel av de grammatiskt korrekta yttrandena i ett språk också är idiomatiska och därmed skulle accepteras som naturliga av infödda talare. De påpekar också att infödda talare inte på något sätt till fullo utnyttjar sin förmåga att skapa språkliga yttranden genom kreativ användning av syntaktiska regler. Det verkar alltså just vara återanvändningen av vissa konventionaliserade strukturer som utmärker en inföddlik – eller idiomatisk – språkanvändning (jfr Ekberg 1999, 2004).

Wray skriver om språkbrukarens kapacitet att hantera nya, kreativt skapade yttranden, både produktivt och receptivt. Hon betonar att man som mest kan hävda att språkbrukarens benägenhet att använda sig av en relativt stor andel färdigbildade, konventionaliserade fraser existerar parallellt med hans eller hennes förmåga att skapa och förstå nya ordförbindelser:

Although we customarily say, *Hi, how are you doing*, or some other idiomatic greeting on meeting a friend, there is nothing at all to stop us from saying, *What a pleasant event it is to see you. Tell me, how is your life progressing at the moment* (2002:12).

Ett återkommande problem i samband med behandlingen av begreppet idiomaticitet är dock att det används med något olika innebörd av olika

forskare. Ofta används det i den betydelse som beskrivs ovan, dvs. avseende den kvalitet i våra språkliga val som gör att de accepteras som naturliga av infödda talare (Yorio 1989:64). I andra fall skriver forskare om idiomaticitet som en egenskap hos ett uttryck som anger dess fixering till sin form och brist på genomskinlighet i betydelse. Med andra ord skulle man kunna hävda att ju mer idiomatiskt ett uttryck är, desto mer fixerat är det till sin form och desto mindre genomskinligt är det i sin betydelse (jfr Coulmas 1979, Gläser 1986, 1988, Fleischer 1997). Fleischer beskriver olika grader av idiomaticitet hos fraseologiska enheter som han kallar för *fraseologismer* (ty. *Phraseologismen*). Han ser graden av idiomaticitet hos ett givet uttryck som ett mått på konstituenternas semantiska förenlighet (kongruens) inom ramen för detta uttryck. Enligt detta synsätt uppvisar uttrycket *dra alla över en kam* 'bedöma alla enligt samma mall' (Svenskt språkbruk 2003:576) en hög grad av idiomaticitet, då konstituenterna inte är semantiskt kongruenta med avseende på deras bokstavliga<sup>1</sup> betydelser. Dessutom bestäms graden av idiomaticitet hos ett uttryck av huruvida en eller flera av konstituenterna används i sin bokstavliga betydelse (se avsnitt 2.2 nedan). Som exempel kan här nämnas *prata i nattmössan* 'inte veta vad man talar om' (Svenskt språkbruk 2003:804f), där verbet *prata* används i sin bokstavliga betydelse, vilket, ur denna synvinkel, innebär att uttrycket uppvisar en lägre grad av idiomaticitet än *dra alla över en kam* (jfr Fleischer 1997:30f). I detta arbete används emellertid termen idiomaticitet i den betydelsen som antyds hos bl.a. Yorio (1989:64); idiomatiskt språkbruk är ett sätt att använda språket, som är naturligt ur den infödda talarens perspektiv.

## 2.2. Konventionalisering, kompositionalitet och genomskinlighet

I likhet med tidigare arbeten (Große 2008, Prentice & Sköldberg und. utg.) använder jag termen *konventionalisering* för att beskriva språkbru-

---

<sup>1</sup> Fleischer använder termerna *wendungsintern* och *wendungsextern*, som man skulle kunna översätta med 'uttrycksintern' och 'uttrycksextern', dvs. de betydelser som konstituenterna har inom respektive utom det idiomatiska uttrycket ifråga.



karnas benägenhet att uttrycka ett visst språkligt innehåll med en viss (mer eller mindre fixerad) språklig form. Tillämpningen av begreppet är, som Svanlund påpekar, inte helt oproblematiskt med tanke på att konventionalisering måste ses som en ständigt pågående, dynamisk process, vilket innebär att uppsättningen av etablerade mönster kontinuerligt förändras (Svanlund 2001:33). Dessutom skriver Svanlund att innehållet i begreppet *språklig konvention* eller *konventionalisering* inte är lika självklart som många språkvetare verkar anta. Vad man menar beror delvis på vilka motsatsrelationer man utgår ifrån. Svanlund nämner bl.a. språklig konvention som ett kollektivt, socialt styrt fenomen i motsats till de psykologiskt betingade, individuellt särpräglade aspekterna av språkanvändningen. Den motsatsrelation som är relevant i samband med detta arbete är språklig konvention som etablerade mönster i språkbruket gentemot nyskapande eller tillfälliga, dvs. ”mer perifert eller hittills inte alls nyttjade uttrycksmöjligheter” (Svanlund 2001:11). Om man anammar det sistnämnda perspektivet, har användningen av begreppet konventionalisering, enligt min åsikt, flera fördelar. Dels riskerar man ingen begreppsförvirring i samband med idiomaticitetsbegreppet, dels framhäver begreppet även den diakrona aspekten som har betydelse för användningen av fraseologiska enheter; konventionaliserade uttryck uppstår genom att språkbrukarna inom en språkgemenskap upprepade gånger gör samma språkliga val som därmed så småningom etableras som ett naturligt sätt att uttrycka ett visst innehåll. I anslutning till detta resonemang väljer jag även att använda termen *konventionaliserat uttryck* för att beteckna föremålet för denna undersökning. Begreppet används i en vid bemärkelse och innefattar olika typer av uttryck som idiom, kollokationer och andra typer av i viss mån fixerade flerordsenheter.

Det är i praktiken knappast meningsfullt att betrakta olika typer av fraseologiska enheter som helt olika, från varandra isolerade fenomen, då skillnaderna emellan dem i många fall måste betraktas som gradueLLa. Olika typer av konventionaliserade uttryck befinner sig snarare längs ett kontinuum; i ena ändan finns det som av vissa forskare har kallats för *rena* eller *prototypiska idiom* (t.ex. Cowie 1988, Sköldberg 2004), som är starkt fixerade i sin form och semantiskt icke *kompositionella* (se vidare i detta avsnitt nedan) – t.ex. *ta tjuren vid hornen* ’direkt och oförskräckt ta itu med något som är svårt eller obehagligt’ (Svenskt språk-

bruk 2003:1211). I andra ändan ligger sådana uttryck som Moon kallar för *frames* (1998:145ff). Dessa har en viss återkommande struktur men kan lexikalt realiseras relativt fritt. Här kan *få kalla kårar*, *få kalla kårar längs ryggen/rygggraden* eller *det går kalla kårar längs ryggen/rygggraden* 'när någon ryser till av rädsla eller obehag' nämnas som exempel (Svenskt Språkbruk 2003:672, jfr Moon 1998:120ff, Sköldberg 2004:174ff). Däremellan finns det bl.a. relativt fasta konstruktioner med ett variabelt led, t.ex. *hålla modet/humöret uppe* 'stå ut med svårigheter, vara vid gott mod' (Luthman 2002:56).

Ett annat relevant begrepp i samband med denna undersökning är *kompositionalitet*. Partee beskriver kompositionalitetsprincipen som följer:

The meaning of an expression is a function of the meaning of its parts and of the way they are syntactically combined (Partee 1984:281).

Svanlund (2009:34) nämner andra, betydligt striktare formuleringar av denna princip, som innebär att betydelsen hos ett flerordsuttryck enbart utgör summan av de ingående elementens betydelse samt de strukturella relationerna emellan dem. Ett exempel på en sådan radikal formulering återfinns hos Costello & Keane:

According to the principle of compositionality, anyone who knows the meaning of each word in a complex expression should need no further information to grasp the meaning of the expression (Costello & Keane 2000:336).

Kompositionalitetsbegreppet ligger nära det som i svensk litteratur ofta beskrivs som grad av semantisk genomskinlighet, dvs. i vilken mån det är möjligt att härleda frasens betydelse utifrån de enskilda konstituenternas betydelse. Svanlund (2009:36f) diskuterar begreppet *genomskinlighet* i förhållande till en sådan strikt syn på kompositionalitet. Genomskinlighet anges ofta som en av de centrala egenskaper hos ett uttryck (i Svanlunds fall sammansättning) som är nära kopplat till adjektiv av typen *förutsägbar* eller *härledbar*. Svanlund själv använder begreppet genomskinlighet i sin undersökning av hur nya sammansättningar etableras i språket på ett, som han anser, mer preciserat sätt:

Om vi uppfattar en sammansättning som genomskinlig så innebär det att vi lätt kan generera en föredragen defaultbetydelse. Detta innebär inte att andra betydelser är otänkbara, men en viss betydelse uppfattas som mer naturlig än andra kandidater och aktiveras därmed först – om ingenting i sammanhanget indikerar annorlunda (Svanlund 2009:36f).

Om ett uttryck är genomskinligt eller inte är alltså, enligt detta synsätt, en fråga om språkbrukarens upplevelse, vilket bl.a. innebär att bedömningen av ett uttrycks genomskinlighet – som är en del av den kvalitativa analysen i detta arbete – inte är helt oproblematiskt. Det är, som Svanlund påpekar, inte möjligt att komma åt den ”genuina upplevelsen” (2009:36) av genomskinlighet som en läsare eller lyssnare kan ha vid första mötet med ett visst uttryck, om man redan vet vad uttrycket betyder. Man måste alltså använda sin språkliga intuition för att avgöra hur pass genomskinligt uttrycket är för någon som läser eller hör det för första gången. Det som ska avgöras är ofta, inte minst i samband med denna undersökning, om läsaren, lyssnaren eller informanten kan förväntas tolka uttrycket på det konventionaliserade sättet. I detta sammanhang bör det också påpekas att genomskinlighet – liksom kompositionalitet – i enlighet med Svanlund bör ses som ett graduellt fenomen. Med andra ord kan den konventionaliserade betydelsen hos ett uttryck vara mer eller mindre uppenbar (Svanlund 2009:37).

## 2.3. Idiom

En typ av konventionaliserade uttryck som har beskrivits och definierats på olika sätt är *idiom*. O’Keefe, McCarthy & Carter(2007:62) påpekar att en stor del av forskningen kring idiom har fokuserat de mest icke-kompositionella idiomatiska uttrycken. Dessa kallas ibland för *prototypiska* idiom (jfr Fernando & Flavell 1981, Sköldberg 2004). Som svenska exempel på sådana idiom nämner Sköldberg *väcka ont blod* och *i ett nötskal* (2004:26). I engelskspråkig litteratur anförs gärna bl.a. *kick the bucket* och *pull someone’s leg* som typexempel för idiom (Moon 1998:4, Wray 2002:4). Prototypiska idiom är sådana som har fler av de egenskaper som man betraktar som utmärkande för just idiom än andra, mindre prototypiska idiom (Sköldberg 2004:26). Även här blir det alltså

tydligt att idiom är ett svåravgränsat fenomen med mer eller mindre prototypiska exempel snarare än en homogen kategori. De för idiomens utmärkande egenskaper som Sköldberg framhäver i sin definition är bl.a. en relativt fixerad form, figurativ betydelse, icke-kompositionalitet och konventionalisering som hon kallar för *institutionalisering* (jfr Moon 1998:7). Vissa forskare har dock använt idiombegreppet i så pass vid bemärkelse att det innefattar olika typer av konstruktioner som inte är fritt formade, som kollokationer (se avsnitt 2.4 nedan), men även icke-kompositionella sammansättningar som *blackbird* (Makkai 1972, jfr Moon 1998:4). O’Keefe, McCarthy & Carter (2007:62) refererar till ”the ordinary idioms of every day” och avser bl.a. öppningsfraser i samtal (eng. *gambits*) av typen *guess what*, och diskurspartiklar (jfr Coulmas 1979).

Jag utgår däremot i samband med denna undersökning, liksom i tidigare arbeten (Prentice & Sköldberg und. utg.), från en definition som återfinns hos Langlotz (2006):

An idiom is an institutionalised construction that is composed of two or more lexical items and has the composite structure of a phrase or semi-clause, which may feature constructional idiosyncrasy. An idiom primarily has an ideational discourse-function and features figuration, i.e. its semantic structure is derivationally non-compositional [or not fully compositional]. Moreover, it is considerably fixed and collocationally restricted (Langlotz 2006:5)

Langlotz kallar denna vida definition för preliminär (ibid.) och även han är noggrann med att påpeka att gränsdragningen mellan idiom och intilliggande fraseologiska enheter på intet vis är skarp (jfr O’Keefe, McCarthy & Carter 2007:81f). Langlotz (2006:6) menar vidare att en sådan definition kan inkludera så pass olika konstruktioner som *red herring* (’villospår’), *take the bull by the horns* (’ta tjuren vid hornen’) och *fall flat* (’misslyckas’) (Prentice & Sköldberg und. utg.).

## 2.4. Kollokationer

En annan typ av fraseologiska enheter som fokuseras i denna undersökning är den som brukar kallas för *kollokation*. Även detta begrepp be-

traktas på olika sätt av forskare inom olika traditioner, vilket refereras i de två följande avsnitten.

### 2.4.1. Kontextualistisk kollokationsforskning

I litteraturen framställs ofta Firth (bl.a. 1957) som pionjär inom kollokationsforskningen, även om Palmer (1933), som Nation påpekar (2001: 317), använde begreppet långt tidigare. Själva kollokationsbegreppet förblir dock något vagt i Firths egna arbeten (jfr 1957:195). Han kan sägas i hand ha inspirerat i första hand den typen av kollokationsforskning som definierar kollokation som frekvent samförekomst (eng. *co-occurrence*) av två eller flera ord i en textkorpus (Malmgren 2003:124, jfr Stubbs 2001:29). Både Halliday, McIntosh & Strevens (1964, se även Halliday 1966) och Sinclair (bl.a. 1966, 1987) har sedan byggt vidare på Firths kollokationsbegrepp i sin forskning. Halliday et al. argumenterar utifrån sannolikheten av ords samförekomst och konstaterar: "Given the item 'chair', we are more likely to find in the same utterance the items 'sit' or 'comfortable' [...] than, say, 'haddock'..." (1964:33). Det lexikala element som studeras med avseende på dess "kollokationsbeteende" kallar Halliday (1966) för *nod* (eng. *node*). De lexikala element som kollokerar med en viss nod kallar han *kollokat* (eng. *collocate*). Sinclair skiljer i sin tur mellan två olika typer av kollokationer: *upward collocations* och *downward collocations*. Distinktionen är relaterad till nodens respektive kollokatets frekvens; om noden är det mer frekventa ordet i en kollokation med ett mindre frekvent kollokat har vi att göra med en *downward collocation*. Om noden däremot kollokerar med ett mer frekvent kollokat talar Sinclair om *upward collocation*. Han illustrerar denna distinktion med en rad kollokationer där han utgår ifrån ordet *back* som nod. Som typiska exempel på downward collocates nämner Sinclair bl.a. verb som *arrive* (to arrive back on thursday), *cut* (to cut back on something) eller *fall* (to fall back on something). Han nämner även substantiv som *hotel* och adjektiv som *normal* (back to the hotel, back to normal), adverb som *straight* och *further* (straight/further back) och prepositioner som *behind* och *towards* (back behind/towards the church) som möjliga downward collocates (Sinclair 1987:328f). Som exempel på *back* (både i betydelsen 'tillbaka' och 'rygg') i kollokation med mer frekventa kollokat (*upward collocates*) nämner han bl.a. *being back at school*, *follow*

*someone back into* the wood och *a slap on the back*. Sinclair ser en systematisk skillnad mellan dessa två kollokationstyper:

Upward collocation of course is the weaker pattern in statistical terms, and the words tend to be elements of grammatical frames, or superordinates. Downward collocation by contrast gives us a semantic analysis of a word (Sinclair 1987:326).

För svenskans del har bl.a. Svanlund (2001) använt sig av frekvensbaserade kollokationsanalyser i sin korpusbaserade studie av konventionella metaforer. Han påpekar dock att lågfrekventa ordkombinationer inte nödvändigtvis är ointressanta:

... ibland kan de spegla en viss semantisk kombinationstyp utan att för den skull alla instantieringar av typen behöver vara särskilt vanliga (Stubbs 1996). Risken att missa väsentliga typer på grund av frekvensspärren är dock liten; sådana typer representeras som regel ändå bland de frekventa kollokaten (Svanlund 2001:107f).

Ett av problemen i samband med kollokationsbegreppet är att frekvens är en relevant aspekt, men att den i sig inte är tillräcklig för att definiera kollokationer som ”meningsfull kategori för en analys av lexikala mönster i språkssystemet” (Lindberg 2004:68, jfr Malmgren 2003:125). Malmgren framhäver de framsteg inom kollokationsforskningen som ledde till en åtskillnad mellan ”grammatiskt välformade och rent tillfälliga kollokationer” (2003:125). Dessa framsteg nåddes redan på 1970- och 80-talet, inte minst med hjälp av stora empiriska studier i svenska (Allén 1975, jfr även Kjellmer 1994 för engelska). Inom den datoriserade korpuslingvistik har man utvecklat statistiska metoder för att kunna skilja mellan visserligen frekventa, men tillfälliga samförekomster av ord, t.ex. *och det är...* (jfr Biber & Conrad 1999 om *lexical bundles*), och ordförbindelser där de ingående orden är starkt associerade med varandra (jfr Ellis, Simpson-Vlach & Maynard 2008), t.ex. *fatta beslut*. Med hjälp av datorer kan man inte bara ta reda på en ordförbindelses frekvens i en korpus, utan också mäta det som kallas för *Mutual Information* (MI). MI är ett statistiskt mått på i vilken mån konstituenterna i en ordförbindelse förekommer oftare tillsammans än man utifrån slumpen kan förutsäga (jfr Church & Hanks 1990); ju högre frasens MI är desto starkare associeras de ingående orden med varandra. På det sättet

kan man alltså skilja kollokationer och andra fraseologiska enheter som har specifika betydelser och funktioner från sådana ordkombinationer som är frekventa endast till följd av att de ingående orden är frekventa i den aktuella korpusen (Ellis, Simpson-Vlach & Maynard 2008:380, jfr Svanlund 2001:108f).

## 2.4.2. Systemorienterad kollokationsforskning

Något som helt saknas i den frekvensbaserade, kontextualistiska traditionen inom kollokationsforskningen är direktionalitetsperspektivet. Detta sätt att tänka kring kollokationer utvecklades av bl.a. Hausmann (1985) inom ramen för den systemorienterade kollokationsforskningen i Tyskland (Malmgren 2003:127). Hausmann, som är en hård kritiker av den kontextualistiska skolan, anser att konstruktioner som *köpa en bok* inte är kollokationer, då de är semantisk triviala. Enligt ett systemorienterat synsätt kännetecknas en kollokation av att en av konstituenterna i viss mån är oförutsägbar och begränsad i sin kombinerbarhet (Malmgren 2003:127). Denna skillnad i konstituenternas status är en viktig förutsättning för tillämpningen av direktionalitetsbegreppet. Hausmann använder begreppen *bas* och *kollokator*, som alltså skiljer sig från kontextualisternas *nod* och *kollokat* just genom betoningen på direktionalitet, dvs. ett slags hierarkiskt förhållande emellan dem; i kollokationen *dra en slutsats* är *slutsats* basen och *dra* kollokator. Direktionaliteten kan, enligt Hausmann, motiveras med att man vid produktion av kollokationer (i text) går från *bas* till *kollokator* (1985:119). Med andra ord kan man som språkbrukare tänkas leta efter det rätta verbet att använda tillsammans med substantivet *slutsats*, om man vill uttrycka innehållet i *dra en slutsats*. Däremot skulle språkbrukaren knappast utifrån verbet *dra* leta efter rätt substantiv att använda i syfte att uttrycka samma innehåll på ett idiomatiskt sätt (jfr Hausmann 1985:119ff, Malmgren 2003:127).

Kollokationsbegreppet i systemorienterad mening introducerades i svensk forskning av Svensén (1987). Huvudsyftet med hans arbete inom detta område har varit att utveckla begreppet för tillämpning inom lexikografisk forskning (Malmgren 2003:127). Inom detta fält har Sköldberg & Toporowska Gronostaj (2008) introducerat en avgränsning av kollokationsbegreppet, som de tillämpar i samband med beskriv-

ningen av kollokationer i ett elektronisk medicinsk lexikon (MedLex, Kokkinakis 2006). Författarna ansluter sig i det väsentliga till den systemorienterade forskningstraditionen. De skriver om prototypiska kollokationer och syftar till att ge en i första hand strukturell och funktionell beskrivning av kollokationer i medicinskt sammanhang (Sköldberg & Toporowska Gronostaj 2008:434f, jfr Bartsch 2004). De konstaterar att det finns många prototypiska kollokationer med liknande strukturer i materialet, bl.a. sådana med samma semantiska förhållande mellan bas och kollokator. Som exempel nämns kollokationer som uttrycker gradförhållanden, t.ex. *hög feber* och *djup depression* (Sköldberg & Toporowska Gronostaj 2008:435). De vanligast förekommande relationerna mellan komponenterna i en kollokation beskriver författarna i termer av lexikala funktioner (se Mel'čuk 1998). De tar även upp direktionaliteten i sin definition och nämner som exempel kollokationen *ta tempen*, som består av basen *tempen* och kollokatorn *ta*. De betonar dock också att det finns kollokationer där det är betydligt svårare att bestämma direktionaliteten. Exempel på sådana kollokationer är, enligt Sköldberg & Toporowska Gronostaj, *ordinera vila* och *ordinera läkemedel* (2008:434).

En tidigare svensk studie med strukturell och funktionell inriktning om en viss typ av just kollokationer är Ekbergs undersökning av det hon kallar för abstrakta övergångsfraser (1989). Med det avses sådana kollokationer som med Ekbergs terminologi utgör ”syntaktisk komplexa predikat som betecknar övergången från ett tillstånd till ett annat tillstånd och är semantiskt ekvivalenta med enkla verb eller adjektivkonstruktioner”. Några av Ekbergs exempel är *falla i sömn*, *gå till attack* och *komma till besinning*.

Om man ska försöka ringa in kollokationsbegreppet på ett sätt som motsvarar denna undersöknings syfte, kan man säga att en kollokation består av en bas som starkt associeras med en, eller ett fåtal kollokatorer. Dessutom kännetecknas kollokationer av ett visst mått av icke-kompositionalitet. Med det menas att det utifrån basens betydelse inte är helt förutsägbart vilka som är de, enligt konventionen, prefererade kollokatorerna. Enligt en sådan definition är *fatta ett beslut* en relevant kollokation, med substantivet *beslut* som bas och verbet *fatta* kollokator. *Läsa en bok* matchar däremot inte kriterierna för en kollokation; för det första finns inga starka restriktioner vad gäller konstituenternas kombinerbar-



het med andra ord – man kan göra olika saker med en *bok* och man kan *läsa* olika saker. För det andra är frasen kompositionell, dvs. utifrån betydelsen av *bok* kan förutsägas vilka ord det kan kombineras med, vilket också gäller för verbet *läsa* (jfr Malmgren 2003:127, Hausman 1985).

Vidare kan man konstatera att frekvens är en relevant, men inte avgörande aspekt i samband med kollokationsbegreppet, som O’Keefe, McCarthy & Carter(2007) påpekar i citatet nedan.

We are all familiar with the situation in class where we fall back on the statement ‘that’s just the way we say it’, when faced with an awkward question from a student about why something is expressed the way it is, and often, what we are really explaining is a strong statistical preference which can be powerfully demonstrated by the use of corpus data. The answer, in the final analysis, is still that collocation shows us ‘the way we say it’, but we can gain considerable confidence as teachers if we can present something as a widespread frequent collocation rather than a one-off occurrence in the particular text we are working with (O’Keefe, McCarthy & Carter2007:59).

Författarna menar att det ofta är de kollokationer vars konstituenten består av banala, vardagliga ord, som är svåra att ta fram och belysa endast utifrån den egna språkliga intuitionen. Dessa fraseologiska enheter kan med fördel tas fram och undersökas med hjälp av kvantitativa, korpuslingvistiska metoder, vilket jag till viss del även har gjort i samband med denna undersökning (se vidare avsnitt 5.3). Frekvensen i en korpus kan dessutom bekräfta intuitionen om ett uttrycks konventionalisering, vilket kan komma till nytta, bl.a. i andra- eller främmandespråksundervisning.

### 3. Lexikal och fraseologisk kunskap – vad kan man när man ”kan” ett uttryck?

Inom kognitiva teorier utgår man ifrån att vår kunskap om världen består av konceptuella strukturer som är lagrade i minnet, och som vid behov kan överföras till fenomen som upplevs i den reella världen för att förstå dem (Langlotz 2006:61). Langlotz nämner som exempel att en människa kan föreställa sig en katt utan att varken se, höra eller röra vid en katt i verkligheten. Detta tyder på att människan har ett begrepp lagrat i minnet, som består i information om bl.a. hur en katt ser ut, låter och känns (jfr även Lamb 2000:183). Dessa begrepp fungerar, enligt Langlotz, som kategorier som hjälper oss att hantera och organisera de enorma mängder information om vår omvärld som vi utsätts för:

Working as a piece of meaningful knowledge, we can project this conceptual CAT-schema as a categorization-standard to make sense of any concretely perceived target-creature that features a certain degree of ‘catness’. Thus, meaningful experience is not only related to our immediate perception of the physical world, but draws upon conceptual routines [...]. Concepts allow us to make sense of our environments [...], conceptual units work as [...] categorisation-standards that can be mapped onto the flow of experience to group and organize it (Langlotz 2006:61f).

Bybee (1998, 2006) är en i en rad funktionellt och kognitivt inriktade lingvister som bedriver språkforskning inom en teoretisk ram som går under benämningen *usage-based approach* (se Greenberg 1966, Hopper & Thomson 1980, Langacker 2000). Hennes utgångspunkt är att lexikonets struktur är baserad på det hon kallar språkliga erfarenheter (jfr Langlotz 2006).

While all linguists are likely to agree that grammar is the cognitive organization of language, a usage-based theorist would make the more specific proposal that grammar is the cognitive organization of one's experience with language. [...] [C]ertain facets of linguistic experience, such as frequency of use of particular instances of constructions, have an impact on representation that we can see in evidence in various ways, for example, in speakers recognition of what is conventionalized and what is not. [...] The proposal presented here is that the general cognitive capabilities of the human brain, which allow it to categorize and sort for identity, similarity, and difference, go to work on the language events a person encounters, categorizing and entering in memory these experiences (Bybee 2006:711).

Detta synsätt på grammatik har bl.a. en del konsekvenser för hur man ser på representationen av konventionaliserade språkliga strukturer i det mentala lexikonet, vilket diskuteras vidare i avsnitt 4.1.1 nedan.

### 3.1. Vad kan man när man kan ett ord?

Av den språkliga kompetensens många olika aspekter är det ordförrådet som utvecklas mest under en människas liv (Viberg 1993:35). Denna utveckling framskrider särskilt snabbt under skolåren. En allmän uppskattning är att ett barn vid skolstarten har ett ordförråd på mellan 8 000 och 10 000 ord på modersmålet och att det lär in ca 3 000 nya ord per år. Här ska dock nämnas att dessa siffror i första hand bygger på läsförståelsetest och att det naturligtvis också finns stor individuell variation (Viberg 1993:35f). Även om forskarna inte är helt överens om de exakta siffrorna innebär skolstarten alltså en drastisk expansion av ordförrådet för förstaspråkselever och samtidigt en stor utmaning för elever som har undervisningsspråket som andraspråk. Andraspråkseleverna behöver komma i kapp de jämnåriga förstaspråkseleverna när det gäller basordförrådet som dessa redan har vid skolstarten, samtidigt som de ska

utöka ordförrådet på andraspråket med flera tusen ord om året (Viberg 1993:36).

Att ha ett ordförråd som räcker till för att hantera de språkliga krav som ungdomar ställs inför, bl.a. i skolan, är dock inte enbart en fråga om mängden ord man hunnit lära in. Lexikonkunskap är en av de mest komplexa och mångfacetterade aspekter av språkbehärskningen, eller med Nations ord:

Words are not isolated units of language, but fit into many interlocking systems and levels. Because of this, there are many things to know about any particular word and there are many degrees of knowing (Nation 2001:23).

Nation utgår ifrån en modell för lexikonkunskap som innebär en generell distinktion mellan receptiv och produktiv lexikonkunskap. Denna modell innefattar en rad olika aspekter av kunskap och språkanvändning (Nation 2001:28) och sammanfattas i tabell 1.

TABELL I: Olika aspekter av lexikonkunskap (efter Nation 2001:27)

		Receptiv/ produktiv kunskap	
Form	tal	R	Hur låter ordet?
		P	Hur uttalas ordet?
	skrift	R	Hur ser ordet ut ?
		P	Hur skrivs och stavas ordet?
	morfologi	R	Vilka delar av ordet kan urskiljas?
		P	Vilka delar behövs för att uttrycka den avsedda betydelsen?
Betydelse	form och betydelse	R	Vilken betydelse förmedlar formen ifråga?
		P	Vilken form kan användas för att uttrycka den avsedda betydelsen?
	begrepp och referenter	R	Vad ingår i det aktuella begreppet?
		P	Vilka föremål i den reala världen kan begreppet referera till?
	associationer	R	Vilka andra ord kommer vi att tänka på i samband med det aktuella ordet?
		P	Vilka andra ord (synonymer) skulle vi kunna använda istället?
Språkanvändning	grammatisk funktion	R	I vilka konstruktioner och mönster förekommer ordet?
		P	I vilka konstruktioner ska ordet användas?
	kollokationer	R	Vilka ord eller typer av ord förekommer i kombination med det aktuella ordet?
		P	Vilka ord eller typer av ord ska användas i kombination med det aktuella ordet
	begränsningar i användning	R	Var, när och hur ofta kan vi förvänta oss att möta ordet?
		P	Var, när och hur ofta kan vi använda ordet?

De flesta beskrivningar av språkinläringen utgår ifrån att receptiv inläring och språkanvändning är enklare än den produktiva behärsknin- gen av språket. Däremot råder det oklarhet över orsakerna till denna skillnad. Nation refererar ett antal möjliga förklaringar som ges i littera-

turen, vilka, enligt honom, inte nödvändigtvis utesluter, utan snarare kompletterar varandra. En förklaring är den som Nation kallar för "the amount of knowledge explanation" (2001:28ff), som bygger på antagandet att produktiv behärskning av lexikonet kräver kunskap utöver den som krävs för receptiv språkbehärskning. Språklig produktion, muntlig eller skriftlig, kräver, enligt detta synsätt, mer exakta kunskaper om ordets eller uttryckets form och betydelse samt dess plats i olika delar av språkssystemet, än receptiv hantering av det språkliga inflödet. I samband med det senare kan det räcka långt att känna till ett mindre antal utmärkande drag hos den lexikala enheten ifråga. En annan möjlig förklaring är enligt Nation "the 'practice' explanation" (2001:29). Enligt denna teori får inlärare vanligtvis fler tillfällen att öva sina receptiva än sina produktiva färdigheter. Nation hänvisar i det sammanhanget till forskning som tyder på att receptiva och produktiva språkfärdigheter måste övas var för sig för att leda till full språkbehärskning (DeKeyser & Sokalski 1996). Denna förklaring strider mot den nyss nämnda synen på produktiv språkbehärskning som summan av alla receptiva, och ytterligare, specifikt produktiva, kunskaper (Nation 2001:29). Ytterligare en förklaring är "the 'access' explanation" som bygger på tanken att ett nytt inlärt ord i andra- eller främmandespråket i ett tidigt inlärningsstadium är direkt länkat till ekvivalenten i inlärarens förstaspråk (Ellis & Beaton 1993, jfr Bialystok 2001, se vidare avsnitt 4.3). Detta innebär att ordet kan hanteras receptivt och förstås med hjälp av förstaspråkskompetensen. På förstaspråket har inlärarens även tillgång till en rad olika associationer i samband med det aktuella ordet (bl.a. synonymer, kollokationer och antonymer). För att kunna använda ordet produktivt på ett korrekt och idiomatiskt sätt måste inlärarens välja rätt bland alla dessa associationer och sedan överföra dem till andraspråket, vilket gör den produktiva användningen mer komplex (Nation 2001:28f).

Även Henriksen (1999), och andra forskare före henne (t.ex. N. Ellis 1995 och R. Ellis 1995) har betonat att lexikonkunskap är sammansatt av flera olika typer av kunskap som tillsammans innebär full behärskning av ett ord (eller ett konventionaliserat uttryck). Hon föreslår en tredimensionell modell för lexikal utveckling, enligt vilken lexikal kompetens består av "(a) *partial to precise knowledge*, (b) *depth of knowledge*, and (c) *receptive to productive use ability*" (1999:303). Modellen är dels tänkt som underlag för fortsatt forskning i form av longitudinella

studier av utveckling av lexikal kompetens, dels för utveckling av olika metoder för att testa inlärares lexikala kompetens utifrån alla tre dimensionerna vid en viss tidpunkt (Henriksen 1999:315). De tre ovannämnda dimensionerna berörs i viss mån även i samband med min undersökning. Dimensionerna (a) och (c) kan på samma sätt appliceras på flerordsenheter som på enskilda ord i lexikonet. Depth-of-knowledge-perspektivet (b) är relevant både utifrån en holistisk syn på flerordsenheter och i relation till uttryckens enskilda konstituent; djupare kunskap om ett ords betydelse handlar inte endast om relationen mellan ordet – det begrepp som det betecknar (se avsnitt 3) – och dess referent i den reella världen, utan även om ordets relationer till andra ord i lexikonet. Härmed avser Henriksen bl.a. paradigmatiske relationer som synonymi eller hyponymi och syntagmatiske relationer som bl.a. innebär restriktioner ifråga om vilka andra ord ett givet ord kan kombineras med (Henriksen 1999:305f, jfr Nation 2001). Det senare är i hög grad relevant för denna undersökning.

I detta avsnitt har ordkunskapsforskningen hittills diskuterats utifrån det enskilda ordet. En anledning till detta är att mycket av ordlärningsforskningen i stor utsträckning fokuserat kunskap på ordnivå (Schmitt 2008:340). Beskrivningarna av olika aspekter av lexikonkunskap som tagits upp här kan dock i princip tillämpas på både ettords- och flerordsenheter, även om det naturligtvis finns ytterligare aspekter som är speciellt relevanta för kunskap om och behärskning av flerordsenheter av olika slag. Utvecklingen av lexikal kunskap på frasnivå och de konventionaliserade uttryckens roll i det mentala lexikonet diskuteras mer utförligt, utifrån tidigare fraseologisk forskning, i avsnitt 4.

En relevant aspekt i samband med kunskap om kollokationers betydelser tas upp av Bartsch (2004), som poängterar att prototypiska kollokationer bär på implicit kulturell eller domänspecifik information:

Beyond influencing and altering the meanings of the constituent words, collocations often carry an additional element of meaning that is not overtly introduced through the individual meanings of their constituents. E.g. the collocation *consenting adult* is defined in the CED [Collins English Dictionary, 1991] as 'a male person over the age of twenty-one, who may legally engage in homosexual behaviour in private'. Taken at face value, all the word combination denotes is an adult who can consent to some unspecified action or event. The additional meaning components are culturally

determined stereotypes implicitly conveyed by this particular collocation in a British English context (Bartsch 2004:75).

Sköldberg & Toporowska Gronostaj (2008:435) nämner *fri abort* och *medicinsk* respektive *kirurgisk abort* som exempel som kan vålla vissa problem, inte minst för inlärare av svenska, som inte enbart kan förlita sig på sina språkliga kunskaper när det gäller att använda eller tolka dessa kollokationer på ett idiomatiskt sätt. I fallet av *fri abort* krävs det kunskap om samhällskontexten, och för att förstå skillnaden mellan *medicinsk* och *kirurgisk abort* måste språkbrukaren använda sig av domänspecifik (medicinsk) metakunskap. Bartsch betonar att "[s]uch implicit or inherent meanings regarding cultural or domain specific stereotypes or encyclopaedic knowledge about the world must be taken as critical for collocations" (2004:76).

### 3.2. Att mäta lexikonkunskap

Lexikonkunskapens komplexitet medför bl.a. att det kan vara svårt att välja rätt testmetod för att mäta inlärares behärskning av lexikonet i målspråket. Vilket testformat som väljs beror givetvis till stor del på vilken aspekt av lexikonkunskapen som ska testas. Nation diskuterar olika typer av ordförrådstest med utgångspunkt i tidigare forskning (2001:344ff). Flervalstester har, enligt Nation (2001:349ff med anförda referenser), en del fördelar framför andra testformat. Inläraren kan t.ex. ha nytta av partiell behärskning av ett ord (2001:349). Han hänvisar också till forskning som tyder på att vissa problem – som t.ex. att inlärarna kan gissa sig fram till det korrekta svaret – inte är av särskilt stor betydelse när det gäller flervalssuppgifter i praktiken:

Paul, Stallman and O'Rourke (1990) interviewed first language learners about the strategies they used to answer particular multiple-choice items [...]. Paul et al. found that for both high-ability and low-ability readers, over 50% of the answers were chosen because of association. [...] The high-ability group guessed only 8% of the answers (with about a 50% success rate) while the low-ability group guessed 21% (with a 35% success rate). This indicates that guessing is not a major problem with multiple choice items and that learners responses are generally not random but largely driven by some knowledge of the words (Nation 2001:349f).



Med andra ord verkar det alltså inte vara möjligt att i någon nämnvärd grad kompensera brist på ordkunskap med gissning i ett flervalstest. Man bör dock i detta sammanhang komma ihåg att den undersökning som gjorts av Paul et al. baseras på självrapporterade data. Forskning som har jämfört flervalstester med andra testformat har visat att flervalstuppgifter verkar vara betydligt lättare att klara av för inlärarna än uppgifter där de t.ex. ombeds att förklara ordet i en mening eller att komplettera en mening med ordet ifråga (Nist & Olejnik 1995). Det är dock möjligt att konstruera flervalstest av varierande svårighetsgrad, som i första hand verkar vara relaterad till hur pass nära distraktorerna ligger den korrekta formen i betydelse (Nagy, Herman & Anderson 1985, se även Nation 2001:349). Denna fråga har även varit relevant att ta ställning till i samband med den enkät som jag konstruerat för min undersökning (se vidare avsnitt 5.4).

Andra testformat som är relevanta i detta sammanhang är det som Nation kallar för *sentence-completion*, vilket i samband med denna undersökning innebär att informanterna får komplettera ett ofullständigt konventionaliserat uttryck i kontext, och *supply item types* (2001:349), där uppgiften består i att komplettera konventionaliserade uttryck utan kontext (se vidare avsnitt 5). Båda dessa testformat har beskrivits som relativt svåra i jämförelse med t.ex. matchningsuppgifter, där ordet (eller uttrycket) ifråga är givet och uppgiften består i att matcha ihop det med dess korrekta betydelse. En skillnad i svårighetsgrad består i att inläraren, utöver att känna till ordens betydelse, också måste ha kunskap om deras kollokationspotential (Nation 2001:352f), vilket i hög grad är relevant för min undersökning.

Golden (2005:88ff) har i sin studie av andraspråkselevs förståelse av metaforiska uttryck i läroböcker använt en form av matchningstest, bestående av flervalstuppgifter som dels testar elevernas förståelse av metaforiska uttryck, dels deras behärskning av konstituenters bokstavliga betydelse. Hennes resultat visar bl.a. att det inte nödvändigtvis är en förutsättning att känna till konstituenternas bokstavliga betydelse för att förstå det metaforiska uttrycket ifråga. Golden förklarar detta resultat med att en del av uttrycken är föremål för automatiserad receptiv användning, dvs. eleverna förstår dem ”direkte uten å gå veien om den konkrete betydningen” (2005:285, jfr Moon 1998:164).

Studier av ordförrådet i svenska läromedel har, med hjälp av en rad olika metoder, genomförts inom ramen för projektet *Ordförrådet i läromedel* (OrdiL). Förutom korpuslingvistiska analyser av hur ordförrådet i läromedel ser ut har man inom ramen för detta projekt också testat elevers behärskning av delar av detta ordförråd med hjälp av olika testformat. Holmegaard (2007:135ff) har bl.a. använt sig av matchning och flervalstest för att testa elevers förståelse av sammansatta ord som *förhandlingsvilja*, *majoritetsbeslut* eller *justitiedepartement*. De infödda eleverna i undersökningen klarade dessa typer av uppgifter betydligt bättre än andraspråkseleverna, vilket måste ses som ett förväntat resultat. Eleverna fick också i uppgift att dela orden i dess huvudled (t.ex. *förhandlings/vilja*) och att bedöma sin egen kunskap om de aktuella orden i ett självskattningstest (ibid.). De ord som var svårast att förstå för samtliga elever var, enligt Holmegaard (2007:158) ”betydelsetunga substantiv som kräver sociokulturell förståelse” (t.ex. *justitiedepartement*). Tittar man på andraspråkseleverna som separat grupp är det däremot abstrakta ”läroboksord” (Holmegaard 2007:158) som *inflytande* och *företräda* samt polysema ord som *rörelse* och *utskott* som verkar vålla störst problem.



## 4. Konventionaliserade uttryck i språkutveckling

### 4.1. Förstaspråksutveckling

Wray (2002:105) påpekar att man inom forskningen kring förstaspråksutveckling sedan länge är överens om att barn ofta verkar hantera flerordskombinationer som enheter i början av sin språkutveckling (Bolinger 1975, Crystal 1997). Hon skriver vidare att forskarna i stort sett är eniga om att barn använder relativt komplexa ordförbindelser innan de behärskar dessa enheters grammatiska struktur. Däremot råder ingen konsensus kring dessa fraseologiska enheters betydelse för förstaspråksutvecklingen (2002:105). Medan vissa forskare inte tillmäter inläringen och användningen av oanalyserade flerordsenheter hos barn någon större betydelse (t.ex. Brown 1973), ser andra helfrasinläringen som en av de grundläggande aspekterna av tidig förstaspråksinläring (t.ex. Cruttenden 1981, se Wray 2002:105f). Cruttenden antar att förstaspråksinläring sker i två stadier, nämligen *item-learning* och *system-learning*. Detta gäller, enligt Cruttenden, alla aspekter av språket, dvs. fonologi, morfologi, syntax och semantik (1981:79ff, se även Wray 2002:106). Cruttenden menar vidare att man kan utgå ifrån att ett yttrande har lärts in som en helhet om ingen av de ingående enheterna används i andra ordkombinationer (1981:84). Locke menar att småbarn i tidiga stadier av sin språkutveckling knappt verkar urskilja enskilda ord som enheter över huvud taget (1997:272, se även Wray 2002:106).

Inom barnspråksforskningen utgår man ifrån att barn i tidiga inlärningsstadier använder två olika typer av flerordskombinationer; (1) ord-

förbindelser som barnet har tagit in via inflödet och som kallas för *underanalyzed strings* och (2) flerordsenheter som barnet själv antas ha skapat och sedan lagrat i minnet som en enhet. Dessa kallas för *fused strings*. Den senare typen kan vara grammatiskt välformad eller inte. Här handlar det om idiosynkratiska ordförbindelser som smälts samman till enheter och på så sätt görs lättillgängliga vid on-line kommunikation (Wray 2002:106ff, jfr Wray & Perkins 2000:2, jfr Peters 1983:82). Eftersom föreliggande arbete fokuserar uttryck som är konventionaliserade i (vuxna) infödda talares språkbruk, är det bara den första typen som är relevant här.

Wray & Perkins presenterar en psykolingvistisk modell för första-språksutveckling, i syfte att förklara den frekventa användningen av (idiomatiska eller idiosynkratiskt bildade) fraseologiska enheter i början av barns språkutveckling. Användningen har funnits minska under en period, för att sedan öka igen, genom en mer automatiserad användning av idiomatiska, konventionaliserade uttryck (2000:19ff). Modellen bygger på Lockes teori (1997) om två olika neurologiska mekanismer som ger upphov till en sådan utveckling. Enligt denna teori är den tidiga språkinläringen stimulerad av barnets sociala omgivning, som sätter ramar för det språkliga inflödet, i vilket komplexa språkliga konstruktioner identifieras som meningsfulla och lagras i minnet. Eftersom barnets kognitiva förmåga ännu inte är tillräckligt utvecklad för en språklig analys av dessa konstruktioner, tas de in som oanalyserade enheter. Den mekanism som ligger till grund för denna process kallas *specialization in social cognition* (SSC) (Wray & Perkins 2000:19). För att språkutvecklingen ska framskrida måste barnet börja analysera de språkliga enheter som samlats i det mentala lexikonet. För att detta ska kunna ske, aktiveras i nästa utvecklingsfas det som Locke kallar för *grammatical analysis module* (GAM). Förklaringen till att användningen av fraseologiska enheter återigen ökar i nästa utvecklingsfas och nu även i högre grad inkluderar konventionaliserade uttryck är, enligt Wray & Perkins (2000:20), att GAM i ökande grad trängs undan av SSC till förmån för en mer effektiv, ekonomisk språkanvändning. Förmågan att tolka och använda metaforiska uttryck som idiom på ett idiomatiskt sätt antas utvecklas i ett ännu senare stadium, då de processer som styrs av GAM och SSC integreras i allt högre grad (Wray & Perkins 2000:19ff, jfr Skehan 1998 om andraspråksutveckling nedan och Bybee 1998). Ef-

fektivitetsskålet, som framstår som mest grundläggande för en ökad användning av flerordsenheter i Wray & Perkins modell, kan dock knappast vara hela sanningen när det gäller användningen av konventionaliserade uttryck. Detta påpekas bl.a. av Wiktorsson (2003), som framhäver uttryckens semantiska funktion (jfr Bartsch 2004:75):

When certain expressions are used repeatedly in the same context, they attract meaning components from that context. These meaning components are part of the meaning of the expressions, and the expressions therefore function as efficient means of communicating quite complex situations (Wiktorsson 2003:149).

## 4.2. Andraspråksutveckling

Kognitivt inriktade andraspråksforskare som Skehan (1998) antar att andraspråksinlärning fungerar med hjälp av två parallella, interagerande system – ett *exempelsystem* och ett *regelbaserat system* – och att det är samspelet mellan dessa system som möjliggör en effektiv och kontinuerlig utveckling av inlärarens språkbehärskning (Skehan 1998:88ff, jfr Cruttenden 1981 och Wray & Perkins 2000 om förstaspråksinlärning). Skehan beskriver utvecklingen av de två systemen utifrån termerna *lexikalisering*, *syntaktikaliserings* och *relexikalisering*. Med hjälp av lexikaliseringsenheten byggs inlärarens exempelsystem, som tjänar till att tillhandahålla färdigbildade fraser i kommunikationssituationer där tid och kapacitet för processande är begränsade. För att inläraren kontinuerligt ska kunna utveckla sitt språk måste så småningom de inlärdade fraseologiska enheterna analyseras genom syntaktikaliseringsenheten, som möjliggör kreativ användning utifrån ett regelbaserat system. Det krävs dock ytterligare ett steg för att så småningom kunna klara av effektiv on-linekommunikation på en mer avancerad nivå, nämligen att de ordförbindelser som genomgått syntaktikaliseringsenheten relexikaliseras för att återigen bli tillgängliga som oanalyserade enheter för inläraren vid automatiserad användning (Skehan 1998:90, se även Lindberg 2004:65, jfr Wray & Perkins 2000). Den samverkan mellan de olika systemen som en effektiv språk-användning kräver sätts, enligt Skehan, vid förstaspråksinlärning automatiskt igång med hjälp av individens *language acquisition devise* (LAD). Andraspråksinlärare med sen startålder antas däremot inte längre ha till-

gång till LAD och riskerar därför att stagnera på lexikaliseringsstadiet (Skehan 1998:90). En liknande utveckling av användningen av konventionaliserade uttryck i ett andraspråk – från holistisk inlärning över en period av analytiskt processande, tillbaka till en mer automatiserad språkanvändning – beskrivs av Wray & Perkins (2000, jfr Bybee 1998, 2006 och Wray 2008 nedan). Som nämndes tidigare, är en anledning till barns frekventa användning av oanalyserade flerordsenheter, enligt Wray & Perkins, att deras kognitiva förmåga inte är tillräckligt utvecklad för att klara av en mer analytisk hantering av det språkliga inflödet. Då detta inte kan förklara liknande mönster i kognitivt mer utvecklade, äldre inlärares andraspråksutveckling, antar författarna, liksom Skehan (1998), att det är kraven på en någorlunda smidig on-linekommunikation som motiverar inlärare till att använda komplexa flerordsenheter innan de har förmågan att analysera deras inre struktur (Wray & Perkins 2000:19ff, se även Wiktorsson 2003:49). Wray (2002:ix) konstaterar att samtidigt som såväl första- som andraspråksinlärare i hög grad förlitar sig på flerordsenheter i början av sin språkutveckling, verkar en idiomatisk användning av konventionaliserade flerordsuttryck vara en av de största utmaningarna på vägen mot en inföddlik språkanvändning (jfr Wray & Perkins 2000:2). I det sammanhanget ska också nämnas att utvecklingen av förstaspråksinlärares förmåga att använda konventionaliserade strukturer på ett idiomatiskt sätt, enligt Wray & Perkins modell, inte avslutas förrän i artonårsåldern (2000:19ff). För föreliggande undersökning innebär detta alltså att även förstaspråkselevernas förmåga att använda konventionaliserade språkliga strukturer på ett idiomatiskt sätt fortfarande befinner sig under utveckling.

Wray (2002, 2008) argumenterar för att det mentala lexikonets innehåll bestäms av det hon kallar för *Needs Only Analysis* (NOA):

The process of analysis which the native speaker child engages in is not that of breaking down as much linguistic material as possible into its smallest components. Rather, nothing is broken down unless there is a specific reason (Wray 2008:17).

Hon betonar vidare att både andraspråksinlärares behov och kognitiva möjligheter skiljer sig från förstaspråksinlärares ifråga om i vilken mån språklig inflöde, och därmed även konventionaliserade flerordsuttryck, analyseras. Hennes hypotes är att sena andraspråksinlärare överlag tende-

rar att i högre grad analysera inlärd uttryck och att lagra mindre lexikala enheter i minnet, vilket innebär ett större behov av regelanvändning (Wray 2008:20). Wray menar att detta kan vara en förklaring till bristen på idiomaticitet i många vuxna inlärares språkbruk (jfr Pawley & Syder 1983, Ekberg 1998, 1999).

This proposal offers an explanation for the difficulties that are routinely reported for adult learners in relation to restricting their output to only the idiomatic patterns that they have encountered [jfr Bybee 1998, 2006] in input – that is, there is a tendency to express messages in alternative ways, which are meaningful but not nativelike (Wray 2008:20).

Frågan om flerordsenheters representation i det mentala lexikonet diskuteras vidare i avsnitt 4.3 nedan.

### 4.3. Mental representation av konventionaliserade uttryck

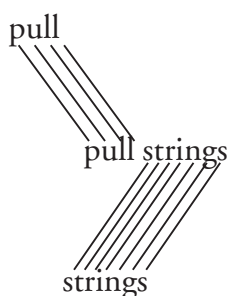
Som tidigare nämnts, diskuterar också Bybee (bl.a. 1998, 2006) frågan på vilket sätt konventionaliserade flerordsenheter är representerade i språkbrukarnas lexikon och hur denna representation utvecklas och påverkas. Som antyds i citatet i avsnitt 3 ovan, innebär Bybees antaganden om grammatik, som inkluderar lexikon, att konventionaliserade flerordsuttryck förs in i minnet och lagras där som lexikala enheter. Denna lexikala representation av idiom och andra konventionaliserade flerordsuttryck är nödvändig på grund av de oförutsägbara aspekterna av deras betydelser (Bybee 2006:713), dvs. på grund av deras brist på kompositionalitet. Hon betonar dock samtidigt att detta inte innebär att dessa typer av uttryck inte är analyserbara för språkbrukaren:

...they are not completely isolated from related words and constructions since many aspects of their meaning and form derive from more general constructions and the meaning of the component words in other contexts (Bybee 2006:713).

Bybee antar alltså att de konventionaliserade uttryck som inlärares ”erfar” lagras som enheter i minnet, samtidigt som de är länkade till de en-



skilda orden som utgör deras lexikala komponenter och andra relaterade lexikala enheter som innehåller delvis samma lexikala, semantiska eller grammatikaliska komponenter (1998:425, 2006:713). Detta innebär alltså att det bildas ett nätverk som integrerar större enheter och deras enskilda komponenter på ett sätt som ger språkbrukaren tillgång till en uppsättning färdigbildade enheter som, inom samma system, kan analyseras i sina beståndsdelar. Bybee illustrerar relationerna mellan idiomatiska uttryck och deras lexikala komponenter som i figur 1 nedan.



FIGUR 1: Exempel på relationer mellan ett idiom och dess lexikala komponenter (Bybee 1998:425, 2006:713)

Att uttala sig om hur lexikala enheter lagras och organiseras i det mentala lexikonet är i samband med föreliggande undersökning ett i hög grad spekulativt företagande. Det finns dessutom en hel del överlappningar mellan de olika modeller som presenterats ovan, samtidigt som de skiljer sig ifråga om vissa grundläggande teoretiska antaganden. Ett antagande som ligger till grund för denna undersökning är att kompositionalitet och därmed gränsen mellan olika typer av konventionaliserade uttryck, är ett graduellt fenomen. Detta förutsätter en kognitiv representation av både enskilda ord (mindre enheter) och större enheter som är sammansatta av dessa enskilda ord i ett slags integrerat system, där de större enheterna är mer eller mindre kompositionella, genomskinliga och fixerade. Med andra ord är Bybees nätverksmodell lättare att förena med övriga grundläggande antaganden som görs i samband med denna undersökning än en modell som ser *exemplarrepresentation* (eng. *exemplar representation*, se Skehan 1998:60ff) och regel användning som helt åtskilda fenomen. Dessutom säger varken Wray &

Perkins (2000) eller Skehan (1998) mycket om hur interaktionen mellan de olika mekanismerna eller de olika systemen skulle kunna se ut.

En viktig faktor för att ett uttryck ska vara representerat i lexikonet som en enhet är, enligt Bybee, frekvensen. Denna påverkar representationen på ett avgörande sätt. Med andra ord har inte alla enheter som är representerade i nätverket samma status: "those that are used frequently have very strong representations, or possibly even multiple representations, but those that are used infrequently, or are potential but have not been used, may have no representation" (1998:423). Hur nätverket ser ut och i vilken mån ett uttryck som lagras som en helhet kan länkas till olika enheter och därmed bli analyserbart beror alltså på inlärares tidigare språkliga erfarenheter, dvs. på i vilken mån dessa enheter i sin tur finns lagrade i minnet.

Liknande antaganden görs av Wray (2008) som argumenterar för ett heteromorft lexikon, där både holistiska fraser och olika typer av mindre komponenter lagras. Även hennes modell kan sägas vara *usage-based*:

The determination of which components are separately stored and, thus, how much flexibility there is for varying the form of the complete string, depends on input evidence and user need (Wray 2008:15).

Wray beskriver flerordsenheter som *morphem equivalent units* (MEUs), men föreslår samtidigt att många av dessa enheter inte är lagrade i helt fixerad form utan som "partly-fixed frames", vilket, enligt henne, är en förutsättning för en effektiv integration av det hon kallar för *formulaicity* och kreativ regelanvändning (2008:16).

Vad gäller den lexikala representationen av två eller flera lexikon hos flerspråkiga individer är forskarna inte helt eniga i frågan om lexikala enheter representeras i ett integrerat eller två oberoende system (Bialystok 2001:102f). Bialystok påpekar att forskningen har gett en del indikationer på att ordförråden är oberoende av varandra men refererar samtidigt forskning som har tagit fram en, som hon anser, trovärdig modell för ett integrerat flerspråkigt ordförråd (Dijkstra & van Heuven 1998). Bialystok påpekar att forskarna däremot är eniga om att båda ordförråden är aktiva när en tvåspråkig individ använder ett av sina språk (2001:102). Ett annat problem i sammanhanget är relationen mellan lexikala enheter på de olika språken och deras betydelser och referenter i den reella

världen. Bialystock betonar att enkla dikotomier som den mellan oberoende och integrerade ordförråd eller den mellan direktlänkar från begrepp till ord i andraspråket å ena sidan och vägen via förstaspråket (jfr Nation 2001) å andra sidan inte styrks av empirin. Det flerspråkiga mentala lexikonet kan i stället antas vara mer komplext.

Adequate descriptions of the organization of mental representations include the effect of such factors as level of proficiency and the circumstances of second-language learning. Moreover it appears likely that multiple arrangements can even coexist in the mind of an individual speaker... (Bialystok 2001:103).

Bialystok utgår vidare ifrån att flerspråkig lexikal utveckling hos barn involverar en lika komplex mental representation som hos vuxna andraspråksinlärare. Språk och begrepp antas vara sammanlänkande på ett dynamiskt sätt och strukturen ses som öppen för förändring i samband med ändrade språkliga behov och ökad språkfärdighet (jfr Wray 2008). Frågan är naturligtvis intressant med tanke på andraspråkinformanterna i undersökningen, vilket är anledningen till att den kommenteras här. Då denna undersökning dock inte behandlar frågor om transfer, lämnas denna diskussion i fortsättningen därhän.

#### 4.4. Konventionaliserade uttryck i andra- och främmandespråksinlärning

Lindberg (2004:69) skriver att det verkar råda stor överensstämmelse mellan infödda talare och mycket avancerade andraspråkstalare av ett språk, när det gäller att bedöma om en ordförbindelse är idiomatisk eller inte. Däremot är andraspråkstalare på mindre avancerad nivå i betydligt mindre grad överens (med infödda talare såväl som sinsemellan) i sina bedömningar av vilka ordförbindelser som känns naturliga i andraspråket. Van Lier (1995) demonstrerar detta när han ger läsaren i uppgift att para ihop de för engelskan mest idiomatiska ordförbindelserna genom att välja från en lista med (förstärkande) adverb och en lista med adjektiv (jfr Granger 1998). Enligt Van Lier blir resultatet kollokationer av typen *solidly built* och *carefully crafted*, vilket han alltså menar att de

allra flesta infödda talare av engelska skulle hålla med om, medan även relativt avancerade andraspråkinlärare uppvisar betydligt större variation i sina svar (jfr Ekberg 1999). Lindberg ger svenska exempel av typen *osedvanligt korkad* och *komplett galen* (2004:81). Denna övning är väl lämpad att illustrera idiomicitetsbegreppet; det finns inget i det grammatiska regelsystemet eller i ordens semantik som hindrar en infödd talare av engelska att t.ex. välja ordkombinationen *solidly crafted* istället. En förstaspråkstalare av svenska skulle på samma sätt kunna tänkas bilda sekvenserna *osedvanligt galen* och *komplett korkad* utifrån Lindbergs lista. Det verkar dock finnas en samstämmighet, som i första hand verkar råda bland infödda talare inom en språkgemenskap, om vilka kollokationer som är de mest etablerade, de mest konventionaliserade och därmed de mest idiomatiska i det aktuella språket (jfr Wiktorsson 2003:56). Varken van Lier eller Lindberg baserar sina resonemang på empiriska studier, vilket gör det desto mer intressant att testa hypotesen om en tydlig skillnad mellan första- och andraspråksanvändares resultat när uppgiften är att sätta ihop de mest idiomatiska kollokationerna utifrån en uppsättning givna ord. Denna hypotes kan ses som utgångspunkt för kollokationstestet i föreliggande undersökning (se vidare avsnitt 5.4).

#### 4.4.1. Studier inom andraspråksforskning

Abrahamsson & Hyltenstam (2009) har undersökt behärsningen av idiom och ordspråk hos andraspråksinlärare på inföddlik nivå. Undersökningen ingår i en större studie om startålder och *uppfattad inföddhet*. Abrahamsson & Hyltenstam använder sammanlagt tio olika testinstrument för att undersöka i vilken mån andraspråksinlärare, som till vardags passerar som infödda talare av svenska, i själva verket behärskar språket på samma nivå som infödda talare. Förutom behärskning av idiom och ordspråk testas informanterna på bl.a. grammatikalitetsbedömning och perception av tal i brus. När det gäller idiom och ordspråk består metoden av två separata datoriserade test (innehållande 50 idiom respektive 50 ordspråk), där informanterna ställs inför uppgiften att inom 10 sekunder muntligt komplettera ett idiom respektive ordspråk som presenteras ofullständigt på datorskärmen. Exempel från testen är *Hon löpte verkligen [linan] ut* och *Ju fler kockar [desto sämre soppa]* (or-

den inom hakparentes är de som utelämnats på skärmen). I analysen jämförs de avancerade inlärnarnas svar med resultaten från en infödd kontrollgrupp. Resultaten för idiomtestet visar tydliga skillnader mellan informanter med olika startålder för andraspråkinläringen. 58 % av informanterna med startålder 1–11 år når samma eller högre resultat än de lägst presterande infödda informanterna, men bara 20 % av informanterna med startålder 13–19 år når upp till dessa resultat. I ordspråkstestet var motsvarande resultat 16 % och 10 % (Abrahamsson & Hyltenstam 2009:279ff).

Ekberg har studerat förhållandet mellan regeltillämpning och lexikonkunskap hos enspråkiga och flerspråkiga elever i årskurserna 5, 6, 7 och 9 i Rosengård (1997, 1998, 1999, se även 2004). En del av hennes arbete består i att kartlägga frekventa syntaktiska och lexikala strukturer som enspråkiga svenska elever använder sig av i en viss språksituation (återgivning av bildberättelser), varefter flerspråkiga elevers bruk av konventionaliserade strukturer i samma språksituation har undersökts (1998:260f). Utöver detta har Ekberg i motsvarande grupper bl.a. undersökt användningen av *komplexa predikat* som pseudosamordningar – t.ex. *han sitter och läser* – samt partikelverb, reflexiva verb och flerordskombinationer som är lexikaliserade på frasnivå – t.ex. nominalfraser som *Vita huset* eller verbfraser som *spela kort* (Ekberg 1998:249f, se även 1997:97ff).

Studierna visar att både enspråkiga och flerspråkiga elever har aktiv kunskap om de undersökta konventionaliserade lexikala och grammatiska konstruktionerna. Skillnaderna mellan de båda grupperna gäller i första hand frekvens och variation i användningen av konventionaliserade strukturer. De enspråkiga eleverna använder fler konventionaliserade strukturer på alla språkliga nivåer (Ekberg 1998:261, se även 1997:106ff). Dessutom är variationen i användningen av dessa frasliknande strukturer större bland de flerspråkiga eleverna, vilket innebär att deras språkbruk i den givna talsituationen är mindre stereotyp, och i många fall även mindre idiomatiskt, än de enspråkiga elevernas. Resultaten kan, enligt Ekberg, vara ett tecken på att de flerspråkiga barnen har en mindre uppsättning färdiga konstruktioner att välja bland, vilket i sin tur kan vara en följd av en relativt låg grad av kontakt med målspråket (1999:94, se även Prentice & Sköldberg und. utg.).

Vad gäller svensk kollokationsforskning på andraspråksområdet har Enström studerat andraspråksinlärares verbanvändning. Hennes material består av uppsatser i svenska som andraspråk, skrivna av gymnasieelever med olika modersmål och varierande språkbehärskning (1996:108f). Enström har bl.a. beskrivit ”de felanvändningar av verb som i första hand [...] rör rent syntagmatiska relationer” (1996:130) och kategoriserat dessa som *kombinatoriska fel* inom ramen för det hon kallar för *lexikalisk-semantisk felkategorisering*. Denna kategorisering innebär en primär uppdelning i *konstruktionsfel* och *primärt semantiska fel*, som sedan indelas vidare i ett antal subkategorier (Enström 1996:108ff). De kombinatoriska felen (som alltså tillhör kategorin konstruktionsfel) innefattar bl.a. det som Enström kallar för *lexikaliska verbkollokationer*. Hon påpekar att de ordkombinationer i hennes material som ”kategoriserats som lexikaliska verbkollokationer utmärks av att de ingående ordens betydelser är förenliga, varför felen inte kan ses som brott mot gällande selektionsrestriktioner utan som att ordkombinationen inte är idiomatisk” (1996:149). Som exempel på icke idiomatiska användningar av verbkollokationer ges bl.a. följande meningar (Enström 1996:150ff):

Lärarna måste **ge** större krav på **läxorna**.  
För att motverka detta måste vi **göra** större **åtgärder**.  
Barnen hade redan gjort snö-gubbar på vår gård och **lekte snöboll**

Enström beskriver också kombinatoriska fel i samband med det hon kallar för *lexikaliserade verbalfraser*. Hon ger exempel på en rad konventionaliserade uttryck som vållar problem för informanterna i hennes undersökning. De flesta exempel från hennes material består av *kontaminationer*, dvs. en sammanblandning av två eller flera uttryck, som i

Om man ligger riktigt stilla och **inte gör** ett väsen **ifrån sig**, då kan man höra hur fåglarna kvittrar,

som, enligt Enström, kan antas vara en kontamination av *inte ge ett ljud ifrån sig* och *inte göra något väsen av sig* (1996:158f).

Även Hyltenstam (1992) har beskrivit avvikelser vid användning av konventionaliserade uttryck hos flerspråkiga och enspråkiga elever. Han framhäver *approximationer* (1992:360) till den konventionaliserade for-

men (som i ”*den fungerar nog inte på längden*”) och kontaminationer av element från två eller flera olika uttryck (som i ”*man beräknar med att...*”) som vanliga feltyper i elevernas muntliga och skriftliga produktion. Analysen visar att approximationer är betydligt vanligare i de flerspråkiga informanternas språkbruk, medan kontaminationer visar sig vara lika frekventa bland enspråkiga elever som i den flerspråkiga gruppen (Hyltenstam 1992:360ff, Große<sup>1</sup> 2008, Prentice & Sköldberg und. utg.). Enströms (1996) och Hyltenstams (1992) beskrivning av avvikelser från konventionaliserade mönster vid andraspråksinlärares användning av konventionaliserade uttryck är relevant för den kvalitativa analysen i samband med min undersökning.

I en studie av receptiv behärskning av konventionaliserade uttryck i *English for Academic Purposes* (EAP) har Ellis, Simpson-Vlach & Maynard (2008) i ett antal experimentella undersökningar visat att det är olika faktorer för första- och avancerade andraspråkstalare av engelska som påverkar hur snabbt de klarar av att processa en fraseologisk enhet. Testresultaten visar att processningstiden för förstaspråksanvändare i första hand påverkas av den aktuella fraseologiska enhetens *MI-värde* (Mutual Information) (se avsnitt 2.4.1) och inte av dess frekvens i inflödet (ibid.). Andraspråksanvändarnas reaktionstider påverkas däremot i första hand av den fraseologiska enhetens frekvens (Ellis, Simpson-Vlach & Maynard 2008:382ff). Den slutsats som författarna drar av sina resultat kan betraktas som högst relevant för forskningen om lexikal kompetens hos avancerade andraspråksinlärare såväl som för andraspråksundervisning:

...native speakers and advanced ESL learners have become sensitive from their usage histories to these [formulaic] expressions so that they process them preferentially. But native speakers and learners are sensitive to different determiners of fluency – learners to [...] frequency, fluent natives to MI (Ellis, Simpson-Vlach & Maynard 2008:389).

En av de avgörande faktorerna för *slutlig nivå* (eng. *ultimate attainment*) i andraspråksinlärning, har visat sig vara startåldern för inlärning (Birdsong 1999, Hyltenstam & Abrahamsson 2000, Abrahamsson & Hyl-

---

<sup>1</sup> Författaren heter numera Julia Prentice

tenstam 2004, 2009). Det finns många kontroverser inom forskningen om startålder och olika aspekter av inlärares språkbehärskning. Dessa handlar bl.a. om att forskare sätter olika åldersgränser och utgår ifrån olika modeller för biologiska mognadsbegränsningar i samband med studier av inlärares möjligheter att nå en inföddlik slutnivå i andraspråket (Hyltenstam & Abrahamsson 2000, Munro & Mann 2005). Denna diskussion kommer inte att fördjupas här. Faktum är dock att man i många studier, liksom även i denna, delar in informanter efter startålder för sin andraspråksinläring (eng. *age of onset*, AO eller *age of immersion*, AOI). Vilka åldersgränser man utgår från vid denna indelning varierar däremot en hel. I den ursprungliga versionen av det som kallas för *the critical period hypothesis* (CPH), dvs. en kritisk period för språkinläring (Lenneberg 1967), sätts gränsen för påbörjande av språkinläring som kan leda upp till infödd slutnivå vid puberteten. Senare studier tyder däremot på att möjligheterna att uppnå en infödd slutnivå på andraspråket avtar successivt långt före puberteten (Hyltenstam & Abrahamsson 2000, jfr Pallier., Dehaene, Poline, LeBihan, Argentié, Dupoux & Mehler 2003:160). Abrahamsson & Hyltenstam (2009) väljer en indelning i fem grupper, baserad på AO: *tidig barndom* (AO ≤ 5 år), *sen barndom* (AO 6–11 år), *ungdom* (AO 12–17 år), *tidigt vuxenliv* (AO 18–23) och *senare vuxenliv* (AO ≥ 24 år). Indelningen kan, enligt författarna, delvis motiveras med utgångspunkt från olika språkutvecklingsfaser, samtidigt som den återspeglar mönster som träder fram i deras analys av sambandet mellan startålder för andraspråksinläring och i vilken mån en inlärare uppfattas som infödd talare (Abrahamsson & Hyltenstam 2009:268). Butler (2000) delar i sin studie av andraspråksbehärskning hos kinesiska andraspråkstalare av engelska in informanterna i tre startåldersgrupper samt en grupp med infödda talare. Grupperna består av informanterna med startålder, 3–5 år, 5–10 år samt 10–15 år (Abrahamsson & Hyltenstam 2004). Den indelning som har valts i samband med denna studie kan i stort sett sägas motsvara indelningarna i de två sistnämnda studierna (se vidare i avsnitt 5.3)



#### 4.4.2. Korpuslingvistiska studier inom främmandespråksinlärning

Betydelsen av konventionaliserade uttryck för andra- och främmandespråksinlärning har också framhävts ur ett kontrastivt perspektiv i en rad korpuslingvistiska undersökningar. Ett relativt stort antal korpusbaserade studier har under de senaste två decennierna visat att även inlärare på avancerad nivå underanvänder idiom och andra konventionaliserade uttryck i jämförelse med infödda talare (jfr O’Keefe, McCarthy & Carter 2007). I det sammanhanget har bl.a. inlärarkorpusen ICLE, the International Corpus of Learner English, spelat en framträdande roll. Korpusprojektet lanserades 1990 av Granger, med målet att bygga upp en databas med texter skrivna i engelska som främmandespråk av inlärare med olika modersmål. Materialet tjänar som underlag för kontrastiva studier i större skala (se Granger 1993, 1996, 1998, 2003, Granger, Hung & Petch Tyson 2002). Granger själv har bl.a. studerat användningen av kollokationer med förstärkande adjektiv som bas (t.ex. *perfectly natural* eller *deeply in love*) hos fransktalande inlärare av engelska (1998:146f). Hennes resultat tyder på en tydlig underanvändning av det hon kallar för *native-like collocations* bland inlärarna. Granger förklarar sina resultat dels med interferens från inlärarnas L1, då de idiomatiska kollokationer som inlärarna använder ofta har en direkt ekvivalent i deras modersmål (1998:151), dels med inlärarnas outvecklade känsla för vilka ordkombinationer som är mer framträdande (eller mer konventionaliserade) än andra i målspråket (Granger 1998:152). Som tidigare nämnts, har Wiktorsson (2003) studerat användningen av konventionaliserade flerordsuttryck (eng. *prefabs*) i svenska inlärares skriftliga produktion i engelska. Även hennes arbete baseras på ICLE, närmare bestämt den svenska subkorpusen SWICLE. Hon jämför data från två olika grupper av inlärare – gymnasieelever och universitetsstudenter i engelska – med infödda skribenters texter. Resultaten visar att skillnaden i användningen av konventionaliserade uttryck mellan den mer avancerade inlärargruppen och de infödda skribenterna snarare är kvalitativ än kvantitativ, dvs. de avancerade inlärarna (svenska universitetsstudenter) använder inte färre, utan andra typer av konventionaliserade uttryck. Den huvudsakliga skillnaden tycks ligga i att inlärarna använder fraser som är utmärkande för ett mer informellt språkbruk, vilket gör att

de i det skriftspråkliga sammanhanget framstår som mindre idiomatiska (2003:108f). Mellan de mindre avancerade inlärnarna (svenska gymnasieelever) och de infödda skribenterna har Wiktorsson däremot funnit en frekvensskillnad i användningen av konventionaliserade uttryck. Inlärnarnas texter innehåller färre konventionaliserade uttryck än de inföddas uppsatser (jfr Ekberg 1997, 1998, 1999). Den kvalitativa skillnaden visar sig vara större mellan de mindre avancerade inlärnarna och de infödda skribenterna än mellan de senare och den mer avancerade inlärnargruppen. Med andra ord använder de mindre avancerade inlärnarna fler uttryck i sitt skriftspråk som bland infödda talare är vanligare i talspråk än de avancerade inlärnarna. Dessutom uppvisar gymnasieelevernas texter en högre grad av interferens från förstaspråket än de mer avancerade inlärnarnas texter (Wiktorsson 2003:145ff)

## 4.5. Konventionaliserade uttryck i utbildningsrelaterat språk

O’Keefe, McCarthy & Carter(2007:92) poängterar att konventionaliserade uttryck, liksom flerordsenheter generellt, kan kopplas till olika register och hänvisar bl.a. till forskning om talat språk i akademiska sammanhang som har visat att idiom utgör ”a not-insignificant feature of the lexical landscape of academic speech” (Simpson & Mendis 2003:427). Som nämndes inledningsvis är vissa typer av konventionaliserade uttryck inte bara vanligt förekommande, utan även högst relevanta ur ett funktionellt perspektiv i det som brukar kallas för akademiskt språk.

Cognitive science demonstrates that knowledge of these formulas is crucial for fluent processing. And current trends in SLA [Second Language Acquisition] and EAP [English for Academic Purposes] demand ecologically valid instruction that identifies and prioritizes the most important formulas in different genres (Ellis , Simpson-Vlach & Maynard 2008:379).

Som antytts tidigare, belyser Ellis, Simpson-Vlach & Maynard frågan om receptiv behärskning av konventionaliserade uttryck i EAP från såväl ett korpuslingvistiskt som ett psykolingvistiskt och didaktiskt perspektiv (2008:375ff). Författarna har med hjälp av korpuslingvistiska

metoder tagit fram fraseologiska enheter som är signifikant mer frekventa i akademisk engelska än i icke-akademiska diskurser (2008:379) och som därmed också antas vara relevanta för undervisning i EAP. För att bekräfta det senare har Ellis, Simpson-Vlach & Maynard, utifrån detta material, låtit ett antal språktexperter och lärare i EAP göra en av följande bedömningar: (1) om de ansåg ordkombinationerna ifråga utgöra fasta fraser, konventionaliserade uttryck eller flerordsenheter (eng. *chunks*), (2) om de ansåg fraserna ha en betydelse och funktion som en enhet (som inte motsvarar de enskilda konstituenternas betydelse och funktion), eller (3) om de ansåg fraserna vara värda att läras ut som idiomatiska uttryck (Ellis, Simpson-Vlach & Maynard 2008:381). Resultatet visar att de enheter som bedöms som mest konventionaliserade och värda att lära ut var de mest frekventa, och de med de högsta värdena för Mutual Information (MI) i korpusen. Lärarnas bedömningar kan alltså på så sätt anses ha bekräftat den korpuslingvistiska analysen.

## 5. Material och metod

Materialet för denna undersökning har samlats in med hjälp av en enkät som delats ut till sammanlagt 10 gymnasieklasser i olika årskurser, på två olika gymnasieskolor. Materialet samlades in under vårterminen 2009. Sammanlagt har 219 elever deltagit i enkätundersökningen. 212 elever har inkluderats i analysen (sju enkäter har alltså av olika anledningar exkluderats). 186 av dessa elever kan sägas studera i en språkligt heterogen klassmiljö. De övriga 26 eleverna tillhör en gymnasieklass (årskurs 1) med enbart elever med helsvensk bakgrund som fungerar som jämförelseklass. Resultaten från den enspråkiga klassmiljön används alltså i första hand som referenspunkt i analysen (se vidare i 5.2).

### 5.1. Etnisk och social bakgrund som påverkande faktorer i samband med elevers skolresultat

Det är knappast förvånande att vistelsetiden i målspråkslandet är en faktor som påverkar andraspråkelevernas prestationer i skolan. Thomas & Collier (2002) har visat att det tar mellan fem och tio år tid för invandrarelever att nå en nivå i sin skolrelaterade andraspråksbehärskning som motsvarar infödda elevers behärskning av skolrelaterat språk (Lindberg 2007:82, jfr Viberg 1993). Skolverket visar i en rapport att varje år senare som en elev anlände i grundskolan i genomsnitt motsvarar ca fem poäng lägre meritvärde i årskurs 9 (Skolverket 2004:47).

Med tanke på att den forskning som har bedrivits sedan mitten på 1960-talet på detta område (t.ex. Coleman 1966) har visat tydliga samband mellan elevers socioekonomiska status och deras skolresultat, kan man anta att även resultaten i min undersökning i viss mån kan påverkas av socioekonomiska faktorer (se vidare i avsnitt 5.2 nedan). Här finns dock ett antal metodologiska problem som gör att en del tidigare forskningsresultat är svåra att tolka. Det har bl.a. att göra med att socioekonomisk status är ett flerdimensionellt begrepp som innefattar bl.a. ekonomiska, sociala och kulturella faktorer. Dessutom är forskningsresultaten också avhängiga de typer av skolresultat som fokuserats, t.ex. vilka ämnen som undersökts och huruvida man tittat på individuella resultat eller resultat på gruppnivå (Skolverket 2009:98f). I senare studier har man dock tagit hänsyn till dessa definitionsproblem. Bl.a. har man börjat skilja mellan ekonomiskt kapital (föräldrarnas inkomst) och kulturellt kapital (som, enligt skolverket, innebär föräldrarnas utbildningsbakgrund) och kommit fram till att det i första hand är det senare som uppvisar tydliga samband med elevers resultat i skolan (t.ex. Gustafsson 1997, Yang & Gustafsson 2003, se Skolverket 2009:100). Även boendesegregationen har pekats ut som en faktor som påverkar skolresultaten negativt för elever med utländsk bakgrund i segregerade områden. Här handlar det ofta också om andra samverkande faktorer som t.ex. ekonomiska problem, etnisk minoritetsstatus, bristfällig behärskning av majoritetsspråket och föräldrarnas utbildningsnivå liksom även skolrelaterade faktorer som skolresurser och lärarkompetens (Skolverket 2004:47, 2009:96). Elever med utländsk bakgrund är också överrepresenterade bland elever med låg socioekonomisk status och i skolor med låg andel elever med utländsk bakgrund är familjernas socioekonomiska status således ofta högre i termer av ekonomiska resurser, föräldrarnas utbildningsbakgrund och arbetsmarknadsanknytning (Skolverket 2009:50). Från Skolverkets sida har man tolkat detta som att det i första hand är socioekonomiska faktorer, snarare än språkliga och andra faktorer i samband med elevernas härkomst, som ligger bakom skillnaderna i skolresultat mellan infödda elever och de med utländsk bakgrund (2005:10). Denna tolkning kan ses som delvis motsägelsefull i relation till den tidigare nämnda rapporten om ankomstålder och skolresultat (2004:47). Lindberg påpekar också problem med att basera prognoser om skolframgång på uppgifter om föräldrarnas utbildningsbakgrund när det

gäller elever med utländsk bakgrund och understryker att skillnaderna i prestation mellan elever som anländer efter skolstarten och infödda elever kvarstår även efter att de socioekonomiska faktorerna har tagits med i beräkningen (Lindberg 2007:78f).

Som framgår ovan är sambanden mellan elevernas sociala bakgrund och deras prestationer i skolan (eller i undersökningar som denna) tämligen komplexa och därför svårbeforskade. Jag har därför i denna undersökning avstått från att relatera resultaten till uppgifter om de individuella elevernas bakgrund i detta avseende och valt att endast kontrollera för deltagarnas språkliga bakgrund. Däremot ges en kort beskrivning av skillnader i skolornas elevsammansättning, bl.a. även med avseende på etnisk och social bakgrund, i nästa avsnitt.

## 5.2. Skolor och elevsammansättning i deltagarklasserna

De två gymnasieskolorna som har deltagit i undersökningen skiljer sig på flera olika sätt från varandra, vilket man bör beakta vid tolkning av resultaten. Det ena gymnasiet, hädanefter kallat Prismagymnasiet (P), ligger i en förortskommun till Göteborg. Den andra skolan, hädanefter kallad Kampusgymnasiet (K), är en innerstadsskola i Göteborg. Även om det inte längre finns fasta upptagningsområden för skolorna, är det inte orimligt att anta att skolans geografiska läge spelar viss roll för elevsammansättningen. Man kan t.ex. tänka sig att en skola som ligger centralt drar till sig elever från flera olika områden i staden, vilket resulterar i en mer heterogen elevgrupp än i en skola som Prismagymnasiet, som ligger i en förortskommun. Som framgår av tabell 2 har Prismagymnasiet också betydligt lägre andel elever med utländsk bakgrund än Kampusgymnasiet på de aktuella gymnasieprogrammen. Detta gäller även om man tittar på skolorna som helhet. Andelen elever med utländsk bakgrund på Kampusgymnasiet är 40 %, medan endast 17 % av eleverna på Prismagymnasiet har utländsk bakgrund (skolverket 2010). Med anknytning till diskussionen kring socioekonomisk bakgrund i föregående avsnitt kan man alltså anta att eleverna på Kampusgymnasiet i högre uträkning kommer från familjer med relativt låg socioekonomisk

TABELL 2: Bakgrundsdata för deltagarklasserna

	KLASS 1	KLASS 2	KLASS 3	KLASS 4	KLASS 5	KLASS 6	KLASS 7	KLASS 8	KLASS 9	Jämförelse- klass
Skola	Prisma	Kampus	Kampus	Kampus	Prisma	Kampus	Kampus	Kampus	Kampus	Prisma
Program och årskurs	SP 1	SP 1	OP 1	SP 1	SP 2	SP 2	OP 2	OP/ NP 2	OP 3	SP/ NP 1
Andel elever med utländsk bakgrund på gymnasieprogram	27 %	51 %	42 %	51 %	27 %	51 %	42 %	62 %/ 42 %	42 %	27 %/ 28 %
Antal elever	28	19	14	24	25	22	16	22	23	26
Antal (andel) flickor	14 (50 %)	11 (58 %)	12 (86 %)	16 (67 %)	10 (40 %)	15 (68 %)	13 (81 %)	16 (73 %)	17 (74 %)	17 (65 %)
Antal (andel) elever med svenska som L2	11 (39 %)	5 (26 %)	14 (100 %)	24 (100 %)	7 (28 %)	14 (64 %)	2 (13 %)	22 (100 %)	14 (61 %)	0
Typ av svensk- undervisning	sv	sv	sva	sva	sv	sv	sv	sva	19 sv/ 4 sva	sv

status än eleverna på Prismagymnasiet. Liknande förhållanden kan tänkas råda när det gäller andelen elever som bor i så kallade segregerade områden, som alltså torde vara högre på Kampusgymnasiet (se föregående avsnitt, jfr Skolverket 2004:47, 2009:96).

Tabell 2 sammanfattar bakgrundsdata för alla de tio gymnasieklasser som deltagit i undersökningen. Bl.a. framgår det att flickor är i majoritet i alla klasser förutom i klasserna 1 och 5. Dessutom visar tabellen hur många andraspråks elever som det finns i varje klass, om klassen läser svenska eller svenska som andraspråk och vilket gymnasieprogram eleverna går på. Skillnaderna mellan olika gymnasieprogram är också en faktor som i viss mån bör beaktas vid tolkning av undersökningens resultat. Informantgruppen består av elever som läser på omvårdnadsprogrammet (OP), samhällsvetenskapliga programmet (SP) och naturvetenskapliga programmet (NP), vilket alltså innebär att informantgruppen är sammansatt av både elever från i huvudsak yrkesförberedande program, liksom från i huvudsak studieförberedande utbildningar (jfr Skolverket 2008:18). Enligt Skolverket (2008:35) når elever som går på studieförberedande utbildningarna i högre grad behörighetskraven. Man betonar dock att skillnaderna mellan olika program är relativt stora. Skillnaderna i studieresultat mellan de här aktuella programmen i årskurs 3 för läsåret 2006/07 (för hela landet) återges i tabell 3. Dessutom anges hur många elever från varje program som ingår i informantgruppen som helhet.

TABELL 3: Studieresultat för relevanta utbildningar i årskurs 3 för hela landet i procent (se Skolverket 2008:25)

Program	Antal elever från respektive program i informantgruppen	Endast samlat betygsdokument (ej slutbetyg)	Slutbetyg men ej behörig	Slutbetyg med grundläggande behörighet
OP	59	ca 21 %	ca 10%	ca 69 %
SP	112	ca 13 %	ca 8 %	ca 79 %
NP	15	ca 8 %	ca 4 %	ca 88 %

Som nämndes i avsnitt 5.1 är det naturligtvis en rad olika faktorer som ligger till grund för dessa skillnader mellan olika gymnasieutbildningars resultat. Sambandet mellan elevernas studieresultat och resultaten av



denna undersökning är heller inte i fokus för detta arbete. Statistiken som redovisas i tabell 3 kan dock antas ha viss betydelse för vilka slutsatser som kan dras av analysen.

### 5.3. Informantgrupperna

Informanterna har delats in i tre huvudgrupper, efter startålder för sin inlärning av svenska. Indelningen visas i tabell 4. Grupperna är grupp *L1* (elever som har angett svenska som sitt förstaspråk och därmed antas att ha börjat inlärningen vid födseln), *L2/tidig* (elever med startålder  $\leq 6$ ) samt *L2/sen* (informanter med startålder  $> 6$ ). Den sistnämnda gruppen har i sin tur delats in i två undergrupper, nämligen informanter med en startålder till och med 10 år och informanter med en startålder efter 10 år<sup>1</sup>. Indelningen motiveras dels av språkutvecklingsmönster som visats i tidigare forskning (Abrahamsson & Hyltenstam 2009, Butler 2000), dels av ålder för skolstarten i Sverige; informanterna i grupp *L2/tidig* har påbörjat sin inlärning av svenska före skolstarten. De har alltså påbörjat sin andraspråksutveckling under den tidiga barndomen och därutöver fått hela sin skolgång på andraspråket. Grupp *L2/sen* består av informanter som påbörjat sin andraspråkutveckling efter sex års ålder, dvs. samtidigt med eller efter skolstarten. Detta innebär en relativt sen startålder, med tanke på biologiska mognadsbegränsningar för språkinlärning (Hyltenstam & Abrahamsson 2000), och därmed heller inte börjat utveckla andraspråket innan de ställts inför skolans språkliga krav. Den sena inlärargruppen har delats in i två undergrupper för att göra det möjligt att säga något om eventuella skillnader mellan elever som har fått hela, eller nästan hela, sin skolgång på andraspråket, och sådana som först i något högre ålder har kommit i kontakt med både vardaglig och skolrelaterad svenska. Startåldersgränsen för den senare gruppen ( $> 10$ ) har alltså valts för att den ligger några år efter skolstarten, men samtidigt före inträdet i puberteten, som i tidigare undersökningar

---

<sup>1</sup> Informanterna med en startålder efter 10 år kan i regel antas att under några år ha gått i skolan på sitt modersmål. Denna faktor har dock inte kontrollerats för i samband med detta arbete.

har visat sig vara en känslig ålder ur språkutvecklingssynpunkt (Abrahamsson & Hyltenstam 2004).

Utöver gruppindelningen visar tabell 4 hur många elever från varje årskurs som ingår i respektive grupp. Som framgår av tabellen är fördelningen av elever från årskurserna 1 och 2 relativt jämn i respektive grupp. Från årskurs 3 finns det däremot bara en klass bland deltagarklasserna och därmed relativt få elever från denna årskurs i varje grupp. Vilken årskurs eleverna går i kan antas ha viss relevans för deras behärskning av konventionaliserat språk då elever som har gått igenom årskurserna 1 och 2 borde ha kommit längre i sin språkutveckling generellt och också ha ett mer utvecklat skolrelaterat språkbruk (jfr avsnitt 4.4.1). Frågan är dock om denna faktor väger tyngre än t.ex. elevernas språkliga bakgrund.

TABELL 4: Antal elever från varje årskurs i grupperna

	GRUPP L1	GRUPP L2/tidig (startålder, ≤ 6)	GRUPP L2/sen			JK
			Informanter med start- ålder ≤10	Informanter med start- ålder >10	Totalt	
ÅK1	28	35	10	8	18	26
ÅK2	38	30	6	9	15	0
ÅK3	8	10	2	2	4	0
Totalt	74	75	18	19	37	26

Utöver de elever som delats in i startåldersgrupper deltog, som tidigare nämnts, även en jämförelseklass (JK) i enkätundersökningen. Som framgår av tabell 4 betår denna klass av enbart infödda svenska elever i årskurs 1 på Prismagymnasiet. I denna klass följer elever från såväl naturvetenskapligt som samhällsvetenskapligt program en utbildning med musikinriktning. Klassen har tagits med i undersökningen i syfte att möjliggöra jämförelser mellan resultaten för informantgrupperna där samtliga elever kan sägas studera i en flerspråkig miljö, med resultat för elever som läser i en mer enspråkig skolmiljö.

## 5.4. Enkäten

Under hösten 2006 genomfördes en pilotundersökning på Kampusgymnasiet för att, utifrån en första version av enkäten, undersöka vilka problem som kan uppstå, dels vid genomförandet av en enkätundersökning som den föreliggande, dels vid analysen av elevernas svar. Pilotundersökningen fokuserade endast förståelsen, dvs. elevernas receptiva behärskning av de konventionaliserade uttrycken ifråga. Analysen av provenkätens genomförande och resultat visade bl.a. att flervalformatet är att föredra framför öppna frågor av typen ”*inte sticka under stol med betyder: \_\_\_\_\_*” när det gäller att testa elevernas kunskap om uttrycks betydelse. Ett återkommande problem visade sig nämligen vara att vissa svar på olika sätt kunde ligga nära den konventionaliserade, efterlysta betydelsen, utan att vara helt korrekta. Om informanterna måste formulera betydelsen självständigt blir det alltså svårt att dra nytta av partiell kunskap om ett uttryck (jfr Nation 2001), då det är omöjligt att analysera ett sådant test i termer av ’mer eller mindre korrekt’. Dessutom blev det tydligt under arbetet med pilotundersökningen, att en undersökning av endast receptiv kunskap i samband med ungdomars behärskning av konventionaliserade uttryck skulle bli för endimensionell. Det finns, som nämnts tidigare, många andra aspekter av lexikonkunskap som är viktiga för en idiomatisk användning av dessa typer av uttryck (jfr bl.a. Nation 2001, Henriksen 1999).

Den slutgiltiga versionen av den enkät som informanterna ombads att fylla i, består av fyra olika delar (se appendix 1). Två av dem, A och D, består av flervalstest, medan delarna B och C har formen av lucktest. På enkätens första sida ombeds informanterna att besvara ett antal frågor om sig själva och sin språkanvändning. Informationen används för analyser av enkätundersökningens resultat i relation till olika bakgrundsvariabler som kön, språklig bakgrund och startålder för inläring av svenska.

I A-delen testas elevernas receptiva kunskap om sammanlagt tio olika konventionaliserade uttrycks betydelse. Till varje uttryck finns fyra svarsalternativ, ett korrekt svar, två distraktorer samt möjligheten att avvisa de övriga tre alternativen och själv föreslå uttryckets korrekta betydelse. Det sistnämnda fungerar dock endast som ytterligare distraktor, då den konventionaliserade betydelsen i samtliga fall finns bland de giv-

na alternativen (se appendix 1). Uppgifterna ser exempelvis ut på följande sätt:

Exempel 1:

ha händerna fulla

- vara rik
- ha mycket att göra
- inte kunna röra sig
- Inget av alternativen stämmer, uttrycket betyder följande:

Som nämndes i avsnitt 3.2 beror svårighetsgraden i ett flervalstest till stor del på hur nära distraktorerna ligger det korrekta svaret i betydelse (Nation 2001). Tanken bakom A-delens utformning är att en av distraktorerna ska kunna kopplas till uttryckets, eller någon av konstituenternas, bokstavliga betydelse medan den andra är tänkt att ligga så långt ifrån det korrekta svaret som möjligt, utan att för den delen kunna uteslutas direkt. Möjligheten att ge ett eget förslag tjänar alltså dels som ytterligare distraktor, dels till att göra det möjligt för eleverna att visa eventuell partiell kunskap om uttryckets betydelse som kan vara relevant för den kvalitativa analysen av materialet.

B-delen består i ett lucktest eller, som Nation (2001) kallar det, ett *sentence-completion test*. Informanterna ställs inför uppgiften att komplettera ett konventionaliserat uttryck i kontexten av en mening genom att fylla i det ord som saknas. Med Nations modell om lexikonkunskap som utgångspunkt (se avsnitt 3.1 ovan), testas på så sätt dels receptiv kunskap om ords och uttrycks betydelse i kontexten, dels produktiv kunskap om kollokationspotential och uttryckets konventionaliserade form. En viktig aspekt när det gäller B-delens utformning har varit frågan om det finns en skillnad i uppgiftens svårighetsgrad beroende på vilket led i frasen som utelämnas. För att kunna utläsa eventuella mönster i detta sammanhang har i vissa fall substantivet och i andra verbet utelämnats, vilket illustreras i exemplen nedan:

Exempel 2:

I Expedition Robinson måste man jobba för att överleva. Det funkar inte att åka *snålskjuts* \_\_\_\_\_ för då blir man uttrötad direkt.

Samma sak gäller dokusåpan Farmen. Vill man vara kvar där så måste man dra sitt strå till stacken.

Del C i enkäten kan, med Nations terminologi, betecknas som ett *supply-item test*. Här består uppgiften i att komplettera sammanlagt femton uttryck utanför en kontext på följande sätt:

Exempel 3:

gå ner i varv

Även här testas alltså produktiv kunskap om bl.a. ords kombinationsmöjligheter, men framför allt om konventionalisering; det finns naturligtvis flera ord som kan kombineras med de givna orden i testet, utan att det uppstår rent grammatiska fel eller semantisk oförenlighet. Informanten måste alltså (receptivt) känna igen ordkombinationen som ett fragment av ett konventionaliserat uttryck, samt (produktivt) välja det ord som fattas för att det konventionaliserade uttrycket ifråga ska bli komplett.

Även i enkätens sista deltest (D) ställs informanterna inför uppgiften att fylla i det utelämnade ordet i ett uttryck. Denna gång ges det däremot ett antal alternativa ord att välja bland. Här handlar det, med andra ord, om ett matchningstest. Sammanlagt ska 20 kollokationer bildas. Kollokationerna är indelade i fyra grupper med fem liknande konstruktioner (t.ex. *verb + substantiv* och *adverb + adjektiv*). Till varje grupp tillhandahålls åtta alternativa ord som kan paras ihop med de fem givna orden, av vilka tre alltså enbart tjänar som distraktorer. Testformatet exemplifieras nedan.

Exempel 4:

Fyll i det ord som saknas i följande uttryck. Välj bland följande alternativ:  
*bilda, dra, föra, hålla, leda, lägga, bryta, ta*

föra ett samtal  
hålla ett tal  
bryta ett beroende  
bilda en union  
dra en slutsats

Informanterna görs uppmärksamma på att alla ord inte behöver användas. Däremot nämns inte i uppgiften att det kan finnas två likvärdiga alternativ att välja bland orden, vilket dock bara är fallet i samband med en kollokation – nämligen *vara/ komma till nytta* (se appendix 1). Informanterna får heller inte veta om samma ord kan användas i flera av de fem kollokationerna (vilket inte är fallet). Återigen är alltså syftet att testa informanternas produktiva kunskap om ords kollokationspotential och konventionaliserade ordförbindelser. Som antytts tidigare har utgångspunkten för denna del av undersökning varit att testa hypotesen om skillnaden mellan första- och andraspråksanvändare (även på avancerad nivå) när det gäller uppgiften att bilda kollokationerna genom att para ihop ett antal givna ord (se avsnitt 4.4).

Urvalet av konventionaliserade uttryck i delarna A, B och C är baserat på en tidigare undersökning av Prentice & Sköldbäck (und. utg.), där materialet består av olika typer av figurativa ordförbindelser hämtade från uppsatser som samlats in inom ramen för projektet *Språk och språkbruk bland ungdomar i flerspråkiga storstadsmiljöer*.<sup>2</sup> Tanken bakom urvalet är att testa konventionaliserade uttryck som förekommer i texter skrivna av gymnasieelever i motsvarande ålder och skolmiljö. Det handlar alltså om uttryck som använts i skolrelaterat språkbruk, även om de inte är av den typen som är mest frekvent i t.ex. läromedel. Kollokationstestet i D är däremot sammansatt med utgångspunkt från två olika korpusar, nämligen *OrdiL-korpusen* (Lindberg & Johansson Kokkinakis 2007) och korpusen *LäSBarT* (Mühlenbock 2009). OrdiL-materialet består av läromedelstexter i NO, SO och matematik för grundskolans senare år. Korpusen innehåller ca 980.000 löpord (Lindberg 2007). LäSBarT är sammansatt av lättlästa nyhetstexter, informationstexter från kommunen, skönlitteratur och barnlitteratur. Materialet består av sammanlagt 1,5 millioner löpord (Mühlenbock & Johansson Kokkinakis 2009). Fyra olika typer av kollokationer har hämtats från de frekvenslistor som bygger på OrdiL och LäSBarT<sup>3</sup>, nämligen *verb + substantiv*, *verb + preposition + substantiv*, *adjektiv + substantiv* samt *adverb +*

---

<sup>2</sup> Projektet genomfördes i samarbete mellan Göteborgs, Stockholm och Lunds universitet och finansierades av Riksbankens jubileumsfond år 2002–2006.

<sup>3</sup> Ett varmt tack till FD Sofie Johansson Kokkinakis för hjälp med framtagning av speciella frekvenslistor ur OrdiL och LäSBarT.

*adjektiv*, i syfte att undersöka om det finns vissa typer av kollokationer som vållar mer problem för informanterna än andra. Då de nämnda korpusarna inte är speciellt stora har även kollokationer som förekommer endast en gång valts ut för undersökningen. Urvalet har alltså snarare varit kvalitativt än kvantitativt. Uttrycken har dock sökts i konkordansmaterialet i Språkbanken (2010), och i sökmotorn Google (2010a) för en bedömning av hur pass vanliga de är i allmänspråket. De flesta kollokationerna som valts ut kan sägas vara relativt frekventa, enligt referensmaterialen. Jag har dock även valt att ta med ett mindre antal ovanliga kollokationer, för att undersöka i vilken mån frekvens i allmänspråket kan tänkas påverka elevernas resultat. Dessa kollokationer är *bilda en union*, *dra en suck*, *hamna i missbruk* och *bryta ett beroende*. Kollokationerna ur korpusmaterialet har sedan kompletterats med ytterligare ett antal uttryck med liknande lexikal och syntaktisk struktur, för att varje typ ska vara representerad med samma antal exempel. Syftet har i detta sammanhang varit att få ihop fem liknande konstruktioner samt åtta alternativa ord att välja bland, där inget skulle kunna uteslutas för enkelt. Tanken är naturligtvis att eleverna ska använda sin lexikala kunskap, snarare än uteslutningsmetoden, för att lösa uppgiften.

Uppgifterna i enkäten har formulerats för att framstå så tydliga och lättbegripliga som möjligt, utan att erbjuda några ledtrådar för hur uppgiften ska lösas. Begreppet *konventionalisering* nämns inte i uppgifterna. I stället betecknas uttrycken som ”vanliga” (del B). Informanterna ombeds fylla i de ord som de tycker saknas (del B och C) alternativt para ihop de ord som de anser ”passar bäst ihop” (del D). Informanterna har inte fått någon tidsbegränsning för att fylla i enkäten. De har däremot uppmanats att inte fundera för länge på de enskilda uppgifterna, då man inte kan resonera sig fram till lösningen, utan helt enkelt måste känna till uttrycket ifråga. Det har tagit mellan 10 och 20 minuter för de flesta informanter att fylla i enkäten. Författaren var själv närvarande vid undersökningens genomförande i samtliga grupper.

## 5.5. Analysmetod

Den jämförande analysen av enkätsvaren är i första hand kvantitativ. Varje svar i enkäterna har antingen bedömts som korrekt eller som avvi-

kande från det konventionaliserade uttrycket ifråga. Varje korrekt svar har givit ett poäng. Informanternas resultat har sedan jämförts på olika sätt och utifrån olika grupperingar (se vidare avsnitt 4). Dessutom har resultaten för de olika deltesten jämförts med varandra, liksom resultaten för de olika uttrycken. I den senare delen av analysen spelar alltså även kvalitativa aspekter en viss roll. Resultatskillnaderna mellan de olika informantgrupperna och enkätens olika komponenter har sedan testats för statistisk signifikans<sup>4</sup>. Skillnaderna har testats med hjälp av *en-vägs variansanalys* (ANOVA) samt *Newman-Keuls-testet* för multipla jämförelser (när endast två grupper jämförs sammanfaller ANOVA med ett två-sample t-test och inget multipeltest behövs). I samband med denna undersökning anses en skillnad vara signifikant om p-värdet är mindre eller lika med 0.05. Normalfördelningsbaserade parametriska metoder har använts, eftersom antalet observationer är så stort att den centrala gränsvärdessatsen kan anses ge approximativt normalfördelade medelvärden och frekvenser. De ovannämnda metoderna samt *Pearson korrelationskoefficient* har dessutom använts för att testa undersökningsmetodens validitet. Frågan är i det sammanhanget i vilken utsträckning de enskilda uppgifterna i enkäten skiljer informanter med bra resultat på hela enkäten ifrån dem med sämre resultat. Detta kan ses som en indikation på hur bra de enskilda uppgifterna (dvs. varje uttryck som skulle kompletteras) har fungerat när det gäller att påvisa skillnader mellan olika elevgrupper (jfr Körner & Wahlgren 2005 ).<sup>5</sup>

En del av materialet har även analyserats kvalitativt. Utgångspunkten här har dels varit svaren från de informanter som har lägst resultat i varje grupp, dels de tio uttryck i enkäten, som genererat lägst antal korrekta svar. Till att börja med har enkätsvaren från de 45 av de lägst presterande informanterna analyserats för de uttryck som visat sig vara svårast för informanterna. För att bekräfta eventuella mönster har sedan ytterligare tio av enkäterna med lägst antal poäng analyserats. Dessutom har enkäterna med högsta resultat, dvs. minst 45 av 55 möjliga poäng

---

<sup>4</sup> Ett varmt tack till FD Tommy Johnsson från Institutionen för filosofi, lingvistik och vetenskapsteori, Göteborgs universitet, för hjälp med de statistiska analyserna.

<sup>5</sup> Pga en felaktig dubbelkodning har de statistiska analyserna gjorts på 187 informanter (75 i grupp L1). Detta har dock ingen större betydelse för resultaten. Siffrorna har rättats till i tabellerna.



analyserats kvalitativt. Syftet har i detta fall varit att undersöka i vilken mån avvikelserna från de konventionaliserade formerna bland de högpresterande informanterna skiljer sig från avvikelserna bland informanterna med lägre resultat. Med utgångspunkt i den tidigare nämnda undersökningen av Ellis et al. (2008) har dessutom eventuella samband mellan uttryckens frekvens och andraspråkselevernas avvikelser från dessa konventionaliserade flerordsenheter undersökts, med avseende på kvantitativa och kvalitativa resultat. PAROLE-korpusen i Språkbanken, som består av 19.4 miljoner löpord (Språkbanken 2010) och innehåller olika typer av textmaterial, har används som referenspunkter när det gäller uttryckens frekvens i det allmänna språkbruket. I viss mån även Google (2010a) använts i samma syfte.

## 6. Resultat

I detta avsnitt redovisas resultaten för en rad olika analyser av materialet. Som ett första steg i den kvantitativa analysen har de deltagande gymnasieklassernas resultat jämförts med varandra, dels som hela klasser (se tabell 5), dels utifrån en indelning i första- och andraspråkselever (se tabell 6). Denna grova indelning förfinas sedan stegvis. Vissa indelningar har dock inte följts upp genom hela analysen då det visat sig inte finnas några relevanta skillnader mellan testresultaten för vissa undergrupper (se vidare avsnitt 6.2). Bl.a. har testresultaten analyserats i förhållande till kön (även här utifrån en indelning i första- och andraspråkstalare). Då det inte finns några större skillnader mellan testresultaten för kvinnliga och manliga elever, ges dessa resultat dock inte något större utrymme här. Vidare har de olika informantgruppernas resultat för varje deltest jämförts, liksom resultaten för de enskilda uttrycken i enkäten. Avslutningsvis redovisas den kvalitativa delen av analysen, där eventuella mönster i avvikelser från de konventionaliserade uttrycken ifråga har undersökts.

### 6.1. Resultaten för de deltagande gymnasieklasserna

Tittar man på resultatet för de sammanlagt 9 gymnasieklasser som har deltagit i undersökningen, samt jämförelseklassen, är det de tre svenska

som andraspråksklasserna på Kampusgymnasiet, klasserna 3, 4 och 8, vars resultat (24 %, 40 % respektive 43 % av maximal poäng) som skiljer sig signifikant från de övriga (p-värde <0.0001). Klass 3 når dessutom ett signifikant lägre resultat än klasserna 4 och 8 (p-värden <0.0001). De övriga klasserna når resultat mellan 61 % (klass 2 på Kampusgymnasiet) och 71% (klass 5 på Prismagymnasiet) av det maximala resultatet (se tabell 5).

Jämförelseklassens resultat ligger med 82 % av det maximala antalet poäng signifikant högre än samtliga övriga deltagarklassers resultat (p-värde <0.0001). Överlag kan de resultat som sammanfattas i tabell 5 alltså anses vara förväntade. Om man bortser från de rena andraspråksklasserna och jämförelseklassen, verkar det däremot inte finnas något direkt samband mellan klassernas resultat och andelen andraspråkselever i klasserna; resultatskillnaderna mellan klasserna 1, 2, 5, 6, 7 och 9 är inte särskilt stora och inte heller statistiskt signifikanta (se tabell 5).

Det verkar inte heller finnas något uppenbart samband mellan årskurs och resultat. Klasserna i årskurs 1 når visserligen något lägre resultat och som framgår av tabell 5, är det klass 3, en andraspråksklass i årskurs 1, som når det signifikant lägsta resultatet. Det finns dock ingen signifikant skillnad i resultat mellan de övriga andraspråksklasserna eller mellan klass 4 (årskurs 1) och 8 (årskurs 2). Klass 9, som är den enda klassen i årskurs 3, skiljer sig heller inte signifikant från de lägre årskurserna. Dessutom går jämförelseklassen, som tidigare nämnts, i årskurs 1 samtidigt som den når det signifikant högsta resultatet.

Det bör i detta sammanhang poängteras att de signifikanta skillnaderna mellan de deltagarklasserna som läser svenska respektive svenska som andraspråk inte bör tolkas som en följd av skillnader i fråga om undervisningsämne. Att elever i de klasser som läser svenska som andraspråk når signifikant lägre resultat än elever som läser svenska, hänger i första hand samman med andelen elever med utländsk bakgrund i klassen, som är 100 % i andraspråksklasserna. Det bör dock noteras att det inte finns några signifikanta skillnader som kan kopplas till andelen andraspråkselever i respektive klass mellan de klasser som läser svenska som modersmål. Däremot är majoriteten av andraspråkseleverna i dessa klasser tidiga inlärare medan det stora flertalet av de elever som har påbörjat sin andraspråksinlärning efter sexårsåldern går i andraspråksklasserna 3, 4 och 8. Skillnaden mellan klasserna som läser svenska och de

TABELL 5: Testresultat per gymnasieklass

	KLASS 3	KLASS 4	KLASS 8	KLASS 2	KLASS 9	KLASS 7	KLASS 6	KLASS 1	KLASS 5	JK
Skola:årskurs	Kampus:1	Kampus:1	Kampus:2	Kampus:1	Kampus:3	Kampus:2	Kampus:2	Prisma:1	Prisma:2	Prisma:1
Typ av svensk- undervisning	sva	sva	sva	sv	19 sv/ 4 sva	sv	sv	sv	sv	sv
Antal elever	14	23	22	17	22	16	23	27	22	26
Antal elever med	14	23	22	5	14	2	14	11	7	0
Svenska som L2	(100 %)	(100 %)	(100 %)	(26 %)	(61 %)	(13 %)	(64 %)	(41 %)	(28 %)	
Resultat	24 %	40 %	43 %	61 %	63 %	68 %	69 %	62 %	71 %	82 %

som läser svenska som andraspråk lär således i första hand bero på startålder för andraspråksinlärning, snarare än på faktorerna utländsk bakgrund eller typ av svenskundervisning i sig.

## 6.2. Resultaten i förhållande till språklig bakgrund

I följande avsnitt redovisas och jämförs testresultaten för hela enkäten per informantgrupp. Dessutom har andraspråkselevernas resultat i de olika startåldersgrupperna jämförts. Som framgår av tabell 6, når eleverna med svenska som förstaspråk (grupp L1), helt enligt förväntningarna, det högsta resultatet bland de sammanlagt fyra informantgrupperna (72 % av maximal poäng). Jämförelseklassens resultat är med 82 % korrekta svar signifikant högre än L1-gruppens (p-värde 0.0031). Av grupperna med svenska som andraspråk når grupp L2/tidig det näst högsta resultatet (50 %), medan resultatet för grupp L2/sen (39 %) ligger tydligt under den förstnämnda. Skillnaderna mellan de tre huvudgrupperna är statistiskt signifikanta (p-värde <0.0001). Tittar man närmare på de två undergrupperna i gruppen L2/sen, ser man ett mönster som överensstämmer med tidigare forskning (jfr bl.a. Thomas & Collier 2002, Hyltenstam & Abrahamsson 2000) som visat att senare startålder och en påbörjad inlärning efter skolstart har relativt stor betydelse för andraspråksbehärsningen såväl generellt som för utvecklingen av ett mer avancerat skolrelaterat språk. Det finns ingen statistisk signifikant skillnad mellan grupp L2/tidig och informanterna med en startålder till och med 10 år i grupp L2/sen. Däremot finns en signifikant skillnad (p-värde <0.0001) mellan informanterna med senare startålder (>10 år) i grupp L2/sen, som når 33 % av maximalt resultat, och de övriga andraspråksgrupperna, dvs. grupp L2/tidigt samt informanter med startålder  $\leq 10$ , som båda når högre resultat (50 respektive 46 %).

Mest relevanta, vad gäller de resultat som sammanfattas i tabell 6 nedan, är alltså skillnaderna

- mellan förstaspråkselever och andraspråkselever med en startålder för inlärningen  $\leq 10$  år,

- mellan förstaspråkselever och andraspråkselever med en startålder >10 år,
- mellan andraspråkselever med en startålder ≤10 år och andraspråkselever med en startålder >10 år.

Det handlar, med andra ord, om en generell skillnad dels mellan första- och andraspråkselever, dels mellan andraspråkselever med olika startålder för inläringen, där 10 år verkar vara en viktig åldergräns.

TABELL 6: Resultat för hela enkäten) per startåldersgrupp

	GRUPP L1	GRUPP L2/tidig (startålder ≤ 6)	GRUPP L2/sen			JK
			start- ålder ≤10	start- ålder >10	totalt	
Antal elever	74	75	19	18	37	26
Resultat	72%	50%	46%	33%	39%	82%

I tabell 6.1 redovisas enbart testresultaten för de informanter som har angett svenska som sitt andraspråk. I denna del av analysen har både grupp L2/tidig och grupp L2/sen delats upp i två undergrupper, beroende på födelseland och startålder. Analysen visar inga nämnvärda skillnader mellan informanter födda i Sverige och informanter födda utomlands i grupp L2/tidig, vilket är anledningen till att denna uppdelning inte följs upp vidare i detta arbete.

TABELL 6.1: Detaljerad redovisning av resultat för informanterna med svenska som L2

	GRUPPL2/tidig (startålder ≤6)			GRUPP L2/sen (start i skolådern)			JK
	födda i Sverige	födda utoml.	totalt	startålder ≤ 6	startålder >10	totalt	
	Antal elever	45	30	75	18	19	
Antal elever som läser sv	29	14 *	43 *	8	1 *	9 *	26
Antal elever som läser sva	16	17 *	33 *	10	17 **	27 *	0
Antal elever som läst modersmål	33	25	58	16	17	33 *	0
Resultat	50%	49%	50%	46%	33%	39%	82%

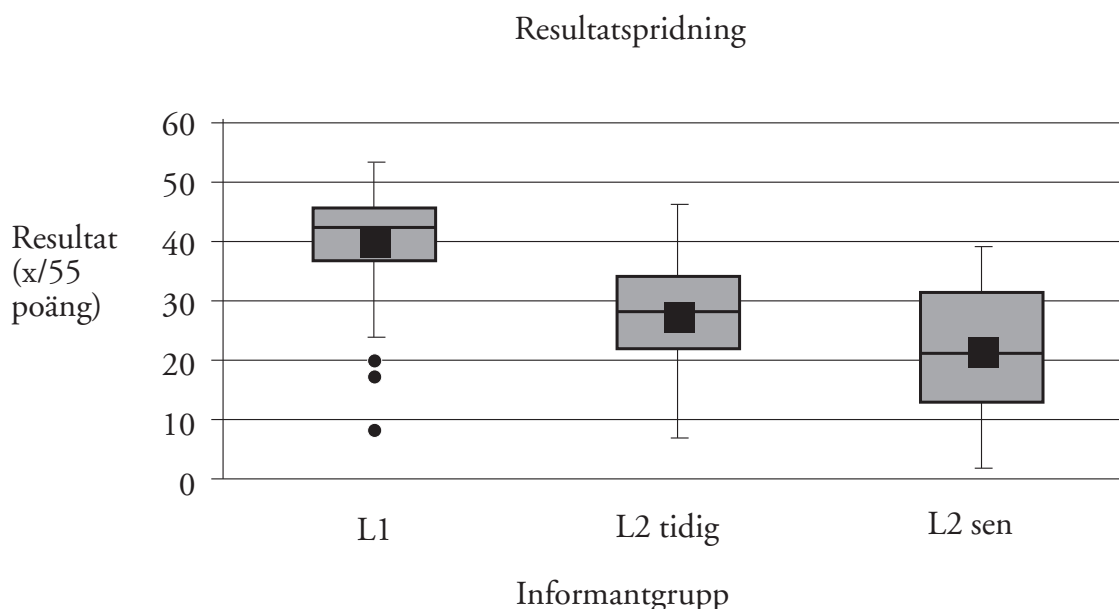
\* 1 av informanterna i gruppen läser både sv och sva

\*\* 1 av informanterna i gruppen läser både sv och sva, 1 har kryssat för ”vet ej” och 1 har inte besvarat frågan.

### 6.2.1. Resultatspridning

Huvudfokus i den kvantitativa analysen ligger av naturliga skäl på resultaten på gruppnivå. Det är likväl intressant att visa hur resultaten är fördelade på individnivå, dvs. hur resultaten sprider sig mellan 0 och 55 i informantgrupperna. Figur 2 visar skillnaderna i resultatspridning mellan informantgrupperna. Lådorna som visas i diagrammet (eller *lådagrammet*, se Wahlgren 2005) visar det spann inom vilken 50 % av alla resultat ligger. Den lodräta linjen visar hur resten av värdena sprider sig ovan och nedanför medianen, som representeras av det vågräta strecket i lådorna. Fyrkanterna i respektive låda representerar medelvärden. Extremt avvikande värden (eng. *outliers*) i förhållande till medianen markeras med egna symboler (se grupp L1 i figur 2). Det framgår tydligt av lådagrammet att den minsta spridningen finns i grupp L1 och den största i grupp L2/sen. Resultatspridningen blir alltså större, ju senare informanterna påbörjat sin inlärning av svenska. Lådagrammet visar också

att medelvärdet sjunker med stigande startålder; detsamma gäller de maximala och minimala resultatvärdena. Det finns dock samtidigt en del andraspråkslever som når betydligt högre resultat än en del av eleverna i grupp L1. Avvikande värden (som markeras med prickar i figur 2) förekommer enbart i grupp L1, där det endast förekommer 3 resultat på 20 poäng eller lägre (se även figur 3.1 nedan).

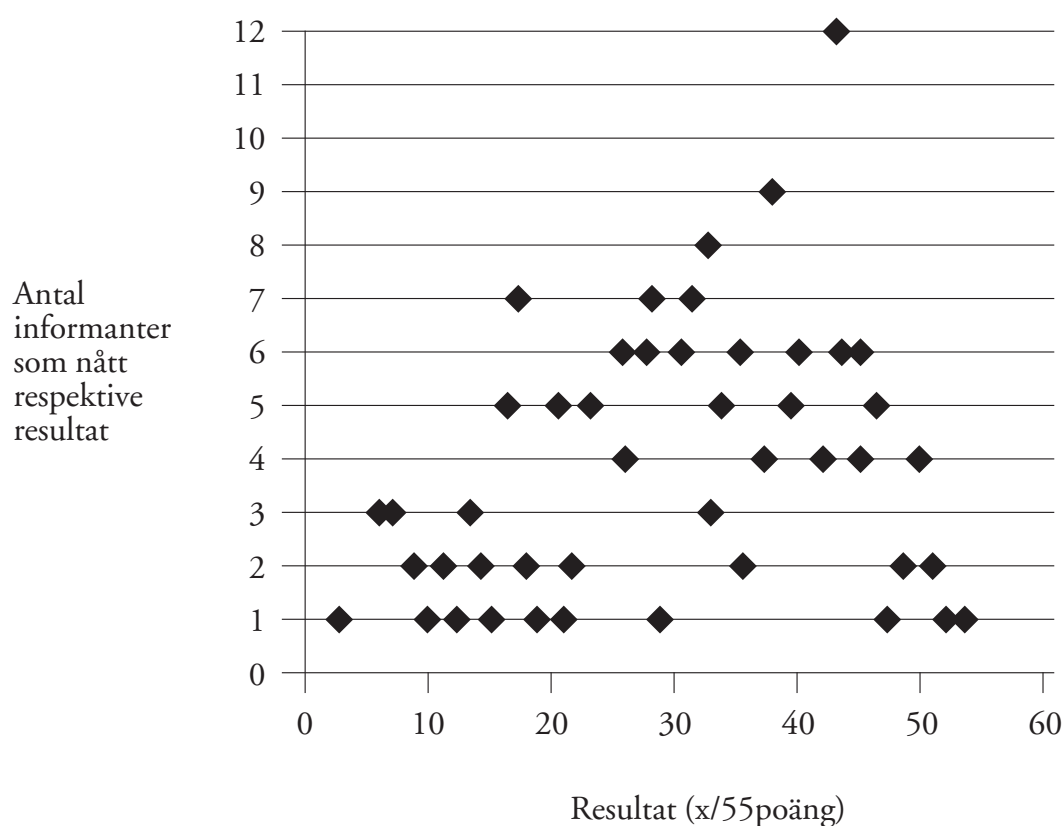


FIGUR 2: Jämförelse av resultatspridningen mellan informantgrupperna

Figurerna 3-3.3 nedan visar något mer detaljerat vilka resultat som förekommer i informantgrupperna och hur många av informanterna som nått ett visst resultat. I figur 3 visas denna resultatspridning för samtliga 186 informanter. Här framgår bl.a. att det vanligaste resultat är 42 av 55 poäng. De resultat som förekommer endast en enstaka gång ligger antingen under 30 eller runt 50 poäng. Det högsta resultatet är 53 av 55 poäng. Detta resultat förekommer endast en gång bland de 186 informanterna. Det lägsta resultatet är 2 av 55 poäng och förekommer också endast en gång i hela informantgruppen.

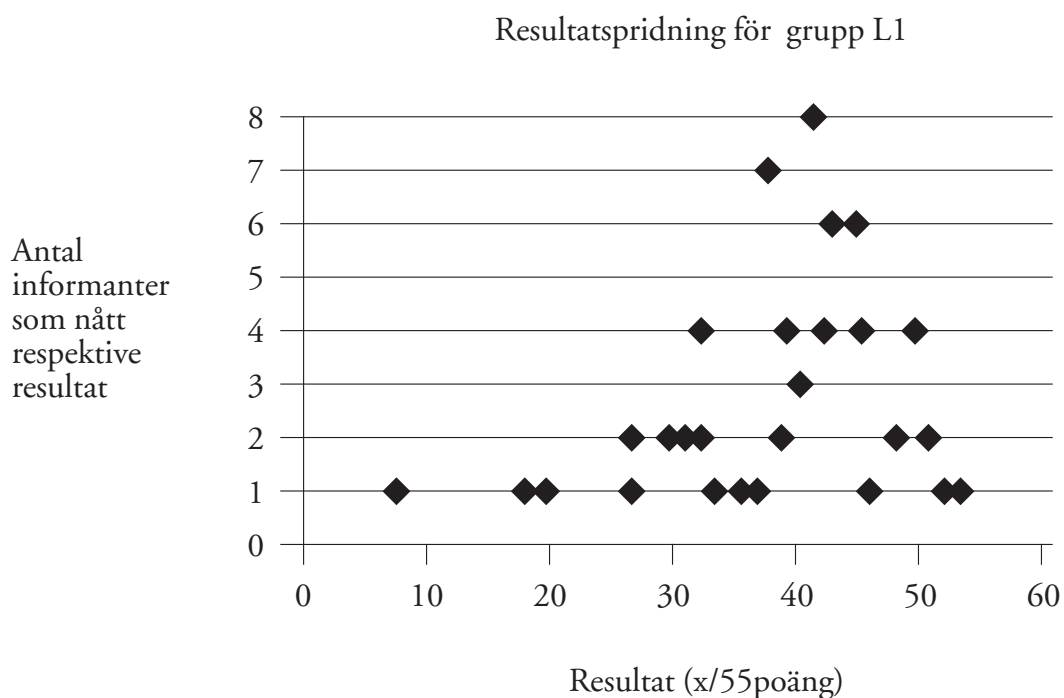


### Resultatspridning för alla informanter



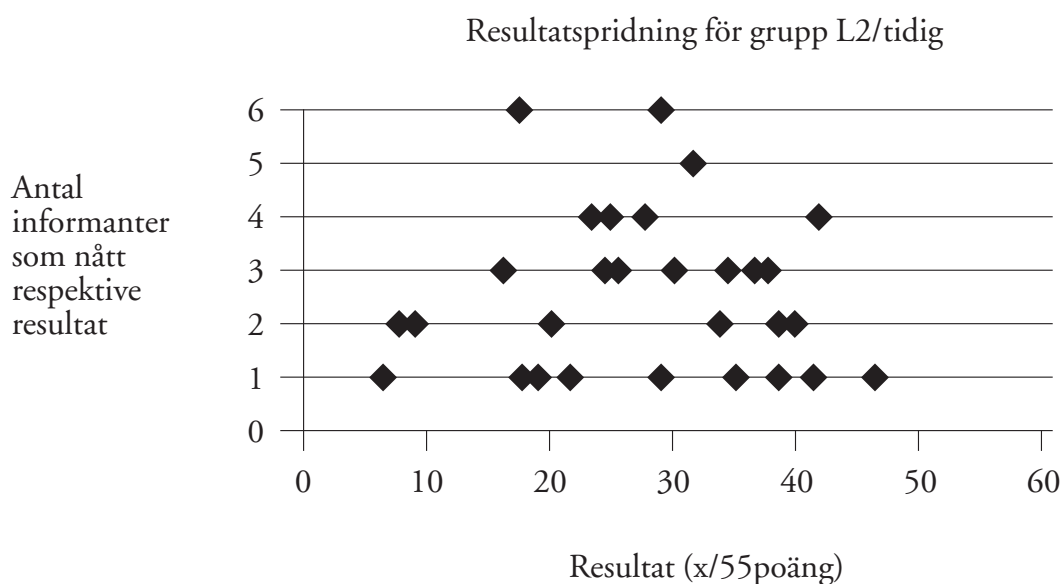
FIGUR 3: Antal informanter som nått olika resultat i hela informantgruppen

Figur 3.1 visar resultatspridningen i grupp L1. Som framgår här är det vanligaste resultatet (som förekommer 8 gånger) i denna grupp 42 av 55 poäng (precis som för hela informantgruppen, jfr figur 2). Majoriteten (42 av 74) av resultaten i grupp L1 ligger mellan 40 och 50 poäng och endast 6 resultat ligger under 30 poäng i denna grupp. Det högsta resultatet (som förekommer en gång) är 53 av 55 poäng och det lägsta resultatet bland dessa elever är 8 av 55 poäng. Även detta resultat förekommer endast en gång. Det finns alltså enstaka resultat under 20 och ett under 10 poäng även i grupp L1. Överlag kan man dock konstatera att spridningen är relativt liten, då det stora flertalet av resultaten ligger mellan 30 och 53 (se figur 3.1).



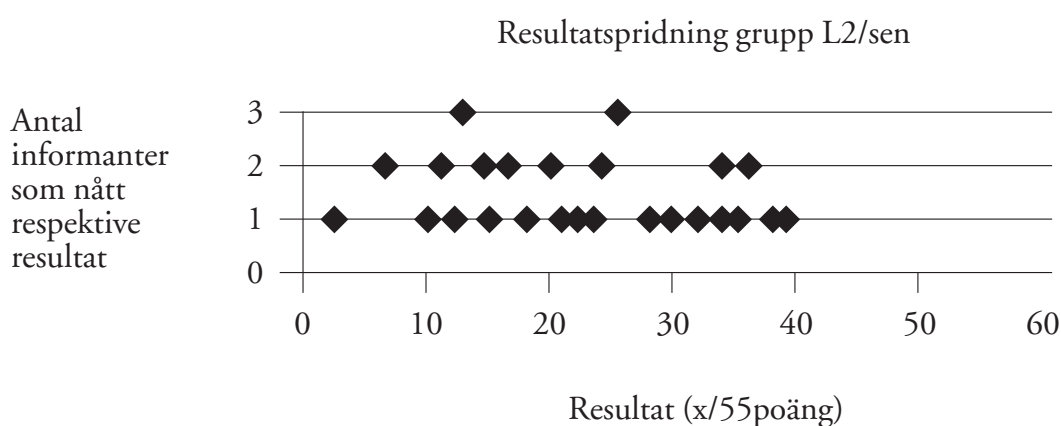
FIGUR 3.1: Antal informanter som nått olika resultat i grupp L1

Av figur 3.2 framgår att de resultat som är vanligast (med 6 förekomster var) i grupp L2/tidig är 17 respektive 28 poäng. Flertalet av resultaten ligger mellan 20 och 46. Här finns alltså en något större spridning än i grupp L1. I grupp L2/tidig är 46 – ett resultat som förekommer endast en gång – det högsta resultatet. Det lägsta resultat i denna grupp är 7 poäng och förekommer också endast en gång.



FIGUR 3.2: Antal informanter som nått olika resultat i grupp L2/tidig

Spridningen i grupp L2/sen skiljer sig tydligt från den i de övriga grupperna (se även figur 2); som framgår av figur 3.3 förekommer samma resultat maximalt tre gånger i denna informantgrupp, och detta gäller endast 2 olika resultat. De allra flesta resultaten förekommer endast en eller två gånger. Som redan framgått av figur 2 ovan, förefaller den individuella variationen alltså vara större hos de sena andraspråksinlärarna än hos övriga informanterna. Större delen av resultaten (21 av 37) ligger mellan 20 och 39. Det högsta resultatet i grupp L2/sen är 39 av 55 poäng och det lägsta 2 av 55 poäng.



FIGUR 3.3: Antal informanter som nått olika resultat i grupp L1

### 6.3. De olika deltesten

Syftet med denna del av analysen är dels att undersöka eventuella mönster i fråga om skillnader mellan delresultaten inom samma grupp, dels att undersöka skillnader i fråga om resultat för de olika deltesten mellan informantgrupperna. Analyserna syftar alltså till att besvara frågan om vilken typ av uppgifter informanterna har klarat bättre respektive sämre. Resultaten sammanfattas i tabell 7.

TABELL 7: Resultat för enkätens olika deltester (% av maximalt resultat)

	GRUPP L1	GRUPP L2/tidig (startålder, ≤ 6)	GRUPP L2/sen		JK	
			start- ålder ≤10	start- ålder >10		
Antal elever	74	75	18	19	37	26
A Förståelse	84%	72%	72%	57%	64%	91%
B Produktion i kontext	72%	41%	33%	16%	24%	83%
C Produktion utanför kontext	64%	34%	24%	14%	19%	74%
D Kollokationer	74%	55%	54%	43%	49%	83%

Som framgår av tabell 7, når alla informantgrupper (inklusive jämförelseklassen) sina högsta resultat i del A. Med tanke på tidigare refererad forskning är detta inget oväntat resultat. Uppgifter som i första hand kräver receptiv behärskning av ordförrådet har i många studier visat sig vara lättare för både första- och andraspråkstalare, än sådana där man måste använda sig av produktiva språkfärdigheter (Nation 2001). Dessutom har Golden (2005), som tidigare nämnts, visat att det inte nödvändigtvis krävs full behärskning av ett uttrycks alla lexikala komponenter för att kunna förstå ett konventionaliserat uttryck. Av tabell 7 framgår vidare att skillnaden mellan grupp L1 och andraspråksgруппerna inte är lika stor i del A som i de övriga deltesterna. Icke desto mindre är skillnaderna mellan grupp L1 och samtliga andraspråksgруппerna statistiskt signifikanta. Detta gäller också skillnaden mellan gruppen med den senaste startåldern (>10) och de övriga informantgrupperna (p-värde <0.0001). Resultatskillnaden mellan jämförelseklassen och grupp L1 är däremot inte statistiskt signifikant för denna del av enkäten.

Resultaten i B-delen är betydligt mer spridda än i A-delen. Resultaten sträcker sig från 72 % i grupp L1 till endast 16% i gruppen med den senaste startåldern (se tabell 7). Mönstret liknar dock det i förståelsetestet i det avseendet att grupp L1 når det högsta resultatet (72 %), följt av grupp L2/tidig (41%) och de tidigare inlärarna i grupp L2/sen (33%). De senare inlärarna i grupp L2/sen når med 16 % det klart lägsta resultatet även här. Jämförelseklassens 83 % är signifikant högre än resultatet för grupp L1 (p-värde = 0.0080). Som i del A är det grupp L1 och de senare inlärarna i grupp L2/sen som skiljer sig signifikant från samtliga övriga grupper i sina resultat (p-värde <0.0001). Skillnaden mellan dessa två informantgrupper är störst i denna del av enkäten. Som nämnades i avsnitt 5.4 har en aspekt vid utformningen av del B varit att undersöka om det finns resultatskillnader beroende på vilket led i uttrycket som utelämnas. Analysen visar dock inga nämnvärda skillnader i resultat mellan de uppgifter där verbet utelämnas och de där informanterna ska komplettera uttrycket med ett substantiv.

Samtliga informantgrupper når sina lägsta resultat i enkätens tredje del (C). Alla resultatskillnader mellan informantgrupperna är statistiskt signifikanta (p-värde > 0.0001) i denna del av enkäten. Mönstret för skillnaderna är detsamma som i de övriga delarna, dvs. grupp L1 når det högsta resultatet (64 %) och de senaste inlärarna i grupp L2/sen når det lägsta resultatet (14 %). Det verkar alltså som om det är svårast för informanterna att komplettera ett ofullständigt konventionaliserat uttryck utan kontext. Även dessa resultat ligger i linje med tidigare forskning (Nation 2001), som har visat att uppgifter av det slag som ställts i del C, kräver en rad olika typer av kunskap som språkbrukaren till fullo måste utnyttja när han eller hon inte har hjälp av kontexten.

Den sista delen (D) i enkäten är ett kollokationstest. Utgångspunkten för detta test har varit en hypotes om att det råder stor enighet bland förstaspråkstalare om hur ett antal givna ord ska paras ihop för att bilda de mest idiomatiska kollokationerna, medan andraspråkstalare uppvisar betydligt mer variation i sina lösningar av en sådan uppgift (van Lier 1995, Lindberg 2004). Resultaten av analysen av den sista delen i enkäten kan sägas ligga i linje med denna hypotes och visar att det finns en signifikant skillnad (p-värde <0.0001) mellan grupp L1 och båda L2-grupperna. Skillnaden mellan grupp L2/tidig och grupp L2/sen är däremot inte statistiskt signifikant. Det gäller även skillnaden mellan de se-

naste inlärarna (>10) i grupp L2/sen och inlärarna i grupp L2/tidig. Även om siffrorna i tabell 7 antyder vissa tendenser till att andraspråksinlärare med tidigare startålder har lättare för kollokationsuppgiften än informanter med mycket senare startålder, är alltså den enda signifikanta resultatskillnaden den mellan första- och andraspråkstalare. Jämförelseklassen når även i det sista deltestet ett signifikant högre resultat än för samtliga övriga grupper (p-värde = 0.0315).

### 6.3.1. Skillnader mellan delresultaten inom varje grupp

En annan intressant fråga i samband med de olika deltesten i enkäten är vilka resultatskillnader som finns mellan deltesten inom respektive informantgrupp. Som tidigare nämnts visar resultaten ett tydligt mönster, nämligen att samtliga grupper når sitt högsta resultat i A-delen och sitt lägsta i C-delen. Därutöver visar sig ett liknande mönster när det gäller de övriga delresultaten i samtliga grupper. Med andra ord är rangordningen mellan delresultaten densamma för varje grupp, vilket framgår i tabell 7.1.

TABELL 7.1: Delresultaten i fallande ordning (% av maximalt resultat) per grupp.

	GRUPP L1	GRUPP L2/tidig (startålder, ≤ 6)	GRUPP L2/sen			JK
			start- ålder ≤10	start- ålder >10	totalt	
Antal elever	74	75	18	19	37	26
A	84%	72%	72%	57%	64%	91%
D	74%	55%	54%	43%	49%	83%
B	72%	41%	33%	16%	24%	83%
C	64%	34%	24%	14%	19%	74%

När det gäller resultatskillnadernas signifikans finns det däremot viss variation grupperna emellan. Jämför man delresultaten för grupp L1, finns det signifikanta skillnader mellan alla deltest, utom mellan del B och D; resultaten för båda dessa deltest är signifikant lägre än delresul-

tatet för A-delen och signifikant högre än delresultatet för C-delen (p-värde <0.0001). I grupp L2/tidig är däremot alla skillnader mellan delresultaten statistiskt signifikanta (p-värde <0.0001). I båda startåldersgrupperna inom grupp L2/sen är alla skillnader, förutom den mellan delresultaten för del B och C, statistiskt signifikanta (p-värde <0.0001). I jämförelsegruppen skiljer sig endast det lägsta delresultatet (C) signifikant från de övriga (p-värde = 0.0001).

Trots att inte alla skillnader mellan delresultaten är statistiskt signifikanta kan man alltså konstatera att det finns en tendens till ett enhetligt mönster bland informantgrupperna, när det gäller vilken typ av uppgifter i enkäten man klarar bättre respektive sämre.

## 6.4. De enskilda uttrycken

För att undersöka vilka typer av uttryck som vållar mer respektive mindre problem för informanterna – i synnerhet med tanke på andraspråkseleverna – har resultaten för de enskilda uttrycken analyserats och jämförts med varandra i informantgrupperna. Dessa resultat redovisas här för de olika delarna i enkäten.

### 6.4.1. Förståelse

Som redan nämnts, när samliga informantgrupper de högsta resultaten i enkätens första del, förståelsetestet A, där uttryckets betydelse ska anges. I tabell 8 redovisas resultaten för de enskilda uttrycken i förståelsetestet i fallande ordning, dvs. de uttryck som genererat högst poäng (om man räknar ihop resultaten från alla informantgrupper) står högst upp i tabellen. En del av resultaten kan betraktas som förväntade. Om man utgår från faktorer som t.ex. genomskinlighet, är det inte särskilt överraskande att fler elever förstår uttrycket *sätta färg på något* än uttrycket *prata i nattmössan*. I andra fall kan man, särskilt med tanke på andraspråksgруппernas resultat, hitta möjliga förklaringar i uttryckets frekvens i språkbruket (jfr Ellis et al. 2008). Uttrycket *innan man vet ordet av*, som kan beskrivas som icke-kompositionellt då de enda semantiskt genomskinliga komponenterna är formorden *innan* och *man*, genererar

exempelvis relativt högt resultat i alla grupper. Här finns heller inga signifikanta skillnader mellan startåldersgrupperna. I PAROLE förekommer detta uttryck 19 gånger (på söksträngen "v. \* ordet av"), medan det endast finns 8 belägg (på söksträngen "i nattmössan"<sup>1</sup>) på *prata i nattmössan* (Språkbanken 2010a). Google visar ca 134 000 träffar på söksträngen *innan man "vet ordet av"* samt ca 66 000 på *innan man "visste ordet av"* – gentemot ca 3 800 belägg av *prata "i nattmössan"* (Google 2010a)<sup>2</sup>. Ser man däremot till den totala bilden som framgår av tabell 8, finns det dock en del resultat som inte enbart kan förklaras med faktorer som hänger ihop med frekvens eller graden av kompositionalitet och genomskinlighet.

Det finns relativt få signifikanta skillnader mellan informantgrupperna när det gäller de uttryck som informantgruppen som helhet klarar bäst. 7 av 10 uttryck i A-delen får sammanlagt mellan 80 % och 90 % av det maximala resultatet. I endast 2 av dessa 7 fall når de senaste inlärarna i grupp L2/sen ett signifikant lägre resultat än de övriga startåldersgrupperna. Detta gäller uttrycken *fastna för något* (p-värde <0.0001) och *ha mage att göra ngt* (p-värde <0.0001). Jämförelsegruppens resultat ligger mellan 96% och 100 % för de sju uttryck som ligger överst i tabell 8. Inget av dessa sju resultat skiljer sig signifikant från resultaten för grupp L1.

Resultaten för uttrycket *något står en upp i halsen* är något iögonfallande, med avseende på resultatskillnader mellan startåldersgrupperna; i detta fall är det startåldersgruppen  $\leq 10$  som med 94 % når det signifikant högsta resultatet (p-värde = 0.0227). Därutöver finns inga signifikanta resultatskillnader mellan grupp L1, som når 61 % av det maximala antalet poäng, och de övriga andraspråksgrupperna eller jämförelsegruppen. Skillnaden mellan grupp L2/tidig (71 %) och informanterna med startålder  $\geq 10$  i grupp L2/sen (53 %) är heller inte signifikant för uttrycket *något står en upp i halsen*. Ett problem i samband med detta resultat är naturligtvis att undergrupperna i L2/sen är relativt små i förhållande till de övriga informantgrupperna. Intressant är dock att det

---

<sup>1</sup> För att minimera risken att missa relevanta träffar har jag i vissa fall valt den minsta möjliga söksträngen och sedan räknat de relevanta träffarna manuellt.

<sup>2</sup> Samtliga frekvenssökningar i PAROLE och Google genomfördes den 11 januari 2010.



fortfarande finns en tendens till ett högre resultat i andraspråksgrupperna än i förstaspråksgruppen, om man jämför hela grupp L2/sen (73 %) och grupp L2/tidig (71 %) med förstaspråkseleverna i grupp L1 (61%). Dessutom är andraspråkgruppernas resultat högre än jämförelsegruppens, som också når 61 % av maximalt antal poäng. Detta förekommer endast i samband med just uttrycket *något står en upp i halsen*. Dessa skillnader är dock inte statistiskt signifikanta.

Uttrycket *något är/ blir en ond cirkel* är det enda i A-delen som genererar ett signifikant högre resultat (86 %) i grupp L1, jämfört med samtliga L2-grupper (p-värde <0.0001). Jämförelseklassens resultat (88 %) skiljer sig inte signifikant från resultatet för grupp L1. Icke-kompositionalitet skulle kunna vara en påverkande faktor, som i detta fall dock verkar ha större betydelse för andraspråkselevernas resultat än för förstaspråkselevernas. Frekvensuppgifter ger inga entydiga svar med avseende på en möjlig förklaring till detta resultat; med 34 förekomster i PAROLE (söksträng: ”*en ond cirkel*” Språkbanken 2010a) kan uttrycket varken beskrivas som särskilt hög- eller lågfrekvent. Googles frekvensuppgifter är i detta fall inte användbara, då sökstängningen ”*ond cirkel*” ger ca 72 000 träffar, ”*en ond cirkel*” däremot drygt 500 000 (Google 2010a). Som påpekats i tidigare arbeten (Große 2008, jfr Prentice & Skölberg und. utg.), är Google inte helt oproblematisk som referensmaterial, bl.a. eftersom man inte har någon kontroll över vilka slags texter som finns representerade där. Dessutom är frekvensuppgifterna inte alltid konstanta och därmed svåra att tolka. Webben kan däremot, med tanke på dess storlek och texternas mångfald, ge en indikation om ett visst uttryck använts vid en viss tidpunkt och, åtminstone relativt sett, i vilken utsträckning (Moon 2008:28, jfr även Lorentzen 2008).

När det gäller uttrycket *prata i nattmössan*, som får det lägsta resultatet i samtliga informantgrupper, finns inga signifikanta resultatskillnader mellan startåldersgrupperna. Jämförelsegruppens resultat (69 %) ligger däremot signifikant högre än alla övriga resultat för detta uttryck (p-värde = 0.0266).

Det kan alltså konstateras att de signifikanta skillnaderna mellan informantgrupperna är få när det gäller resultaten för de enskilda konventionaliserade uttrycken i förståelsetestet. I de sammanlagt 4 fall där det finns en signifikant skillnad är det dessutom endast en av startåldersgrupperna som skiljer sig signifikant från de övriga. Jämförelsegruppens

resultat, som i 7 av 10 fall ligger på en jämn, hög nivå mellan 96 % och 100 %, skiljer sig i de flesta fall inte heller signifikant från L1-gruppens resultat.

Det finns 2 fall där eleverna med den högsta startåldern i grupp L2/sen som enda grupp skiljer sig signifikant från de övriga informanterna; för uttrycken *fastna för något* och *ha mage att göra något* når de sena inlärarna med 53 % respektive 42 % de signifikant lägsta resultaten. Orsakerna till detta kan här bara spekuleras i. *Fastna för något*, med 109 träffar (söksträngar ”fastna.\* för” och ”fastna.\* [1 eller 2 ord emellan] för” Språkbanken 2010a) respektive ca 550 000 träffar (söksträngar ”fastna [samtliga verbformer] för” och ”fastna [samtliga verbformer] du/hon/han/vi/de för”, Google 2010a), kan ses som relativt högfrekvent. *Ha mage att göra något* är däremot mindre frekvent i det använda referensmaterialet (22 förekomster i PAROLE på ”*mage att*” och ca 51 000 träffar på Google på *ha ”mage att*”). Skillnaden i frekvens skulle kunna anföras som förklaring till att hela grupp L2/sen når ett lägre resultat för det mer lågfrekventa uttrycket, medan förstaspråkselevernas resultat ligger på ungefär samma höga nivå för båda uttrycken (jfr Ellis et al. 2008). Skillnaden mellan grupp L1 och gruppen med den högsta startåldern lär däremot bero på andra faktorer. Både *fastna för något* och *ha mage att göra något* är semantisk relativt ogenomskinliga. Icke-kompositionalitet kan alltså tänkas vara åtminstone en del av förklaringen till att just dessa uttryck verkar vara svårare för andraspråkselever med sen startålder än för förstaspråkselever och andraspråkselever med en tidigare startålder. Ett sätt att försöka komma något närmare en förklaring är genom en kvalitativ analys, som i avsnitt 6.5 nedan. Motsvarande gäller för det tidigare nämnda uttrycket *något är en ond cirkel*. Vissa av dessa resultat kommer dock att förbli svåra att förklara, inte minst det faktum att andraspråkseleverna når högre resultat än förstaspråkseleverna (inklusive jämförelsegruppen) för just uttrycket *något står en upp i halsen*.

## 6.4.2. Lucktest i kontext

I jämförelse med resultaten i förståelsetestet, finns det större variation i resultaten mellan och inom informantgrupperna i B-delen i fråga om de enskilda konventionaliserade uttrycken. Ser man på summan av resultaten för samtliga startåldersgrupper, varierar resultaten mellan 90 % av

TABELL 8: Resultat på förståelsetest i del A per uttryck (i fallande ordning)

	SUMMA alla informanter	GRUPP L1	GRUPP L2/tidig (startålder ≤ 6)	GRUPP L2/ sen (startålder > 6)		JK
				startålder ≤10	startålder > 10	
antal elever	186	74	75	18	19	37
UTTRYCK						
sätra färg på något	167 (90%)	70 (95%)	67 (89%)	15 (83%)	15 (79%)	30 (81%)
hålla huvudet högt	166 (89%)	67 (91%)	68 (91%)	15 (83%)	16 (84%)	31 (84%)
fastna för något	161 (87%)	67 (91%)	67 (90%)	17 (94%)	10 (53%)	27 (73%)
ha fria tyglar	160 (86%)	69 (93%)	61 (81%)	17 (94%)	13 (68%)	30 (81%)
innan man vet ordet av	152 (82%)	70 (95%)	58 (77%)	13 (72%)	11 (58%)	24 (65%)
stå i rampljuset	150 (81%)	68 (92%)	60 (80%)	11 (61%)	11 (58%)	22 (59%)
ha mage att göra ngt	148 (80%)	68 (92%)	58 (77%)	14 (78%)	8 (42%)	22 (59%)
ngt är/blir en ond cirkel	118 (63%)	64 (86%)	37 (49%)	7 (39%)	10 (53%)	17 (46%)
ngt står en upp i halsen	125 (67%)	45 (61%)	53 (71%)	17 (94%)	10 (53%)	27 (73%)
prata i nattnössan	55 (30%)	32 (43%)	15 (20%)	3 (17%)	5 (26%)	8 (22%)
						18 (69%)

maximalt resultat, för uttrycket *på egen hand* och 19 % för *sätta käppar i hjulen*. Ett annat iögonfallande resultat som framgår av tabell 8.1 är att det finns flera nollresultat i gruppen med startålder >10, samt ett i hela grupp L2/sen, något som inte förekommer i förståelsetestet. Även jämförelsegruppen uppvisar större variation i resultaten för de enskilda uttrycken i detta *sentence-completion-test* (mellan 46 % och 100 %) än i den första delen av enkäten. Det finns få (2 av tio) signifikanta skillnader mellan jämförelsegruppen och grupp L1 även i den detaljerade analysen av B-delen.

Samtliga startåldersgrupper (förutom grupp L1, som når ett marginellt högre resultat för uttrycket *ta första steget*) når alltså sitt högsta resultat i B-delen för uttrycket (*göra något*) *på egen hand*. När det gäller skillnaderna grupperna emellan, uppvisar de senare inlärarna i grupp L2/sen (53 %), ett signifikant lägre resultat (p-värde <0.0001). Övriga resultat ligger mellan 89 % och 96 % och skillnaderna här är inte statistiskt signifikanta. Ett liknande mönster gäller för uttrycket *ta första steget*, som ligger på andra plats i tabell 8.1. Som framgår av tabellen är variationen mellan gruppernas resultat större för de övriga uttrycken i detta deltest. I den nedre delen av tabell 8.1 framträder en allt tydligare skillnad mellan grupp L1 och andraspråksgrupperna. Ser man på resultatskillnaderna mellan startåldersgrupperna för de sex uttrycken som står nederst i tabellen, är resultatet för grupp L1 det signifikant högsta i alla sex fallen (p-värden <0.0001). De sex konventionaliserade uttrycken är (i fallande ordning) *komma till kritan*, *falla för frestelsen*, *hålla modet/humöret uppe*, *få någon/något på halsen*, *klättra på väggarna* och *sätta käppar i hjulen*. För de senare 4 av dessa 6 uttryck uppnår de senare inlärarna i grupp L2/sen 0 % av det maximala resultatet. Det är med andra ord ingen informant i denna grupp som kompletterar dessa konventionaliserade uttryck på det förväntade sättet. För uttrycket *sätta käppar i hjulen* gäller detta för hela grupp L2/sen. Det är också detta uttryck som har fått lägst poäng i B-delen överlag. Intressant i samband med detta uttryck är också att jämförelsegruppens resultat (81 %) skiljer sig tydligt och signifikant från L1-gruppens, som endast når 43% av maximalt resultat. Med tanke på faktorn kompositionalitet, är det inte helt oväntat att ett semantiskt ogenomskinligt uttryck som *sätta käppar i hjulen* är bland dem som relativt få elever behärskar i en undersökning som denna. Mot bakgrund av de resultat som presenteras i Ellis et al.

TABELL 8.1: Resultat på lucktestet i del B per uttryck (i fallande ordning)

## SUMMA alla informanter

	SUMMA alla informanter	GRUPP L1	GRUPP L2/tidig (startålder ≤ 6)	GRUPP L2/ sen (startålder > 6)		JK
				startålder ≤10	startålder > 10	
antal elever	186	74	75	18	19	37
UTTRYCK						
på egen hand	168 (90%)	70 (95%)	72 (96%)	16 (89%)	10 (53%)	26 (70%)
ta första steget	160 (86%)	71 (96%)	67 (89%)	15 883%	7 (37%)	22 (59%)
dra alla över en kam	117 (63%)	66 (89%)	40 (53%)	8 (44%)	3 (16%)	11 (30%)
ta sig i kragen	107 (58%)	66 (89%)	33 (44%)	5 (28%)	3 (16%)	8 (22%)
komma till kritan	105 (56%)	63 (85%)	30 (40%)	8 (44%)	4 (21%)	12 (32%)
falla för frestelsen	79 (42%)	49 (66%)	24 (32%)	3 (17%)	3 (16%)	6 (16%)
hålla modet/ humöret uppe	61 (33%)	39 (52%)	20 (27%)	2 (11%)	0 (0%)	2 (25%)
få ngt/ngn på halsen	50 (27%)	41 (55%)	8 (11%)	1 (0,5%)	0 (0%)	1 (3%)
klättra på väggarna	43 (23%)	34 (45%)	7 (9%)	2 (11%)	0 (0%)	2 (5%)
sätta käppar i hjulen	36 (19%)	32 (43%)	4 (5%)	0 (0%)	0 (0%)	0 (0%)
						21 (81%)

(2008) kan även uttryckets jämförelsevis låga frekvens tänkas påverka framför allt andraspråkselevernas behärskning av *sätta käppar i hjulen* negativt; uttrycket förekommer sammanlagt 30 gånger i PAROLE (på ”*käppar i hjulen*” och *käppar i hjulet*” Språkbanken 2010a), medan det finns 470 belägg på högstplacerade *på egen hand* (söksträng ”*på egen hand*”) i korpusen. Använder man Google som referenspunkt får man ca 290 000 träffar på uttrycket *sätta käppar i hjulen*, (sammanlagt på söksträngarna *sätta ”käppar i hjulen*” och *sätta ”käppar i hjulet*”) och ca 758 000 på söksträngen ”*på egen hand*” (Google 2010a). Ser man i första hand på frekvenserna i PAROLE, finns det vissa tendenser som ligger i linje med de resultat som Ellis et al. (2008) presenterar. Med andra ord uppnår grupp L2/sen låga resultat för de flesta uttryck med relativt låg frekvens i korpusen; bland de uttryck, för vilka någon eller båda undergrupperna i grupp L2/sen når resultat nära eller lika med noll, förekommer *klättra på väggarna* 5 gånger (”*klättr.\* [1 eller 2 ord emellan] på väggarna*”) och *hålla modet/humöret uppe* sammanlagt 8 gånger (en gång på söksträngen ”*humöret uppe*” och 7 gånger på ”*modet uppe*”) på 19.4 miljoner ord. *Få något/någon på halsen* förekommer däremot 28 gånger (”*på halsen*”) och *sätta käppar i hjulen*, som tidigare nämnts, 30 gånger. Dessa frekvenser är alltså något högre, men fortfarande betydligt lägre än frekvensen för uttrycket *på egen hand*, där grupp L2/sen når sina högsta resultat. Frekvensen framstår dock inte som en påverkande faktor i samband med alla uttryck. Uttrycket *ta första steget*, där de sena andraspråksinlärarna når sina näst högsta resultat, förekommer endast 27 gånger i PAROLE (på söksträng ”*första steget*”). Man kan här alltså som mest hävda att uttryckets frekvens kan ses som en av flera möjliga faktorer som kan tänkas påverka elevernas produktiva behärskning av de konventionaliserade uttrycken i B-delen.

### 6.4.3. Lucktest utanför kontext

Som redan nämnts i avsnitt 6.4, når samliga grupper sitt lägsta delresultat i C-delen, där uppgiften består i att komplettera ett ofullständigt konventionaliserat uttryck utan kontext. Frågan som ska besvaras i detta avsnitt är i vilken mån resultaten för de enskilda uttrycken varierar inom grupperna och om det finns vissa uttryck som är speciellt svåra för en viss grupp. Ser man på tabell 8.2 som helhet, kan man konstatera att

variationen bland resultaten för de enskilda uttrycken inom varje grupp är relativt stor. Även när det gäller signifikanta resultatskillnader mellan startåldersgrupperna, finns det en del variation. Ett allmänt mönster som kan utläsas är dock att grupp L1 når de högsta resultaten för samtliga uttryck, följt av grupp L2/tidig. Grupp L2/sen når i samtliga fall de lägsta resultaten. Resultatskillnaderna mellan grupperna är dock inte alltid statistiskt signifikanta.

Vad gäller uttrycket *ta saken i egna händer*, som återfinns överst i tabell 8.2, når 3 av 4 grupper relativt höga resultat (mellan 78 % och 97 % av maximala antalet poäng). Endast de senare inlärarna i grupp L2/sen skiljer sig signifikant från övriga grupper med 47 % (p-värde <0.0001). Längre ner i tabellen blir de signifikanta skillnaderna mellan startåldersgrupperna fler. Resultatet för uttrycket *av hela mitt hjärta* är 92 % för grupp L1, medan grupp L2/tidig når ett signifikant lägre resultat än förstaspråkseleverna med 71 %. Det finns ingen signifikant skillnad mellan grupp L2/tidig (72 %) och gruppen med startålder ≤10 i grupp L2/sen (61 %). Däremot når de senare inlärarna i grupp L2/sen det signifikant lägsta resultatet för uttrycket *av hela mitt hjärta* (42 %, p-värde <0.0001). I samband med nästa uttryck i tabell 8.2, *gå bakom ryggen på någon*, skiljer sig resultatet för båda undergrupperna i grupp L2/sen (50 % respektive 42 %) signifikant från resultatet för grupp L2/tidigt (71 %). Den senare når i sin tur ett signifikant lägre resultat än grupp L1, som når 96 % av maximal poäng för detta uttryck. Även om det alltså finns en del variation när det gäller resultaten för de enskilda uttrycken i C-delen, är det i samband med nästan hälften av uttrycken (7 av 15) endast förstaspråkselevernas resultat som skiljer sig signifikant från resultaten för samtliga andraspråksgrupper. Det gäller uttrycken *som gräddde på moset*, *rinna ut i sanden*, *ha ett finger med i spelet*, *med ngt i högsta hugg*, *på rak arm*, *få upp ögonen för något* (p-värden <0.0001) och *hålla någon i schack* (p-värde = 0.0002). Informanterna med startålder >10 i grupp L2/sen når nollresultat för de senare 6 av dessa 7 uttryck. Uttrycket *hålla någon i schack* är dessutom det uttryck som genererar lägst poäng av samtliga uttryck i enkäten och är med andra ord det uttryck som informantgruppen som helhet verka behärska sämst av dem som testats i undersökningen. Ett annat uttryck som sticker ut är *livet leker för någon*, där både grupp L1 och jämförelsegruppen endast når 27 % av maximalt resultat. Här finns heller ingen signifikant skillnad mellan någon

TABELL 8.2: Resultat på lucktestet i del C per uttryck (i fallande ordning)

	SUMMA alla informanter	GRUPP L1	GRUPP L2/tidig (startålder, ≤ 6)	GRUPP L2/ sen (startålder > 6)		JK	
				startålder ≤10	startålder > 10		totalt
				18	19		
antal elever	186	74	75	18	19	26	
<b>UTTRYCK</b>							
ta saken i egna händer	158 (85%)	72 (97%)	63 (84%)	14 (78%)	9 (47%)	23 (62%)	
av hela mitt hjärta	141 (76%)	68 (92%)	54 (72%)	11 (61%)	8 (42%)	19 (51%)	
gå bakom ryggen på ngn	141 (76%)	71 (96%)	53 (71%)	9 (50%)	9 (42%)	17 (46%)	
på löpande band	135 (73%)	65 (88%)	56 (75%)	11 (61%)	3 (16%)	14 (38%)	
ingen dans på rosor	106 (57%)	62 (83%)	36 (48%)	6 (33%)	2 (10%)	8 (22%)	
som gräddde på moset	81 (43%)	57 (76%)	19 (25%)	3 (16%)	2 (10%)	5 (13%)	
tid är pengar	77 (41%)	43 (57%)	21 (28%)	5 (28%)	8 (42%)	13 (35%)	
rinna ut i sanden	73 (39%)	58 (78%)	15 (20%)	0 (0%)	0 (0%)	0 (0%)	
ha ett finger med i speleet	57 (30%)	43 (57%)	12 (16%)	2 (11%)	0 (0%)	2 (5%)	
inte för fem öre	54 (29%)	40 (54%)	13 (17%)	1 (5%)	0 (0%)	1 (3%)	
med ngt i högsta hugg	41 (22%)	35 (47%)	5 (7%)	1 (5%)	0 (0%)	1 (3%)	
på rak arm	39 (21%)	32 (43%)	7 (9%)	0 (0%)	0 (0%)	0 (0%)	
livet leker för ngn	40 (21%)	20 (27%)	18 (24%)	2 (11%)	0 (0%)	2 (5%)	
få upp ögonen för ngt	37 (20%)	28 (38%)	9 (12%)	0 (0%)	0 (0%)	0 (0%)	
hålla ngn i schack	20 (11%)	17 (23%)	2 (3%)	1 (5%)	0 (0%)	1 (3%)	



av grupperna. Även detta uttryck hör till de tio uttryck i enkäten som har fått de lägsta poängen (se vidare avsnitt 6.6 om den kvalitativa analysen) men skillnaden i förhållande till *hålla någon i schack* är alltså att jämförelsegruppens resultat inte skiljer sig signifikant från L1-gruppens resultat.

Frågan om vilka faktorer som påverkar resultatet för de enskilda uttrycken är inte heller enkel att besvara för denna del av enkäten. Frekvensuppgifter i PAROLE och Google för de högst och lägst placerade uttrycken ger inga entydiga svar. Uttrycket *ta saken i egna händer*, som har fått högst poäng i C-delen, förekommer 25 gånger i PAROLE (söksträng: ”saken i egna händer”) och ca 550 000 gånger på Google (”saken i egna händer). Uttrycket *hålla någon i schack*, som har fått lägst poäng i hela materialet, förekommer 56 gånger i PAROLE och får ca 46 000 träffar på Google (”i schack”, Språkbanken 2010a, hålla ”i schack” Google 2010a). De låga resultaten för de två ovannämnda uttrycken kan tänkas relateras till kompositionalitet och genomskinlighet; *ta saken i egna händer* är semantiskt mer genomskinligt och i viss mån även mer kompositionellt än uttrycket *hålla någon i schack*. Ordet *schack* är dessutom polysemt och kan tolkas bokstavligt som ’schackspel’, om informanten inte identifierar den ofullständiga frasen (\_\_\_\_\_ *någon i schack*) som ges i enkäten som just det konventionaliserade uttrycket *hålla någon i schack*. Det senare tas även upp i samband med den kvalitativa analysen i avsnitt 6.6 nedan.

Det finns sammanlagt 3 uttryck som utmärker sig genom att hela grupp L2/sen når nollresultat i samband med dem. Uttrycken är *rinna ut i sanden*, *på rak arm* och *få upp ögonen för något*. Alla tre tillhör alltså de uttryck för vilka det finns en signifikant resultatskillnad mellan första- och andraspråkselever. När det gäller *rinna ut i sanden*, är skillnaden mellan grupp L1 (78 %) och andraspråksgrupperna L2/tidig och L2/sen (20 % respektive 0 %) tydlig. Jämförelsegruppen når 88 % av maximalt resultat för detta uttryck. Uttrycket *rinna ut i sanden* verkar alltså vara betydligt svårare för andraspråkseleverna än för eleverna med svenska som modersmål. De andra två uttrycken får däremot relativt låga resultat även bland grupp L1 och jämförelsegruppen; *på rak arm* får 43 % respektive 50 % och *få upp ögonen för något* får 38 % respektive 58 % av maximalt resultat. Ser man på frekvensen för uttrycket *rinna ut i sanden*, hittar man 23 förekomster i PAROLE (söksträng: ”*ut i sanden*”) och

drygt 1,1 miljon träffar på Google (på samma söksträng). *På rak arm* förekommer endast 11 gånger i PAROLE ("på rak arm"). Däremot får man drygt 3,5 miljoner träffar på Google på söksträngen "på rak arm". Uttrycket *få upp ögonen för något* förekommer 67 gånger i PAROLE (söksträng: "upp ögonen för") och ca 930 000 gånger på Google (samma söksträng) (Språkbanken 2010a, Google 2010a). Frekvensuppgifterna är alltså även i detta fall svårtolkade. Inget av uttrycken är så lågfrekvent i referensmaterialen att man, mot bakgrund av Ellis et al. (2008), skulle kunna tolka det som att frekvensen påverkar andraspråkselevernas resultat för dessa uttryck. När det gäller uttrycken *på rak arm* och *få upp ögonen för något* är frekvenserna – i alla fall på Google – relativt höga men resultaten för alla informantgrupperna relativt låga. Här lär alltså andra faktorer ligga till grund för resultaten. *På rak arm* kan ses som icke-kompositionellt och ogenomskinlig medan *få upp ögonen för något* är semantiskt förhållandevis mer genomskinligt och till viss del mer kompositionellt. En annan faktor som skulle kunna ha betydelse är informanternas ålder. Man kan tänka sig att dessa uttryck helt enkelt inte är speciellt frekventa i den aktuella åldersgruppens språkbruk.

Sammanfattningsvis framgår det av de resultat som redovisats i detta avsnitt att det finns en signifikant skillnad i resultat mellan första- och andraspråkstalare för nästan hälften av uttrycken i C-delen. Man kan tolka detta som att förstaspråkstalare överlag har betydligt lättare för uppgiften att komplettera ett konventionaliserat uttryck utan hjälp av kontexten, vilket naturligtvis inte är ett oväntat resultat. Resultaten för de enskilda uttrycken varierar dock rätt mycket, även för de elever som har svenska som modersmål.

#### 6.4.4. Kollokationer

I den sista delen av enkäten testas elevernas behärskning av sammanlagt 20 olika kollokationer. Av tabell 8.3 framgår att skillnaderna mellan grupperna inte är lika stora i denna del av materialet som i C-delen. Intressant är att skillnaden mellan grupp L1 och grupp L2/tidig i 16 av 20 fall inte är statistisk signifikant. Här skiljer sig alltså resultatet från den detaljerade analysen från det sammanlagda resultatet för hela del D (se avsnitt 6.3), där skillnaden mellan grupp L1 och grupp L2/tidig är den enda som är statistiskt signifikant i en jämförelse mellan alla informant-

TABELL 8.3: Resultat på matchningstestet i del D per uttryck (i fallande ordning)

	SUMMA alla informanter	GRUPP L1	GRUPP L2/tidig (startålder ≤ 6)	GRUPP L2/ sen (startålder > 6)		JK	
				startålder ≤10	startålder > 10		totalt
				18	19		
antal elever	186	74	75	18	19	26	
<b>UTTRYCK</b>							
allvarlig sjukdom	157 (84%)	69 (93%)	59 (79%)	15 (83%)	14 (76%)	29 (78%)	
vara/komma till nytta	149 (80%)	67 (91%)	58 (77%)	13 (72%)	11 (57%)	24 (65%)	
kasta en blick	147 (79%)	68 (92%)	54 (72%)	14 (78%)	11 (58%)	25 (67%)	
svårt sjuk	137 (74%)	62 (84%)	55 (73%)	12 (66%)	8 (48%)	20 (54%)	
få i uppgift	131 (70%)	65 (88%)	47 (63%)	10 (55%)	9 (48%)	19 (51%)	
olika slags	130 (70%)	61 (82%)	51 (68%)	8 (44%)	10 (48%)	18 (49%)	
djupt andetag	127 (68%)	49 (65%)	52 (69%)	13 (72%)	13 (67%)	26 (70%)	
full fart	127 (68%)	61 (81%)	48 (64%)	11 (61%)	7 (43%)	18 (49%)	
evigt tacksam	123 (66%)	64 (86%)	41 (55%)	12 (66%)	6 (38%)	18 (49%)	
hamna i missbruk	123 (66%)	55 (74%)	46 (61%)	12 (66%)	10 (57%)	22 (59%)	
knappast trolig	119 (64%)	64 (86%)	40 (53%)	9 (50%)	7 (43%)	16 (43%)	
stor betydelse	116 (62%)	56 (76%)	43 (57%)	8 (44%)	9 (52%)	17 (46%)	
ge ett exempel	124 (66%)	57 (76%)	41 (55%)	14 (78%)	12 (67%)	26 (70%)	
bryta ett beroende	106 (57%)	51 (69%)	38 (51%)	8 (44%)	9 (52%)	17 (46%)	
politisk aktiv	96 (52%)	50 (68%)	32 (43%)	9 (50%)	5 (33%)	14 (38%)	
bilda en union	97 (52%)	43 (58%)	38 (51%)	8 (44%)	8 (43%)	16 (43%)	
gå i arv	82 (44%)	44 (59%)	28 (37%)	6 (33%)	4 (21%)	10 (27%)	
dra en suck	71 (38%)	38 (51%)	21 (28%)	6 (33%)	6 (38%)	12 (32%)	
ligga till grund	60 (32%)	33 (45%)	18 (24%)	6 (33%)	3 (14%)	9 (24%)	
lyckligt ovetande	53 (28%)	36 (48%)	12 (16%)	2 (11%)	3 (14%)	5 (13%)	

grupper resultat. Däremot finns det även i denna del av analysen en tydlig tendens, nämligen att grupp L1 alltid når högre resultat än samtliga andraspråksgrupper (se tabell 8.3). Skillnaden mellan förstaspråkseleverna och grupp L2/tidig är statistiskt signifikant för uttrycken få i uppgift, evigt tacksam, knappast troligt och lyckligt ovetande (p-värden <0.0001). Ett annat intressant mönster som framgår av tabell 8.3 är att skillnaden mellan grupp L2/tidig och de tidigare inlärnarna i grupp L2/sen endast är signifikant i samband med kollokationen *olika slags* (p-värde = 0.0023). Ser man närmare på dessa två gruppers resultat i tabell 8.3, kan man konstatera att de tidigare inlärnarna i grupp L2/sen dessutom når högre resultat än grupp L1 i vissa fall, t.ex. när det gäller *evigt tacksam* och *ge ett exempel*. Dessa skillnader är dock, som sagt, inte signifikanta.

En tendens som är konstant över samtliga deltester är att gruppen med startålder >10 når lägst resultat i jämförelse med övriga startåldersgrupper (samt jämförelsegruppen). I den detaljerade analysen av kollokationstestet skiljer sig resultaten för de sena inlärnarna i L2/sen dock endast i 7 av 20 fall signifikant från resultaten för de övriga startåldersgrupperna. Detta gäller uttrycken *vara/komma till nytta* (p-värde = 0.0061), *kasta en blick* (p-värde = 0.0015), och *svårt sjuk* (p-värde = 0.0021), överst i tabell 8.3, samt *full fart* (p-värde = 0.0012) och *evigt tacksam* (p-värde <0.0001) i mitten av tabellen, och *politiskt aktiv* (p-värde = 0.0013) samt *gå i arv* (p-värde = 0.0046), som återfinns i nedre halvan av tabell 8.3. En skillnad i förhållande till resultaten för de enskilda uttrycken i delarna B och C är att det, i likhet med del A, inte finns några nollresultat för någon av kollokationerna.

I nästan hälften av fallen (9 av 20) finns det inga signifikanta skillnader i resultat mellan grupperna över huvud taget. Så likartade resultat förekommer inte i lika stor utsträckning i de övriga deltesten. Detta gäller också kollokationen *allvarlig sjukdom* som får den högsta poängen sammanlagt och därmed hamnar högst upp i tabell 8.3. Alla grupper når relativt höga resultat för detta uttryck, mellan 76 % och 93 %. När det gäller det uttryck som ligger näst sist i tabellen, *ligga till grund*, finns det heller inga signifikanta resultatskillnader mellan startåldersgrupperna. Något som är anmärkningsvärt med detta resultat är att jämförelsegruppen, vars resultat i övrigt ligger mellan 69 % och 96 %, endast når 38 % av maximalt resultat för denna kollokation.

Orsakerna till den relativt stora resultatskillnaden för dessa två kollokationer kan man återigen bara spekulera om. Ser man på uttryckens frekvens i PAROLE och på Google, kan man konstatera att den inte framstår som en avgörande faktor, då frekvensen för det näst sist placerade uttrycket *ligga till grund* är betydligt högre (193 på ”*ligga till grund*” i PAROLE, drygt 380 000 miljoner på *ligga ”till grund*” på Google) än den för förstplacerade *allvarlig sjukdom*, som endast förekommer 7 gånger i PAROLE och ca 45 000 gånger på Google (”*allvarlig sjukdom*”, Språkbanken 2010a, ”*allvarlig sjukdom*”, Google 2010a). Det sist placerade uttrycket *lyckligt ovetande* får ungefär lika låga frekvenser som *allvarlig sjukdom* i referensmaterialen (4 i PAROLE och ca 28 000 på Google, söksträng: ”*lyckligt ovetande*”). Kompositionalitet kan däremot förmodas påverka resultaten, speciellt i fråga om *ligga till grund* som kan sägas vara både semantiskt betydligt mindre genomskinligt och mindre kompositionellt än kollokationen *allvarlig sjukdom*. Detta gäller till viss del även *lyckligt ovetande*, även om skillnaden inte är lika tydlig i det fallet. Betydelsen hos *ligga till grund* kan dessutom sägas vara tämligen abstrakt.

Utgår man, som i avsnitt 4.4, ifrån van Liers (1995) hypotes om skillnaden mellan första- och andraspråksbrukare, när det gäller att bilda de mest idiomatiska kollokationerna av ett antal givna ord, är resultatet av den detaljerade analysen som beskrivits här något oväntat. Signifikanta skillnader mellan grupp L1 och någon av andraspråksgrupperna föreligger endast i drygt hälften (11 av 20) av fallen. I 7 av dessa handlar det dessutom enbart om en signifikant skillnad mellan grupp L1 och andraspråksleverna med den högsta startåldern bland informanterna. Signifikanta skillnader mellan de olika startåldersgrupperna bland andraspråksleverna är få, vilket ligger i linje med den hypotes som redovisas i avsnitt 4.4. Även om det för många av uttrycken alltså inte finns några signifikanta skillnader mellan grupp L1 och andraspråksgrupperna, visar tabell 8.3 en relativt tydlig tendens till att grupp L1 konstant

---

<sup>3</sup> I detta fall har jag sökt på samtliga verbformer för att undvika träffar på ”*lägga till grund*”. Sökuttrycket har dessutom trunkerats för att även få träff på former med ett eller två ord emellan verbet och prepositionen. Denna trunkering gav sammanlagt 9 (av 193) träffar.

presterar bättre för de enskilda uttrycken än de elever som har ett annat modersmål än svenska.

Det kan också konstateras att det, i likhet med de detaljerade resultaten för de övriga deltesten, finns få signifikanta resultatskillnader mellan jämförelsegruppen och grupp L1 i den detaljerade analysen av kollokationstestet; det är endast i samband med uttrycken *dra en suck* (p-värde = 0.0263), *stor betydelse* (p-värde = 0.0234) och *lyckligt ovetande* (p-värde <0.0001) som jämförelsegruppen når ett signifikant högre resultat (77 %, 96 % respektive 92 %) än grupp L1 (51 %, 76 % respektive 48 %). Även på detta sätt skiljer sig alltså resultatet av den detaljerade analysen från analysen av de samlade delresultaten för del D där jämförelsegruppen når ett signifikant högre resultat än grupp L1 i alla fyra delarna av materialet. Som i de övriga delarna kan man dock även i analysen av de enskilda uttrycken i D se en tendens till att jämförelsegruppen når konstant högre resultat än grupp L1 (förutom när det gäller resultaten för *ligga till grund*).

Slutligen kan noteras att de mer lågfrekventa kollokationer som tagits med i undersökningen hamnar relativt långt ned i tabellen, om än inte längst ned. Kollokationen *dra en suck* ligger med 71 poäng bland de sista tre uttrycken i tabellen, *bilda en union* och *bryta ett beroende* ligger med 97 respektive 106 poäng något högre upp. *Hamna i missbruk ligger* med 123 högst upp av de fyra lågfrekventa uttrycken. Den låga frekvensen i referensmaterialet framstår alltså knappast som den avgörande faktorn för resultaten i samband med de aktuella uttrycken – även om man kan tänka sig att den har viss betydelse – då det finns mer frekventa uttryck som fått lägre poäng.

## 6.5. Validitet

Som nämntes i avsnitt 5.5 har testmetodens validitet testats med hjälp av Pearsins korrelationskoefficienten. I detta syfte har korrelationer mellan resultaten för respektive uttryck och det totala resultatet samt delresultaten (A–D) för hela informantgruppen beräknats. Dessutom har medelvärden för resultaten på hela enkäten mellan de informanter som behärskar, och de som inte behärskar några av de svåraste respektive lättaste uttrycken (dvs. de uttryck som få respektive många informanter

har kompletterat på det förväntade sättet), jämförts varpå skillnaderna har testats för signifikans. Båda dessa analyser ger statistiskt sett samma p-värden och tjäna alltså till att synliggöra vissa statistiska samband; om resultatet för uttryck x korrelerar med hela informantgruppens resultat dels på hela enkäten, dels på de olika deltesterna, innebär det att många av informanterna som har nått höga resultat totalt har rätt på den enskilda uppgiften ifråga. Likaså, om de informanter som har kompletterat ett svårt uttryck på det förväntade sättet sammantaget når ett signifikant högre resultat på hela enkäten samt på de olika deltesterna än den mindre gruppen av informanter som avviker från det förväntade svaret, visar det på att det finns ett samband mellan informanternas positiva resultat för det enskilda uttrycket och deras högre resultat totalt. Samma gäller för de elever som inte har kompletterat ett lätt uttryck på det förväntade sättet; om deras resultat är signifikant lägre än resultaten för den större elevgruppen som besvarat uppgiften på det förväntade sättet, tyder det på att det finns ett samband mellan det avvikande svaret på den enskilda uppgiften och lågt resultat på hela enkäten respektive de olika deltesterna.

De samband som beskrivs ovan visar sig vara starkt signifikanta i det stora flertalet av fallen (i de flesta av dessa fall är r-värdet  $>0.25$  och p-värdet  $<0.0001$ ). Det uttrycket som sticker ut i detta sammanhang är *få något på halsen* där korrelationerna mellan resultatet för uttrycket och det totala resultatet samt delresultaten är svaga (r-värde totalt = 0.05833) och det inte kan påvisas en signifikant skillnad mellan medelvärdet av totalt resultat för de 126 informanter som behärskar uttrycket och medelvärdet för de 61 informanterna som inte gör det (p-värde totalt = 0.4278, jfr även tabell 8). Överlag kan dock konstateras att det finns en genomgående tendens resultaten emellan och att testuppgifterna verkar uppfylla syftet att diskriminera mellan de informanter som behärskar många av de konventionaliserade uttrycken ifråga och de som inte gör det.

## 6.6. Kvalitativ analys

Hittills har i första hand kvantitativa resultatskillnader mellan informantgrupperna diskuterats. I detta avsnitt kommer en del av materialet

(se avsnitt 5.5) beskrivas kvalitativt. En av de frågor som ska besvaras i detta sammanhang är vilka typer av avvikelser från de konventionaliserade konstruktionerna i fråga som finns i materialet. Efter analys av några av enkäterna med lägst resultat i varje grupp – sammanlagt 45 – har samtliga avvikande svar klassificerats på ett av följande sätt:

- *inget svar* – när informanten inte har kompletterat uttrycket i fråga
- *vanlig kontamination* – när kontaminationen i fråga förekommer hos flera (minst 5) informanter
- *idiosynkratisk kontamination* – när kontaminationen endast förekommer hos enskilda (högst 4) informanter
- *icke-idiomatiskt* – när informanten fyllt i ett ord som är grammatiskt och semantiskt möjligt, men gör att frasen ifråga avviker från det förväntade konventionaliserade uttrycket
- *ogrammatiskt* – när ordet som informanten kompletterat med orsakar ett grammatiskt eller semantiskt problem i den aktuella kontexten

Alla dessa kategorier är inte lika relevanta för alla de tio uttrycken. Skillnaden mellan *vanliga* och *idiosynkratiska kontaminationer* görs då det visat sig finnas kontaminationer i materialet som är vanligt förekommande. Gränsen för vad som kan anses som vanligt har satts vid 5 förekomster. ”Vanlig” är i detta fall i hög grad att betrakta som ett relativt begrepp och den valda gränsen kan naturligtvis diskuteras. Vid mindre än 5 förekomster bör dock ett uttryck knappast kunnas anses som vanligt i det här aktuella materialet. Därtill kommer att man i vissa fall kan misstänka ett samarbete mellan två eller möjligtvis flera elever, något som har lett till identiska lösningar av uppgiften ifråga. Det är alltså heller inte rimligt att se alla kontaminationer som förekommer mer än en gång som en egen kategori. Dessutom bör man komma ihåg att det alltid finns en spekulativ aspekt i att klassificera en fraseologisk enhet som en kontamination av två eller flera olika konventionaliserade uttryck. Man kan knappast bevisa att informanten har associerat till andra befintliga uttryck och därvid blandat ihop flera enheter i samma fras. Man kan dock argumentera för att sannolikheten att man av ren slump väljer en eller flera komponenter från ett visst uttryck och blandar dem med



komponenter från ett annat konventionaliserat uttryck, som uppvisar likheter i form och/eller betydelse, inte borde vara så stor.

Kategorin *icke-idiomatisk* är bred och inte helt lätt att avgränsa. Kontaminationer kan sägas vara specialfall av icke-idiomatisk avvikelse. Mer problematisk är däremot gränsdragningen mellan icke-idiomatiska och ogrammatiska avvikelser, särskilt i samband med abstrakta och i relativt hög grad lexikaliserade konventionaliserade uttryck som t.ex. *hålla någon i schack*. Här är det i många fall inte självklart om användningen av ett annat verb än *hålla* (som i samband med detta uttryck kan betraktas som tämligen betydelsetomt) bör ses som icke-idiomatisk eller ogrammatisk. De ovannämnda kategorierna diskuteras närmare i följande avsnitt.

Av de 45 informanterna vars enkäter har analyserats kvalitativt ingår 14 i grupp L1, 15 i grupp L2/tidig och 16 i grupp L2/sen.

### 6.6.1. De svåraste uttrycken

De tio konventionaliserade uttryck som har fått lägst poäng i materialet, och därmed kan sägas visat sig vara svårast för informanterna i denna undersökning, presenteras nedan i stigande ordning, efter det resultat som informantgruppen som helhet har uppnått för uttrycket:

- *hålla någon i schack* (20/186)
- *sätta käppar i hjulen* (36/186)
- *få upp ögonen för något* (37/186)
- *på rak arm* (39/186)
- *livet leker för någon* (40/186)
- *med något i högsta hugg* (41/186)
- *klättra på väggarna* (43/186)
- *få något/någon på halsen* (50/186)
- *lyckligt ovetande* (53/186)
- *inte för fem öre* (54/186)

Bortsett från *lyckligt ovetande* (del D) återfinns alltså samtliga uttryck i delarna B (4 av tio) och C (6 av tio) i enkäten. A-delen är inte representerad bland de uttryck som fått lägst resultat. Detta kan sägas återspegla det mönster som framgår av den kvantitativa analysen.

Genom att titta närmare på de avvikande svaren för de ovannämnda uttrycken kan man tänkas komma något närmare en förklaring till, varför just dessa uttryck vållar störst problem för de eleverna som deltagit i undersökningen.

### 6.6.1.1. Hålla någon i schack

TABELL 9: Antal olika typer av avvikelser i samband med uttrycket *hålla någon i schack* per informantgrupp

	Inget svar	Vanliga kontaminationer	Idio-synkratiska kontaminationer	Icke-idiomatiska svar	Ogrammatiska svar	Förväntade svar
L1 (15)	1	–	–	9	5	–
L2/tidig (14)	9	–	–	3	2	–
L2/sen (16)	10	–	–	1	5	–
Totalt	20	–	–	13	12	–

*Hålla någon i schack* är det uttryck som har fått lägst poäng i materialet som helhet (20 av 186 möjliga). Bland de 45 informanterna som inkluderats i den kvalitativa delen av analysen har 20 avstått från att komplettera fragmentet \_\_\_\_\_ *någon i schack*. 19 av dessa är andraspråkselever (9 i grupp L2/tidig, 10 i grupp L2/sen). Svarsbortfallet är bland det högsta för de uttryck som diskuteras här.

Det förekommer inga kontaminationer i samband med uttrycket *hålla någon i schack* i den delen av materialet som analyserats kvalitativt. Däremot finns det 13 kompletteringar som har klassificerats som icke-idiomatiska. Dessa fördelar sig samtliga på exemplen *slå någon i schack* (L2/tidig, L2/sen), *sätta någon i schack* (L1), *ta någon i schack* och *dra någon i schack*. Alla dessa exempel, förutom *dra någon i schack*, förekommer flera gånger. Ett exempel som sticker ut något i sammanhanget är användningen av verbet *slå*. Det förekommer sammanlagt 5 gånger i denna del av materialet, vilket inte är helt oväntat, med tanke på att frasen *slå någon i schack* är såväl grammatiskt som semantiskt-pragmatiskt

fullt möjlig att använda, även om det inte kan betraktas som ett konventionaliserat uttryck i den här avsedda bemärkelsen.

Endast 4 av de icke-idiomatiska svaren återfinns bland andraspråkseleverna (3 i grupp L2/tidig och 1 i grupp L2/sen). Eleverna i grupp L1 står alltså för 9 avvikande svar i denna kategori. Man kan tolka de hittills nämnda resultaten som att de andraspråkelever som inte behärskar uttrycket snarare avstått från att besvara uppgiften, medan fler elever i grupp L1 har försökt att komplettera uttrycket, i många fall dock utifrån en annan betydelse än den som avses i enkäten.

12 kompletteringar har klassats som ogrammatiska. 5 av dessa svar förekommer i grupp L1. Förstaspråkseleverna har producerat formerna \**spela någon i schack* (2 gånger), \**explodera någon i schack*, \**göra någon i schack* och \**som någon i schack*. I grupp L2/tidig förekommer 2 ogrammatiska svar, nämligen \**sätt någon i schack* och \**spela någon i schack*. I grupp L2/sen förekommer 5 ogrammatiska svar, nämligen \**hela någon i schack*, \**är någon i schack* (2 gånger) och \**vinna/vann någon i schack*.

Ingen av de 45 informanterna har kompletterat det ofullständiga uttrycket på det förväntade sättet.

En faktor som verkar vålla problem för både första- och andraspråkseleverna är polysemin i samband med substantivet *schack*. Valet av verb som *vinna*, *slå* eller *spela*, istället för det konventionaliserade *hålla*, kan antas bero på att informanterna associerar till 'schackspel'. I dessa fall ligger problemet alltså på det receptiva planet; då eleverna utgår ifrån en annan betydelse av konstituenten *schack*, avviker följaktligen även de ord som de associerar med denna konstituent från den efterlysta formen (jfr Nation 2001, se avsnitt 3.1). Den ogrammatiska formen \**spela någon i schack* (istället för *spela mot någon*) skulle dessutom kunna ses som influerad av engelskan, där formen *to play someone at chess* är grammatiskt möjlig. Det förekommer också exempel där man kan förmoda ett avsiktligt nonsenssvar (förmodligen i brist på kunskap om det förväntade svaret). Hit kan t.ex. användningen av verbet *explodera* räknas.

### 6.6.1.2. Sätta käppar i hjulen

TABELL 9.I: Antal olika typer av avvikelser i samband med uttrycket *sätta käppar i hjulen* per informantgrupp

	Inget svar	Vanliga kontaminationer	Idio-synkratiska kontaminationer	Icke-idiomatiska svar	Ogrammatiska svar	Förväntade svar
L1 (15)	3	–	1	8	2	1
L2/tidig (14)	10	–	1	2	1	–
L2/sen (16)	5	–	3	3	5	–
Totalt	18	–	5	13	8	1

Sammanlagt 20 av de 45 informanterna har avstått från att komplettera fragmentet *sätta \_\_\_\_\_ i hjulen*. Även i detta fall har flertalet noll-svar, dvs. 19 av dem, lämnats av andraspråkseleverna (9 i grupp L2/tidig och 10 bland de sena inlärarna).

Det finns sammanlagt 5 svar som har klassats som kontaminationer i detta material. I dessa fall har informanterna kompletterat det ofullständiga uttrycket med *stopp* (L2/sen), *eld* (L2/tidig och L1), *spiken* (L2/sen). När det gäller de två förstnämnda lösningsförslagen kan man förmoda att kontaminationer med de konventionaliserade uttrycken *sätta stopp för något*, *sätta eld på något* kan vara en förklaring. Här är alltså verbet *sätta* den gemensamma nämnare som verkar orsaka kontaminationerna. Användningen av substantivet *spiken* skulle kunna bero på en kontamination med *slår (sista) spiken i kistan för något*, som i viss mån även förekommer i varianten *sätta spiken i kistan för något* (se Google 2010b, Språkbanken 2010b).

Det finns 13 svar i materialet som har klassats som icke-idiomatiska. 5 av dessa har producerats av andraspråkselever (2 i grupp L2/tidig och 3 i grupp L2/sen) och 8 av elever i grupp L1. Bl.a. har ordformerna *sig*, *pinnen*, *pinnar*, *bromsen* och *in* använts. Dessutom har formerna *käppen* (L1) och *käpparna* (L2/sen) används. Även dessa svar har klassats som icke-idiomatiska. Detta kan dock diskuteras; Sköldberg (2004) har t.ex. visat att idiomat *sätta käppar i hjulen* varierar en hel del i språkbruket.

Hennes material uppvisar dock endast variation i numerus av lexemen *käpp* och *hjula*. Formerna som förekommer här är alltså *sätta en käpp i hjulen/hjulet* och *sätta käppar i hjulet/hjulen*. De bestämda formerna av substantivet *käpp* (*käppen, käpparna*) använts däremot inte (Sköldberg 2004:207). I anslutning till Sköldbergs resultat och för att inte vålla ytterligare gränsdragningsproblem i samband med analysen, har jag alltså valt att räkna användningen av bestämd form som avvikelser från den konventionaliserade formen (enligt Svenkt språkbruk 2003).

8 av svaren i samband med *sätta käppar i hjulen* har klassats som ogrammatiska. 2 av dessa förekommer i grupp L1, 1 i grupp L2/tidig och 5 i grupp L2/sen. Det handlar om användningen av formerna *\*snur* och *\*tjappen* (grupp L1); partiklarna *dit* (L2/tidig), *upp*, *ned* och *fast* (L2/sen) samt substantivet *pinne* och verbet *snurra* (grupp L2/sen). *Pinne* kan sägas ligga ganska nära det efterlysta ordet i betydelse, men är inte en idiomatisk lösning på uppgiften. Avgörande för kategoriseringen är att användningen av denna ordform utan obestämd artikel bryter mot svenskan grammatiska regler. Även *\*sätta snurra i hjulen* är oförenligt med svenskans grammatik, oavsett om man betraktar *snurra* som ett substantiv eller ett verb. *\*Sätta tjappen i hjulen* skulle kunna tänkas vara ett resultat av stavfel och därmed ett försök att komplettera det ofullständiga uttrycket med ordformen *käppen*, vilket alltså i så fall snarare skulle vara ett exempel på ett icke-idiomatiskt svar. Eftersom ordformen dock i så fall skulle innehålla flera stavfel samtidigt, ligger det närmare till hands att tolka svaret som att informanten inte behärskar den korrekta formen och därför har producerat ett ogrammatiskt svar. Intressant nog förekommer just detta fall bland förstaspråkseleverna i grupp L1<sup>4</sup>.

Endast 1 av 45 informanter har svarat på det förväntade sättet. Informanten ingår i grupp L1. Uttrycket verkar alltså vålla problem för både första- och andraspråkselever. Svartsbortfallet är relativt stort, inte minst bland andraspråkseleverna. Man kan konstatera att det även förekommer en hel del avvikelser från den konventionaliserade formen

---

<sup>4</sup> Här bör dock nämnas att informanten ifråga är speciell på det sättet att hon har angett svenska som enda modersmål och som det enda språk hon talar, men samtidigt skriver att hon inte är född i Sverige och att hon börjat lära sig svenska vid ettårsåldern.

bland förstaspråkseleverna. Det förekommer också ett antal svar som visar att informanten i fråga har partiell kunskap om uttrycket. Härtill kan svar som *sätta pinnar/pinnen i hjulen* och kanske även *sätta tjappen i hjulen räknas*. När det gäller svar som *sätta käppen/käpparna i hjulen* måste man utgå ifrån att informanterna i fråga känner till uttrycket, även om de använder en från den ”mest konventionaliserade” något avvikande form. Det senare är, som antytts tidigare, inte förvånansvärt med tanke på den variation som förekommer i samband med *sätta käppar i hjulen* (jfr Sköldberg 2004).

### 6.6.1.3. Få upp ögonen för något

TABELL 9.2: Antal olika typer av avvikelser i samband med uttrycket *få upp ögonen för något* per informantgrupp

	Inget svar	Vanliga kontaminationer	Idiosynkratiska kontaminationer	Icke-idiomatiska svar	Ogrammatiska svar	Förväntade svar
L1 (15)	8	–	4	1	1	1
L2/tidig (14)	10	–	2	1	1	–
L2/sen (16)	10	–	1	4	1	–
Totalt	28	–	7	6	3	1

Det förekommer sammanlagt 28 nollsvaret i samband med uttrycket *få upp ögonen för något*. Detta är det största svarsbortfallet bland de tio uttrycken som tas upp i den kvalitativa analysen. Även i detta fall står andraspråksgrupperna för större delen av svarsbortfallet (tio i grupp L2/tidig och 10 i L2/sen mot 8 i L1).

7 kompletteringar av *få upp \_\_\_\_\_ för något* har klassificerats som kontaminationer. 4 av dessa finns i grupp L1, nämligen exemplen *få upp huvudet för något*, *få upp hoppet för något*, *få upp modet för något*, *få upp hakan för något* samt *få upp den för något*. Det första exemplet kan ses som en kontamination med uttrycket *hålla huvudet högt*, som förekommer i A-delen av enkäten, vilket skulle kunna ha bidragit till att informanten blandat in en komponent från detta konventionaliserade ut-

tryck. Bakom det andra exemplet kan man förmoda en kontamination med uttrycken *ge upp hoppet* eller *hålla hoppet uppe*. Det tredje exemplet har analyserats som kontamination av *få upp ögonen för något* och det konventionaliserade uttrycket *hålla modet uppe*, som också förekommer i enkäten (del B). *Få upp hakan för något* kan ses som en kontamination med frasen *upp med hakan!* När det gäller det sistnämnda exemplet, finns det visserligen ingen kontext som kan indikera vad pronomenet *den* syftar på, vilket gör följande tolkning om möjligt ännu mer spekulativ än de övriga kontaminationsanalyserna. Då informanten är en 16-17-årig gymnasieelev, är det dock inte helt orimligt att förmoda en kontamination med frasen *få upp den* (som ofta används negerat i vardagligt språk, där *inte få upp den* betyder 'ha erektionsproblem'). Vid denna typ av svar ligger det nära till hands att utgå ifrån en medveten avvikelse från den konventionaliserade formen. Huruvida informanten behärskar det konventionaliserade uttrycket i fråga eller inte kan man i sådana fall bara spekulera om.

I grupp L2/tidig förekommer 2 svar som har klassats som kontaminationer. Eleverna har använt ordformerna *saken* och *humöret* för att komplettera det ofullständiga uttrycket ifråga. Formerna *få upp saken för något* och *få upp humöret för något* skulle kunna ses som kontaminationer med uttrycken *ta upp saken (med någon)* respektive *hålla humöret uppe*. Det senare uttrycket förekommer i B-delen av enkäten.

Ett av svaren i grupp L2/sen har klassats som kontamination. Även i detta fall har eleven använt substantivet *hakan* som komplettering.

6 av 45 svar i samband med uttrycket *få upp ögonen för något* har klassats som icke-idiomatiska. Ett av dessa förekommer i grupp L1. Frasen som eleven har producerat är *få upp \*u\*en för något* (som har tolkats som *få upp [kuken] för något*). Exemplet liknar det tidigare nämnda exemplet *få upp den för något*, som har räknats som kontaminerad form, då *få upp den*, i alla fall i vardagligt språk, kan ses som konventionaliserat uttryck (jfr Google 2010a). Även i fallet *få upp kuken för något* kan man förmoda en medvetet producerat avvikelse från den konventionaliserade formen ifråga. I grupp L2/tidigt har 1 informant använt pronomenet *det* för att komplettera det ofullständiga uttrycket. De övriga 4 icke-idiomatiska svaren återfinns hos informanterna i grupp L2/sen. Här förekommer formerna *få upp glöd för något*, *få upp någon för något* (2 gånger) och *få upp det för något*.

3 av de 45 svaren i föreliggande del av materialet har klassificerats som ogrammatiska. En av informanterna i grupp L1 har kompletterat fragmentet med verbet *stå*, vilket resulterar i formen *\*få upp stå för något*. En alternativ kategorisering skulle i detta fall kunna vara den som kontamination av *få upp ögonen för något* med uttrycket *stå för något*.

I de övriga 2 fallen har informanterna använt reflexivt personligt pronomen (L2/tidigt och L2/sen), vilket resulterar i *\*få upp sig/mig för något*. De exempel som redovisats här tyder på att flertalet informanter har fokuserat på konstituenten *upp* vid försöket att hitta rätt komplettering till det aktuella uttrycksfragmentet. Det sista ledet i frasen, *för något*, verkar däremot inte ha beaktats i särskilt hög utsträckning i en del av exemplen.

Endast 1 av 45 informanter har kompletterat fragmentet *få upp \_\_\_\_\_ för något* på det förväntade sättet. Även i detta fall ingår informanten i grupp L1.

#### 6.6.1.4. På rak arm

TABELL 9.3: Antal olika typer av avvikelser i samband med uttrycket på *rak arm* per informantgrupp

	Inget svar	Vanliga kontaminationer	Idio-synkratiska kontaminationer	Icke-idio-matiska svar	Ogrammatiska svar	Förväntade svar
L1 (15)	1	10	–	–	2	2
L2/tidig (14)	6	4	1	2	1	–
L2/sen (16)	9	4	–	–	3	–
Totalt	16	18	1	2	6	2

16 av 45 informanter har inte kompletterat fragmentet *på rak \_\_\_\_\_*. I grupp L1 förekommer endast ett nollsvar. Andraspråkeleverna i grupp L2/tidig står för 6, och de i grupp L2/en för 9 av de övriga nollsvaren.

Den kontaminerade formen *på rak sak* förekommer sammanlagt 9 gånger i denna del av materialet (5 i grupp L1, 2 i grupp L2/tidig samt 2 i grupp L2/sen). Det ligger nära till hands att en kontamination av de



konventionaliserade uttrycken *på rak arm* och *rakt på sak* ligger till grund för denna avvikelse, då dessa uttryck dels har två lexikala komponenter gemensamma, dels till viss del ligger nära varandra i betydelsen. En annan vanlig kontamination, som förekommer 6 gånger (5 i grupp L1, 1 i grupp L2/tidig) är den mellan *på rak arm* och uttrycket *på egen hand* (det senare förekommer också i B-delen av enkäten), som resulterar i formen *på rak hand*. I detta fall kan det även tänkas spela roll att substantivet *hand* ligger nära *arm* på ett semantiskt plan. Här kan man, med Hyltenstams terminologi, tala om en approximation (eng. *close approximation*) till målformen (Hyltenstam 1992, se avsnitt 4.4.1). En kontaminerad form som förekommer 5 gånger och därmed också klassats som vanlig, är *på rak väg* (2 gånger i L1, en gång i L2/tidig, en gång i L2/sen). I detta fall förmodas en kontamination med uttrycket *raka vägen* ligga till grund för avvikelsen. I samband med uttrycket *på rak arm* förekommer alltså tre olika typer av vanliga kontaminationer, vilket är ovanligt i det kvalitativt analyserade materialet. En komplettering som har räknats som idiosynkratisk kontamination är *på rak sikt* (L2/tidig). Denna form kan ses som en kontamination mellan *på rak arm* och *på sikt*.

Det förekommer 2 icke-idiomatiska svar i de 45 enkäter som inkluderats i denna del av analysen, båda i grupp L2/tidig. Exempelen är *på rak rygg* och *på rak sida*.

Sammanlagt 6 av svaren har klassats som ogrammatiska. I grupp L1 har substantivet *kurva* använts av 2 informanter (som förmodas att ha samarbetat kring svaren) som komplettering. Formen *\*på rak kurva* har klassats som grammatiskt inkorrekt då substantivet *kurva* vanligtvis inte kan föregås av prepositionen *på*. I grupp L2/tidig förekommer den inkongruenta formen *\*på rak tal*. I grupp L2/sen har bl.a. formen *\*på rak spår* bildats. Även här är den valda kompletteringen alltså inkongruent med den givna formen *rak*. I detta fall kan man förmoda en kontamination med uttrycket *gå raka spåret*. I L2/sen förekommer också formerna *\*på rak saken* och *\*på rak fram*. Bakom det senare exemplet kan man förmoda en kontamination med adverbialen *rakt fram*. Vid klassificeringen har dock den ogrammatiska formen fokuserats (jfr analysen av *få upp stå för ngt* i 6.6.1.4).

2 informanter, som båda ingår i grupp L1, har svarat på det förväntade sättet. Som för de uttryck som diskuterats tidigare, är det alltså en-

dast ett mindre antal förstaspråkselever som klarar av att bilda det konventionaliserade uttrycket ifråga.

### 6.6.1.5. Livet leker för någon

TABELL 9.4: Antal olika typer av avvikelser i samband med uttrycket *livet leker för någon* per informantgrupp

	Inget svar	Vanliga kontaminationer	Idiosynkratiska kontaminationer	Icke-idiomatiska svar	Ogrammatiska svar	Förväntade svar
L1 (15)	4	–	–	3	4	4
L2/tidig (14)	7	–	–	2	4	1
L2/sen (16)	9	–	1	1	5	–
Totalt	20	–	1	6	13	5

När det gäller uttrycket *livet leker för någon*, har 20 av de 45 informanterna avstått från att komplettera det givna fragmentet *livet \_\_\_\_\_ för någon* (4 i grupp L1, 7 i L2/tidig och 9 i L2/sen).

Det förekommer endast 1 (idiosynkratisk) kontamination i samband med det aktuella uttrycket, nämligen formen *livet faller för någon* (L2/sen), som kan förmodas vara en sammanblandning av *livet leker för någon* och uttrycket *falla för någon/något*.

Däremot finns det 6 svar i denna del av materialet som klassats som icke-idiomatiska (3 i grupp L1, 2 i L2/tidig och 1 i L2/sen). I grupp L1 har informanterna bildat formerna *livet klarnar/exploderar/föds för någon*. Alla dessa kompletteringar är grammatiskt möjliga. När det gäller användningen av verbet *explodera*, kan man däremot diskutera om det är möjligt på det semantiska planet. Man skulle dock kunna tänka sig en slags överförd betydelse av *explodera* och formen har därmed inte klassats som ogrammatisk. Liknande gäller för användningen av *föds*. I grupp L2/tidig har formerna *livet rycker* respektive *är viktigt för någon* bildats. I det förstnämnda fallet är det något oklart vilken betydelse av ordet *rycker* informanten avser. Även här lär det handla om en överförd användning av verbet. Det andra exemplet är grammatiskt och seman-

tiskt fullt möjligt, men däremot inte konventionaliserat. Det förekommer även ett icke-idiomatiskt exempel i grupp L2/sen, nämligen frasen *livet kämpar för någon*. I detta fall skulle man kunna argumentera för en klassificering som ogrammatisk, då verbet *kämpa* vanligtvis kräver animat subjekt. Utgår man från en användning av *kämpa* i överförd betydelse, är exemplet dock inte mindre grammatiskt än det efterlysta konventionaliserade uttrycket *livet leker för någon*, då även verbet *leka* i bokstavlig betydelse kräver ett animat subjekt. Exemplet kan däremot inte anses som idiomatiskt i föreliggande form (jfr användningen av *explodera* och *födast* ovan).

Sammanlagt 13 av 45 svar i samband med uttrycket *livet leker för någon* har klassificerats som ogrammatiska. 4 av dessa har producerats av elever i grupp L1. Formerna som bildats här är *\*livet är för någon* (3 gånger) och *\*livet för för någon*. I grupp L2/tidig förekommer 4 ogrammatiska svar. Eleverna har bildat fraserna *\*livet är för någon* (2 gånger), *\*livet viktigt för någon* och *\*livet hör för någon*. De sammanlagt 5 ogrammatiska exempel som producerats i grupp L2/sen är samliga förekomster av *\*livet är för någon*. Användningen av verbet *vara* är alltså relativt vanlig. Även här resulterar kompletteringen i ett frasfragment och anses därför ogrammatisk.

5 av informanterna har besvarat uppgiften på det förväntade sättet genom att använda verbet *leka*, 4 i grupp L1 och 1 i grupp L2/tidig. *Livet leker för någon* tillhör alltså de relativt få konventionaliserade uttrycken i den kvalitativa analysen, som någon av andraspråkseleverna har bildat på det förväntade sättet.

### 6.6.1.6. Med något i högsta hugg

TABELL 9.5: Antal olika typer av avvikelser i samband med uttrycket *med något i högsta hugg* per informantgrupp

	Inget svar	Vanliga kontaminationer	Idiosynkratiska kontaminationer	Icke-idiosynkratiska svar	Ogrammatiska svar	Förväntade svar
L1 (15)	3	3	1	1	5	2
L2/tidig (14)	8	4	1	–	1	–
L2/sen (16)	7	4	2	1	2	–
Totalt	18	11	4	2	8	2

När det gäller kompletteringen av fragmentet *med något i högsta \_\_\_\_\_* har 18 av 45 informanter avstått från att besvara uppgiften. I grupp L1 finns 3 icke-kompletterade fragment, samt 8 i grupp L2/tidig och 7 i L2/sen.

Vidare finns det en typ av vanlig kontamination som förekommer sammanlagt 11 gånger i denna del av materialet, formen *med något i högsta grad*, som här antas vara resultatet av en kontamination av *med något i högsta hugg* och det konventionaliserade uttrycket *i (allra) högsta grad*. Denna kontaminerade form har bildats av 3 informanter i grupp L1, 4 i grupp L2/tidig och 4 i grupp L2/sen. Dessutom förekommer 4 idiosynkratiska kontaminationer (en i grupp L1, 1 i grupp L2/tidig och 2 i grupp L2/sen). I grupp L1 återfinns formen *med något i högsta beredskap*, som kan vara en kontamination med uttrycket *vara/ligga i högsta beredskap*. I detta fall kan man notera att det finns viss semantisk likhet mellan de kontaminerade formerna. I grupp L2/tidig har formen *med något i högsta domstolen* bildats, bakom vilken man kan förmoda en kontamination med den konventionaliserade nominalfrasen *högsta domstolen*. De två kontaminerade formerna i grupp L2/sen är *med något i högsta tid* och *med något i högsta kvalitet*. Här förmodas att kontaminationer med uttrycken *det är hög tid* och *vara av högsta kvalitet* ligger till grund för valet av respektive ord.

2 av svaren har klassificerats som icke-idiomatiska (ett i grupp L1 och ett i grupp L2/sen). 1 av eleverna i grupp L1 har använt substantivet *höjd* för att komplettera fragmentet ifråga. I grupp L2/sen har den icke-idiomatiska frasen *med något i högsta skolan* bildats.

8 av 45 svar i samband med uttrycket *med något i högsta hugg* har klassats som ogrammatiska. I grupp L1 har dessa kompletteringar resulterat i formerna *\*med något i högsta bomb*, *\*med något i högsta sker*, *\*med något i högsta hönset* och *\*med något i högsta strå* (2 gånger, informanterna antas ha samarbetat). När det gäller användningen av ordformen *hönset* skulle man kunna argumentera för en klassificering som kontamination med den konventionaliserade nominalfrasen *högsta hönset*. Eftersom formen är nonsensartad och inte heller grammatiskt välformad har den dock klassats som ogrammatisk (jfr ovan). I grupp L2/tidig har ett svar klassats som ogrammatiskt, nämligen formen *\*med något i högsta betyg*. Denna förekommer även i grupp L2/sen, tillsammans med den ogrammatiska frasen *med \*med något i högsta skolan*.

Det finns endast 2 bland de 45 informanterna som har använt substantivet *hugg* och därmed bildat det konventionaliserade frasen *med något i högsta hugg*. Båda dessa informanter ingår i grupp L1.

#### 6.6.1.7. Klättra på väggarna

TABELL 9.6: Antal olika typer av avvikelser i samband med uttrycket *klättra på väggarna* per informantgrupp

	Inget svar	Vanliga kontaminationer	Idio-synkratiska kontaminationer	Icke-idiomatiska svar	Ogrammatiska svar	Förväntade svar
L1 (15)	2	–	–	6	5	2
L2/tidig (14)	9	–	–	2	2	1
L2/sen (16)	9	–	–	5	2	–
Totalt	20	–	–	13	9	3

I likhet med de uttryck som diskuterats tidigare är den vanligaste ”avvikelsen” i samband med uttrycken *klättra på väggarna* ett uteblivet svar;

i 20 av 45 fall har informanterna alltså inte kompletterat detta uttryck alls. 18 av dessa 20 icke-svar förekommer i andraspråksgrupperna (9 i grupp L2/tidig, 9 i grupp L2/sen). Detta verkar alltså vara ett uttryck som vållar större problem för andraspråkseleverna än för elever med svenska som förstaspråk.

13 av de 50 svaren i samband med *klättra på väggarna* har klassificerats som icke-idiomatiska (6 i grupp L1, 2 i grupp L2/tidig och 5 i grupp L2/sen). Några exempel är formerna *hänga, glo* och *äta* (L1), *stirra, slå* (L2/tidig) samt *måla* och *rita* (L2/sen) *på väggarna*.

Det förekommer inga kontaminationer i samband med *klättra på väggarna* i det kvalitativt analyserade materialet. Däremot finns 2 avvikelser som har klassats som ogrammatiska, som kan förmodas vara resultatet av något som liknar en kontamination. Två informanter i grupp L1 (som troligtvis har samarbetat) har kompletterat uttrycket med egennamnet *Karlsson*. Man kan tänka sig att det är Astrid Lindgrens barnboksfigur *Karlsson på taket* som ligger till grund för denna avvikelse, vilket innebär att svaret kan klassas som ett slags kontamination. *Karlsson på taket* är visserligen ett egennamn och skiljer sig därför från de konventionaliserade uttryck som är föremål för denna undersökning. Ordförbindelsen kan dock definitivt betraktas som konventionaliserad. Då svaret resulterar i nonsens-meningen

Det är mer underhållande än att se människor isolerade i ett hus \*som *Karlsson på väggarna* för att de är uttråkade

har svaren dock klassats som ogrammatiska. I detta fall ligger det nära till hands att förmoda ett avsiktligt nonsens-svar. Om orsaken till detta sätt att svara är att eleverna inte behärskar det efterlysta konventionaliserade uttrycket, är däremot omöjligt att bedöma. Andra exempel på ogrammatiska svar i samband med uttrycket *klättra på väggarna* (sammanslagt 9) är ...*att se människor isolerade i ett hus \*som att/hål* (L2/tidig), \**klister/tavlor/bläddrar på väggarna* (L1). Användningen av verbet *bläddra* har klassats som ogrammatiskt då den resulterar i ett brott mot svenskans selektionsregler i den aktuella kontexten. Utgår man från ordets betydelse, kan valet verka helt obegripligt. Däremot finns en viss fonologisk och grafisk likhet med det efterlysta verbet *klättra*, vilket kan vara en möjlig förklaring till denna avvikelse.

Det finns 3 bland de 45 informanterna som har kompletterat uttrycket *klättra på väggarna* på det förväntade, konventionaliserade sättet. 2 av dem finns i grupp L1 och 1 i grupp L2/tidig.

### 6.6.1.8. Få någon/något på halsen

TABELL 9.7: Antal olika typer av avvikelser i samband med uttrycket *få någon/något på halsen* per informantgrupp

	Inget svar	Vanliga kontaminationer	Idiosynkratiska kontaminationer	Icke-idiosynkratiska svar	Ogrammatiska svar	Förväntade svar
L1 (15)	4	–	4	2	–	5
L2/tidig (14)	6	–	1	4	3	–
L2/sen (16)	8	–	3	–	5	–
Totalt	18	–	8	6	8	5

Även när det gäller uttrycket *få någon/något på halsen* är svarsbortfallet relativt högt. 18 av 45 informanter har inte kompletterat fragmentet *få en riktig pratkvart på \_\_\_\_\_*. 14 av dessa är andraspråkselever (6 i grupp L2/tidig, 8 i L2/sen).

Det förekommer sammanlagt 8 idiosynkratiska kontaminationer i samband med *få någon/något på halsen* (4 i grupp L1, 1 i grupp L2/tidig och 3 i grupp L2/sen). Ett exempel från grupp L1 är *få en riktig pratkvart på nacken*, som kan ses som kontamination med *ha (några år) på nacken*. Dessutom har ordformerna *köpet* och *hjärnan* använts i grupp L1. I dessa fall ligger det, på grund av uttryckens likartade struktur, nära till hands att utgå ifrån kontaminationer med uttrycken *få något på köpet* och *få något [t.ex. en låt] på hjärnan*. En av informanterna i grupp L1 har endast fyllt i bokstaven *g* för att komplettera uttrycksfragmentet ifråga. Detta skulle kunna ses som en möjlig kontamination av *få något på halsen* och uttrycket *ha något på g[ång]*. Vidare finns det 1 idiosynkratisk kontamination som förekommer sammanlagt 3 gånger i de båda andraspråksgруппerna. 2 elever i grupp L2/sen och 1 i grupp L2/tidig har använt ordformen *käften* för att komplettera det aktuella ut-

trycket. Här är alltså en kontamination med *få på käften* en möjlig förklaring. Dessutom förekommer 1 kontamination med uttrycket *få någon/något på kroken* i grupp L2/sen.

6 av svaren i samband med *få någon/något på halsen* har klassificerats som icke-idiomatiska. I grupp L1 har ordformerna *axlarna* och *axeln*, och i grupp L2/tidig *munnen*, *sig*, och *mig* (2 gånger) använts för att komplettera uttrycksfragmentet i fråga.

Ogrammatiska lösningar av den aktuella uppgiften förekommer endast bland andraspråkseleverna (sammanlagt åtta). Exempel på denna typ av avvikelse är svar som *\*få något på allmänt/att* (grupp L2/tidig) och *\*få något på människor/tungal kommentar* (grupp L2/sen).

5 informanter, som samtliga ingår i grupp L1, har bildat det konventionaliserade uttrycket på det förväntade sättet. Det förekommer alltså inga förväntade svar bland andraspråkseleverna i föreliggande del av materialet.

### 6.6.1.9. Lyckligt ovetande

TABELL 9.8: Antal olika typer av avvikelser i samband med uttrycket *lyckligt ovetande* per informantgrupp

	Inget svar	Vanliga kontaminationer	Idiosynkratiska kontaminationer	Icke-idiomatiska svar	Ogrammatiska svar	Förväntade svar
L1 (15)	3	–	–	7	1	4
L2/tidig (14)	8	–	–	3	3	–
L2/sen (16)	6	–	–	5	4	1
Totalt	17	–	–	15	8	5

Uttrycket *lyckligt ovetande* är den enda kollokation som hamnar bland de tio uttryck som fått lägst poäng i hela materialet. Sammanlagt 17 av 45 informanter har avstått från att komplettera fragmentet \_\_\_\_\_ *ovetande*. 3 av dessa nollsvar förekommer i grupp L1, 8 i grupp L2/tidig och 6 i grupp L2/sen.



Som tidigare nämnts består uppgiften i att para ihop sammanlagt 5 adjektiv med ett av 8 givna adverb, som i detta fall är *evigt*, *exakt*, *hemligt*, *knappast*, *lyckligt*, *negativt*, *politiskt* och *svårt*. Adjektiven, bortsett från *ovetande*, är *troligt*, *sjuk*, *aktiv* och *tacksam* (se även appendix 1). Det är följaktligen adverbena *negativt*, *exakt* och *hemligt* som inte bör användas alls i denna del av uppgiften.

Beskrivningen av olika typer av avvikelser är något vanskelig i samband med tvåordskollokationer, då vissa av kategorierna som använts hittills är föga eller inte alls relevanta med tanke på hur uppgiften har ställts. Detta gäller i första hand kontaminationer. I många fall är det dock heller inte rimligt att klassificera avvikelser från kollokationen i fråga som ogrammatiska; när det gäller tvåordsuttryck av liknade struktur är det i de flesta fall en fråga om idiomacitet och språklig konvention när det gäller vilka ord som anses passa ihop. Med detta sagt, finns det dock några fall i materialet, där den komponent som informanten har valt är så oförenlig med adjektivet i fråga på ett semantiskt plan, att svaren klassats som ogrammatiska.

Resonemanget innebär att de flesta avvikelserna i samband med kollokationen *lyckligt ovetande*, bortsett från nollsvaren, har klassats som icke-idiomatiska. Av de sammanlagt 15 avvikelserna i denna kategori förekommer 7 i grupp L1, 3 i grupp L2/tidig och 5 i grupp L2/sen. Informanterna i grupp L1 har bildat formerna *svårt/politiskt/hemligt/evigt/knappast ovetande* (det sistnämnda exemplet förekommer 2 gånger). I grupp L2/tidigt förekommer *hemligt* och *politiskt* (2 gånger) *ovetande*. I grupp L2/sen har adverbena *knappast*, *politiskt* och *hemligt* (2 gånger) används för att komplettera den ofullständiga kollokationen.

De svar som innehåller adverbena *exakt* eller *negativt* har klassats som ogrammatiska då dessa adverb enligt svenskans semantiska regler och språkliga konventioner inte ingår i kollokationspotentialen för adjektivet *ovetande*. Dessa avvikelser förekommer sammanlagt 8 gånger. I grupp L1 är det bara *exakt* som använts som ogrammatisk komplettering. Informanterna i grupp L2/tidig har använt *exakt* en gång och *negativt* 2 gånger. I grupp L2/sen förekommer formen *\*negativt ovetande* 3 gånger och *\*exakt ovetande* en gång.

5 av 45 informanter har bildat den konventionaliserade formen *lyckligt ovetande* (4 i grupp L1 och 1 i grupp L2/sen). Det gör uttrycket till ett av dem som genererat flest förväntade svar bland de nio uttryck som

diskuterats hittills, vilket bekräftar det allmänna mönstret; ser man på hela materialet, har den aktuella kollokationen fått näst högsta antal poäng (53/186) bland de tio uttrycken som tas upp i den kvalitativa analysen.

#### 6.6.1.10. Inte för fem öre

TABELL 9.9: Antal olika typer av avvikelser i samband med uttrycket *inte för fem öre* per informantgrupp

	Inget svar	Vanliga kontaminationer	Idio-synkratiska kontaminationer	Ickeidiomatiska svar	Ogrammatiska svar	Förväntade svar	
				<i>ett femtio</i>			
L1 (15)	1	–	–	2	2	4	6
L2/tidig (14)	5	–	–	6	1	1	1
L2/sen (16)	5	–	–	2	6	3	–
Totalt	11	–	–	10	9	8	7

När det gäller uttrycket *inte för fem öre* har 11 av 45 informanter avstått från att besvara uppgiften. 1 av dessa ingår i grupp L1, 5 i grupp L2/tidig och 5 i grupp L2/sen.

Det förekommer inga kontaminationer i samband med detta uttryck. Däremot förekommer ett antal avvikelser som klassats som icke-idiomatiska och som alla är av samma typ. Denna typ av avvikelse består i användningen av ett annat räkneord än *fem* och förekommer hela 19 gånger i föreliggande del av materialet (4 i grupp L1, 7 i L2/tidig och 8 i L2/sen). De räkneord som används är *ett* (2 gånger i L1, 6 i L2/tidig, 2 i L2/sen) och *femtio* (2 gånger i L1, ett i L2/tidig, 6 i L2/sen). De avvikande räkneorden är den enda typen av icke-idiomatiska kompletteringar som förekommer i samband med uttrycket *inte för fem öre*. Man kan naturligtvis inte med säkerhet uttala sig om varför det är just *ett* eller *femtio öre* som används. Att man väljer en ännu mindre valör än *fem* är, med tanke på frasens betydelse, inte orimligt. Uttrycket används för att förstärka en negering, t.ex. för att betona att man 'inte alls tror på något, inte ens så lite som för det låga värdet av fem öre'. Om man endast

behärskar uttrycket partiellt, dvs. känner till dess betydelse men inte den konventionaliserade formen, kan det ligga nära till hands att ta till det lägsta värdet man kan tänka sig i öre. Här kan det dock noteras att en alternativ förklaring skulle kunna vara en kontamination med uttrycken *inte ett ruttet/rött öre*. En faktor som kan ligga till grund för användningen av *femtio* istället för *fem* är att *fem* och *femtio* är ganska lika till formen. En annan förklaring skulle kunna vara att informantens tankegång går via den bokstavliga betydelsen av konstruktionen *x öre*; då det numförtiden är femtio öre som är det lägsta myntet, är det logiskt att använda det i den aktuella kontexten.

Sammanlagt 8 av svaren i samband med *inte för fem öre* har klassats som ogrammatiska (4 i grupp L1, ett i L2/tidig och 3 i L2/sen). I grupp L1 har formerna *\*inte för korv öre*, *\*inte för ett korv öre*, *\*inte för mycket öre* och *\*inte för rätta öre* bildats. De två första exemplen kan förmodas beror på en kontamination med uttrycket *inte ett korvöre* (även här har dock klassificeringen baserats på den ogrammatiska formen). Exemplet *\*inte för en öre* i grupp L2/tidig avviker från den konventionaliserade formen både med avseende på idiomacitet och genuskongruens. I grupp L2/sen förekommer fraserna *\*inte för enda öre*, *\*inte för ända öre* och *\*inte för en ända öre*, som samtliga kan antas vara ett försök att bilda frasen *inte för ett enda öre*.

7 av 45 informanter har använt räkneordet *fem* och därmed kompletterat fragmentet *inte för \_\_\_\_\_ öre* på det förväntade sättet. 6 av dessa ingår i grupp L1 och 1 i grupp L2/tidig. Även här visar sig alltså samma tendens som i materialet som helhet; bland de tio uttrycken som ingår i den kvalitativa analysen har *inte för fem öre* fått högst poäng (54/186), sett till hela informantgruppen. Detta är alltså också fallet när det gäller de 45 enkäter som analyserats här.

## 6.6.2. De vanligaste avvikelserna

TABELL 10: Antal olika typer av avvikelser för samtliga tio uttryck per informantgrupp

	Inget svar	Kontaminationer			Ickeidiomatiska svar	Ogrammatiska svar	Förväntade svar
		vanliga	idiosynkratiska	alla			
L1 (15)	30	13	10	26	41	29	27
L2/tidig (14)	78	8	6	14	27	19	3
L2/sen (16)	78	8	10	18	29	35	1
Totalt	186	29	26	55	97	83	31

Som framgår av tabell 10, är den klart vanligaste ”avvikelsen” från de förväntade svaren i enkäten, att informanten har avstått från att komplettera uttrycksfragmentet i fråga. Det totala svarsbortfallet bland de 45 enkäterna som har analyserats är 186 av 450. Kontaminationer är betydligt mindre vanliga. Det finns sammanlagt 31 exempel på det som rubricerats som vanliga kontaminationer och 27 idiosynkratiska kontaminationer. De båda kontaminationstyperna är alltså ungefär lika vanliga i materialet (se tabell 10). De icke-idiomatiska svaren är, med sammanlagt 97 exempel, relativt vanligt förekommande. De ogrammatiska svaren hamnar med 81 förekomster på tredje plats. Att det finns relativt få exempel (sammanlagt 31) på de förväntade konventionaliserade uttrycken i detta material är givet, då lågt antal poäng har varit urvalskriteriet för de uttryck som analyserats här.

Jämför man antalet olika avvikelser i de tre informantgrupperna, visar sig några intressanta tendenser. I båda andraspråksgrupperna är svarsbortfallet betydligt högre än i grupp L1. Skillnaderna är däremot inte lika tydliga för övriga kategorier av avvikelser. Intressant är att kontaminationerna i det material som analyserats här verkar vara ungefär lika vanliga (t.o.m. något vanligare) bland förstaspråkseleverna som bland andraspråkseleverna, oavsett grupp. Detta ligger i linje med Hyltens stams tidigare refererade resultat (1992, jfr även Prentice & Sköldberg und. utg.). Som nämnts ovan är de icke-idiomatiska svaren de näst van-

ligaste efter nollsvaren. Antalet icke-idiomatiska svar i grupp L1 (41) är högre än i båda andraspråksgrupperna (26 i L2/tidig och 28 i L2/sen). När det gäller ogrammatiska svar är dessa, intressant nog, ungefär lika många i grupp L1 och L2/sen (29 respektive 35) men relativt få (19) i grupp L2/tidig.

I fråga om förväntade svar är resultatet tydligt och kan sägas ligga helt i linje med resultaten i den kvantitativa analysen (se tabellerna 8 – 8.3); nästan samtliga dessa svar (27 av 31) har producerats av eleverna i grupp L1. I grupp L2/tidigt förekommer endast tre, och i grupp L2/sen ett svar där eleverna har kompletterat uttrycksfragmentet på det förväntade sättet.

För att bekräfta de mönster som framträder i tabell 10 har ytterligare 10 enkäter av dem med lägsta resultat analyserats på samma sätt som de första 45. Tabell 10.1 visar att mönstret är i stort sett detsamma efter denna utökning av materialet.

TABELL 10.1: Antal olika typer av avvikelser för samtliga tio uttryck per informantgrupp – utökat material.

	Inget svar	Kontaminationer	Ickeidiomatiska svar	Ogrammatiska svar	Förväntade svar
L1 (19)	43	29	49	38	30
L2/tidig (16)	81	21	35	20	3
L2/sen (20)	100	23	36	41	1
Totalt	224	73	120	99	34

Resultaten av den kvalitativa analysen visar alltså på några intressanta aspekter med relevans för frågan varför vissa uttryck får låga resultat i undersökningen, såväl i hela materialet som i vissa av informantgrupperna. Det kan bl.a. konstateras att de låga resultaten hos de sena andraspråksinlärarna till stor del verkar bero på att eleverna helt enkelt har avstått från att komplettera de ofullständiga uttrycken i fråga. När det gäller förstaspråkseleverna verkar avvikelserna från den konventionaliserade formen i många fall vara en fråga om idiomacitet. Intressant nog förekommer det dock också relativt många svar som klassats som ogrammatiska i grupp L1, ungefär lika många som i grupp L2/sen och betyd-

ligt fler än i L2/tidig. I det sammanhanget måste det relativt låga antalet ogrammatiska lösningsförslag i den sistnämnda gruppen betraktas som ett något oväntat resultat, som dock ska ses i relation till det stora antalet nollsvär.

Överlag kan man säga att avvikelserna i grupp L1 är relativt jämnt fördelade över de olika kategorierna, till skillnad från andraspråksgrupperna, där kategorin *inget svar* är mycket mer markant.

### 6.6.3. Enkäterna med de högsta resultaten

En kvalitativ analys av enkäterna med högst resultat ger naturligtvis inte lika stort utslag som analysen av de enkäterna med lägst resultat i varje grupp. Men även om det inte finns lika många avvikelser att analysera i dessa enkäter, är det intressant att se i vilken mån mönstret i fråga om avvikelser från de konventionaliserade formerna bland de elever som klarar av större delen av uppgifterna skiljer sig ifrån mönstret som beskrivits i föregående avsnitt. Därför har enkäterna med högst resultat, dvs. minst 45 poäng, också analyserats kvalitativt (se avsnitt 5.5). 19 av de 20 informanterna som har nått 45 poäng eller mer ingår (inte helt oväntat, med tanke på de resultat som presenterats hittills) i grupp L1. Dessutom finns det 1 informant i grupp L2/tidig som har nått 46 poäng (jfr avsnitt 6.2.1).

Tabell 11 visar hur avvikelserna från de konventionaliserade formerna i samband med de tio svåraste uttrycken fördelar sig på de olika avvikelse typer som nämnts i föregående avsnitt. Det framgår av tabellen att samtliga 10 uttryck som har visat sig svårast för informanterna i materialet som helhet även genererar avvikelser bland de elever som har nått höga resultat i undersökningen. Det uttryck som visat sig generera flest avvikande svar i materialet, med andra ord det uttryck som verkar vara svårast av samtliga konventionaliserade uttryck i materialet, nämligen *hålla någon i schack*, genererar också flest avvikande svar bland eleverna med högst resultat (se tabell 11). Endast 7 av 20 informanter har fått poäng för uttrycket. Förutom detta uttryck är det uttrycket *livet leker för någon* som får betydligt lägre poäng än de övriga bland de tio svåraste uttrycken. Med sammanlagt 8 poäng i de 20 enkäterna med högst resultat är *livet leker för någon* alltså ett av två uttryck med ett resultat under 14 poäng. I de 45 enkäterna med lägst resultat får uttrycket sammanlagt

5 poäng, vilket innebär ett av de högsta resultaten bland de tio analyserade uttrycken i den delen av materialet.

För de övriga 8 uttrycken ligger resultaten i de 20 enkäterna med högst resultat på en ganska jämn nivå (mellan 14 och 17 av 20 poäng). Detta mönster skiljer sig alltså något ifrån analysen av de 45 enkäterna med låga resultat, där uttrycken får resultat som är spridda över alla siffror mellan 0 och 5 poäng.

En annan tydlig skillnad jämfört med resultaten för den ursprungliga kvalitativa analysen är att nollsvaren är betydligt färre i enkäterna med högst resultat. 6 av de 10 uttrycken har inte fått några nollsvaret alls. Bland dessa finns även *hålla någon i schack*, som i övrigt har genererat de flesta avvikelserna i den aktuella delen av materialet. Än en gång är det uttrycket *livet leker för någon* som sticker ut något, då uttrycket med 6 nollsvaret har genererat ett större antal sådana avvikelser än de övriga uttrycken. I det avseendet liknar resultatet det för den kvalitativa analysen av enkäterna med lågt resultat, där *livet leker* har genererat 20 nollsvaret.

En skillnad i jämförelse med den första kvalitativa analysen är att de högpresterande informanterna producerar få ogrammatiska svar. Av de sammanlagt 5 avvikelser som har klassats som ogrammatiska (jämfört med 81 ogrammatiska svar i den första kvalitativa analysen) förekommer 4 för uttrycket *livet leker för någon* och 1 för *med något i högsta hugg*.

Även när det gäller kontaminationer skiljer sig resultaten något från den första kvalitativa analysen; i de 45 enkäterna med lågt resultat förekommer sammanlagt 58 kontaminationer i samband med de 10 uttryck som har inkluderats i analysen. Dessa är jämfördelade på vanliga och idiosynkratiska kontaminationer. I de 20 enkäterna med högst resultat förekommer endast 9 kontaminationer. De kontaminationer som har visat sig vara vanligt förekommande i den första kvalitativa analysen verkar inte lika vanliga bland de högpresterande informanterna. Endast 2 av dessa kontaminationer förekommer i den aktuella delen av materialet, nämligen kontaminationen av *med något i högsta hugg* och *i högsta grad* samt den av *på rak arm* med *på egen hand*.

Icke-idiomatiska svar är med 31 förekomster den klart vanligaste kategorin av avvikelser som producerats av de 20 informanterna med högst resultat överlag. Här finns ytterligare en skillnad jämfört med resultaten för den kvalitativa analysen av enkäterna med lågt resultat, där nollsvaren är vanligast. Dessutom kan det konstateras att uttrycket *hål-*

TABELL II: Typer av avvikelser i de 20 enkäterna med högst resultat

	Hålla i schack	Käppar i hjulen	Få upp ögonen	På rak arm	Livet leker	Högsta hugg	Klättra på väggarna	Få ngt/ngn på halsen	Lyckligt overtande	Inte för fem öre	Totalt
avvikelser totalt	13	4	5	4	12	5	5	3	3	6	60
förväntade svar	7	16	15	16	8	15	15	17	17	14	140
inget svar	-	-	3	-	6	1	1	-	-	-	11
ogrammatiska svar	-	-	-	-	4	1	-	-	-	-	5
kontaminationer	-	1	1	-	2	-	-	2	-	2	8
vanliga kontaminationer	-	-	-	4	-	2	-	-	-	-	6
icke-idiomatiska svar	13	3	1	-	-	1	4	1	3	4	31



*la någon i schack* genererar betydligt fler icke-idiomatiska svar bland de högpresterande informanterna (13) än de övriga uttrycken, där antalet ligger mellan 0 och 5 förekomster av den aktuella typen av avvikelse. I likhet med de kvalitativa resultaten för enkäterna med lågt resultat är dock orsaken till avvikelsen även hos de högpresterande informanterna i många fall en bokstavlig tolkning av komponenten *schack* (som i 'schackspel').

I anslutning till detta kan man konstatera att skillnaden mellan resultaten av den första kvalitativa analysen och de avvikelser som de 20 högstpresterande informanterna producerar i första hand är kvantitativ. De avvikelser som de informanter som nått högst resultat ändå producerar, i samband med de tio aktuella uttrycken, är överlag av liknande karaktär som de avvikelser som produceras av informanterna med lågt resultat. I många fall används även samma ord för att komplettera uttrycksfragmenten i enkäten. Detta gäller bl.a. för de avvikande kompletteringar i samband med uttrycket *inte för fem öre*, där den vanligaste avvikelsen, även i denna del av materialet, är användningen av räkneordet *ett* istället för *fem*. Andra avvikelser som förekommer hos både låg- och högpresterande informanter är t.ex. *slå/sätta någon i schack*, *sätta stopp/pinnen/käppen i hjulen*, *stirrar på väggarna*, *\*livet är för någon*, *knappast ovetande* och, som tidigare nämnts, de kontaminerade formerna *på rak hand* och *med något i högsta grad*.

## 7. Sammanfattning och slutsatser

I föreliggande arbete har behärsknigen av ett antal svenska konventionaliserade uttryck bland ungdomar i flerspråkiga miljöer undersökts med hjälp av en enkätundersökning som innefattat både kvantitativa och kvalitativa analyser. En central fråga för undersökningen har varit huruvida behärsknigen av de aktuella uttrycken skiljer sig mellan första- och andraspråksinlärare av svenska, samt mellan andraspråksinlärare med olika startåldrar för inläringen av svenska.

Ytterligare frågeställningar för undersökningen har varit om det finns skillnader mellan olika typer av uttryck vad gäller informanternas behärskning av dessa och på vilket sätt informanterna avviker från de aktuella konventionaliserade former som efterlyses i enkäten. I de följande avsnitten sammanfattas och tolkas undersökningens viktigaste resultat.

### 7.1. Kvantitativ analys

Som framgår av den kvantitativa delen av detta arbete, finns det en rad signifikanta skillnader mellan startåldergrupperna, vilket tyder på att språklig bakgrund, och inte minst startåldern för inläring av ett andraspråk, är en relevant faktor i samband med behärsknigen av konventionaliserade uttryck. Resultaten som presenterats här kan sägas ligga i linje med tidigare studier som har visat att konventionaliserat språk är

en av de stora stötestenarna på väg mot en inföddlik språkbehärskning (jfr bl.a. Wray 2002, Abrahamsson & Hyltenstam 2009). I denna undersökning bekräftas detta genom signifikanta resultatskillnader mellan grupp L1 och grupp L2/tidig i varje deltest. Skillnaderna tyder på att även andraspråkselever på avancerad nivå med en relativt tidig start för inläringen har större svårigheter än förstaspråkselever när de ställs inför uppgiften att tolka och producera konventionaliserade uttryck på det sätt som krävs i föreliggande studie.

De signifikanta skillnaderna mellan grupp L1 och jämförelseklassen tyder på att klassmiljöernas heterogenitet i fråga om elevernas språkliga bakgrund kan ha viss betydelse när det gäller elevernas resultat i denna undersökning. Här måste man dock understryka att dessa resultatskillnader med hög sannolikhet också kan bero på en rad olika faktorer. I detta sammanhang kan skillnader i elevernas studiemotivation och/eller socioekonomiska faktorer (se avsnitt 5.1) vara minst lika avgörande för testresultaten som en språkligt homogen respektive heterogen skol- och klassrumsmiljö. Även om det alltså inte kan fastställas att resultatskillnaderna mellan förstaspråkseleverna och jämförelseklassen enbart beror på skillnaderna i klassernas språkliga sammansättning, kvarstår det faktum att den språkligt homogena jämförelseklassen når signifikant högre resultat än de förstaspråkselever som studerar i flerspråkiga klassmiljöer i alla deltest förutom del A.

### 7.1.1. Skillnader mellan deltesten

Formaten för de fyra deltesten är olika, vilket gör det något vanskligt att jämföra dem. Med andra ord är det naturligtvis inte helt överraskande att eleverna presterar bättre på ett flervals- och ett matchningstest än på ett lucktest, där de inte har givna lösningar att välja emellan. Man kan alltså inte dra slutsatsen att de konventionaliserade uttrycken i delarna A och D är lättare att förstå respektive bilda än dem i del B och C. Det innebär däremot inte att de signifikanta skillnaderna deltesterna emellan är ointressanta; dessa resultat säger något om svårighetsgraden i olika testuppgifter som i sin tur kan kopplas till olika aspekter av lexikal kunskap och behärskningen av konventionaliserade uttryck. Det finns två tydliga mönster när det gäller jämförelsen av deltesternas resultat inom de olika informantgrupperna; resultaten blir lägre med stigande

startålder för andraspråksinlärning i samtliga delar. Dessutom når samtliga grupper sina bästa resultat i A-delen och sina sämsta resultat i B-delen (se tabeller 7 och 7.1). Uppgifterna i del A – som verkar vålla minst problem för samtliga informantgrupper – kräver i första hand receptiv, och inte nödvändigtvis full behärskning av de konventionaliserade uttrycken i fråga (jfr Golden 2005). Även i D-delen kommer informanterna långt med sina receptiva kunskaper om kollokationerna, då konstituenterna är givna. Då orden ska matchas ihop, krävs det däremot även viss produktiv kunskap i fråga om de aktuella ordens kollokationspotential (Nation 2001). I lucktesten (B och C) är informanterna däremot tvungna att i högre grad använda produktiva kunskaper om de aktuella formerna, då det inte finns färdiga lösningar att välja emellan. Detta innebär att resultaten i dessa test säger något om det produktiva ordförrådets storlek och djup, vilket återspeglas i det höga svarsbortfallet i den sena startåldersgruppen i dessa delar av enkäten. När det gäller del C, måste informanterna behärska uttryckens form som helhet, då det inte är möjligt att sluta sig till det förväntade svaret med hjälp av kontexten eller utifrån de övriga konstituenternas betydelser. Detta visar sig, inte helt oväntat, vara svårast för samtliga informantgrupper.

Förhållandet mellan de olika delresultaten är inte bara intressant med tanke på uppgifternas svårighetsgrad och olika aspekter av lexikal och fraseologisk kunskap, utan även med tanke på undersökningsmetodens reliabilitet. Det enhetliga mönster som beskrivits för delresultaten ovan är en indikation på kontinuitet och koherens mellan enkätens olika delar (jfr även avsnitt 6.4).

## 7.2. De enskilda uttrycken

Även när det gäller resultaten för de enskilda uttrycken i enkäten visar undersökningen på en rad intressanta skillnader, både mellan informantgrupperna och de olika uttrycken. Det kan dock konstateras att det är betydligt svårare att beskriva några tydliga mönster för sambandet mellan ett visst uttrycks egenskaper och det resultat det genererar i undersökningen. Det är med andra ord svårt att dra några generella slutsatser om möjliga orsaker till att vissa uttryck behärskas av fler informanter än andra. Faktorerna genomskinlighet och kompositionalitet

däremot tänkas ligga till grund för ett mer allmänt mönster som framgår av tabellerna 8 – 8.3. De uttryck som kan beskrivas som mest genomskinliga och kompositionella får i allmänhet högre resultat än de minst genomskinliga och minst kompositionella. Denna skillnad kan illustreras med uppgiftsparen *sätta färg på något* och *prata i nattmössan* (A), *ta första steget* och *sätta käppar i hjulen* (B), *ta saken i egna händer* och *hålla någon i schack* (C) samt *allvarlig sjukdom* och *ligga till grund*. Ett tydligt undantag från denna regel är *dra alla över en kam*, som definitivt kan betraktas som ett av de minst kompositionella och relativt ogenomskinliga i materialet, men ändå ligger på tredje plats i tabell 8.1. Liknande kan sägas för *komma till kritan*, som också genererar ett relativt högt resultat medan ett mer genomskinligt uttryck som *hålla modet/humöret uppe* i samma tabell ger ett relativt lågt resultat. Som nämnts i avsnitt 2.2 är bedömningen av ett uttrycks genomskinlighet emellertid inte oproblematiserad, då den helt bygger på, i föreliggande fall författarens, intuition. Det finns inget objektiva sätt att bedöma hur genomskinligt ett uttryck är för någon som aldrig har mött det förut och begreppet i sig och bedömningarna som jag gör i samband med denna undersökning är att betrakta som relativa. Det finns alltså ingen skarp gräns mellan genomskinliga och ogenomskinliga uttryck och bedömningen handlar bl.a. om hur invecklad metaforiken inom ett uttryck är och om faktorer som betydelsetyngd eller polysemi hos de enskilda konstituenterna (jfr Svanlund 2009).

Det finns ytterligare möjliga påverkansfaktorer att diskutera i detta sammanhang, som t.ex. frekvens och i viss mån även sociala faktorer som informanternas ålder. När det gäller åldersfaktorn blir resonemangen dock ganska spekulativa, eftersom det inte har gjorts någon jämförelse med andra åldersgrupper. Inte heller frekvensdiskussionen är oproblematiserad mot bakgrund av de problem som diskuterats i samband med frekvenssökningar på Google. Som tidigare nämnts är Googlesökningarna i första hand att betrakta som en referenspunkt och som komplement till sökningarna i PAROLE, som inte alltid räcker till för att ge en bild av hur vanliga relativt lågfrekventa språkliga fenomen av den typ som behandlas här faktiskt är i det allmänna språkbruket. Även om webben alltså inte lämpar sig som språkkorpus i vanlig mening med tanke på det ständigt föränderliga materialet (Moon 2008), måste man ändå diskutera i vilken mån man kan bortse från språkbruket på webben i un-

dersökningar som denna. Det handlar trots allt om en språklig källa som används alltmer av allmänheten – både produktivt och receptivt – och som tillhandahåller ett stort, aktuellt och varierat språkligt material. Sammanfattningsvis kan man konstatera att det finns vissa tendenser som pekar i samma riktning som resultaten i Ellis, Simpson-Vlach & Maynard (2008), men att det utifrån denna analys inte kan dras några entydiga slutsatser om frekvensens roll för informanternas behärskning av de aktuella konventionaliserade uttrycken.

### 7.3. Den kvalitativa analysen

Den kvalitativa beskrivningen utifrån olika typer av avvikelser från konventionaliserade konstruktioner har visat sig något vanskelig, då skillnaderna mellan de olika typerna i vissa fall är svåravgränsade. Detta gäller särskilt kategorierna *icke-idiomatisk* och *ogrammatisk*. En annan vanskelig kategori är kontaminationerna. Problemet här är att bedömningen alltid blir mer eller mindre spekulativ. Icke desto mindre visar de kvalitativa resultat som presenterats här en rad intressanta tendenser bl.a. när det gäller vilka strategier informanterna använder för att lösa uppgifterna och hur de associerar mellan olika lexikala enheter.

Att förstaspråkseleverna står för de flesta förväntade svaren visades redan i den kvantitativa analysen. I den kvalitativa analysen av det reducerade materialet framträder däremot en del tydliga och intressanta tendenser vad gäller sambandet mellan typ av avvikelser och informanternas språkliga bakgrund. Ett viktigt resultat är att andraspråkseleverna i större uträkning än eleverna med svenska som modersmål väljer att avstå från att lösa uppgifterna. Det verkar som att förstaspråkselever, när de inte behärskar ett uttryck, är mer benägna att pröva olika typer av lösningar med hjälp av olika strategier. Detta är naturligtvis inget förvånande resultat; förstaspråkseleverna kan antas vara betydligt tryggare i sin språkanvändning än i synnerhet eleverna med sen startålder för inläring. De sena andraspråksinlärarna kan heller inte förväntas ha kommit ikapp eleverna i grupp L1, bl.a. när det gäller det ordförråd de har att laborera med. Med anknytning till Bybee (1998, 2006) finns det här alltså skäl att anta att de helt enkelt inte har utvecklat representationer för orden och uttrycken ifråga, inte heller sådana som kan associeras till

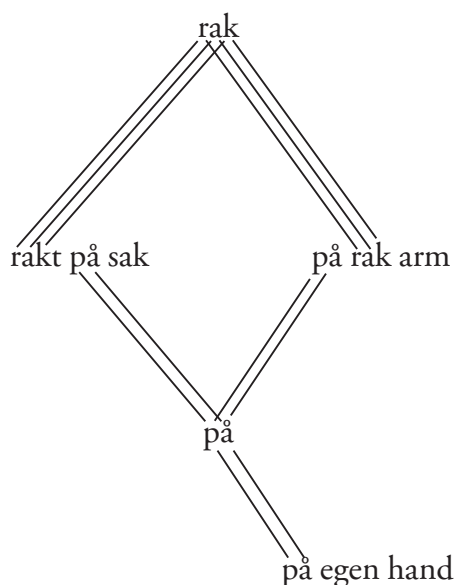
relaterade uttryck, vilket kan tänkas i högre grad vara fallet bland första-språkseleverna, som har fler och rikare språkliga erfarenheter när det gäller svenska och därmed en större uppsättning mentala representationer att välja bland (jfr Ekberg 1999).

I vissa fall kan den kvalitativa analysen ge en bättre bild av varför uttrycket i fråga har fått låga resultat i den kvantitativa analysen. Ett exempel är *inte för fem öre*, där det visar sig finnas en osäkerhet kring räkneordet, som dock inte verkar väljas helt godtyckligt. Det verkar snarare finnas ett tydligt mönster i fråga om vilka avvikande räkneord som kommer ifråga, nämligen den lägsta tänkbara valören i öre (*ett*) eller den för närvarande lägsta valören för mynt (*femtio*). När det gäller användningen av *ett* skulle man dock, som tidigare nämnts, också kunna analysera avvikelsen som en kontamination av uttrycket *inte för fem öre* med andra konventionaliserade former som *inte ett ruttet öre* eller *inte ett rött öre*.

### 7.3.1. Lexikala komponenter, kontext och mental representation

En annan aspekt som blir tydlig i den kvalitativa analysen av de 45 enkäterna med lågt resultat är att informanterna i många fall tar fasta på en av konstituenterna i uttrycksfragmentet och utifrån denna letar efter ord att komplettera med. Exempel på sådana konstituenten är partikeln *upp* i *få upp ögonen för något*, *högsta* i *med något i högsta hugg* och *sätta* i *sätta käppar i hjulen*, som verkar väcka associationer till en rad olika konventionaliserade konstruktioner, som i vissa fall bara har en konstituent gemensam med det efterlysta uttrycket (t.ex. *upp med hakan!*, *ta upp ett spår*, *ta upp saken*; *högsta hönset*, *högsta betyg*, *i högsta grad*; *sätta eld på något*, *sätta stopp för något* eller *sätta/slå spiken i kistan för något*). Detta är en tendens som visserligen i större utsträckning gäller för de lågpresterande informanterna, men som märks även bland de med de 20 högsta resultaten i undersökningen (se bl.a. kontaminationen av *med något i högsta hugg* och *i högsta grad*). Även om man utifrån föreliggande undersökning inte kan dra några generella slutsatser, är det intressant att fundera på vad detta kan säga om hur konventionaliserade flerordsuttryck lagras i det mentala lexikonet. Det verkar som att även enstaka lexikala

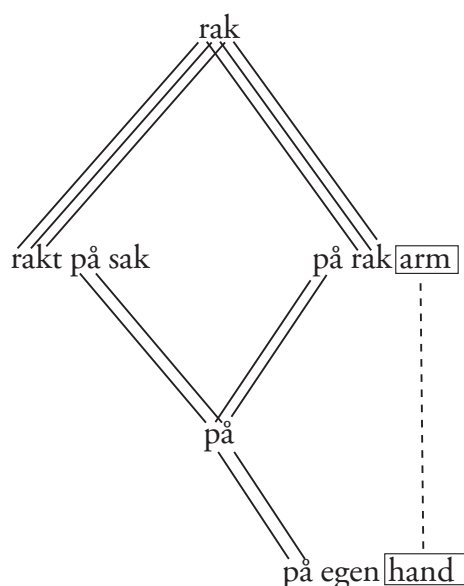
komponenter som är gemensamma för vissa uttryck kan ge upphov till kontamination av två uttryck som innehåller just denna komponent. Detta skulle kunna tyda på att olika konventionaliserade uttryck är sammanlänkade genom lexikala komponenter som är gemensamma för dem och därmed ge visst stöd för den lexikala modell som presenteras av Bybee (1998, 2006). Enligt Bybees modell skulle nätverket mellan olika uttryck med delvis samma komponenter utifrån föreliggande material kunna beskrivas som i figur 4.



FIGUR 4: Nätverk mellan de kontaminerade uttrycken och deras lexikala komponenter (efter Bybee 1998, 2006)

Samtliga slutsatser om den exakta karaktären hos dessa relationer i det mentala lexikonet är naturligtvis spekulativa. Man skulle även kunna hävda att gemensamma semantiska komponenter i olika uttryck med liknande struktur kan ge upphov till kontaminationer. Som nämnts i avsnitt 6.5.1.4, skulle man, med tanke på den kontaminerade formen *på rak hand*, även kunna förmoda att det finns en länk mellan komponenterna *arm* och *hand*, som tillhör samma semantiska fält. Enligt en sådan analys skulle relationerna mellan uttryck och lexikala komponenter kunna illustreras som i figur 4.1.





FIGUR 4.1: Nätverk mellan olika konventionaliserade uttryck och deras lexikala och semantiska komponenter

Det kan vidare noteras att det varierar kraftigt i vilken utsträckning informanter som nått lägsta resultat i sin grupp tar hänsyn till den övriga kontexten, dvs. till frasens övriga konstituenten och/eller den omgivande meningen. Detta gäller både form och betydelse och utgör i många fall skillnaden mellan icke-idiomatiska och ogrammatiska lösningsförslag. Det senare kan illustreras genom att jämföra formen *med något i högsta beredskap* med *\*med något i högsta hönset*, *på rak sak* med *\*på rak saken* eller *\*på rak fram* samt satserna *...att se människor isolerade i ett hus som stirrar på väggarna för att de är uttråkade* och *...att se människor isolerade i ett hus \*som Karlsson på väggarna för att de är uttråkade*. När det gäller informanterna med högst resultat överlag, kan det konstateras att de avvikelser som produceras i mindre utsträckning är ogrammatiska och att dessa informanter också producerar färre kontaminationer, vilket kan vara ett tecken på att de i större uträkning tar hänsyn till kontexten (i de fallen där det finns en sådan) vid kompletteringen av uttrycksfragmenten i enkäten. Detta är naturligtvis en tentativ slutsats och det finns, som tidigare påpekats, en rad olika faktorer som kan orsaka de resultatskillnader mellan hög- och lågpresterande informanter som påtalats här.

Det förekommer också ett relativt stort antal avvikande exempel i materialet som liknar den efterlysta konventionaliserade formen på olika sätt. Dessa likheter kan ligga på ett semantiskt plan (*sätta pinnen i hjulen, få upp blicken för något*), på ett fonologiskt (*sätta tjappen i hjulen, bläddra på väggarna*) och/eller grafiskt plan (*med något i högsta högg*). Dessa typer av avvikelser tyder på att informanterna kan ha partiell kunskap om det konventionaliserade uttryck som efterlyses, dvs. de känner till minst en av uttryckets konstituenten och har en ungefärlig uppfattning om vad den betyder, hur den låter eller hur den ser ut i skrift. Detta kan även vara en möjlig förklaring till att två uttryck som liknar varandra i ett eller flera avseenden kontamineras. Ett sådant exempel är den tidigare nämnda formen *med något i högsta beredskap*, där de kontaminerade uttrycken *med något i högst hugg* och *varalligga i högsta beredskap* till viss del liknar varandra, både semantiskt och med tanke på de ingående lexikala komponenterna. De kvalitativa resultat som sammanfattas ovan tyder alltså på att informanterna använder sig av en rad olika strategier för att komma fram till en lösning när de inte behärskar ett uttryck till fullo.

I båda analysdelarna har informanterna beskrivits som grupp. Även om det framträder tydliga mönster i analysen, finns det naturligtvis även en del individuell variation i materialet som helhet såväl som inom startåldersgrupperna. Resultaten i grupp L2/sen varierar mellan 2 och 39 av 55 möjliga poäng. I grupp L1 finns en spridning mellan 8 och 53 (se avsnitt 6.2.1). I den kvalitativa analysen framträder också en del individuella, oväntade lösningsförslag. Överlag finns det dock tydliga tendenser och förvånansvärt många identiska, eller i alla fall likartade svar bland dem som på olika sätt avviker från den förväntade, konventionaliserade formen, även om man tar med i beräkningen att det i vissa fall troligtvis har förekommit samarbete mellan informanterna.



## 8. Avslutande diskussion

Studien som presenterats här bekräftar en del tidigare forskningsresultat angående skillnader i behärskning av konventionaliserade uttryck mellan första- och andraspråksinlärare (jfr bl.a. Ekberg 1997, 1998, Hyltenstam 1992 och Abrahamsson & Hyltenstam 2009). Undersökningen visar signifikanta skillnader mellan första- och andraspråks elever i behärskning av konventionaliserade uttryck och bekräftar dessutom på ett tydligt sätt att startålder för inläring är en avgörande faktor när det gäller behärsningen av konventionaliserade uttryck på ett andraspråk. Utöver detta bidrar studien till forskningen kring konventionaliserat språk med en undersökning av både receptiv och produktiv behärskning av olika typer av konventionaliserade uttryck bland elever med varierande språklig bakgrund samt en genomgång av enskilda konventionaliserade uttryck med hjälp av både kvantitativa och kvalitativa analyser. Arbetet kan på det sättet bidra till ökad kunskap om vilka svårigheter särskilt andraspråks elever ställs inför i samband med olika typer av konventionaliserade uttryck. Studien kan dessutom ge en bättre bild av vilka mönster som finns i avvikelser från konventionaliserade strukturer i informantgrupper med olika startålder för inläringen av svenska och olika prestationsnivåer på tester liknande de som har används här. De kvalitativa resultaten kan dessutom ge vissa indikationer på hur både första- och andraspråksinlärare associerar mellan lexikala enheter som har gemensamma komponenter eller som liknar varandra på andra sätt.

Det finns naturligtvis andra faktorer än de som redovisats i denna studie som kan tänkas påverka förståelsen och produktionen av kon-

ventionaliserade uttryck på ett andraspråk. Bl.a. hade en analys av sambandet mellan olika modersmål och olika typer av avvikelser från de konventionaliserade formerna på svenska varit intressant. En sådan undersökning hade kunnat ge svar på frågan vilken betydelse transfer från modersmålet (se Ringbom 2006) eller andra språk har för andraspråksanvändares resultat i den typen av tester som genomförts inom ramen för detta arbete. Det hade emellertid bl.a. krävts ett mer systematiskt informanturval med avseende på modersmål samt en hel del kunskap om konventionaliserade uttryck i de aktuella språken. Syftet med denna undersökning har dock inte varit att studera skillnader mellan olika språkgrupper, utan att undersöka skillnader, dels mellan första- och andraspråkselever i olika språkliga miljöer, dels mellan andraspråkselever med olika startålder för inläringen.

En annan möjlig analys av föreliggande material – som inte rymts inom ramen för detta arbete men som skulle vara särskilt intressant i anknytning till tidigare forskning av Ellis, Simpson-Vlach & Maynard (2008) – skulle vara en undersökning av eventuella samband mellan elevernas behärskning av konventionaliserade uttryck och uttryckens frekvens och Mutual Information i en större korpus. Detta skulle bl.a. kunna ge en bättre bild av orsakerna till en del av de resultatskillnader som framträder i denna undersökning och på så sätt vara ytterligare ett bidrag till forskningen om inläringen av konventionaliserade språkliga mönster. Dessutom skulle en sådan undersökning kunna vara en utmärkt utgångspunkt för fortsatt forskning, inte minst som underlag för utarbetandet av en fraslista för svenska i t.ex. akademiska sammanhang (jfr Ellis, Simpson-Vlach & Maynard 2008).

Ett metodiskt val som kan diskuteras är formen för uttrycksfragmenten i de delarna av enkäten där fragmenten ska kompletteras utan kontext. De former som har valts kan sägas motsvara ordboksformer, vilket inte är helt oproblematiskt då dessa sällan används i naturlig produktion. Detta gäller bl.a. användningen av *något* och *någon*; det är inte särskilt sannolikt att eleverna har mött formerna *med något i högsta hugg*, *livet leker för någon* eller *hålla någon i schack* många gånger i allmänt språkbruk. Användningen av dessa former kan alltså bidra till ytterligare dekontextualisering av de efterlysta konventionaliserade uttrycken. Det är däremot svårt att tänka sig vad man skulle kunna använda istället utan att ge en kontext till fragmenten ifråga.

Det bör också poängteras att de metoder som används i denna undersökning har sina klara begränsningar när det gäller att utröna i vilken utsträckning informanterna behärskar ett uttryck eller inte (jfr Henriksen 1999). Det är naturligtvis också svårt att dra några säkra slutsatser om vad enstaka elevers resultat kan tänkas bero på. Bl.a. är det för försöksledaren omöjligt att veta i vilken mån informanterna har varit motiverade att lösa uppgifterna på bästa sätt. Vissa svar som framträder i den kvalitativa analysen tyder på att informanterna ibland medvetet gett ett svar som skiljer sig från det efterlysta svaret. Detta kan man emellertid inte med säkerhet veta utan att t.ex. intervjua eleverna. Samma osäkerhet gäller för de fall där informanterna har avstått från att lösa uppgiften; om nollsvaren i det individuella fallet beror på bristande motivation, bristande självförtroende eller bristande kunskap om uttrycket ifråga, kan här inte med säkerhet fastställas. Det är heller inte möjligt att veta om informanterna kan ha haft partiell kunskap om uttrycket, men ändå varit så pass osäkra att de valt att avstå ifrån ett lösningsförslag. Det finns naturligtvis metoder som möjliggör en djupare analys av olika grader av behärskning av lexikala enheter. Ett exempel är så kallade *think-a-loud-protocols*, där informanterna ges möjlighet att högt för sig själv resonera kring ett ord eller en fras, medan de spelas in (jfr Spöttl & McCarthy 2004). Ett annat alternativ är att använda en blandning av flera olika metoder, t.ex. test av olika slag i kombination med intervjuer (jfr Ellis, Simpson-Vlach & Maynard 2008). Dessa metoder är dock ofta tidskrävande både för försöksledaren och deltagarna, vilket gör det svårt att genomföra en sådan undersökning med en större informantgrupp, i synnerhet när det handlar om ungdomar som kan vara svåra att motivera för uppgiften. En enkät har fördelen att den är relativt lätt att distribuera till ett större antal informanter och genererar ett material som är relativt lätthanterligt. Även om denna undersökning inte kan tillhandahålla en djupare analys av elevernas behärskning av olika typer av konventionaliserade uttryck, visar den att det finns tydliga och signifikanta skillnader i elevers behärskning av olika typer av konventionaliserade uttryck och att dessa skillnader kan kopplas till faktorerna språklig bakgrund och startålder för andraspråksinläring (även om dessa naturligtvis inte är de enda bidragande faktorerna). Genom en kvalitativ beskrivning av olika typer av avvikelser från de efterlysta konventionaliserade formerna erhålls dessutom en bättre (om än inte heltäckande) bild av

tänkbara förklaringar till de låga testresultaten för vissa uttryck och till hur olika typer av avvikelser kan tänkas uppstå. Resultaten visar dessutom att det finns vissa samband mellan språklig bakgrund och olika typer av avvikelser i materialet. Undersökningen kan härigenom höja medvetenheten om olika typer av svårigheter som i synnerhet elever med flerspråkig bakgrund kan ställas inför i samband med tolkningen och användningen av konventionaliserade uttryck i svenskan. Resultaten som presenterats här kan också tjäna som underlag för fortsatt forskning, t.ex. vid en djupare analys av vissa uttryck som på olika sätt visat sig intressanta i denna undersökning.

# Referenser

- Abrahamsson, Niclas & Hyltenstam, Kenneth (2004). Mognadsbegränsningar och den kritiska perioden för andraspråksinläring. I: Hyltenstam, Kenneth & Lindberg, Inger (red.), *Svenska som andraspråk – i forskning undervisning och samhälle*. Lund: Studentlitteratur. S. 221–258
- Abrahamsson, Niclas & Hyltenstam, Kenneth (2009). Age of Onset and Nativelikeness in a Second Language: Listener Perception Versus Linguistic Scrutiny. *Language Learning*, 59/2: 249–306.
- Allén, Sture (1975). *Nusvensk frekvensordbok 3. Ordförbindelser/Frequency Dictionary of Present-Day Swedish 3. Collocations*. (Data Linguistica 10.) Stockholm: Almqvist & Wiksell.
- Bartsch, Sabine (2004). *Structural and Functional Properties of Collocations in English*. Tübingen: Gunter Narr Verlag.
- Bialystok, Ellen (2001). *Bilingualism in Development. Language Literacy & Cognition*. Cambridge: Cambridge University Press.
- Biber, Douglas & Conrad, Susan (1999). Lexical bundles in conversation and academic prose. I: Hasselgard, Hilde & Oksefjell, Signe (eds.), *Out of Corpora: Studies in Honor of Stig Johansson*. Amsterdam: Rodopi. S. 181–189.
- Birdsong, David (1999). *Second Language Acquisition and the Critical Period Hypothesis*. Mahwah, N.J., London : Erlbaum
- Bolinger, Dwight (1975). *Aspects of Language*. (2nd edition.) New York: Harcourt Brace Jovanovich.



- Brown, Roger (1973). *A First Language*. Cambridge, Massachusetts: Harvard University Press.
- Butler, Yuko G. (2000). The age effect in second language acquisition: Is it too late to acquire native-level competence in a second language after the age of seven? I: Oshima-Takane, Y., Shirai, Y. & Sirai, H. (eds.), *Studies in Language Science 1*. Tokyo: The Japanese Society for Language Science. S. 159–169.
- Bybee, Joan (1998). The Emergent Lexicon. I: Gruber, M. Catherine, Higgins, Derrick, Olson, Kenneth S. & Wysocki, Tamra (eds.), *CLS 34: The Panels*. Chicago: Chicago Linguistic Society. S. 421–439.
- Bybee, Joan (2006). From Usage to Grammar: The Mind's Response to Repetition. *Language* 82/4: 711–733.
- Church, Kenneth W. & Hanks, Patrick (1990). Word association norms, mutual information, and lexicography. *Computational Linguistics* 16/1: 22–29.
- Coleman, J.S. (1966). *Equality of Educational Opportunity*. Washington, D.C.: US Government Printing Office.
- Costello, Fintan J. & Keane, Mark T. (2000): Efficient creativity: constraint-guided conceptual combination. *Cognitive Science* 24/2: 299–349
- Coulmas, Florian (1979). On the sociolinguistic relevance of routine formulae. *Journal of Pragmatics* 3: 239–266.
- Cowie, Anthony P. (1988). Stable and creative aspects of vocabulary use. I: Carter, Ronald & McCarthy, Michael. *Vocabulary and language teaching*. London: Longman. S. 126–139.
- Cruttenden, Alan (1981). Item-Learning and System-Learning. *Journal of Psycholinguistic Research*, 10/1: 79–88.
- Crystal, David (1997). *The Cambridge encyclopedia of language*. Cambridge: Cambridge University Press.
- DeKeyser, Robert M. & Sokalski, Karl J. (1996). The differential roles of comprehension and production practice, *Language Learning* 46: 613–642.
- Dijkstra, Ton & van Heuven Walter J. B. (1998). The BIA model and bilingual word recognition. I: Grainger, Jonathan & Jacobs, Arthur M. (eds.), *Localist connectionist approaches to human cognition*. Mahwah, NJ: Erlbaum. S. 189–225.

- Ekberg, Lena (1989). *Gå till anfall och falla i sömn. En strukturell och funktionell beskrivning av abstrakta övergångsfraser*. Lund: Lund University Press.
- Ekberg, Lena (1997). Diskursiva och syntaktiska mönster i återberättelser hos invandrabarn i Rosengård. I: Håkansson, Gisela et al. (utg.), *Svenskans beskrivning 22*. Lund: Lund University Press. S. 97–110.
- Ekberg, Lena (1998). Regeltillämpning kontra lexikonkunskap i svenskan hos invandrabarn i Rosengård. I: Møller, Janus, Quist, Pia, Holmen, Anne & Norrman Jørgensen, Jens (utg.), *Nordiske sprog som andetsprog. Københavnerstudier i tosprogethed 30*. København: Københavns universitet. S. 247–263.
- Ekberg, Lena (1999). Användningen av komplexa predikat hos invandrabarn i Rosengård. I: Andersson, Lars-Gunnar, Lundqvist, Aina, Norén, Kerstin & Rogström, Lena (utg.), *Svenskans beskrivning 23. Förhandlingar vid Tjugotredje sammankomsten för svenskans beskrivning*. Lund : Lund University Press. S. 86–95.
- Ekberg, Lena (2004): Grammatik och lexikon i svenska som andraspråk på nästan infödd nivå. I: Hyltenstam, Kenneth & Lindberg, Inger (red.), *Svenska som andraspråk – i forskning, undervisning och samhälle*. Lund: Studentlitteratur. S. 259–276.
- Ellis, Nick C. (1995). The psychology of foreign language vocabulary acquisition: Implications for CALL. *Computer Assisted Language Learning* 8: 103–128.
- Ellis, Nick C. & Beaton, Alan (1993). Factors affecting foreign language vocabulary: imagery keyword mediators and phonological short-term memory. *Quarterly Journal of Experimental Psychology* 46: 533–558.
- Ellis, Nick. C., Simpson-Vlach, Rita & Maynard, Carson (2008). Formulaic Language in Native and Second Language Speakers: Psycholinguistics, Corpuslinguistics and TESOL. *TESOL Quarterly* 42/3: 375–396.
- Ellis, Rod, (1995). Modified oral input and the acquisition of word meanings. *Applied Linguistics* 16: 409–441.
- Enström, Ingegerd (1996). *Klara Verba. Andraspråksinlärares verb-användning i svenskan*. (Nordistica Gothoburgensia 18.) Göteborg: Acta Universitatis Gothoburgensis.

- Fernando, Chitra & Flavell, Roger (1981). *On idiom: Critical views and perspectives*. (Exeter Linguistic Studies 5.) Exeter: University of Exeter.
- Firth, John R. (1957). *Papers in Linguistics 1934-1951*. London: Oxford University Press.
- Fleischer, Wolfgang (1997). *Phraseologie der deutschen Gegenwartssprache*. (2., durchgesehene und ergänzte Auflage.) Tübingen: Niemeyer.
- Gläser, Rosemarie (1986). *Phraseologie der englischen Sprache*. Tübingen: Niemeyer.
- Gläser, Rosemarie (1988). The grading of idiomaticity as a presupposition for a taxonomy of idioms. I: Hüllen, Werner & Schulze, Rainer (eds.), *Understanding the lexicon. Meaning, sense and world knowledge in lexical semantics*. Tübingen: Niemeyer. S. 264–279.
- Golden, Anne (2005). *Å gripe poenget. Forståelse av metaforiske uttrykk fra lærebøker i samfunnskunnskap hos minoritets elever i ungdomsskolen*. Oslo: Unipub forlag.
- Granger, Sylviane (1993). The International Corpus of Learner English. I: Aarts, Jan, de Haan, Pieter & Oostdijk, Nelleke, *English Language Corpora: Design, Analysis and Exploitation*. Amsterdam: Rodopi. S. 57–71.
- Granger, Sylviane (1996). Learner English around the world. I: Greenbaum, Sidney (ed.), *Comparing English World-Wide*. Oxford: Clarendon Press. S. 13–24.
- Granger, Sylviane (1998). Prefabricated Patterns in Advanced EFL Writing: Collocations and Formulae. I: Cowie, Anthony P. (ed.), *Phraseology, Theory, Analysis and Applications*. Oxford: Oxford University Press. S. 145–160.
- Granger, Sylviane (2003). The International Corpus of Learner English: a new resource for foreign language learning and teaching and second language acquisition research. *TESOL Quarterly* 37/3: 538–546.
- Granger, Sylviane, Hung, Joseph & Petch Tyson, Stephanie (eds.) (2002). *Computer Learner Corpora, Second Language Acquisition and Foreign Language Teaching*. Amsterdam: John Benjamins.
- Greenberg, Joseph (1966). *Language universals: With special reference to feature hierarchies*. The Hague: Mouton.

- Große [Prentice], Julia (2008). Fraseologiska enheter på flerspråkig grund. En jämförelse mellan verklighet och fiktion. *Nordand. Nordisk tidskrift för andrespråksforskning* 3/2: 103–132.
- Gustafsson, J.-E. (1997). Measurement Characteristics of the IEA Reading Literacy Scales for 9- and 10-Year-Olds at Country and Individual Levels. *Journal of Educational Measurement* 34/3: 233–251.
- Halliday, M. A. K., McIntosh, Angus & Stevens, Peter D. (1964). *The Linguistic Sciences and Language Teaching*. London: Longmans.
- Halliday, M. A. K. (1966). Lexis as a linguistic level. I: Bazell, Charles E., Catford, John C., Halliday, M. A. K. & Robins, Robert H. (eds.), *In Memory of J.R. Firth*. London: Longmans. S. 148–162.
- Hausmann, Franz Josef (1985). Kollokationen im deutschen Wörterbuch. Ein Beitrag zur Theorie des lexikographischen Beispiels. I: Bergenholtz, Henning & Mugdan, Joachim (Hrsg.), *Lexikographie und Grammatik. Akten des Essener Kolloquiums zur Grammatik im Wörterbuch 28.-30.6. 1984*. Tübingen: Niemeyer. S. 118–129.
- Henriksen, Birgit, (1999). Three Dimensions of Vocabulary Development. *Studies in Second Language Acquisition* 21: 303–317.
- Holmegaard, Margareta (2007). Långa ord – en svårighet för flerspråkiga studerande? I: Lindberg, Inger & Johansson Kokkinakis, Sofie, *OrdiL. En korpusbaserad kartläggning av ordförrådet i läromedel för grundskolans senare år. Rapporter om svenska som andraspråk (ROSA) 8*. Göteborg: Institutet för svenska som andraspråk, Göteborgs universitet. S. 135–170.
- Hopper, Paul, J. & Thompson, Sandra A. (1980). Transitivity in grammar and discourse. *Language* 56: 251–99.
- Hyltenstam, Kenneth (1992). Non-native features of Near-native speakers: On the Ultimate Attainment of Childhood L2 Learners. I: Harris, Richard (ed.), *Cognitive processing in Bilinguals*. Amsterdam; London, New York, Tokyo: North Holland. S. 351–368.
- Hyltenstam, Kenneth & Abrahamsson, Niclas, (2000). Who can become a native speaker in a second language? All or none? On the maturational constraints controversy in second language acquisition. *Studia Linguistica* 54/2: 150–166.
- Kjellmer, Göran (1994). *A Dictionary of English Collocations. Based on the Brown Corpus. I-III*. Oxford: Clarendon Press.

- Kokkinakis, Dimitrios (2006). Collection, Encoding and Linguistic Processing of a Swedish Medical Corpus – The MEDLEX Experience. *Proceedings of the 5th Language Resources and Evaluation Conference (LREC), May 2006*, Genoa, Italy.
- Körner, Svante & Wahlgren, Lars (2005). *Statistiska metoder* (2. uppl.). Lund: Studentlitteratur.
- Lamb, Sidney M. (2000). Neuro-cognitive structure in the interplay of language and thought. I: Pütz, Martin & Verspoor, Marjolijn, *Explorations in linguistic relativity*. Amsterdam: John Benjamins. S. 173–196.
- Langacker, Ronald (2000). A dynamic usage-based model. I: Barlow, Michael & Kemmer, Suzanne (eds.), *Usage Based Models of Language*. Stanford, California: CSLI Publications. S. 1–64.
- Langlotz, Andreas (2006). *Idiomatic Creativity. A cognitive-linguistic model of idiom representation and idiom-variation in English*. Amsterdam & Philadelphia: John Benjamins.
- Lenneberg, Eric (1967). *Biological Foundations of Language*. New York: Wiley & Sons.
- Lindberg, Inger 2004. Språkanvändning – kreativt skapande eller förutsägbara mönster? I: Olofsson, Mikael (red.), *Symposium 2003, Arena Andraspråk*. Stockholm: HLS Förlag. S. 61–81.
- Lindberg, Inger, (2007): Multilingual Education. A Swedish Perspective. I: Carlson, Marie, Rabo, Annika & Gök, Fatma (eds.), *Education in 'Multicultural' Societies. Turkish and Swedish Perspectives*. (Swedish Research Institute in Istanbul Transactions, vol. 18.) Istanbul: Svenska Forskningsinstitutet i Istanbul. S. 71–90.
- Lindberg, Inger & Johansson Kokkinakis, Sofie (2007). *OrdiL. En korpusbaserad kartläggning av ordförrådet i läromedel för grundskolans senare år*. Rapporter om svenska som andraspråk (ROSA) 8. Göteborg: Institutet för svenska som andraspråk, Göteborgs universitet.
- Locke, J., (1997). A theory of linguistic development. *Brain and Language* 58: 265–326.
- Lorentzen, Henrik (2008): Korpusbaseret lemma-eksklusion eller ”hvor gik ordene hen da de gik ud?”. I: Jóhannesson, Kristinn, Landquist, Hans, Lundqvist, Aina, Rogström, Lena, Sköldberg, Emma & Wallgren Hemlin, Barbro (utg.), *Nog ordat? Festskrift till Sven-Göran Malmgren*. (Meijerbergs Arkiv för svensk ordforskning 34.) Göteborg. S. 259–267.

- Luthman, Hans (2002). *Svenska idiom. 3500 vardagsuttryck*. Lund: Folkuniversitetets förlag.
- Makkai, Adam (1972). *Idiom structure in English*. The Hague & Paris: Mouton.
- Malmgren, Sven-Göran (2003). *Begå eller ta självmord?* Om svenska kollokationer och deras förändringsbenägenhet. *Språk och stil* 13: 123–168.
- McCarthy, Michael, (1998). *Spoken language and applied linguistics*. Cambridge: Cambridge University Press.
- Mel'čuk, Igor, (1998). Collocations and lexical Functions. I: Cowie, Anthony, P. (ed.), *Phraseology, Theory, Analysis and Applications*. Oxford: Oxford University Press. S. 23–53.
- Moon, Rosamund (1998). *Fixed Expressions and Idioms in English. A Corpus-Based Approach*. Oxford: Oxford University Press.
- Moon, Rosamund, (2008). Conventionalized *as*-similes in English. A problem case\*. *International Journal of Corpus Linguistics* 13/1: 3–37.
- Mühlenbock, Katharina (2009). Readable, legible or plain words - Presentation of an easy-to-read Swedish corpus. Multilingualism, Proceedings of the 23rd Scandinavian Conference of Linguistics, *Studia Linguistica Upsaliensia* 8 s. 325–327.
- Mühlenbock & Johansson Kokkinakis (2009). LIX 68 revisited. An extended readability measure. Konferensbidrag på Corpus Linguistics Conference (CL) 2009. University of Liverpool, UK20-23 July 2009.
- Munro, Miles & Mann, Virginia (2005). Age of immersion as a predictor of foreign accent. *Applied Psycholinguistics* 26: 311–341.
- Nagy, William E., Herman, Patricia A. & Anderson, Richard C. (1985). Learning words from Context. *Reading Research Quarterly* 20: 233–253.
- Nation, Paul (2001). *Learning Vocabulary in Another Language*. Cambridge: Cambridge University Press.
- Nist, Sherrie L. & Olejnik, Stephen (1995). The role of context and dictionary definitions on varying levels of word knowledge. *Reading Research Quarterly* 30: 172–193.
- O'Keefe, Anne, McCarthy, Michael & Carter, Ronald (2007). *From Corpus to Classroom: language use and language teaching*. Cambridge. Cambridge University Press.

- Pallier, C., Dehaene, J., Poline, J.-B., LeBihan, D., Argentié, A.-M., Dupoux, E. & Mehler, J. (2003). Brain Imaging of Language Plasticity in Adopted Adults: Can a Second Language Replace the First? *Cerebral Cortex* 2003 13: 155–161.
- Palmer, Harold E. (1933). *Second Interim Report on English Collocations*. Tokyo: Kaitakusha.
- Partee, Barbara H. (1984). Compositionality. I: Landman, Fred & Veltman, Frank (eds.), *Varieties of Formal Semantics. Proceedings of the fourth Amsterdam Colloquium, September 1982*. Dordrecht: Foris. S. 281–311.
- Paul, Peter V., Stallman, A. C. & O'Rourke, Joseph P. (1990). *Using three test formats to assess good and poor readers' word knowledge*, Technical Report No. 509 of the Center for the Study of Reading. Urbana-Champaign: University of Illinois.
- Pawley, Andrew & Syder, Frances H. (1983). Two puzzles for linguistic theory: nativelike selection and nativelike fluency. I: Richards, Jack & Schmidt, Richard (eds.), *Language and Communication*. London: Longman S. 191–221.
- Peters, Ann M. (1983). *Units of Language Acquisition*. Cambridge: Cambridge University Press.
- Prentice, Julia & Emma Sköldbäck (und. utg.) *Klättra på väggarna eller bara vara ett med soffan? Om figurativa ordförbindelser bland ungdomar i flerspråkiga skolmiljöer*. Språk och stil 20.
- Ringbom, Håkan (2006). On L1 Transfer in L2 Comprehension and L2 Production. *Language Learning* 42/1: 85–112.
- Schmitt, Nobert (2008). Instructed second language vocabulary learning. *Language Teaching Research* 12/3: 329–363
- Sinclair, John (1966). Beginning the Study of Lexis. I: Bazell, Charles E., Catford, John C., Halliday, M. A. K. & Robins, Robert H. (eds.), *In Memory of J.R. Firth*. London: Longmans. S. 410–430.
- Sinclair, John (1987). Collocation: A progress report. I: Steele, Ross & Treadgold, Terry (eds.), *Language Topics. Essays in honour of Michael Halliday*. Amsterdam: John Benjamins. S. 319–331.
- Skehan, Peter (1998). *A Cognitive Approach to Language Learning*. Oxford: Oxford University Press.
- Skolverket (2004). *Elever med utländsk bakgrund. Rapport till regeringen, 1 oktober 2004*. (Dnr 75 – 2004:545.) Stockholm: Skolverket.

- Skolverket (2005). *Elever med utländsk bakgrund. En sammanfattande bild*. Stockholm: Skolverket.
- Skolverket (2008). Studieresultat i gymnasieskolan – en statistisk beskrivning av ofullständiga gymnasiestudier. (Skolverkets aktuella analyser.) Stockholm: Skolverket.
- Skolverket (2009). *Vad påverkar resultaten i svensk grundskola? Kunskapsöversikt om betydelsen av olika faktorer*. Stockholm.
- Sköldbberg, Emma (2004). *Korten på bordet. Innehålls- och uttrycksmässiga variation hos svenska idiom*. (Meijerbergs Arkiv för svensk ordforskning 31.) Göteborg: Göteborgs universitet.
- Sköldbberg, Emma & Toporowska Gronostaj, Maria (2008). Modell för beskrivning av kollokationer i ett medicinskt lexikon (MedLex). I: Svavarsdóttir, Ásta, Kvaran, Guðrun, Ingólfsson, Gunnlaugur & Jónsson, Jón Hilmar (red.), *Nordiske Studier i Leksikografi 9. Rapport fra konference om leksikografi i Norden, Akureyri 22. – 26. maj 2007*. Reykjavík: Nordisk Forening for Leksikografi. S. 433–446.
- Spöttl, Carol and Michael McCarthy (2004). Comparing knowledge of formulaic sequences across L1, L2, L3, and L4. I: Schmitt, Norbert (ed.), *Formulaic Sequences*. Amsterdam & Philadelphia: John Benjamins. S. 191–225.
- Stubbs, Michael (2001). *Words and Phrases. Corpus Studies of Lexical Semantics*. Oxford & Malden, Massachusetts: Blackwell.
- Svanlund, Jan (2001). *Metaforen som konvention. Graden av bildlighet i svenskans vikt- och tyngdmetaforer*. (Stockholm Studies in Scandinavian Philology. New series. 23.) Stockholm: Almqvist & Wiksell.
- Svanlund, Jan (2009). Lexikal etablering. *En korpusundersökning av hur nya sammansättningar konventionaliseras och får sin betydelse*. (Acta Universitatis Stockholmiensis. Stockholm Studies in Scandinavian Philology. New series. 52.) Stockholm. Stockholms universitet.
- Svensén, Bo (1987). *Handbok i leksikografi*. Stockholm: Norstedts ordbok.
- Thomas, Wayne & Collier, Virginia (2002). A National Study of School Effectiveness for Language Minority Students' Long-Term Academic Achievement. Center for Research on Education, Diversity & Excellence (CREDE).
- Van Lier, Leo (1995). *Introducing Language Awareness*. London: Penguin Books.



- Viberg, Åke (1993). Andraspråksinläring i olika åldrar. I: Cerú, Eva (red.), *Lärarybok 2. Sevnska som andraspråk. Mera om språket och inläringen*. Stockholm: Natur och Kultur, Utbildningsradion. S. 13–83.
- Wahlgren, Lars (2005). *SPSS steg för steg*. Lund: Studentlitteratur.
- Wiktorsson, Maria (2003). *Learning Idiomaticity. A corpus-based study of idiomatic expressions in learners' written production*. Lund: Lund University.
- Wray, Alison, 2002: *Formulaic Language and the Lexicon*. Cambridge: Cambridge University Press.
- Wray, Alison (2008). *Formulaic Language: Pushing the Boundaries*. Oxford: Oxford University Press.
- Wray, Alison & Perkins, Michael R. (2000). The functions of formulaic language: an integrated model. *Language and Communication* 20: 1–28.
- Yang, Kajsa & Gustafsson, Jan - Eric (2003). Measuring Socioeconomic Status at Individual and Collective Levels. *Educational Research and Evaluation* 10/3: 259–288.
- Yorio, Carlos A. (1989). Idiomaticity as an indicator of second language proficiency. I: Hyltenstam, Kenneth & Obler, Loraine K. (eds.), *Bilingualism across the lifespan*. Cambridge: Cambridge University Press. S. 55–72.

## Övriga källor

- Google (2010a). <[www.google.se](http://www.google.se)>, hämtat 11 januari 2010.
- Google (2010b). <[www.google.se](http://www.google.se)>, hämtat 6 februari 2010.
- Skolverket 2010, <<http://siris.skolverket.se>>, hämtat 20 juni 2010.
- Språkbanken (2010a). <<http://spraakbanken.gu.se/parole/>>, hämtat 11 januari 2010.
- Språkbanken (2010b). <<http://spraakbanken.gu.se/konk/>>, hämtat, 6 februari 2010.
- Svenskt språkbruk. Ordbok över konstruktioner och fraser (2003)*. Stockholm: Norstedts ordbok.

# Appendix

**Först några korta frågor om dig och din språkanvändning.  
Kryssa i de alternativ som beskriver dig!**

- Tjej
- Kille

Gymnasieprogram:

---

- Är född i Sverige och har svenskfödda föräldrar
- Är född i Sverige och har utlandsfödda föräldrar
- Är född i Sverige och har **en** utlandsfödd förälder
- Är inte född i Sverige. Kom till Sverige vid \_\_\_\_\_ års ålder
- Började lära mig svenska när jag var \_\_\_\_\_ år gammal

Vilket/vilka är ditt/dina modersmål? \_\_\_\_\_

Vilket/vilka språk talar ni hemma i familjen?  
\_\_\_\_\_

Vilket/vilka språk talar du mest med dina kompisar?  
\_\_\_\_\_  
\_\_\_\_\_

Har du under din skoltid läst svenska som andraspråk och i så fall när och hur länge?

Nej

Ja

När? \_\_\_\_\_

Hur många år? \_\_\_\_\_

Har du under din skoltid läst ämnet modersmål?

Nej

Ja

När? \_\_\_\_\_

Hur många år? \_\_\_\_\_

Vilket typ av svenskundervisning deltar du i just nu?

svenska

svenska som andraspråk

båda

vet ej

## A

**Vad tror du att följande uttryck betyder? Välj det alternativ som du tycker stämmer bäst. Markera bara ett alternativ!**

Exempel:

ha händerna fulla

- vara rik
- ha mycket att göra
- inte kunna röra sig
- Inget av alternativen stämmer, uttrycket betyder följande:

---

---

1) prata i nattmössan

- prata i sömnen
- prata otydligt, mumla
- säga dumheter, inte veta vad man talar om
- Inget av alternativen stämmer, uttrycket betyder följande:

---

---

2) fastna för något

- inte kunna ta sig loss från något, sitta fast
- bli förtjust i något och bestämma sig för att t.ex. köpa det
- inte förstå något
- Inget av alternativen stämmer, uttrycket betyder följande:

---

---

3) något är/blir en ond cirkel

- när människor med onda avsikter bildar en grupp
- när en tråkig händelse leder till nästa
- när någon skadar sig gång på gång
- Inget av alternativen stämmer, uttrycket betyder följande:

---

---

4) något står en upp i halsen

- när man är tvungen att ta hand om något
- när man inte kan svälja något
- när man har fått nog, när man är mycket trött på något
- Inget av alternativen stämmer, uttrycket betyder följande:

---

---

5) hålla huvudet högt

- gå upprätt
- se ner på andra människor
- behålla sin stolthet
- Inget av alternativen stämmer, uttrycket betyder följande:

---

---

6) ha fria tyglar

- vara lösaktig
- få göra precis som man vill
- strunta i något
- Inget av alternativen stämmer, uttrycket betyder följande:

---

---

7) ha mage att göra något

- ha den fysiska kraften att göra något
- våga göra något som andra kan uppfatta som fräckt
- vara lagom tjock för att göra något
- Inget av alternativen stämmer, uttrycket betyder följande:

---

---

8) sätta färg på något

- måla om
- göra något trevligare, gladare eller mer intressant
- dela in, kategorisera något med hjälp av olika färger
- Inget av alternativen stämmer, uttrycket betyder följande:

---

---

9) stå i rampljuset

- vara i centrum för allmänhetens intresse
- spela teater
- vara bländad av ljus
- Inget av alternativen stämmer, uttrycket betyder följande:

---

---

10) innan man vet ordet av

- när man inte kommer på vad man ska säga
- när något händer så snabbt att man inte hinner reagera
- när man aldrig hört talas om något
- Inget av alternativen stämmer, uttrycket betyder följande:

---

---

## B

**Här kommer exempel på några vanliga uttryck och fraser. Fyll i det ord som du tycker saknas i de olika uttrycken.**

Exempel:

I Expedition Robinson måste man jobba för att överleva. Det funkar inte att åka snålskjuts för då blir man utrustad direkt.

Samma sak gäller dokusåpan Farmen. Vill man vara kvar där så måste man dra sitt strå till stacken.

1) När jag misslyckas försöker jag alltid att hålla \_\_\_\_\_ uppe och inte deppa i onödan.

2) Det är år 2003 och alla borde vara jämlika.

Jag vill inte \_\_\_\_\_ *alla män över en kam* men de flesta vet nog vad jag pratar om.

3) Jag erkänner, om jag hade makten att ändra det förflutna eller mitt eller ditt öde, så skulle jag nog \_\_\_\_\_ *för frestelsen*.

4) På Fame Factory får deltagarna lära sig hur man agerar på scen och skriver egna låtar. Det är mer underhållande än att se människor isolerade i ett hus som \_\_\_\_\_ *på väggarna* för att de är uttråkade.

5) När jag sedan blev äldre och skulle gå i skolan, fick tiden helt plötsligt betydelse. Från att ha blivit körd av föräldrarna till dagis, skulle jag nu ta mig till skolan *på egen* \_\_\_\_\_ och vara där i tid.

6) Ibland låter man helt enkelt bli att prata med folk, av rädsla att *få en riktig pratkvart på* \_\_\_\_\_.

7) Gör din röst hörd! Få politikerna att förstå att det är vi, nästa generation, som ska styra samhället i framtiden, som blir lidande. Något måste hända och det är vi ungdomar som måste ta det *första* \_\_\_\_\_ !



8) Kan det vara så att vi skjuter upp saker i onödan och när det sedan verkligen \_\_\_\_\_ till kritan klagar vi på att vi har haft för lite tid för att åstadkomma ett bra resultat?

9) En tolkning av scenen från filmen ”Moderna tider”, där Charlie Chaplin med våld vrider tillbaka tiden, skulle kunna vara att Chaplin inte vill följa med i utvecklingen. Därför försöker han bokstavligen att sätta \_\_\_\_\_ i hjulen.

10) Ryck upp er alla latmaskar! *Ta er i* \_\_\_\_\_ och börja träna!

C

**Fyll i det ord som fattas i uttrycken nedan. Fundera inte för länge på varje uttryck utan skriv det som du först kommer att tänka på.**

Exempel:

gå ner i varv

- 1) tid är \_\_\_\_\_
- 2) ta saken i egna \_\_\_\_\_
- 3) som \_\_\_\_\_ på moset
- 4) av hela mitt \_\_\_\_\_
- 5) \_\_\_\_\_ någon i schack
- 6) livet \_\_\_\_\_ för någon
- 7) på rak \_\_\_\_\_
- 8) på löpande \_\_\_\_\_
- 9) ha ett \_\_\_\_\_ med i spelet

- 10) gå bakom \_\_\_\_\_ på någon  
11) inte för \_\_\_\_\_ öre  
12) ingen dans på \_\_\_\_\_  
13) med något i högsta \_\_\_\_\_  
14) rinna ut i \_\_\_\_\_  
15) få upp \_\_\_\_\_ för något

## D

**Vilka ord passar bäst tillsammans? Välj det ord som du tycker passar bäst bland de olika alternativen. Tänk på att alla de alternativa orden inte behöver användas!**

Exempel:

Fyll i det ord som saknas i följande uttryck. Välj bland följande alternativ: *bilda, dra, föra, hålla, leda, lägga, spela, ta*  
dra en slutsats

Fyll i det ord som saknas i följande uttryck. Välj bland följande alternativ: *hamna, hoppa, komma, leva, ligga, sitta, stå*  
leva i celibat

Fyll i det ord som saknas i följande uttryck. Välj bland följande alternativ: *djup, evig, full, negativ, politisk, stor, svensk, typisk*  
svensk medborgare

Fyll i det ord som saknas i följande uttryck. Välj bland följande alternativ: *djupt, evigt, exakt, fullt, mycket, stort, särskilt, typiskt*  
exakt likadan

Fyll i det ord som saknas i följande uttryck. Välj bland följande alternativ: *bilda, bryta, dra, föra, ge, kasta, leda, lägga*

- 1) \_\_\_\_\_ en blick
- 2) \_\_\_\_\_ ett exempel
- 3) \_\_\_\_\_ ett beroende
- 4) \_\_\_\_\_ en union
- 5) \_\_\_\_\_ en suck

Fyll i det ord som saknas i följande uttryck. Välj bland följande alternativ:  
*få, gå, hamna, komma, ligga, sträva, stå, vara*

- 6) \_\_\_\_\_ i uppgift
- 7) \_\_\_\_\_ till nytta
- 8) \_\_\_\_\_ i missbruk
- 9) \_\_\_\_\_ till grund
- 10) \_\_\_\_\_ i arv

Fyll i det ord som saknas i följande uttryck. Välj bland följande alternativ:  
*allvarlig, djup, full, lätt, mycket, olika, stark, stor*

- 11) \_\_\_\_\_ slags
- 12) ett \_\_\_\_\_ andetag
- 13) \_\_\_\_\_ fart
- 14) en \_\_\_\_\_ sjukdom
- 15) \_\_\_\_\_ betydelse

Fyll i det ord som saknas i följande uttryck. Välj bland följande alternativ:  
Välj bland: *evigt, exakt, hemligt, knappast, lyckligt, negativt, politiskt, svårt*

16)\_\_\_\_\_troligt

17)\_\_\_\_\_sjuk

18)\_\_\_\_\_ovetande

19)\_\_\_\_\_aktiv

20)\_\_\_\_\_tacksam



**ROSA.** (*Rapporter Om Svenska som Andraspråk*) utkommer vid Institutet för svenska som andraspråk, Institutionen för svenska språket, Göteborgs universitet. Redaktörer för serien är Sofie Johansson Kokkinakis och Sofia Tingsell.

**DET FINNS** många sätt att uttrycka ett visst innehåll på ett grammatiskt korrekt sätt i ett språk, men relativt få av dessa accepteras som naturliga uttrycksätt av infödda talare, då de genom upprepad användning har blivit konventionaliserade inom språkgemenskapen i fråga. Med andra ord kännetecknas ett inföddligt eller idiomatiskt språkbruk bl.a. av användningen av konventionaliserade språkliga strukturer. På ett semantiskt plan utmärker sig dessutom konventionaliserade flerordsuttryck som idiom och kollokationer genom att deras betydelse inte helt motsvarar summan av de ingående komponenternas betydelse.

I denna rapport, som är en del i en doktorsavhandling i svenska som andraspråk, redovisas en studie av behärskningen av ett femtiotal konventionaliserade svenska uttryck bland gymnasieungdomar med varierande språklig bakgrund. Med såväl kvantitativa som kvalitativa analyser av elevernas receptiva och produktiva behärskning av olika typer av uttryck som *prata i nattmössan*, *på rak arm* eller *lyckligt ovetande* utforskar Julia Prentice på vilket sätt dessa konventionaliserade uttryck kan sätta käppar i hjulen för en idiomatisk språkanvändning hos gymnasieungdomar med olika startålder för inläringen av svenska.