

GÖTEBORGS UNIVERSITET
Institutionen för svenska språket

Akademiskt ordförråd under utveckling:
med fokus på substantiv i Tisustexter

Ylva Olausson

Magisteruppsats, SSA220, 15 hp
Svenska som andraspråk
Vt 2017
Handledare: Sofie Johansson

Sammandrag

Den här uppsatsen undersöker huvudsakligen kvantitativa aspekter av samband mellan lexikala särdrag, med fokus på substantiv, och bedömd färdighetsnivå på Tisus (test i svenska för universitets- och högskolestudier) skriftliga del. Eftersom ett godkänt resultat på Tisus ger grundläggande behörighet i svenska för högskolestudier så kan bedömningen få mycket stora konsekvenser för den enskilda testtagaren och därför är det här ett angeläget område att utforska.

Studien genomförs på 32 Tisustexter, 16 godkända och 16 underkända, hämtade från Tisuskorpusen som är en inlärlarkorpus inom projektet Swedish Learner Language (SweLL). Delområden i den kontrastiva undersökningen är nominalkvot, substantiviska sammansättningar samt jämförelser med innehållet i *Svensk akademisk ordlista* (AO) och i frekvensbaserade ordlistor från bloggkorpus. Syftet är att beskriva hur lexikala särdrag, med fokus på substantiv, skiljer sig mellan godkända och underkända texter och därmed eventuellt påverkar bedömningen av Tisus skriftliga del.

Studien av nominalkvot visar att det verkar finnas ett samband mellan andelen substantiv i texten och bedömningen. Vidare är substantiviska sammansättningar samt de nominaliseringar, och övriga ord som sammanfaller med AO mer frekventa i de 16 godkända texterna jämfört med de underkända. Dessutom har de underkända texterna en högre täckningsgrad i jämförelsen med de frekvensbaserade ordlistorna från bloggkorpusen vilket tyder på att de innehåller fler talspråkliga drag än de som är godkända.

Underlaget är inte tillräckligt stort för att dra generella slutsatser, men det finns mycket som tyder på att testtagarens kunskap om och användning av de innehållsord som tillhör ordklassen substantiv, vilka bidrar till texternas informationstäthet, spelar roll för bedömningen. Det verkar dock förekomma en hel del avvikelser från medelvärdena i enskilda texter vilket pekar mot att andra bedömningskriterier som exempelvis helhetsbedömningen av texten väger tungt.

Nyckelord: Korpuslingvistik, inlärlarkorpus, SweLL, akademiskt skrivande, bedömning, frekvensordlistor, *Svensk akademisk ordlista*, sammansättning, Tisus, sva-inlärlare, ordförråd, ordinlärlning, skrivutveckling

Innehållsförteckning

| | |
|--|----|
| 1. Inledning | 1 |
| 1.1. Syfte och forskningsfrågor | 1 |
| 1.2. Bakgrund | 2 |
| 2. Tidigare forskning | 8 |
| 2.1. Lexikal språkutveckling | 9 |
| 2.2. Korpuslingvistik | 11 |
| 2.3. Ordlängd och sammansatta ord | 12 |
| 2.4. Frekvensbaserade ordlistor | 13 |
| 2.5. Bedömning av elevtexter | 15 |
| 3. Material | 16 |
| 4. Metod | 20 |
| 4.1. Kartläggning och jämförelser | 21 |
| 4.2. Undersökningens 32 Tisustexter | 22 |
| 4.3. Substantiven i de 32 Tisustexterna | 23 |
| 4.4. Nominalkvot | 25 |
| 4.5. <i>Svensk akademisk ordlista (AO)</i> | 25 |
| 4.6. Substantiviska sammansättningar | 28 |
| 4.7. Frekvensbaserade ordlistor | 29 |
| 4.8. Metoddiskussion | 29 |
| 4.9. Etiska överväganden | 31 |
| 5. Resultat | 32 |
| 5.1. Nominalkvot | 32 |
| 5.2. <i>Svensk akademisk ordlista (AO)</i> | 33 |
| 5.3. Substantiviska sammansättningar | 38 |
| 5.4. Frekvensbaserade ordlistor | 40 |
| 6. Diskussion | 41 |
| 7. Avslutning | 47 |
| Litteraturförteckning | 49 |

Bilagor

Tabellförteckning

- Tabell 1. Tre lexikala dimensioner (efter Hellspång & Ledin, 1997:78 f.)
- Tabell 2. Exempel från Tisustexterna på substantiviska sammansättningar med förled från olika ordklasser
- Tabell 3. Olika aspekter av lexikonkunskap (efter Nation 2013:49)
- Tabell 4. 105 texter i Tisuskorpusen – fördelning utifrån färdighetsnivå
- Tabell 5. 105 texter i Tisuskorpusen – fördelning utifrån färdighetsnivå samt redovisade med och utan skribenter som angivit svenska som modersmål
- Tabell 6. Texter i Tisuskorpusen – skribenter som angivit svenska som modersmål, alla variabler
- Tabell 7. Undersökningens 32 Tisustexter – fördelning antal tokens och types/färdighetsnivå, medelvärden antal tokens/types, samt TTR
- Tabell 8. Antal substantiv, types i Tisustexterna + antal i faktabladet
- Tabell 9. Nominalkvot – medelvärden för G5, G3 och U2
- Tabell 10. Medelvärden – types substantiv i G5, G3 och U2-texterna som återfinns i AO
- Tabell 11. Spännvidd inom G5, G3 och U2 – från lägst till högst andel substantiv som återfinns i AO räknat i %
- Tabell 12. Substantiv i de 32 texterna som innehåller substantivbildande suffix och som också återfinns i AO
- Tabell 13. Användning av *fall*, *form*, *roll* och *verk*
- Tabell 14. Medelvärden, substantiviska sammansättningar i texterna i relation till tokens
- Tabell 15. Medelvärden – antal stavelser per substantivisk sammansättning / types (särskrivningar medräknade)
- Tabell 16. Jämförelse – sammansatta ord med tre led, fyra led, dubbel syftning samt särskrivning
- Tabell 17. Medelvärden procent substantiv i de undersökta texterna som sammanfaller med de frekvensbaserade ordlistorna samt medelvärde procent av alla ord i de undersökta texterna, token coverage

1. Inledning

Att i en testsituation skriva en formell resonerande text på avancerad språklig nivå om ett ämne som testtagaren inte är förberedd på och med en provtid på två och en halv timme är inte en enkel uppgift ens för den som skriver på sitt förstaspråk. Men så ser förutsättningarna ut för de andraspråksinlärare som vid två olika provtillfällen per år genomför den skriftliga delen av Tisus (test i svenska för universitets- och högskolestudier [www]). 32 Tisustexter från provtillfället i maj 2006, varav hälften godkända och hälften underkända, utgör underlag för den här uppsatsen.

Inlärarspråk skiljer sig på många punkter från L1 och andraspråksinlärning är ett brett forskningsfält i ständig utveckling. En hel del svensk forskning grundar sig på språkutvecklingsteorier som beskriver att inlärare oberoende av språklig bakgrund har liknande problem på vägen mot en målspråksnära kompetens, till exempel Pienemanns processbarhetsteori (1998). Ett exempel är Heinonen Eklund (2009) som i sin avhandling studerar testtagarnas färdigheter i Tisus muntliga del och där använder processbarhetsteorin i den grammatiska analysen.

Under senare decennier har allt fler forskare ägnat sig åt andraspråksinlärares lexikala utveckling och i kvantitativa frekvensbaserade studier undersökt exempelvis ordval, lexikal variation, samt överanvändning och felanvändning av ord. De möjligheter som den digitala tekniken numera erbjuder har i snabb takt öppnat för nya möjligheter till framtagning av frekvensbaserade ordlistor, jämförelser mellan andraspråksinlärare i olika korpussamlingar och mycket annat.

1.1. Syfte och forskningsfrågor

Syftet med studien är att utifrån Tisuskorpusen (*Språkbanken*, Göteborgs universitet [www]), en enspråkig synkron korpus innehållande 105 texter (59 639 löpord/tokens) från provtillfället i maj 2006, beskriva kvantitativa aspekter av lexikaliska särdrag som eventuellt påverkar bedömningen i godkända respektive underkända texter skrivna av avancerade sva-inlärare.

Tisus ges på sex olika universitet/högskolor i Sverige och på ett antal orter utomlands och testtagarna har ofta som mål att studera på svenska universitet och

högskolor. Detta betyder att bedömningen av Tisus kan få stora konsekvenser för den enskilda testtagaren vilket i sin tur innebär att det här är ett angeläget område att utforska.

Enligt Enström (2013:172) har den här kategorin inlärare ett stort behov av att utöka sitt ordförråd inom tre kategorier: ämnesrelaterade facktermer, ämnesneutrala ”akademiska” ord samt mindre frekventa vardagliga ord. Sva-inlärares lexikaliska kompetens har med andra ord avgörande betydelse för resultatet på Tisus. Min undersökning strävar efter att utifrån syftet besvara följande forskningsfrågor:

Skiljer sig lexikala särdrag åt vid en jämförelse mellan de texter i Tisus-korpusen som har bedömts med betyg G respektive betyg U? Om skillnader finns mellan G- och U-texter: Vilka skillnader kan man urskilja i skribenternas användning av substantiv vid jämförelser utifrån nominalkvot, innehållet i *Svensk akademisk ordlista* (AO), i frekvensbaserade ordlistor samt substantiverade sammansättningar?

En hypotes i min undersökning är att de godkända texterna innehåller ett mer utvecklat ordförråd vad gäller substantiv än de som är underkända. Att språket inte känns ”begränsat lexikalt eller syntaktiskt” (bilaga 1), är ett kriterium för att bli godkänd på Tisus skriftliga del. Det innebär att i bedömningen ingår, mer eller mindre medvetet, testtagarens användning av de viktiga innehållsord som substantiven representerar. Följaktligen kan vi anta att exempelvis mindre vanliga substantiv, ämnesneutrala ”akademiska” substantiv och sammansatta substantiv förekommer mer frekvent i de texter som är godkända jämfört med de underkända.

1.2. Bakgrund

I det här kapitlet presenteras inledningsvis en bakgrund till lexikala särdrag. Därpå följer en kort presentation av projektet Swedish Learner Language (SweLL) och Tisus, samt av de fyra olika delområden som har fungerat som utgångspunkt för jämförelser av lexikala särdrag i godkända och underkända Tisustexter nämligen: nominalkvot, innehållet i *Svensk akademisk ordlista* (AO) och i frekvensbaserade ordlistor samt substantiviska sammansättningar.

1.2.1. Lexikala särdrag

Det nuvarande interna bedömningsbladet för Tisus utgår från det holistiska perspektivet, och där formuleras nivån för godkänt som:

Texten är klar och sammanhängande och kan i regel följas och förstås utan ansträngning. Tankar har utvecklats och språket känns inte nämnvärt begränsat lexikalt eller syntaktiskt och har en sådan formell kvalitet att inga nämnvärda missförstånd uppstår (bilaga 1).

Vilka lexikala särdrag som gör att språket inte känns begränsat och som därmed inte hindrar förståelsen specificeras inte i bedömningsbladet (se bilaga 1). Men vad som skiljer ett planerat skriftspråk från spontant tal har många språkvetare beskrivit på olika sätt. Ett ofta citerat exempel på detta är Hellspong & Ledin (1997:78 f.) som presenterar tre lexikala dimensioner, se tabell 1 nedan. Den första dimensionen är skalan *nominal-verbal* där många substantiv i relation till verb och pronomen utmärker ett genomtänkt skriftspråk.

Den andra dimensionen som Hellspong & Ledin (1997:78 f.) presenterar är skalan *specifik-allmän* där till exempel långa ord, sammansatta substantiv och många fackord ingår i det som de menar är typiskt för skriftspråket.

Den tredje dimensionen, slutligen, är *abstrakt-konkret* och dit hör att många abstrakta substantiv, exempelvis avledda substantiv, utmärker det genomtänkta skriftspråket, något som illustreras i tabell 12 i min undersökning.

Tabell 1. Tre lexikala dimensioner (efter Hellspong & Ledin, 1997:78 f.)

| Skriftspråk | ↔ | Talspråk | |
|--------------------|----------|-----------------|---------------------------------|
| Nominal | ↔ | Verbal | = grad av informationstäthet |
| Specifik | ↔ | Allmän | = grad av specialisering |
| Abstrakt | ↔ | Konkret | = grad av närvaro i tid och rum |

Sva-inlärarens lexikala kompetens, som den här studien fokuserar på, har enligt många SLA-forskare stor betydelse för språkutvecklingen i sin helhet, inklusive utvecklingen av skrivförmågan. Följande citat från Milton (2009:251) får illustrera detta:

...the more we understand of the vocabulary learning process, the more important it seems to be to the whole of the language learning process.

1.2.2. SweLL och Tisuskorpusen

År 2016 presenterades projektet SweLL där man samlar in och digitaliserar texter skrivna av sva-inlärare (Volodina et al. 2016:1). I korpussamlingen SweLL som alltså är under utveckling, och som innebär en ny resurs för forskning på inlärarspråk, ingår Tisuskorpusen och därifrån är materialet till den här studien hämtat.

Den första testomgången med Tisus genomfördes 1997. Testet består av tre delar: muntlig examination, läsförståelse samt, för att testa den skriftliga förmågan, en text i den resonerande genren. Tisuskorpusen innehåller 105 texter (59 639 löpord/tokens) från Tisus skriftliga del i Stockholm i maj 2006. Ett godkänt resultat på Tisus ger som redan nämnts allmän behörighet i svenska till akademiska studier i Sverige. Att med fokus på substantiven undersöka vilka lexikala särdrag som kan tänkas ha påverkat bedömningen av texterna i korpusen är angeläget dels eftersom vi behöver fördjupa våra kunskaper inom området och också för att det i nuläget råder brist på korpusstudier på Tisus.

1.2.3. Nominalkvot

Substantiv är innehållsord och enligt Lagerholm (2008:132) är beräkning av nominalkvoten ett grovt kvantitativt mått på textens informationstäthet. Hellspång & Ledin (1997:78 f.) nämner nominalkvoten i den första lexikala dimensionen, *nominal/verbal* (se tabell 1). Man kan beräkna nominalkvot på olika sätt och i den här studien har valet av praktiska skäl (se 4.8) fallit på den enkla nominalkvoten som enligt Lagerholm (2008:133) är textens totala antal substantiv dividerat med textens totala antal verb.

Nominalkvoten säger givetvis ingenting om vilka substantiven är och deras funktion i texten där de uppträder, eller om de ens är korrekt använda i sitt sammanhang. Men andelen substantiv ger ändå en fingervisning om textens informationstäthet och kan avslöja skillnader mellan olika texter och därmed utgöra en utgångspunkt för undersökningen. Talspråk karakteriseras av en låg nominalkvot och man kan tänka sig att graden av mer formell skriftspråklighet åtminstone till viss del avspeglas i den enkla nominalkvoten. Följaktligen är det logiskt att i utgångsläget förvänta en lägre nominalkvot i de underkända texterna jämfört med de godkända.

Lagerholm (2008:108) betonar att för att få grepp om textens stil behöver man, förutom att undersöka nominalkvoten, gå vidare och undersöka vilken sorts substantiv det handlar om. Lagerholm beskriver, på liknande sätt som Hellspong & Ledin (1997:78 f.) i tabell 1, att det har stor betydelse för texttypen om substantiven är konkreta eller abstrakta, enkla eller sammansatta samt högfrekventa eller lågfrekventa. Och förutom den enkla nominalkvoten är detta också faktorer som undersöks i den här studien.

1.2.4. Svensk akademisk ordlista (AO)

Svensk akademisk ordlista (AO) är tänkt för att möta universitetsstudenters behov och den har arbetats fram av forskare vid Göteborgs universitet ur en korpus på mer än 25 miljoner ord. AO presenterar vanligt förekommande icke-ämnesspecifika akademiska ord i frekvensordning. Därmed är AO också användbar som ett verktyg vid analyser av lexikala aspekter i inlärttexter på avancerad språklig nivå som till exempel Tisustexter.

AO är tillgänglig på *Språkbanken* [www] och innehåller 652 lexikala enheter vilka är utvalda ur en akademisk korpus i *Språkbanken*. Ord definieras i det här sammanhanget som lemmatiserade¹ lexikala enheter, och presenteras i AO med grundform samt böjningsformer. 42 % av orden i AO är substantiv, 26 % verb och 14 % adjektiv (Ribeck et al. 2014). Enström (2013:172) beskriver som sagt den typ av ämnesneutrala ord som AO innehåller som vanliga i alla typer av akademiska texter och därmed nödvändiga för inlärare på avancerad språknivå.

¹ Lemmatiserar: De böjda formerna förses med uppgift om grundformen och det är grundformens frekvens som beräknas.

En jämförelse mellan substantiv i undersökningens 32 texter och AO kan ge en uppfattning om Tisusskribenternas akademiska språkbruk och det kan också vara ett sätt att upptäcka skillnader mellan godkända och underkända texter, och därmed vad som kan tänkas påverka bedömningen. Med tanke på skrivuppgiftens karaktär och därmed förväntad stilnivå i texterna så borde dessa innehålla en hel del abstrakta ord, bland annat nominaliseringar. De sistnämnda gör texten mer koncentrerad och ger plats för mer tankeinhåll enligt många forskare, exempelvis Hultman & Westman (1977:110).

Hypotesen i min undersökning innebär att ordförrådet i de underkända texterna förväntas vara på en mindre avancerad nivå och till exempel innehålla färre ord som sammanfaller med orden i AO jämfört med de godkända texterna.

1.2.5. Substantiviska sammansättningar

Nordenfors (2011:105–111) visar i sin avhandling om svenska högstadielävers skriftspråksutveckling att orden blir längre och att antalet sammansättningar ökar ju äldre eleverna blir. Elevernas språkbruk förändras i samband med ett ökat behov av specifik terminologi, men ordlängden är också starkt relaterad till texttyp, där långorden är mer frekventa i faktatexter jämfört med i berättande texttyper.

I den här undersökningens hypotes ingår att även vuxna sva-inlärares användning av sammansättningar (en typ av långord) i den resonerande texttyp som Tisus skriftliga del innebär, förväntas öka i relation till att den skriftliga förmågan utvecklas.

Här följer en kort morfologisk/semantisk beskrivning av substantiviska sammansättningar som i huvudsak överensstämmer med Enström (2016:96–104). Därpå presenteras några exempel på problem för inlärares vad gäller reception, det vill säga förståelse för sammansättningens betydelse. Enligt Nation (2013:271) föregås produktiv ordkunskap av receptiv, det vill säga att för att kunna producera exempelvis sammansatta ord i en text måste inlärares först ha den receptiva kunskapen om ordet, mer om detta i 2.1 nedan och tabell 3.

Alla sammansatta ord består alltid av förled och efterled där efterleden är såväl det semantiska som det grammatiska huvudordet och alltså bestämmer vilken ordklass sammansättningen tillhör samt sammansättningens huvudbetydelse, till exempel *lärare + lön* → *lärarlön*. Det kallas determinativ sammansättning. Såväl förled som

efterled kan innehålla mer än ett rotmorfem: *gymnasium + lärarlön* → *gymnasielärarlön* respektive *lärarlön + lyft* → *lärarlönelyft*.

Sammansättningar kan vara lexikaliserade som exempelvis *soffbord* eller mer eller mindre tillfälliga som till exempel *uppsatsångest*. Substantiviska sammansättningar kan bildas med samtliga ordklasser som förled. I tabell 2 nedan ser vi exempel på förleder från sju olika ordklasser i substantiviska sammansättningar i ord som är excerperade ur mitt undersökningsmaterial.

Tabell 2. Exempel från Tisustexterna på substantiviska sammansättningar med förled från olika ordklasser²

| Förled | Substantivisk sammansättning |
|-------------|--------------------------------------|
| Substantiv | <i>valfrihet</i> (G3.2) ¹ |
| Adjektiv | <i>fritid</i> (G3.6) |
| Verb | <i>fikarast</i> (U2.33) |
| Adverb | <i>framgång</i> (U2.19) |
| Preposition | <i>mellanrum</i> (G5.84) |
| Räkneord | <i>4-åring</i> (U2.16) |
| Pronomen | <i>självförtroende</i> (G5.23) |

¹Inom parentes anges bedömning (G5, G3, U2) samt textens nummer i Tisuskorpusen.

²Jag har inte hittat några exempel i mitt material på konjunktioner eller interjektioner som förled. Enström (2016:97) har *som-bisats* respektive *nejröst* som exempel på dessa.

Förutom de substantiviska sammansättningarna som kan bildas med samtliga ordklasser som förled, och som utgör huvuddelen av svenskans sammansatta ord, finns också adjektiviska och verbala, men dessa ingår inte i den här undersökningen.

Enström (2016:112) konstaterar att sammansättning är en vanlig och produktiv ordbildningsprocess i svenskan och att det faktum att inte många av världens språk har lika rika möjligheter att bilda sammansättningar kan vara en orsak till att dessa innebär svårigheter för många inlärare.

Sammansatta substantiv kan ofta uttryckas med en nominalfras: *soffbord* → *bord vid soffan* och *uppsatsångest* → *ångest för uppsatsen*. Betydelsen förändras inte, men Reichenberg (2008:36) menar att nominalfrasen kan vara lättare för inläraren att förstå. Källström (2012:176) pekar å andra sidan på att sammansättningar ofta underlättar ordinlärningen eftersom inläraren utifrån de ingående orden kan dra slutsatser om sammansättningens betydelse.

Vidare påvisar Källström (2012:176) och Enström (2013:187) hur även de genomskinliga sammansättningarna, där alltså inläraren förstår betydelsen av de ingående delarna, kan vålla problem eftersom relationen mellan förled och efterled kan vara oklar. Som exempel på det är *lykta* huvudordet i både *ljuslykta* och *snölykta*, men *ljus* i *ljuslykta* innebär en lykta med ett ljus som är placerat inuti, medan *snö* i *snölykta* innebär att lyktan är gjord av materialet snö (och har ett ljus inuti). Till problematiken kommer dessutom att de genomskinliga sammansättningarna till skillnad från de ogenomskinliga ofta inte finns med i ordböcker och lexikon.

Som vi har sett av exemplet *ljuslykta/snölykta* räcker kunskap om de ingående ledens betydelse i ett sammansatt ord långt ifrån alltid till för att förstå sammansättningens betydelse. Detta gäller i högsta grad ogenomskinliga sammansättningar där delarna ger sken av att vara genomskinliga utan att vara det, exempelvis *nyckelpiga/nyckelhål* (Enström 2013:187), där ju varken förledet *nyckel* eller efterledet *piga* hjälper till med att förklara vad en *nyckelpiga* är.

1.2.6. Frekvensbaserade ordlistor

Med den digitala teknikens utveckling har det blivit enkelt att ta fram frekvensbaserade ordlistor ur olika textkorpusar. Varje frekvensbaserad ordlista som genereras digitalt avspeglar källan som den genererades från (Nation 2013:27). Det innebär exempelvis att en enskild forskare kan välja källa till de frekvensbaserade ordlistor hon eller han vill använda utifrån syftet med undersökningen.

Ordens frekvens är ofta ett huvudspår vid utforskning av det lexikala innehållet i inlärotext. Det är då underförstått att användning av hög- respektive lågfrekventa ord påverkar textens kvalitet, och följaktligen att inlärares användning av lågfrekventa ord signalerar en mer avancerad skriftlig färdighet.

2. Tidigare forskning

Här följer först en allmän bakgrund till forskning om inlärares lexikala språkutveckling och vad det innebär att kunna ett ord, och därpå något om korpuslingvistik. Vidare presenteras forskning om ordlängdens (bland annat sammansatta ords) betydelse för

textkvaliteten, och sedan en bakgrund till frekvensbaserade ordlistor. Och slutligen presenteras kortfattat något om forskning kring bedömning av texter.

2.1. Lexikal språkutveckling

Under 1960-talet började forskarna se andraspråket som ett system under utveckling och termerna *interimspråk* och *inlärarespråk* började användas. Från 1970-talet har utvecklingen av andraspråksforskning gått snabbt, och i Sverige var den ökade invandringen en bakomliggande orsak till det intresset för det svenska inlärarespråket.

Ett flertal studier visar att det tar fem till tio år för andraspråksinlärare att utveckla en sådan nivå på det akademiska språket, bland annat ordförrådet, att de klarar gymnasiestudier (Lim Falk & Holmberg 2016:166). Den tidsåtgången kan relateras till att godkänt på Tisus innebär behörighet till eftergymnasiala studier.

Pienemann & Håkansson (1999) går igenom ett stort antal svenska studier av utvecklingen av morfologi och syntax i avsikt att söka empiriskt stöd för processbarhetsteorin. Och Heinonen Eklund (2009) använder alltså, som nämndes i inledningen, processbarhetsteorin (Pienemann, 1998) i den grammatiska analysen av Tisus muntliga del och finner en samstämmighet. Men processbarhetsteorin har också begränsningar bland annat därför att den inte lägger vikt vid den lexikala utvecklingen.

Nation (2013), Milton (2009) och Henriksen (1999) är tongivande forskare på den internationella arenan när det gäller L2-inlärares lexikala utveckling. En av de definitioner av det komplexa begreppet lexikalisk kompetens som det ofta refereras till inom forskningen är Nations (2013:49), där han i sin modell i tabell 3 summerar vad det innebär att kunna ett ord, och delar in kunskapen om ett ord i tre aspekter: kunskap om form, betydelse samt användning. Varje aspekt i Nations modell innehåller receptiv (R) respektive produktiv (P) kunskap och är i sin tur uppdelad i underavdelningar.

Tabell 3. Olika aspekter av lexikonkunskap (efter Nation 2013:49)

| | | | |
|-----------------|---|---|--|
| 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 förmedla den avsedda betydelsen? |
| | begrepp och referenter | R | Vad ingår i begreppet? |
| | | P | Vad kan begreppet referera till? |
| | associationer | R | Vilka andra ord får detta ord oss att tänka på? |
| | | P | Vilka andra ord kan vi använda i st.f. det här? |
| Språkanvändning | grammatisk funktion | R | I vilka konstruktioner och mönster förekommer ordet? |
| | | P | I vilka konstruktioner används ordet? |
| | kollokationer | R | Vilka ord/typer av ord finns i kombination med det aktuella ordet? |
| | | P | Vilka ord/typer av ord används i kombination med det aktuella ordet? |
| | begränsningar användning (register, frekvens ...) | R | Var, när och hur ofta kan vi förvänta oss att möta ordet? |
| | | P | Var, när och hur ofta kan vi använda ordet? |

Enligt Nation (2013:271) finns det många forskningsresultat som visar att det är mer krävande för L2-inläraren att erövra den produktiva kunskapen om ett ord jämfört med den receptiva kunskapen, något som är relevant för min undersökning som ju fokuserar på den produktiva ordkunskapen. Nation menar att det framför allt är två faktorer som påverkar om ord används produktivt: kunskap och motivation. Dessutom anses högfrekventa ord vara lättare att erövra för inläraren än de lågfrekventa. Lågfrekventa ord är för övrigt den enskilt viktigaste faktorn när man ska förutse ords inläringssvårigheter anser Nation (2013:44 f.). Han presenterar i det sammanhanget begreppet *learning burden*, det vill säga att ju ovanligare ett ord är desto svårare är det att lära sig.

2.2. Korpuslingvistik

Korpus kommer från latinets *corpus* och används sedan 1960-talet i betydelsen ”en textmängd som är föremål för språkvetenskapligt studium” (SO 2009). Empiriska språkvetenskapliga undersökningar av inlärarespråk i korpussamlingar är sedan 1960-talet ett stort och aktivt forskningsområde och Sverige var tidigt ute sett ur ett internationellt perspektiv.

Ett exempel på en allmänt tillgänglig tidig svensk resurs är Cross-check-projektet som avslutades 2004 och som bland annat innehåller en annoterad autentisk inlärarespråk i form av sfi-korpusen (Eriksson & Lindberg, 2004 [www]).

SUF-korpusen (Källström & Lindberg 2011), som har sitt ursprung i det stora forskningsprojektet *Språk och språkbruk bland ungdomar i flerspråkiga storstadsmiljöer* vid Göteborgs, Lunds och Stockholms universitet, är en av de största svenska resurserna med en omfattande mängd inspelat tal, samt skrift i form av uppsatser från nationella prov i gymnasiet av 222 elever. Ett stort antal studier har använt sig av SUF-korpusen.

En annan svensk inlärarespråk, ASU, är longitudinell (Hammarberg 2013). Enligt Hammarberg används flertalet inlärarespråk internt av enskilda forskare eller av en forskningsgrupp och de blir därmed oftast inte allmänt tillgängliga. ASU-korpusen är dock sedan 2010 tillgänglig för forskning och utbildning, men den ligger inte online på det sätt som exempelvis korpussamlingarna i *Språkbanken* gör (Hammarberg 2013:29–30). Tisuskorpusen är en skyddad korpus som ingår i *Språkbankens* korpussamling, och studenter och forskare kan utan större problem få tillgång till den.

Möjligheterna till forskning på digitaliserade inlärarespråk, en del av korpuslingvistik, har påverkat språk teori och språkbeskrivning i hög grad trots att dessa endast har existerat sedan slutet av 1980-talet (Mackay & Gass 2012:7). Doró & Pietilä (2015:12) beskriver också hur andraspråkforskarens intresse för lexikala aspekter vid bedömning av L2-inlärares skriftliga färdigheter har ökat under de senaste tre decennierna eftersom många numera anser att ordförrådet är ett av de starkaste måtten på textkvalitet. Forskningen på olika lexikala aspekter, bland annat den produktiva sidan av ordkunskapen, har enligt Doró & Pietilä (2015:12) sin plats där språkundervisning, diskursanalys, korpuslingvistik och skrivpedagogik möts.

Tidigare korpusbaserade studier av lexikala aspekter i inlärares akademiska skriftspråk har exempelvis fokuserat på verb (Bloch, 2010) och kollokationer (Laufer & Waldman 2011). Vidare har Crossley & McNamara (2011) undersökt lexikaliska drag i fyra korpusar innehållande argumenterande texter skrivna på engelska av studenter med olika L1 och jämfört dessa med texter skrivna av studenter med engelska som L2. De kom fram till att L2-skriventernas texter i jämförelsen innehöll ett mindre sofistikerat ordförråd, till exempel använde L1-skriventerna fler ord med specifik betydelse.

Korpuslingvistik är det område inom vilket min undersökning kommer att röra sig, närmare bestämt lexikala särdrag som påverkar bedömningen av texterna i Tisus-korpusen.

2.3. Ordlängd och sammansatta ord

Svenska skolelevers produktiva ordkunskap i samband med uppsatsskrivande har exempelvis Grundin (1975), Hultman & Westman (1977) och Nordenfors (2011) studerat. Grundins (1975) och Nordenfors (2011) undersökningsresultat visar bland mycket annat att ordlängden ökar med elevernas ålder. Ordlängd uppfattas ofta som en svårighet också för andraspråksinlärare, men med stöd av olika forskningsresultat menar Laufer (1997:144) att många andra faktorer förutom ordlängd samverkar när man ska avgöra om ett ord är svårt eller lätt för inlärares, till exempel homonymi och polysemi.

Holmegaard (2007:135 f.) har undersökt flerspråkiga inlärares förståelse av långa ord, huvudsakligen substantiviska sammansättningar, som förekommer i SO-ämnets läromedel. Holmegaard drar slutsatsen att de långa orden kan hindra läsförståelsen av orsaker som att det handlar om informationstäta abstrakta substantiv som dessutom kräver sociokulturell förståelse. Att ett led i en sammansättning utgörs av ett polysemt ord, som ofta när det används separat har en annan betydelse, utgör enligt Holmegaard (2007:158) en komplikation för inlärares. Ett exempel hon nämner är *rörelse* som förekommer i hennes undersökningsmaterial i orden *fackföreningsrörelse* och *reformationsrörelse*.

Som redan nämnts så menar Nation (2013:51 f.) att den receptiva ordkunskapen kommer före den produktiva. Holmegaard (2007) kommer fram till att samman-

sättningar är ett problem för inlärare på den enligt Nation för inläraren lättare receptiva nivån, och bland annat mot den bakgrunden bedömer jag det som relevant att göra jämförelser mellan den produktiva kunskapen och användningen av substantiviska sammansättningar i studiens godkända och underkända Tisustexter.

Hultman & Westman (1977:116–7) jämför svenska gymnasieelevers användning av substantiviska sammansättningar med förekomsten av dessa i bruksprosa och kommer fram till att gymnasieeleverna ligger på en oväntat låg nivå vilket de antar beror på brister vad gäller förmåga att uttrycka sig med precisa termer. De menar också att en förklaring kan vara att eleverna inte bildar nya, tillfälliga sammansättningar. Av bland annat den anledningen kan man förvänta att andelen substantiviska sammansättningar i den här undersökningens underkända texter är lägre i jämförelse med de godkända.

Felaktiga särkrivningar av sammansatta ord är vanligt förekommande bland L1-skribenter och det finns enligt Hallencreutz (2002:101–2) många olika orsaker till det. Eftersom särkrivning kan tänkas ha en koppling till den lexikala utvecklingen av inlärarspråket jämför jag förekomsten av särkrivningar i de substantiviska sammansättningarna i de godkända respektive underkända Tisustexterna. För sva-inlärare kan åtminstone en ytterligare tänkbar orsak till felaktiga särkrivningar läggas till, nämligen att de oftare än L1-skribenter inte kan ta hjälp av betoning/intonation för att utröna om ordet ska skrivas ihop.

2.4. Frekvensbaserade ordlistor

Frekvensbaserade ordlistor, och ordtest som är baserade på frekvens, har en lång tradition inom EFL (English as a Foreign Language). I Sverige var vi tidigt ute med frekvensbaserade ordlistor. 1970-1980 publicerades Sture Alléns *Nusvensk frekvensordbok baserad på tidningstext* i fyra volymer, och 1972 kom hans frekvensbaserade ordlista *Tiotusen i topp*. Forskning om och framtagande av frekvensbaserade ordtest som motsvarar exempelvis de ofta i den engelskspråkiga världen använda *Vocabulary Levels Test* (Nation 2013:44 f.) har pågått en tid i Sverige men fortfarande har inget svenskt frekvensbaserat ordtest fått någon större spridning i vårt land (Bokander 2016:40). Däremot är det lätt att ta fram frekvensbaserade ordlistor ur olika korpussamlingar, en möjlighet som den här undersökningen tar vara på i jämförelserna mellan godkända och underkända texter.

När det gäller framtagning av akademiska frekvensbaserade ordlistor som innehåller allmänna icke-ämnesspecifika ord, så har det länge funnits sådana på engelska där de har använts flitigt på kurser i skriftlig framställning. Behovet uppstod som en följd av att forskare som till exempel Nation (2013) visade forskningsresultat som pekar mot att ordförrådet är den viktigaste enskilda framgångsfaktorn för skolelever. Enligt Nation (2013:19) är *The Academic Word List* (Coxhead 2000) den mest kända av dem som finns på engelska, men även *A New Academic Vocabulary List* (Gardner & Davies 2014) är värd att nämnas i sammanhanget.

Den frekvensbaserade *Svensk akademisk ordlista* (AO), som används i min undersökning, utvecklades relativt nyligen (Ribeck et al. 2014). Lim Falk & Holmberg (2016) använder AO i en jämförande studie mellan gymnasieelevers skrivande på svenska i gymnasieskolor där mer än hälften av undervisningen sker på engelska med elever i den reguljära gymnasieskolan. Där analyserade de täckningsgraden från AO i varje text, och för att stärka undersökningens validitet gjorde de också en kvalitativ undersökning av ett mindre antal texter för att se hur orden från AO fungerade i sammanhangen.

Inlärarens val och användning av ord kan naturligtvis bero på andra faktorer som påverkar den lexikala språkutvecklingen än ordens frekvens. Den inlärartyp som föredrar att hålla sig till redan kända ord, bland annat för att undvika fel, nämns av Jiménez Catalán & Fitzpatrick (2014:87). De pekar dessutom exempelvis på att språkundervisningen och ordförrådet i läroböcker och annat material kan styra ordinläringen.

Min undersökning avser inte att studera inlärartyper, och vad gäller material som styr lärandet kan en kanske något långsökt koppling göras till det faktablad (bilaga 3) som tillhör Tisus 2006 och som bland annat innehåller en ordlista som testtagaren får, men inte måste, använda. Det kan hända att testtagaren redan har produktiv kunskap om orden i listan, men det kan också hända att hon eller han enbart har receptiv kunskap om ett eller flera av orden när Tisus inleds, men att hon eller han påbörjar en process mot produktiv kunskap under skrivandet av Tisustestet. För att studera detta närmare behövs en annan typ av undersökning, till exempel tester på ordförrådets djup.

2.5. Bedömning av elevtexter

Bedömning av texter och textanalys är breda och komplexa forskningsområden och de exempel som presenteras nedan utgör bara en ytterst liten del av de undersökningar som gjorts.

Textlängd är ett kvantitativt mått som självklart inte isolerat kan säga något om textens kvalitet, men som har använts i en del studier på grundskole- och gymnasieelever med svenska som förstaspråk. Exempelvis Hultman & Westman (1977:53) och Nordenfors (2011:100) konstaterar att textens omfång har ett direkt samband med bedömningen på så sätt att en längre text får ett högre betyg. Hultman & Westmans (1977:185) resultat visar också att IG²-texterna har den största spridningen i längd. Texternas omfång mäts också i min undersökning och även där kan en relation mellan omfång och bedömning ses, och även här verkar spridningen vara störst bland de underkända texterna (bilaga 7).

Både Hultman & Westman (1977:185) och Östlund-Stjärnegårdh (2002:52 f.) har genom lärarenkäter undersökt vad som skiljer bedömningen av godkända respektive inte godkända gymnasieelevers texter. Där framkommer att helhetsbedömning av texten är ett av de viktigaste kriterierna. I Tisus interna bedömningsblad (bilaga 1) lägger man också vikt vid den holistiska bedömningen.

Hultman & Westman (1977:185 f.) undersöker ordklassfördelningen i 1970 års standardprov för årskurs 3 på gymnasiet och konstaterar en relation mellan betyg och nominal stil på så sätt att texter som får högre betyg har en högre nominalkvot. En liknande studie har gjorts av Johansson Kokkinakis & Magnusson (2011:120) på gymnasieelever med svenska som andraspråk och där var medelvärdet för nominalkvoten något lägre jämfört med Hultman & Westman.

Vidare vad gäller det lexikala perspektivet vid bedömning, och mer specifikt vilka lexikala särdrag som skiljer godkända texter från underkända, så har Magnusson (2011) i sin avhandling undersökt elevtexter i årskurs 9 och på gymnasiet. Hon har då studerat en typ av långa ord, grammatiska metaforer (GM, en term från systemisk-funktionell grammatik) och deras andel i förhållande till textlängden (Magnusson 2011:105 f.). Resultatet visar ett signifikant samband mellan andel grammatiska

² IG = inte godkänd

metaforer och betyg, och den vanligaste kategorin inom GM är nominalisering. Resultatet visar dessutom att enspråkiga elevers texter var mer GM-täta än flerspråkiga elevers. Magnusson genomför också en kvalitativ undersökning av den funktion som GM har i ett urval elevtexter (Magnusson 2011:231).

I tabell 12 redovisas de nominaliseringar, avledda substantiv, i min undersökning som återfinns i AO oavsett typ av nominalisering. En mer ingående studie av GM, nominalisering, funktion i sammanhanget etc. ryms inte inom ramen för den här undersökningen.

3. Material

Undersökningsmaterialet är hämtat från Tisuskorpusen (2006-05-01–2006-05-31) som är en specialiserad synkron andraspråkskorpus som innehåller autentiska texter från Tisus, ett test som ger behörighet för högskolestudier. Provorten för samtliga texter i Tisuskorpusen är Stockholm, temat för Tisus i maj 2006 är ”Stress” och texttypen är resonerande. Korpusen består av 105 texter, 3 484 meningar och 59 639 tokens (med tokens menas det totala antalet ord, löpord), och den ingår i *Språkbankens* skyddade korpussamlingar. Tisuskorpusen är inte öppen för allmänheten utan det krävs tillstånd för att få tillgång till den och då är den sökbar i Korp-verktyget i *Språkbanken* (Göteborgs universitet [www]).

Av Tisus-korpusens totalt 59 639 tokens är cirka 19 % substantiv. En hög andel substantiv i en text är som nämnts ovan i 1.2.1 utmärkande för koncentrerat skriftspråk. Det finns läromedelsstudier, exempelvis en tidig kvantitativ studie av huvudsakligen samhällsvetenskapliga läromedelstexter gjord av Margareta Westman (1974), som visar att substantiv och adjektiv dominerar i de i studien ingående läromedlen. Substantiv är en öppen ordklass, och den ordklass som innehåller mest information per ord (Lagerholm 2008:108). Därmed utgör substantiven en central del av L2-inlärarens ordförråd. I det sammanhanget är undersökningens fokus på substantiv relevant.

Informationen på Stockholms universitets hemsida om Tisus, avsedd för eventuella framtida testdeltagare, innehåller några exempelprov, svar på vanliga frågor samt följande:

Tisus är ett färdighetstest, inget kunskapstest. Det förekommer alltså inga krav på att du ska ha läst viss litteratur eller ha genomgått vissa kurser. Däremot måste språkfärdigheterna i svenska språket ligga på en tillräckligt hög nivå för att klara akademiska studier på svenska. (Tisus, Stockholms universitet, [www])

Beprövad erfarenhet pekar mot att den del av Tisus som i de flesta fall innebär den största språkliga utmaningen för testtagarna är den som testar den skriftliga färdigheten. Att använda texterna i Tisuskorpusen för att undersöka ordförrådet, närmare bestämt den produktiva kunskapen och användningen av substantiv i avancerade sva-inlärares skrivna texter, innebär att man har ett material där inlärnarna tar sig an språkligt och kognitivt komplexa uppgifter.

Tabell 4 visar fördelningen av korpusens texter utifrån färdighetsnivå, dels i antal och dels i procent:

Tabell 4. 105 texter i Tisuskorpusen - fördelning utifrån färdighetsnivå

| Färdighetsnivå | Antal texter | % |
|-----------------------|---------------------|----------|
| U2 | 31 | 29,52% |
| G3 | 39 | 37,14 % |
| G4 | 23 | 21,90 % |
| G5 | 12 | 11,43 % |
| Totalt antal | 105 | 100,00 % |

Den bedömning som testtagaren meddelas är betygen G (godkänd) respektive U (underkänd). Men dessutom finns en intern holistisk bedömning, en skala som anger en stigande färdighet från 1 till 5: 1-2 = U och 3-5 = G (se bilaga 1). Den skalan används i intern statistik, vid sambedömning och diskussioner om gränsfall, till exempel där man tar in en tredjebedömare. För betyget G ska kriterierna för textens struktur och sammanhang, ordförråd och grammatik vara uppfyllda på ett tillfredställande sätt. Det är alltid två bedömare som granskar Tisustexterna var för sig

och sätter betyg. Betygen sammanvägs sedan och i tveksamma fall kallas tredjebedömaren in.

Den holistiska bedömningen finns förutom G och U-betygen annoterad i Tisuskorpusen. År 2006 då texterna i Tisuskorpusen skrevs användes en annan bedömningsskala, 1–15, (bilaga 2), men i Tisuskorpusen är den gamla bedömnings-skalan omvandlad till den skala som används idag. Detta för att underlätta jämförelser med Tisustexter som skrivs i nutid, något som inte är aktuellt i den här undersökningen.

I ett Excelark som tillhör dokumentationen kring Tisuskorpusen finns korpusens alla texter listade i nummerordning med olika kolumner som bland annat anger resultat på testets olika delar, antal år i Sverige och modersmål. Uppgifterna i Excelarket har använts i tabell 4 som visar Tisuskorpusens texter fördelade utifrån färdighetsnivå.

Nedan i tabell 5 som förutom informationen från tabell 4 innehåller en separat redovisning av skribenter som angivit svenska som enda, eller ett av flera, modersmål, ser vi att medelvärdena för de som har angivit svenska som modersmål avviker en hel del. Skillnaden är exempelvis 22,1 procent för G5 när vi jämför de 13 texterna med svenska som modersmål (30,8 %) med de 92 övriga (8,7 %).

Tabell 5. 105 texter i Tisuskorpusen - fördelning utifrån färdighetsnivå samt redovisade med och utan skribenter som angivit svenska som modersmål

| Färdighetsnivå | Alla texter | L1/sv | exkl. L1/sv |
|-----------------------|--------------------|--------------|--------------------|
| U2 | 31 / 29,52 % | 2 / 15,4 % | 29 / 31,6 % |
| G3 | 39 / 37,14 % | 4 / 30,8 % | 35 / 38 % |
| G4 | 23 / 21,90 % | 3 / 23 % | 20 / 21,7 % |
| G5 | 12 / 11,43 % | 4 / 30,8 % | 8 / 8,7 % |
| Totalt antal | 105 | 13 | 92 |

De tretton texter i tabell 6 där svenska uppgivits som enda modersmål eller som ett av flera modersmål, ingår inte i undersökningen. Det beror dels, som nämnts ovan, på att fördelningen mellan färdighetsnivåerna som vi sett avviker väsentligt från medelvärd-

dena i korpusen, speciellt vad gäller U2 och G5, och dels därför att den här undersökningen har fokus på sva-inlärare.

I tabell 6 presenteras de utgallrade 13 uppsatserna och där anges Tisuskorpusens alla variabler bland annat för att ge en tydlig bild av vilken information som finns att tillgå i korpusen.

Tabell 6. Texter i Tisuskorpusen – skribenter som angivit svenska som modersmål, alla variabler

| | | | | | | | | | | | | | |
|---------------|----------|------|-------|-------|-------|-------|-------|-------|-------|------|------|-------|------|
| Ålder | 17 | 18 | 18 | 18 | 20 | 20 | 21 | 21 | 22 | 22 | 23 | 23 | 31 |
| Tid i | | | | | | | | | | | | | |
| Sv. | 3 år | 4 år | 13 år | 6 mån | 10 år | 11 år | 5 mån | 1 år | 15 år | 22år | 7 år | 13 år | 5 år |
| Bedöm. | U2 | G4 | G5 | U2 | G5 | G4 | G3 | G3 | G5 | G5 | G4 | G3 | G3 |
| L1 | sp/en/sv | sv | sv | sv/en | sv | sv | sv/ty | sv/ty | sv | sv | sv | sv/it | sv |
| Kön | M | K | K | K | K | K | M | K | K | M | M | K | M |

Några närmare upplysningar om de som uppger svenska som sitt modersmål, eller som ett av sina modersmål i Tisuskorpusen, vars provort är Stockholm, finns inte förutom det som framgår av tabell 6. Vi vet till exempel inte om testtagaren är utlandssvensk på tillfälligt besök i Stockholm, eller permanentboende i Sverige. Det kan också finnas testtagare som har en IB-examen från en svensk gymnasieskola, men som saknar behörighet i svenska för högre studier och därför väljer att göra Tisus (Antagningsservice [www]).

Tisus omfattas av sekretess och jag får därför inte publicera texterna i sin helhet i den här uppsatsen. Författarna är anonyma i Tisuskorpusen, men det finns information om tid i Sverige, skolbakgrund, förstaspråk, kön och betyg på Tisus. Jämförelser mellan texterna utifrån andra inlärarvariabler än bedömningen skulle alltså vara möjliga att genomföra men faller utanför den här undersökningens ram.

I Tisus provinstruktion för den skriftliga delen i maj 2006 framgår att texten ska bestå till cirka en tredjedel av en inledande faktadel och till två tredjedelar av resonerande text. Där uppmanas också testtagaren att använda några av de fakta som finns på faktabladet och som kan vara en relevant bakgrund till det han eller hon tänker diskutera i texten. Faktabladet (bilaga 3) innehåller bland annat en lista med nyckelord ”att inspireras av” varav de flesta är substantiv och av dem är många sammansatta ord.

4. Metod

Metodkapitlet inleds med en allmän presentation av undersökningens val av metod och fokus samt kartläggningen och jämförelserna. Därpå presenteras de i undersökningen ingående 32 texterna, och vidare något om hur excerperingen av de 32 texternas substantiv har gått till. Efter det följer fyra avsnitt om metod vid undersökning av nominalkvot, av substantiv som återfinns i AO, av substantiviska sammansättningar samt av substantiv som återfinns i frekvensbaserade ordlistor.

Korpuslingvistiska metoder är i huvudsak kvantitativa eftersom storleken på materialet gör exempelvis frekvensstudier möjliga och dessutom kan resultaten ofta grundas på god representativitet (Macay & Gass 2012:9). Den här undersökningen av Tisuskorpusen utgår ifrån tidigare forskning och är kvantitativ i och med att den söker svar på forskningsfrågorna genom kartläggning och kvantitativ analys av användningen av substantiv i 32 inlärtartexter i en jämförelse mellan godkända och underkända texter.

I korpusen är skrivuppgiften en och samma för alla testtagare och texterna är skrivna vid samma tillfälle vilket är ett bra utgångsläge för jämförelser mellan godkända och underkända texter. I undersökningen studeras nominalkvoten och sammansatta ord kontrastivt. Där ingår också jämförelser med de frekvensbaserade ordlistorna 1-5 000 från en bloggkorpus och med AO. I AO undersöks en mer specifik kategori, som Enström (2013:172) kallar ämnesneutrala ”akademiska” ord.

Temat för Tisustexten 2006 är ”Stress”, och därmed ett exempel på de allmänna ämnen med anknytning till samhällsvetenskap som en Tisus-testtagare förväntas kunna resonera kring utan att ha tagit del av en specifik kurs eller kurslitteratur. Situationsvariabeln i studien är Tisus, maj 2006, och endast en av de inlärtarvariabler som redovisas i Tisuskorpusen, nämligen färdighetsnivå (G3-5/U2), är aktuell, och faktorer som ålder, kön, L1, tid i Sverige, eller om provtagaren är bosatt utanför Sverige etc. ingår inte i studien.

Vidare har de provtagare som anger svenska som modersmål³ uteslutits ur undersökningen eftersom undersökningens fokus ligger på sva-inlärtare. Andra

³ *Modersmål* är den term som används i Tisuskorpusen. I uppsatsen förekommer L1 och förstaspråk som synonymer.

inlärarvariabler förutom färdighetsnivån kan säkerligen ha betydelse för undersökningen, kanske framför allt tid i Sverige, men avgränsningen är nödvändig med tanke på den här undersökningens begränsade omfattning.

4.1. Kartläggning och jämförelser

Alla substantiv, oavsett om de förekommer i grundform eller i olika böjningsformer, excerpas från de tre ur Tisuskorpusen valda grupperna av inlärartexter, det vill säga de som har fått betyg G5, G3 respektive U (tabell 7). Textutdrag från Tisuskorpusen används för att illustrera till exempel problem vid excerperingen och olika ords funktion i sammanhanget.

Kartläggningen och jämförelserna innebär en kontrastiv, strikt synkron analys av substantiven i interimsspråket. Diakroniska aspekter som till exempel lexikal utveckling ingår alltså inte.

Contrastive Interlanguage Analysis (CIA) är sedan slutet av 1980-talet en ofta använd metod inom forskning på inlärarkorpusar och den har enligt Mackay & Gass (2012:18) kritiserats då man gjort jämförelser mellan inlärarspråk och målspråksnormen, eftersom man då inte undersöker interimsspråket i dess egen rätt. Det är inget problem i den här studien när jämförelser görs inom inlärarkorpusen, det vill säga mellan texter som fått G och U i betyg. Möjligen kan det vara ett problem då man analyserar approximationer (= nästan korrekta substantiv för ett visst innehåll) eftersom det då är målspråket som är normen. Men här kan man se jämförelsen med målspråket som ett nödvändigt metodiskt verktyg för att komma åt den produktiva kunskapen om substantivens betydelse och användning i inlärarspråket i Tisuskorpusens texter.

Substantiven i de 32 Tisustexterna sammanställs, kvantifieras och jämförs ur olika perspektiv utifrån en huvudsakligen ”hypothesis-driven approach” – vilket innebär en explorativ metod för att hitta mönster (Mackay & Gass 2012:13). Det innebär i praktiken att man exempelvis till att börja med undersöker hur stor andelen substantiv i de i undersökningen ingående texterna är, och nominalkvoten beräknas samt analyseras kontrastivt. Vidare undersöks och jämförs substantiven i Tisustexterna med frekvensbaserade ordlistor och AO, och andelen sammansatta substantiv jämförs mellan G3, G5 och U2-texterna, för att senare presenteras text, tabeller och bilagor.

Kartläggningen av substantiven och sökandet efter mönster i jämförelsen mellan godkända och underkända texter i Tisuskorpusen har skett så förutsättningslöst som möjligt.

4.2. Undersökningens 32 Tisustexter

Bland de 92 texterna som återstår i Tisuskorpusen när de som har uppgivit svenska som modersmål har tagits bort har 16 godkända och 16 underkända texter valts ut för undersökningen (se tabell 7). Antalet beror på att syftet med undersökningen är att jämföra godkända och underkända texter skrivna av sva-inlärare. Det finns totalt 8 texter som har bedömts som G5 och med utgångspunkt från det valdes de första åtta texterna i korpusen bedömda som G3 samt de första sexton texterna bedömda som U2. Det betyder att jämförelser också är möjliga att göra mellan G5, den högsta färdighetsnivån, och G3 förutom att jämföra de godkända texterna med de underkända.

Tabell 7. Undersökningens 32 Tisustexter – fördelning antal tokens¹ och types²/färdighetsnivå, medelvärden antal tokens/types, samt TTR³

| Färdighetsnivå | Antal | Antal ord tokens | Antal ord types | Medelvärde / text tokens/types | TTR |
|----------------|-------|------------------|-----------------|--------------------------------|------|
| U2 | 16 | 7691 | 1542 | 481/96 | 0.20 |
| G3+G5 | 16 | 8891 | 2532 | 556/158 | 0,28 |
| G3 | 8 | 4429 | 1221 | 554/153 | 0,28 |
| G5 | 8 | 4462 | 1311 | 558/164 | 0,29 |

¹ token = totala antalet ord ² type = unika ord ³type/token ratio = procentuell andel unika ord

Av tabell 7 framgår att medelvärdet för de godkända texternas omfång är större än för de underkända, 556 jämfört med 481, medan det mellan G5 och G3 bara skiljer fyra tokens. Att textlängden har en relation till bedömningen stämmer med tidigare forskning, se 2.5, med det är inget som den här studien fördjupar sig i. Vidare är type/token ratio (TTR) ett omdiskuterat mått på lexikal variation, och Hultman & Westman

(1977:48) betonar att man inte kan ha det som det enda måttet i en undersökning av ordvariation bland annat eftersom ord repeteras mer i en längre text vilket då leder till ett lägre TTR. Numera använder man därför ordvariationsindex (OVIX) som också väger in textens längd i beräkningen. Men trots att G3/G5-texterna är längre än U2-texterna, och alltså borde innehålla fler upprepningar och därmed få ett lägre TTR, så är TTR betydligt högre i de godkända (0,28) jämfört med de underkända (0,20), vilket alltså tyder på högre grad av lexikal variation i de godkända texterna.

4.3. Substantiven i de 32 Tisustexterna

För att få fram de i undersökningen ingående Tisustexternas substantiv har en textfil med respektive text använts samt konkordansprogrammet *AntConc* 3.4 (Anthony 2014) och *Språkbankens* verktyg *Korp* (*Språkbanken*, Göteborgs universitet [www]). Ur de frekvensbaserade ordlistorna som man får fram på det här sättet har sedan substantiven extraherats och listats, dels som tokens och dels som types.

En komplikation har varit att en del av orden i de frekvensbaserade ordlistorna kan tillhöra olika ordklasser beroende på funktionen i sammanhanget. Därför har jag i en hel del fall kontrollerat orden i sitt sammanhang i korpusen. Dessutom kan till exempel formfel leda till att ord sorteras fel. Ett exempel på det sistnämnda är att jag från den frekvensbaserade ordlistan till text G5.102 först excerperade *orsaker* som ett substantiv i plural, men senare visade det sig, se nedan, att det fungerar som ett verb i presens i textsammanhanget och *orsaker* fick därför strykas från substantivlistan för text G5.102.

Å andra sidan kan det också vara bristen på ett meningsfullt arbete som **orsaker** stress. G5.102

En annan komplikation vid framtagning av ordlistor i *AntConc* 3.4 (Anthony 2014) är att vid den digitala överföringen av testtagarnas handskrivna texter så markerades skribenternas avstavningar med (-) oberoende av var på raden texten hamnade. Det innebär att avstavade ord och eventuella sammansättningar med bindestreck inte kan särskiljas. I exemplet nedan betyder det att *AntConc* 3.4 (Anthony 2014) uppfattar ordet *verk* som ett substantiv när det egentligen ska vara *verkligheten*. Jag har alltså

behövt undersöka texterna närmare för att i största möjliga mån undvika felaktigheter i ordlistorna som tagits fram.

Enligt min åsikt berättar resultaten de olika forskarna när en hel del om forskarna själva och deras syn på **verk-ligheten**. G3.10

De flesta grammatikor räknar egennamn som en undergrupp till substantiv enligt Josefsson (2005:30 f.). *Svenska Akademiens Grammatik* (SAG1999) skiljer dock ut egennamnen som en egen ordklass. I SAG, del 1:165, lyder definitionen: ”Det typiska egennamnet saknar betydelse och pekar ut en (i sammanhanget) unik referent.” Jag har valt att inte ta med några egennamn i min undersökning bland annat av den anledningen att många av de egennamn som förekommer i Tisustexterna också finns med i faktabladet (bilaga 3), och att det då inte går att avgöra om användningen av dessa tillhör testtagarens produktiva ordförråd.

För jämförelser med de frekvensbaserade ordlistorna, AO och listan med substantiv som återfinns i faktabladet för Tisus, maj 2006, har *AntWordProfiler* (Anthony 2013) använts. Och för att få en uppfattning om användningen av substantiven i det i materialdelen beskrivna faktabladet i bilaga 3, görs en jämförelse nedan i tabell 8 mellan totalt antal types i texterna respektive antal types som återfinns i faktabladet.

Tabell 8. Antal substantiv, types i Tisustexterna + antal i faktabladet

| Medel- värden | Totalt antal subst. types | % subst.av antal types i texterna | varav i fakta- bladet, types | % i faktabl. av alla subst. |
|--------------------------|--------------------------------------|--|---|--|
| G5 | 388 | 29,6 % | 14 | 3,6 % |
| G3 | 324 | 26,5 % | 16 | 4,9 % |
| U2 | 401 | 26,0 % | 18 | 4,5 % |

I tabell 8 ser vi att G5 har den högsta andelen substantiv (29,6 %) i texterna vid en jämförelse av types, 3,1 % högre än G3 och 3,6 % högre än U2 (26 %). En möjlig tolkning av jämförelsen vad gäller substantiv från faktabladet skulle kunna vara att G5-skribenterna, som har lägst andel, 3,6 %, är mer säkra och att de därför lutar sig

mindre mot det stöd de får i faktabladet än skribenterna av U2/G3. Men för att säkerställa den tolkningen behövs en närmare studie där man exempelvis tittar närmare på tokens, studerar enskilda texter samt undersöker hur substantiven fungerar i sammanhanget i respektive text, något som faller utanför ramen för den här undersökningen.

De substantiv som återfinns i faktabladet sammanfaller inte med orden i AO. Däremot är många av orden i faktabladets ordlista substantiverade sammansättningar, och dessa har uteslutits i statistiken över texternas substantiviska sammansättningar (bilaga 7) eftersom det faktum att de finns i stödmaterialet i bilaga 3 innebär att man inte kan utgå ifrån att de ingår i skribentens produktiva ordförråd.

4.4. Nominalkvot

Tisustexterna studeras kontrastivt genom beräkning av den enkla varianten av nominalkvot. Uträkningen av den enkla nominalkvoten i de 32 texterna innebär antalet substantiv dividerat med antalet verb (Lagerholm 2008:132). Verben i respektive text har extraherats från frekvensbaserade ordlistor i *AntConc* 3.4 (Anthony 2014) på liknande sätt som i 1.5.3. redovisas för substantiven. Man kan förvänta sig att finna en högre nominalkvot i de godkända texterna, se 1.2.1.

4.5. Svensk akademisk ordlista (AO)

Den frekvensbaserade *Svensk akademisk ordlista* (AO) är som tidigare nämnts tillgänglig i *Språkbanken* (*Språkbanken*, Göteborgs universitet [www]). Valet har fallit på att närmare undersöka de substantiv i texterna som sammanfaller med AO eftersom dessa förväntas uppvisa skillnader i godkända respektive underkända texter.

I bilaga 5 redovisas listorna på substantiv från AO som återfinns i undersökningens 32 Tisustexter och som har tagits fram med *AntWordProfiler* (Anthony 2013). Dessutom har en genomgång gjorts av texterna bland annat med hjälp av sökfunktionen i Word. Detta för att kontrollera upprepningar av substantivet i texten (tokens), om särkrivningar förekommer (vilket gör att ord tappas bort eller felaktigt läggs till), samt om substantiven används på ett för läsaren begripligt sätt i sammanhanget.

Bland de justeringar av ordlistan som detta medförde kan nämnas att substantiv som hittades av *AntWordProfiler* men som visade sig vara felaktigt särskrivna för- eller efterled i ett sammansatt ord togs bort. Ett exempel på detta är det särskrivna *prestations krav* (U2.49) där *krav* finns med i AO, men inte sammansättningen *prestationskrav*. Antal substantiv från AO som återfinns i var och en av de 32 texterna redovisas sedan som types i bilaga 5.

AO innehåller till övervägande del abstrakta substantiv som alltså ligger till höger på skalan i den andra lexikala dimensionen, *specifik/allmän*, i 1.2.2, tabell 1 (Hellspång & Ledin 1997:78 f.). En genomgång av listorna i bilaga 5 gjordes för att undersöka om det förekom några substantiv i konkret betydelse eftersom detta säger något om textens stilnivå. Det finns givetvis ingen exakt gräns mellan abstrakt och konkret och många substantiv kan användas i såväl abstrakt som konkret betydelse (Hultman 2008:45). Det gör att man i en del fall behöver gå in i texten och titta närmare på sammanhanget för att kunna avgöra om ordet i sin kontext fungerar i konkret eller abstrakt betydelse. Ett exempel från undersökningen är det polysema substantivet *gräns* som i Tisustexterna enbart återfinns i abstrakt betydelse, till exempel:

Gränserna mellan arbete och fritid suddas ut. (G5.43)

Avledningar är enligt Källström (2012:173) en princip för ordbildning som antagligen finns i de allra flesta av världens språk, och i svenskan finns såväl ett antal produktiva avledningssuffix som lexikaliserade avledningar.

En analys av alla avledda substantiv i de 32 Tisustexterna är inte möjlig inom ramen för den här studien trots att det skulle vara av intresse med tanke på syftet. Därför begränsade jag mig till att redovisa de suffixavledda substantiven i tabell 12 och kvantitativt jämföra godkända respektive underkända texter vad gäller det totala antalet nominaliseringar (types) som återfinns bland AO-orden (bilaga 5) utan att närmare undersöka typ av nominalisering, funktion i respektive text, upprepningar i samma text och så vidare.

4.5.1. Fem flerordsenheter

Kvalitativa studier är ofta induktiva och genomförs förutsättningslöst, alltså utan att först presentera en hypotes eller en teori (Mackay & Gass 2012:181). Den här kvalitativa delundersökningen kom till då det under sammanställningen av ordlistorna upptäcktes att några av substantiven i Tisustexterna och som också återfinns i AO, är enstaviga rotmorfem som i sitt sammanhang i texterna ibland ingår i flerordsenheter som exempelvis *roll* i *spelar roll*. Med *rotmorfem* avses här Josefssons (2005:173) definition: ”morfem som uppbär betydelse och som inte måste haka på något annat morfem eller ord”. Och *flerordsenheter* är den överordnade term som Prentice & Sköldberg (2013:197) väljer när de behandlar olika typer av flerordskombinationer, där fokus inte ligger på betydelsen av enskilda ord utan på hela ordkombinationen.

Det förekommer alltså i några fall att ord som vid en första anblick på listorna i bilaga 5 ser ut som rotmorfem med egen betydelse förekommer i vissa texter i flerordskombinationer där de ingående orden tillsammans har en specifik betydelse. Mot den bakgrunden valde jag att gå in i texterna och titta närmare på sammanhanget där substantiven *fall*, *form*, *roll* och *verk* förekommer eftersom jag i några fall hade observerat dem i de konventionaliserade uttrycken *i alla fall*, *i så fall*, *i form av*, *spelar roll* och *i själva verket*.

Problem kring definitionen av konventionalisering diskuterar Prentice (2010:28) i sin avhandling där hon formulerar språkliga konventioner som: ”etablerade mönster i språkbruket gentemot nyskapade eller tillfälliga sätt att uttrycka ett visst innehåll”. Vid en kontroll i *Svenskt språkbruk* (2003) av de fem flerordskombinationer som undersöks återfinns alla där, något som förstärker uppfattningen att det handlar om konventionaliserade uttryck.

Enligt Prentice & Sköldberg (2013:215-6) kan även vanligt förekommande flerordsenheter som inläraren enkelt förstår i ett sammanhang innebära svårigheter att producera. Därför kan det vara intressant att se om några skillnader mellan godkända och underkända texter kan upptäckas när det gäller de ovan nämnda fem ordkombinationernas korrekta eller nästan korrekta användning i sitt sammanhang.

4.6. Substantiviska sammansättningar

Ordlistor innehållande alla substantiviska sammansättningar i de 32 texterna har först framtagits med hjälp av *AntConc* 3.4 (Anthony 2014). Men då det efterhand blev tydligt att alla inte gick att få fram den vägen har sammansättningarna markerats i respektive text och sedan listats. Det som komplicerade excerperingen av sammansättningarna var att de inte alltid var hopskrivna. *AntConc* redovisade då de sammansatta orden som två och i något fall även som tre olika ord. Dessutom hade, som tidigare nämnts i 5.3, testtagarnas avstavningar markerats med (-) under digitaliseringen av de handskrivna texterna, oberoende av var på raden de hamnade något som *AntConc* 3.4 (Anthony 2014) inte klarade av att analysera då det rörde sig om sammansättningar.

I följande exempel vet vi alltså inte om sammansättningen har avstavats och alltså är en korrekt sammanskrivning, eller om den har skrivits med bindestreck inne i den löpande texten:

... det andra förstör ens koncentration, **själv-förtroende** och gör en så orolig och ibland fysisk sjuk att man kan inte göra sitt jobb. U2-31

I min undersökning skiljer jag inte mellan lexikaliserade och tillfälliga sammansättningar. Jag räknar även med felaktigt särskrivna respektive felaktigt hopskrivna sammansättningar och sammansättningar som innehåller felaktiga bindestreck under förutsättning att de fungerar i sitt sammanhang och att texten är begriplig för läsaren.

Förutom att jämföra andelen substantiviska sammansättningar i godkända respektive underkända texter (tabell 14) har sammansättningar med tre och fyra led samt med dubbel syftning (till exempel *flex- eller övertid*) plockats ut ur listorna och jämförts (tabell 16). Dessutom har antalet stavelser i sammansättningarna räknats (tabell 15). Detta för att få grepp om huruvida det är möjligt att urskilja skillnader mellan godkända och underkända texter vad gäller de substantiviska sammansättningarnas komplexitet och längd. Vidare redovisas några exempel på hur de längre substantiviska sammansättningarna fungerar i sitt sammanhang i texterna. Det sistnämnda i syfte att illustrera eventuella skillnader i stilnivå mellan godkända och underkända uppsatser.

4.7. Frekvensbaserade ordlistor

De frekvensbaserade ordlistorna, 1000-5000, som används i studien har tillhandahållits av min handledare Sofie Johansson som är verksam vid Institutionen för svenska språket vid Göteborgs universitet. De består av graford⁴ och är uppdelade på 1000-frekvensnivåer. Listorna representerar 5 000 av de vanligaste graforden i en bloggkorpus från 2012. Bloggkorporus har valts eftersom bloggtexter kan anses vara en blandning mellan skriftspråk och talspråksliknande text. Eftersom testtagarna för Tisus förväntas skriva på en formell stilnivå kan det tänkas att orden i undersökningens 16 underkända texter representerar en lägre stilnivå och därmed i högre grad samstämmer med de frekvensbaserade ordlistorna, åtminstone på de lägre frekvensnivåerna.

En undersökning av de substantiv som finns i de 32 Tisustexterna men som inte sammanfaller med de frekvensbaserade ordlistorna skulle vara intressant bland annat med tanke på att en högre grad av lexikal variation kan förväntas i de godkända texterna. Fördjupning i lexikal variation faller dock till största delen utanför ramen för den här undersökningen.

4.8. Metoddiskussion

En utgångspunkt för undersökningen är att digitaliseringen inklusive annoteringen etc. av Tisuskorporus har skett på ett reliabelt sätt. Inom ramen för den här studien finns exempelvis ingen möjlighet att kontrollera kodningen av materialet genom att jämföra två olika personers kodning på det sätt som Mackay & Gass (2012:147) beskriver.

En annan faktor som vi inte kommer åt i analysen är om skribenterna i testsituationen verkligen ger en rättvis bild av sitt ordförråd. Exempelvis kan en och samma skribents prestation variera från dag till dag. Skrivprocessen är dessutom komplex och vi har ingen information om huruvida olika val av substantiv beror på att skribenten har en inläraresprofil som undviker det som känns osäkert, eller är risktagare, eller befinner sig någonstans däremellan. Det finns bland annat en

⁴ graford = löpord, word token

undersökning av Norrby & Håkansson (2007:62 f.) där det konstateras att olika inlärarprofiler är en faktor som har betydelse för språkutvecklingen.

Vid urvalet av de 32 uppsatserna valde jag bort de som i Tisuskorpusen hade angivit svenska som modersmål. Den ofta självskrivna indelningen i infödda och icke-infödda språkbrukare (på engelska ofta förkortat NS/NNS) har varit och är fortfarande central i forskning inom alla språkvetenskapens områden, men den börjar ifrågasättas och är kanske inte helt självklar. Boyd & Fraurud (2011:67 f.) sätter distinktionen under lupp inom ramen för SUF-projektet (se 2.2) där 222 ungdomar i flerspråkiga svenska miljöer ingår och får fram en mängd olika språkprofiler bland informanterna. En av deras slutsatser är att indelningen NS/NSS inte är användbar i undersökningar av språkets användning, variation och förändring i flerspråkiga samtida miljöer.

Ytterligare ett problem är att det är tillåtet att använda svensk-svensk ordbok under Tisus-provets skriftliga del, men det finns inte dokumenterat någonstans om den enskilde skribenten har använt sig av den möjligheten, och i så fall hur mycket. Det finns heller inget sätt att undersöka om, och i så fall hur, det påverkar den här studiens resultat.

Ett problem med validiteten kan vara att färdighetsnivån är den enda inlärarvariabel som används i studien. Dessutom är den bakomliggande bedömningen inte helt oproblematisk med tanke på att det handlar om fri textskrivning inom ett givet tema. Visserligen finns det alltid minst två bedömare per Tisustext, men ett antal texter är alltid gränsfall, och vilka dessa är, och om de slutligen bedömdes som godkända eller underkända, framgår inte av Tisuskorpusen.

Beräkningen av enkel nominalkvot i stället för full är inte helt tillfredsställande, men det valet var nödvändigt med tanke på förutsättningarna för den här studien. Tisuskorpusen är ordklasstaggad så man kan snabbt få fram ordlistor från respektive ordklass som ingår i korpusen och man kan anta att samma precision i taggningen gäller jämnt över alla texter, och till exempel homonymer hamnar i de flesta fall automatiskt på rätt plats. Men det gäller endast korpusen i sin helhet. Det går däremot inte att på ett enkelt sätt få fram listor på vilka adverb, prepositioner, particip och verb som förekommer i en enskild text, eller i exempelvis de åtta G3-uppsatser som ingår i undersökningen, något som skulle vara nödvändigt för att beräkna full nominalkvot (= antal substantiv + prepositioner + particip/antal pronomen + verb + adverb).

När det gäller valet att undersöka substantiviska sammansättningar så menar Lagerholm (2008:219) att långa ord i allmänhet är ovanligare än korta. Vidare menar han att ord som är långa, och ofta sammansatta, är nödvändiga i en mer formell stil. Men ordlängden är ett omdiskuterat mått och Lagerholm understryker att den inte har något förklaringsvärde i sig. Vidare menar Laufer (1997:45) exempelvis att om de olika leden i långa ord är välkända var för sig så kan det innebära att inläsaren inte upplever dem som svåra. Trots problemen med att använda ordlängd som måttstock så finns det enligt Lagerholm (2008:219) statistiska belegg för att det finns ett samband mellan en högre frekvens av långa ord och texter med högre svårighetsgrad. Det innebär att de underkända Tisustexterna borde innehålla en lägre andel substantiviska sammansättningar än de godkända.

I det stora hela utgör kvantitativa beräkningar av andelen substantiv, typer av substantiv etc. naturligtvis grova mått med tanke på det komplexa fenomen som en Tisustext utgör. Laufer & Nation (1995:309) ifrågasätter till exempel om många innehållsord överhuvudtaget är ett tillförlitligt mått på en texts informationstäthet. De kvantitativa resultat som studien presenterar kan därmed inte betraktas som något annat än indikationer och tendenser. Dock är beräkningarna grundade på 32 olika skribenters texter vilket väl får anses som ett rimligt underlag för en begränsad studie av det här slaget.

4.9. Etiska överväganden

Vetenskapsrådet gav 2002 ut dokumentet ”Forskningsetiska principer inom humanistisk-samhällsvetenskaplig forskning”. Där finns fyra huvudkrav för individskydd: informationskravet, samtyckeskravet, konfidentialitetskravet och nyttjandekravet.

Tisuskorpusen är en delkorpus i pilotprojektet till SweLL. Under pilotprojektet som pågick 2013-16 etablerades kontakt med Etikprövningsnämnden och man utformade tillståndsblanketter utifrån deras rekommendationer (Volodina et al. 2016). I den här studien förutsätts att SweLL:s projektgrupp har uppfyllt det grundläggande individskyddskravet och därmed behandlat personuppgifterna i Tisuskorpusen utifrån Forskningsetiska principer för humanistisk-samhällsvetenskaplig forskning (Vetenskapsrådet, 2002). Det innebär bland annat att det i projektgruppen ska finnas god kännedom om personuppgiftslagen (1998:204) (PUL), personuppgiftsförordningen

(1998:1191) samt arkivlagen (1990:782) och arkivförordningen (1991:446). Dessutom förutsätts att projektledningen har inhämtat de tillstånd som krävs, fullgjort de anmälningsplikter som föreskrivs, samt tagit ansvar för att alla som på något sätt deltar i arbetet med projektet har god kunskap om forskningsetiska principer och om hur personuppgifter ska hanteras.

I undersökningen och redovisningen av undersökningens materialet från Tisuskorpusen har jag tagit nödvändiga hänsyn vad gäller konfidentialitetskravet och nyttjandekravet. Till exempel redovisas inte texterna i sin helhet och textexemplen har valts så att ingenting finns med som kan avslöja vem skribenten är. De i undersökningen ingående 32 texterna från Tisuskorpusen har enbart använts för forskningsändamål och alla texter i korpusen är avidentifierade.

5. Resultat

Här i resultatkapitlet redovisas undersökningens delområden var för sig, det vill säga nominalkvot, AO, substantiverade sammansättningar och frekvensbaserade ordlistor.

5.1. Nominalkvot

Medelvärdena i beräkningarna av nominalkvot visar nedan, i tabell 9, en tydlig skillnad mellan de 16 godkända och de 16 underkända texterna då medelvärdet för G-texterna är 1,10 och för U-texterna 0,86. Där är skillnaden alltså 0,24, medan skillnaden mellan G5 och G3 är 0,1.

Tabell 9. Nominalkvot - medelvärden för G5, G3 och U2

| Medelvärden | Tokens subst. | % av alla tokens i texten | Tokens verb | % verb | Totalt antal ord | Nominalkvot (n/v) |
|-------------|---------------|---------------------------|-------------|--------|------------------|-------------------|
| G3 | 123 | 22 | 111 | 20 | 554 | 1,15 |
| G5 | 117 | 21 | 114 | 20 | 558 | 1,05 |
| G3+G5 | 120 | 22 | 113 | 20 | 556 | 1.10 |
| U2 | 89 | 19 | 107 | 22 | 481 | 0,86 |

I bilaga 4 finns nominalkvoterna presenterade i mer detaljerade tabeller. Där framgår att i G5-texterna varierar nominalkvoten mellan 0,91 och 1,21, i G3-texterna mellan 0,85 och 1,55 och i U2-texterna mellan 0,65 och 1,19. Det betyder att enskilda texter kan avvika avsevärt från medelvärdet.

5.2. Svensk akademisk ordlista (AO)

I tabell 10 nedan kan vi se att medelvärdet för den procentuella andelen substantiv i de godkända texterna (20 procent) som återfinns i AO är 3,2 procent högre i jämförelse med de underkända texterna (16,8 procent). Vi ser också att det skiljer 2,7 procent mellan G3 (21,4 procent) och G5 (18,7 procent), där G3 har den högre andelen.

Tabell 10. Medelvärden – types substantiv i G5, G3 och U2-texter som återfinns i AO

| Medelvärdet | Types subst. | Antal types subst. i AO | % subst. types i AO |
|-------------|--------------|-------------------------|---------------------|
| G3 | 75 | 16 | 21,4 |
| G5 | 79 | 14 | 18,7 |
| G3+G5 | 77 | 15 | 20 |
| U2 | 55 | 9 | 16,8 |

I tabell 11 nedan jämförs spännvidden, lägsta och högsta procentandel substantiv som återfinns i AO. Spännvidden visar generellt en ganska hög procentandel, och den är i stort sett densamma när vi jämför U2 och G5: 11,7–25,3 procent respektive 11,7–25,6 procent. Däremot är spännvidden högre i G3: 11,7–39,2 procent.

Tabell 11. Spännvidd inom G5, G3 och U2 – från lägst till högst andel substantiv som återfinns i AO räknat i %

| G5 | | G3 | | U2 | |
|---------|--------|-------|--------|-------|--------|
| G5.87 | 11,7 % | G3.9 | 11,7 % | U2.17 | 11,1 % |
| G5. 102 | 25,3 % | G3.10 | 39,2 % | U2.19 | 25,6 % |

I bilaga 6 redovisas andelen substantiv som återfinns i AO i varje enskild uppsats och därmed ser man bakgrunden till den spännvidd som redovisas i tabell 11. Där framgår att G3.10 avviker anmärkningsvärt från resten. Om vi utesluter G3.10 så blir

spännvidden mindre för G3-texterna: G3.6 och G3.13 som kommer närmast efter G3.10 har båda en andel på 23,3 %.

5.2.1. Konkreta substantiv

De substantiv i de 32 texterna som sammanfaller med substantiv i AO är som väntat, med tanke på den typ av ordlista som AO representerar, med få undantag abstrakta, till exempel *teori* och *möjlighet*. Vid en genomgång av listorna (bilaga 5) och kontroll med sammanhanget i de texter där de förekommer finner man endast följande konkreta substantiv: *artikel* (G5.84), *individ* (G3.1, 10,13, G5.95, U2.55), *läsare* (G3.10) och *kollega* (G3.13, G5.72, G5.84, G5.95, G5.102, U2.30).

5.2.2. Nominaliseringar

En bred sökning efter nominaliseringar bland substantiven i de 32 texterna som återfinns i AO (bilaga 5), det vill säga substantivbildande suffix som Josefsson (2005:118 f.) beskriver dem, visar i tabell 12 nedan att totalt 38 substantiv utgörs av olika typer av nominaliseringar. Av dessa 38 nominaliseringar återfinns 22 i G5-texterna, 24 i G3-texterna och 19 i U-texterna. En sammanräkning av undersökningens 16 godkända texter (G3+G5) visar på att 34 av de 38 nominaliseringarna finns i dessa. Detta ska jämföras med 19 i de underkända texterna.

Tabell 12. 38 substantiv i de 32 texterna som innehåller substantivbildande suffix och som också återfinns i AO

| | 8 texter, G3 | 8 texter, G5 | 16 texter, U2 _(types) |
|----------------------------------|--------------|--------------|----------------------------------|
| aktivitet | 2 | - | 1 |
| allmänhet | 2 | - | 1 |
| ambition | 1 | - | - |
| bedömning | - | 1 | - |
| bemärkelse | - | 1 | - |
| definition | 1 | - | - |
| diskussion | 2 | - | - |
| egenskap | - | 1 | - |
| engagemang | 1 | - | - |
| erfarenhet | - | 1 | 3 |
| forskning | 3 | 3 | 5 |
| funktion | - | 1 | - |
| företeelse | - | 1 | - |
| förhållande | 1 | - | 1 |
| förklaring | - | 1 | 3 |
| förutsättning | - | 1 | - |
| förväntning | 1 | 1 | - |
| förändring | - | - | 1 |
| kompetens | 1 | 1 | 1 |
| konsekvens | 2 | 2 | 1 |
| lösning | 3 | 6 | 4 |
| mening | 1 | 1 | - |
| möjlighet | 3 | 1 | 6 |
| omgivning | - | 1 | - |
| omständighet | 2 | 1 | 1 |
| osäkerhet | - | - | 2 |
| påverkan | 1 | - | - |
| reaktion | - | - | 1 |
| sanning | 1 | - | - |
| skillnad | 2 | 1 | 2 |
| ställning | 1 | 1 | - |
| svårighet | - | - | 1 |
| uppfattning | 1 | - | - |
| uppmärksamhet | 3 | - | 1 |
| uttalande | 1 | - | - |
| utveckling | 2 | 1 | 3 |
| verklighet | 1 | 1 | 1 |
| värdering | - | 1 | - |
| Totalt antal¹: | 24/39 | 22/30 | 19/39 |

¹ Den första siffran anger hur många av de 38 substantiven som förekommer i G3, G5 resp. U2. Den andra siffran är types som använts i enskilda texter inom resp. G3, G5 och U2. T.ex. har 3 olika U2-texter använt ordet *utveckling*. Upprepningar av ett ord i samma text (tokens) finns inte med i tabellen.

Avledda substantiv, nominaliseringar, är enligt både Lagerholm (2008:108) och Hellspong & Ledin (1997:78 f.) en typ av abstrakta substantiv som är vanliga i formell stil. Tabell 12 ovan kan kopplas till Hellspong & Ledins (1997:78 f.) tredje dimension, dit avledda substantiv hör, se 1.2.1.

Ett exempel på när ett substantiv inte representerar en sak, ett ting, är den typ av nominalisering där en händelse, process, uttrycks som ett substantiv i stället för ett verb, vilket gör texten mer abstrakt. Exempelvis ”ständig **utveckling** av sina talanger” skulle kunna uttryckas med verb: *att ständigt **utveckla** sina talanger* som i följande exempel:

Om vi pratar om kvinnor, då måste jag påpeka att kvinnor har verkligen stressande liv. Moderna kvinnor vill hinn allt – karriär, bra utbildning, ständig **utveckling** av sina talanger och till på köpet skaffa en lycklig familj med barn. G3.6

När det gäller nominaliseringar i undersökningens texter finns mycket kvar att analysera och kontrastera såväl kvantitativt som kvalitativt eftersom jag tvingades avgränsa mig här.

5.2.3. *Fem flerordsenheter*

Resultatet av den kvalitativa undersökningen av *fall*, *form*, *roll* och *verk* presenteras i tabell 13 nedan. I den övre delen av tabellen visas det totala antalet förekomster av *fall*, *form* och *roll* och *verk*, alltså både när dessa fungerar som rotmorfem med egen betydelse i textsammanhanget och i den nedre delen av tabellen när de ingår i en flerordsenhet. alltså antal förekomster av *fall*, *form*, *roll*, och *verk* i flerordsenheterna *i alla fall*, *i så fall*, *i form av*, *spelar roll* och *i själva verket*.

Tabell 13. Användning av *fall*, *form*, *roll* och *verk*

| | <i>fall</i> | <i>form</i> | <i>roll</i> | <i>verk</i> |
|----|-------------|-------------|-------------|-------------|
| G3 | 3 | - | 3 | - |
| G5 | 2 | 2 | - | 1 |
| U2 | 2 | - | 5 | - |

varav korrekt eller nästan korrekt användning av *fall*, *form*, *roll* och *verk* i flerordsenheter:

| | <i>i alla fall</i> | <i>i så fall</i> | <i>i form av</i> | <i>spelar roll</i> | <i>i själva verket</i> |
|----|--------------------|------------------|------------------|--------------------|------------------------|
| G3 | 1 | - | - | 1 | - |
| G5 | 1 | - | 1 | - | 1 |
| U2 | - | 2 | - | 4 | - |

De två exemplen nedan illustrerar användningen av *fall* som ett rotmorfem med egen betydelse, respektive som en del av en flerordsenhet:

I det andra **fallet** – när man har ett arbete man inte trivs med – handlar det väldigt ofta om bristande motivation och bekräftelse. G5.102

Men att inte arbeta är ingen lösning heller. **I alla fall** inte om man ofrivilligt avstår från arbetet. G5.84

Ordkombinationen *i alla fall* förekommer i två godkända texter och får många träffar i *Språkbankens* korpus från sociala medier (*Språkbanken*, Göteborgs universitet [www]) vilket tyder på att den är frekvent i mer informell stil. I exemplet ovan från en godkänd text där *i alla fall* ingår, är dock stilen formell med stildrag som nominaliseringen *lösning*, användningen av *man* och formuleringen ”ofrivilligt avstår från”.

Två av de fem flerordsenheterna som undersöks återfinns i de underkända texterna: *spelar roll* och *i så fall*. Det bör tilläggas att av de fyra förekomsterna av *spelar roll* så återfinns tre av dem i en och samma text, U2.16. Ordkombinationen *i så fall* förekommer i två underkända texter, U2.19 och U2.29.

5.3. Substantiviska sammansättningar

I tabell 14 nedan presenteras och jämförs medelvärden för substantiviska sammansättningar i godkända och underkända texter. Där kan vi utläsa att skillnaden i medelvärde är relativt stor när det gäller antal substantiviska sammansättningar i relation till det sammanlagda antalet tokens i texterna där godkända uppsatser har 3,2 och underkända 2,1 procent.

Tabell 14. Medelvärden, andel substantiviska sammansättningar i texterna i relation till tokens

| Medelvärden | Totalt antal tokens | antal substantiv tokens | medelvärde antal subst. sms /text tokens | % subst. sms av totalt antal tokens | medelvärde % subst. sms av totalt antal subst/text tokens |
|-------------|---------------------|-------------------------|--|-------------------------------------|---|
| 8xG3 | 554 | 123 | 17 | 3 | 13,8 |
| 8xG5 | 558 | 117 | 19 | 3,4 | 16,2 |
| 16xG3+G5 | 556 | 120 | 18 | 3,2 | 15 |
| 16xU2 | 481 | 89 | 9 | 2,1 | 10,1 |

Medelvärdet för antal substantiviska sammansättningar per text skiljer sig också åt. Det är 18 för de godkända texterna och 9 för de underkända, men då får man komma ihåg att textlängden, alltså medelvärdet för totalt antal tokens är betydligt lägre i de underkända texterna, 481 jämfört med de godkändas 556. Medelvärdet för andelen substantiviska sammansättningar i relation till alla substantiv ger 2,4 % mer för G5 jämfört med G3, och hela 6,1 % skiljer mellan G5 och U2.

I bilaga 7 kan vi se att några texter, till exempel G3.2 och G5.48 avviker en hel del från de medelvärden som redovisas ovan i tabell 14.

I tabell 15 jämförs medelvärdet för antalet stavelser i de substantiviska sammansättningarna (types) i godkända och underkända uppsatser. De finns redovisade per uppsats i bilaga 10. Här framkommer inga tydliga skillnader mellan godkända och underkända uppsatser i den procentuella fördelningen.

Tabell 15. Medelvärden - antal stavelser per substantivisk sammansättning/types (särskrivningar medräknade)

| | G3 | | G5 | | G3+G5 | | U2 | |
|-------------|-------|-----|-------|-----|-------|-----|-------|-----|
| | antal | % | antal | % | antal | % | antal | % |
| 2 stavelser | 26 | 20 | 35 | 23 | 62 | 21 | 34 | 23 |
| 3 stavelser | 44 | 33 | 54 | 35 | 98 | 34 | 51 | 35 |
| 4 stavelser | 29 | 21 | 25 | 16 | 54 | 18 | 38 | 24 |
| 5 stavelser | 27 | 20 | 27 | 17 | 54 | 18 | 18 | 12 |
| 6 stavelser | 6 | 4 | 13 | 8 | 19 | 7 | 8 | 5 |
| 7 stavelser | 2 | 2 | - | - | 2 | 2 | 2 | 1 |
| 8 stavelser | - | - | 1 | 1 | - | 1 | - | - |
| Summa | 134 | 100 | 155 | 100 | 289 | 100 | 151 | 100 |

Nedan presenteras tre olika sexstaviga substantiviska sammansättningar i sina respektive textsammanhang:

Man måste hitta en mellanväg där arbetet varken tar för stor del av ens liv eller en för liten del. För att balansera detta krävs att vi är både fysiskt aktiva samt att vi tar regelbundna pauser i arbetet. En arbetsgivare kan styrka detta genom att lägga mera vikt i de anställdas friskvård samt bättre **arbetsförhållanden**. G5.94

Inom forskarsamhället råder stor **meningsskiljaktighet** över hur fenomenet ”stress” egentligen skulle värderas. Diskussionen, som under närmaste tiden skapat sig en betydlig uppmärksamhet genom att äger rum i pressen, har hittills framför allt kretsat om begreppsparet ”stress och arbete”. G3.10

Ett exempel som bäst beskriver detta situation är lärare inom Vux. De har inte samma arbetsrättighet som andra **kommunallärare**. De får arbeta 24/7 hela tiden, eller gör som många andra lärare: Undervisar endast för pengar. Var ligger gränsen? Vems ansvar. U2.53

I exemplen ovan ser vi till att börja med att alla tre textutdragen är begripliga. Men U2.53 innehåller vad som förmodligen ska tolkas som en felaktig sammanskrivning. Troligen menar skribenten *kommunalt anställda lärare*. Dessutom ger stilnivån i U2.53 intryck av att vara generellt lägre än de två exemplen från godkända uppsatser.

Nedan i tabell 16 jämförs antalet substantiverade sammansättningar bestående av tre eller fyra led samt dubbel syftning. Där ser man ingen märkbar skillnad mellan godkända och underkända texter, förutom att dubbel syftning inte förekommer i U2-texterna. Däremot ser vi en relativt stor skillnad när vi jämför antalet särskrivningar. 17 % av de undersökta sammansättningarna i de sexton U2-texterna är felaktigt särskrivna, jämfört med 11 % i de åtta G3-texterna och bara 2 % i de 8 G5-texterna.

Tabell 16. Jämförelse – substantiverade sammansättningar med tre led, fyra led, dubbel syftning, samt särskrivning

| | 16xU2 | 8xG3 | 8xG5 | Exempel |
|-----------------------|-------|------|------|--------------------------|
| tre led | 8 | 9 | 7 | fritidsintresse |
| fyra led | - | 1 | - | fritidssysselsättning |
| dubbel syftning | - | 3 | 1 | alkohol- och drogmisbruk |
| särskrivning, antal | 26 | 14 | 3 | privat livet |
| särskrivning, procent | 17 % | 11 % | 2 % | |

5.4. Frekvensbaserade ordlistor

Av tabell 17 framgår att på frekvensnivåerna 1000 och 2000 (alltså de 1000 mest frekventa orden samt 1000-nivån därefter) har substantiven i de underkända texterna högre samstämmighet med de frekvensbaserade ordlistorna från bloggkorporusen jämfört med de godkända.

Tabell 17. Medelvärden procent substantiv i de undersökta texterna som sammanfaller med de frekvensbaserade ordlistorna, samt medelvärde procent av alla ord i de undersökta texterna, token coverage

| Medelvärden | subst 1000 % | subst 2000 % | subst 3000 % | subst 4000 % | subst 5000 % | subst totalt % 1- 5000 | Subst % inte 1- 5000 | TC ¹ % hela texten 1- 5000 |
|-------------|--------------------|--------------------|--------------------|--------------------|--------------------|------------------------------------|-------------------------------|---|
| G5 | 24,6 | 13,6 | 12,3 | 4,1 | 7,2 | 62 | 38 | 84,1 |
| G3 | 24,8 | 16,7 | 13,9 | 5 | 7,5 | 68 | 32 | 85 |
| G3+G5 | 24,7 | 15,2 | 13 | 4,6 | 7,3 | 65 | 33,2 | 84,6 |
| U2 | 28,8 | 18,2 | 12,9 | 4,3 | 7,4 | 71,7 | 28,3 | 86,8 |

¹ TC = token coverage (ordens täckningsgrad)

På 1000-nivån skiljer det 4,1 % mellan godkända och underkända uppsatser vad gäller substantiv som återfinns i de frekvensbaserade ordlistorna. Och på 2000-nivån är den procentuella skillnaden 3 %. På frekvensnivåerna 3000-5000 kan ingen nämnvärd skillnad urskiljas. Vid jämförelse mellan de undersökta texterna i sin helhet i relation till bloggkorpusens fem frekvensnivåer så har de underkända texterna 2,2 % högre täckningsgrad. I bilaga 10 redovisas siffrorna för alla 32 texterna.

6. Diskussion

Hypotesen att lexikala särdrag, med fokus på användningen av substantiv, påverkar bedömningen av Tisustexter verkar stämma i det undersökta empiriska materialet i uppsatsen. Men underlaget är inte tillräckligt stort för att vi ska kunna dra mer generella slutsatser.

I det här kapitlet diskuteras resultaten i undersökningens olika delområden och dessa relateras till tidigare forskning och framtida forskningsbehov. Sedan följer en kort diskussion om konsekvenser för undervisning och inläring av akademisk svenska. Och slutligen mer allmänna reflektioner kring framtida forskning inom området.

Ett första intressant resultat är att den enkla beräkningen av nominalkvot i studien har gett utslag i en jämförelse mellan de godkända och de underkända texterna där medelvärdet var 1.10 för de 16 godkända texterna respektive 0,86 för de 16 underkända (tabell 9). Dessa resultat kan sättas i relation till det Lagerholm (2008:133) anger, nämligen att enklare texter har en nominalkvot på drygt 1, och bruksprosatexter mellan 1,5 och 2, medan spontant tal har en nominalkvot på endast några tiondelar. Därmed har ingen av U2-texterna en nominalkvot i närheten av spontant tal vilket tyder på en viss grad av informationstäthet även i dessa (tabell 1). I stort sett alla undersökningens texter ligger i närheten av det Lagerholm anger som nominalkvot för enklare texter. Beräkningen av den enkla nominalkvoten pekar mot att de godkända texterna är mer informationstäta och har en mer skriftspråklig stil, vilket i sin tur tyder på att andelen substantiv i texten påverkar bedömningen. I kapitlet metodiska diskussioner beskrivs bakgrunden till att ingen beräkning har gjorts av full nominalkvot trots att den har gett stort utslag i nyligen gjorda kvantitativa undersökningar och

är mer tillförlitlig och används av bland annat Hultman & Westman (1977) samt i Johansson Kokkinakis & Magnussons (2011) kvantitativa undersökning av det nationella provets elevuppsatser.

Den kontrastiva analysen av de godkända respektive underkända texterna rörande substantiv som återfinns i AO blev undersökningens mest omfattande delområde delvis beroende på att kartläggningen ledde vidare till en begränsad kvalitativ del. Som förväntat hade de godkända texterna enligt medelvärdet i tabell 10 en högre andel substantiv som sammanföll med AO. Men vid en närmare titt på spännvidden inom G5, G3 och U2 som redovisas i tabell 11 framkommer stora skillnader mellan enskilda uppsatser. Här skulle exempelvis en systematisering och granskning av de abstrakta substantiv som finns i texterna och AO, samt av de substantiv som inte återfinns i AO, behövas för att få en tydligare bild av skillnaderna mellan G5, G3 och U. Spännvidden som redovisas i tabell 11 kan även tyda på att andra bedömningskriterier, som till exempel den i 2.5 nämnda helhetsbedömningen av texten har större inverkan på betyget än textens lexikala särdrag. Men framförallt behövs en närmare granskning av hur substantiven används och fungerar i Tisusskribenternas texter innan några slutsatser kan dras.

Sammanställningen av avledda substantiv, nominaliseringar, i tabell 12 visar som förväntat att dessa är mer frekventa i de godkända texterna jämfört med de underkända, men här återstår mycket att systematisera och utforska vidare. Forskningen på grammatiska metaforer (GM) är relativt ung och där spelar avgränsningar och definitioner en viktig roll enligt Magnusson (2011:91). Hon utgår i sin avhandling från den systemisk-funktionella lingvistik (SFL) och nominalisering av process är den vanligaste GM hon finner i gymnasieelevers utredande uppsatser (2011:107). Exempelvis *Utveckling* i tabell 12 kan vara ett exempel på nominalisering av process i alla texter där det förekommer, men för att fastställa detta och undersöka andra grammatiska metaforer behöver vi gå in i texterna och utforska hur nominaliseringarna fungerar i sina sammanhang, något som alltså kräver en separat studie.

Den kvalitativa undersökningen av fem flerordsenheter tyder på att vilken typ av flerordsenhet som används och hur den infogas i textsammanhanget kan ha betydelse för bedömningen av textens kvalitet. Flerordsenheterna *spelar roll* och *i så fall* kan

givetvis, liksom många andra konventionaliserade ordkombinationer, förekomma på olika stilnivåer beroende på sammanhanget. Men med tanke på att sökningar på *det spelar ingen roll* och *i så fall* resulterar i ett mycket stort antal träffar i *Språkbankens* korpus från sociala medier bör de tillhöra de mer frekventa flerordskombinationerna i vardagligt svenskt skriftspråk. Och i och med det förväntas nybörjaren möta dem på ett tidigt stadium i språkutvecklingen och som en konsekvens av det lära sig dem receptivt. Därmed verkar det logiskt att just dessa frekventa exempel används produktivt även av underkända skribenter.

De underkända testtagarnas användning av flerordsenheterna *spelar roll* och *i så fall* kan också tyda på att texterna i sin helhet ligger på en mer informell, talspråklig stilnivå något som kan ha påverkat bedömningen (se 1.2.1) och som följande exempel får illustrera:

Man lever där på ett annat tempo, man lever långsammare, man kan säga och jag anser att det beror på att invånare i min lilla by tittar inte på TV eller media *spelar* inte så stort *roll* som det gör här. U2.16

I form av och *i själva verket* med varsin förekomst i två olika G5-texter, och med betydligt färre träffar i *Språkbankens* korpussamling från sociala medier, används av dessa två testtagare på en mer formell stilnivå vilket exemplen nedan visar:

I själva verket beror stress på en brist på kontroll över livet, över ens tillvaro. När man förstår det då förstår man varför människor blir stressade och vad man kan göra för att åtgärda situationen. G5.23

Utvecklingen har gett oss en ny miljö med fördelen *i form av* mer flexibilitet, rörlighet, delad ansvar och fler möjlighet att utvecklas som människa. G5.48.

Det låga totala antalet förekomster av de fem flerordsenheterna i materialet gör att vi givetvis inte kan dra några långtgående slutsatser. Det är dock inte ett oväntat lågt antal eftersom även frekventa flerordsenheter som L2-inläraren har receptiv kunskap

om enligt Prentice & Sköldberg (2013:215–6) används sparsamt i produktion. Den kvalitativa delundersökningens resultat tyder alltså som förväntat på att det finns stilistiska skillnader i de godkända texterna jämfört med de underkända, se 1.2.1 och tabell 1. För övrigt återstår det mycket att undersöka när det gäller flerordsenheter i Tisuskorpusen.

Andelen substantiviska sammansättningar var som förväntat högre i de godkända texterna än i de underkända. Det stämmer också med det som tidigare nämnts om Grundins (1975) och Nordenfors (2011) undersökningsresultat i studier av svenska skolelevers skrivande som bland mycket annat visar att ordlängden ökar med elevernas ålder. Ordlängd och användning av sammansatta ord verkar alltså ha ett samband med utvecklingen av sva-inlärares skriftspråk och kan tänkas påverka bedömningen av inläratexter. När det gäller sammansättningarnas antal led och antal stavelser framkommer inga markanta skillnader mellan godkända och underkända texter. Däremot kan man ana stilistiska skillnader när man tittar på exempel med sammansättningarna i sina sammanhang, där exemplen från godkända uppsatser tyder på en mer formell text. Vidare var andelen felaktiga särkrivningar som förväntat högre i de underkända texterna. Skillnaden var där markant, och det tyder på att särkrivningar har ett samband med den lexikala utvecklingen och möjligtvis också med bedömning.

Resultatet av jämförelserna med de frekvensbaserade ordlistorna från bloggkorpus visar att de underkända texterna har en högre täckningsgrad totalt. Dessutom sammanfaller fler substantiv från de underkända texterna med de 2000 första substantiven från de frekvensbaserade ordlistorna, vilket var förväntat eftersom dessa bör ha fler talspråkliga och informella drag än de godkända texterna. Här skulle en undersökning av vilka substantiv som enligt tabell 17 återstår i G5 (38 %), G3 (32 %) respektive U2 (28,3 %) vara mycket intressant, men det får bli en fråga för framtiden.

Sammanfattningsvis kan vi alltså se relativt starka samband mellan bedömd färdighetsnivå, G/U, och resultaten av undersökningens fyra olika delområden, vilket skulle tyda på att testtagarens kunskap om och användning av de aspekter av lexikala särdrag som undersökningen fokuserat på har betydelse för bedömningen. Några tydliga skillnader mellan G5 och G3 har inte kunnat påvisas. Hur stor påverkan olika lexikala

särdrag har på bedömningen är dock omöjligt att fastställa enbart utifrån den här begränsade undersökningens resultat.

Tisus, som alltså är ett språkligt färdighetstest, är en av vägarna till behörighet i svenska (Antagningservice [www]). En annan möjlighet är att få godkänt på behörighetsgivande kurs i svenska, en kurs som ges vid de flesta svenska universitet och högskolor. Där sammanfaller målet för den skriftliga färdighetsnivån med gränsen för godkänt på Tisustexten. På behörighetsgivande kurs vid Stockholms universitet får studenterna bland annat följande information:

Kursen förbereder studenten för akademiska studier på svenska. Godkänt resultat på hela kursen ger behörighet i svenska för högskolestudier i Sverige enligt föreskrifter från Högskoleverket (HSVFS 2009:1) där denna utbildning i behörighetshänseende jämföras med Svenska som andraspråk kurs B/Svenska som andraspråk 3 (Gy11) i svenskt gymnasium eller kommunal vuxenutbildning, (Stockholms universitet, 2017 [www]).

Erfarenheten visar att det är viktigt för undervisande lärare på behörighetsgivande kurs, och på andra liknande kurser, att redan i kursstarten få en uppfattning om variationen i studentgruppen vad gäller det mer formella ordförrådet för att, i den mån det är möjligt, anpassa undervisning och eventuella stödinsatser.

Milton (2009:79 f.) presenterar undersökningar som visar på att framgångsrika miljöer för andraspråksinläring av ord karakteriseras av regelbunden uppbyggnad av ordförrådet. Han poängterar också de individuella variationerna som läraren bör ta i beaktande. Som tidigare påpekats anser Enström (2013:172) att studenter som siktar på att bedriva akademiska studier har ett stort behov av att utveckla sitt ordförråd inom kategorierna ämnesspecifika fackord, mindre frekventa vardagliga ord samt icke-ämnesspecifika akademiska ord.

Lärare på behörighetsgivande kurs, och andra liknande kurser, samt alla studenter som har behörighet i svenska som mål har alltså ett behov av lättillgängliga och varierade frekvensbaserade ordtest liknande de som finns för engelska som målspråk (Nation 2013:514 f.). Inom ramen för projektet BIGFOT har Sofie Johansson vid

Göteborgs universitet utvecklat FOT, en svensk variant av Nations (2013:36 f.) Vocabulary Level Test (VLT), alltså ett frekvensbaserat ordtest. Duek (2012) som använder FOT i sin magisteruppsats skriver att ” förhoppningsvis kommer FOT att [...] finnas tillgängligt för lärare inom en relativt snar framtid”. Och eftersom FOT nu finns med i Skolverkets nya bedömningsstöd, *Bygga svenska* (Skolverket [www]), som lanseras vårterminen 2017 så får det mycket snart en större spridning. Förutom behovet av olika typer av ordtest behöver också didaktiskt material förnyas och utvecklas med genomtänkt och varierad träning på ordförrådet för studenter på avancerad språklig nivå (Nation 2013:569 f.).

Som framgår av diskussionen ovan kring undersökningens delresultat återstår mycket att utforska i och utanför Tisuskorpusen vad gäller den lexikala utvecklingen hos avancerade andraspråksinlärare och relationen till bedömning av färdighetsnivå. En jämförande undersökning vad gäller förändringar i ordförrådet sedan 2006 skulle exempelvis vara möjlig med en egen insamlad Tisuskorpus i nutid. Utformingen av det faktablad som hör till Tisustexten har visserligen förändrats något sedan 2006 och bytt namn till inspirationsblad, och bedömningsbladet (bilaga 1) har också förändrats något. Men i det stora hela är förutsättningarna för skrivandet desamma. Textens tema vid det specifika provtillfället visar sig naturligtvis i ordförrådet i texterna, speciellt innehållsorden. Dock rör sig alltid temat för respektive Tisustillfälle kring allmänt samhällsvetenskapliga ämnen som en testtagare förväntas kunna resonera kring utan att ha tagit del av en specifik kurs eller kurslitteratur.

Fler inlärarvariabler i Tisuskorpusen skulle kunna utforskas. Det vore exempelvis ett naturligt steg att gå vidare med en undersökning av hur vistelsetiden i Sverige påverkar textkvaliteten och därmed bedömningen av Tisustexterna.

Kvalitativa studier av ett litet urval texter skulle också vara möjliga att genomföra ur olika perspektiv, till exempel undersökningar av olika typer av flerordsenheter. Prentice & Sköldberg (2013:216) betonar att det behövs mer såväl undervisning som forskning på det här området. Det var mot den bakgrunden som jag valde att undersöka och jämföra godkända och underkända texter vad gäller förekomsterna av de konventionaliserade ordkombinationerna *i alla fall, i form av, spelar roll* och *i själva verket*.

Det finns naturligtvis även en mängd individuella avvikelser i materialet som det skulle vara intressant att analysera i olika typer av kvalitativa fallstudier. I min undersökning, som i huvudsak är kvantitativ, har dessa mest fungerat som en påminnelse om att de lexikala särdrag som jag studerat sannolikt påverkar, men inte behöver vara en avgörande faktor vid bedömning av texter skrivna av sva-inlärare på en relativt avancerad språklig nivå.

Tisus genomförs samtidigt på cirka 60 provorter i världen (Tisus utland, [www]) och forskning om svenska som andraspråk undersöker i stort sett alltid språket hos personer som är bosatta i Sverige (Rydenvald, 2014:4). En undersökning av de svenskspråkiga Tisustesttagarna på utländska provorter såväl vad gäller testtagarnas lexikala särdrag i tal och skrift som kopplingen mellan språk och identitet skulle vara intressant. Då skulle en kartläggning av inlärarespråket i texterna behöva kompletteras med till exempel den muntliga examinationen samt intervjuer och det skulle krävas ett betydligt större projekt än en magisteruppsats.

7. Avslutning

En säker slutsats man kan dra utifrån den här och andra undersökningar av inlärarespråkets lexikala särdrag är att inlärarespråket behöver ett både brett och djupt produktivt ordförråd för att uttrycka sig väl på avancerad formell svenska. Hultman & Westman (1977:125) sammanfattar kapitlet om substantiv i den numera klassiska boken *Gymnasistsvenska* med en Lars Forsell-dikt som är fullspäckad av icke-frekventa substantiv. Jag bedömer det som att den dikten passar väl som avslutning på den här uppsatsen trots att det i dikten mest handlar om muntlig kommunikation och mest om ord som inte är relevanta i texter skrivna på akademisk svenska.

Ni måste lära er ord

Jag måste använda ord
när jag talar till er!

Tänk er kaptenen
på en galeas, en hukarejakt
en skonare, en brigantin,
en fregatt, en korvett, en snau
med molnet av segelduk över sig:

gaffel- och bomsegel,
toppsegel, stagsegel,
skothorn, fallhorn, halshorn

När det blåser till storm
kan han väl inte bara peka med
handen
och ropa: Ta ner dom där!
Det skulle se ut.

Snickaren har en låda med fackord
Muraren sina
och den som ska stat
kan väl inte bara gasta i stormen:
Dom där sitter fel! *Reva dom där!*

Jag måste använda ord
När jag talar till er
Ni måste lära er ord

Litteraturförteckning

- Allén, Sture 1970-1980. *Nusvensk frekvensordbok baserad på tidningstext*. Stockholm: Almqvist & Wiksell.
- Allén, Sture 1972. *Tiotusen i topp: ordfrekvenser i tidningstext*. Stockholm: Almqvist & Wiksell.
- Antagningservice, Universitets- och högskolerådet, <<https://antagning.se/sv/Det-har-galler-for-dig-som-gatt/Utlandsk-gymnasieutbildning/Det-har-behover-du-ha/Behorighet-i-svenska/>>. Hämtat 2017-03-15.
- Anthony, Laurence 2014. *AntConc*. <<http://www.laurenceanthony.net/software.html>>. Hämtat 2017-03-12.
- Anthony, Laurence 2013. *AntWordProfiles*. Hämtat 2017-03-12.
- AO = *Svensk akademisk ordlista*. <<https://spraakbanken.gu.se/ao/om.html>>. Hämtat 2016-12-01.
- Bloch, Joel 2010. A Concordance-based study of the use of reporting verbs as rhetorical devices in academic papers. *Journal of Writing Research* 2 (2): 219-244.
- Bokander, Lars 2016. SweLT 1.0 – konstruktion och pilottestning av ett nytt svenskt frekvensbaserat ordförrådstest. I: *Nordand – Nordisk tidskrift för andraspråksforskning* 11 (1): 39–60.
- Boyd, Sally & Kari Fraurud 2011. The native–non-native speaker distinction and the diversity of linguistic profiles of young people in multilingual urban contexts in Sweden. In: Källström, Roger & Inger Lindberg (eds.), s. 67–88.
- Coxhead, Averil 2000. A New Academic Word List. In: *TESOL Quarterly*, 34 (2): 213-238.
- Crossley, Scott & Danielle MacNamara 2011. Shared features of L2-writing: Intergroup homogeneity and text classification. *Journal of Second Language Writing* 20: 271–285.
- Duek, Susanne 2012. “Ordförråd och läromedel”. En studie om ordförråd hos högstadielärover och ord i läromedel. Magisteruppsats i svenska som andraspråk. Göteborg: Institutionen för svenska språket. Göteborgs universitet.
- Doró, Katalin & Päivi Pietilä, Researching Vocabulary in L2 Writing. In: Pietilä, Päivi, Katalin Doró & Renata Pípalová (eds.), 2015. *Lexical Issues in L2 Writing*. Cambridge: Cambridge Scholars Publishing, s. 11–26.

- Enström, Ingegerd 2016. *Ordens värld. Svenska ord – struktur och inläring*. Uppsala: Hallgren & Fallgren.
- Enström, Ingegerd 2013. Ordförråd och ordinläring med särskild fokus på avancerade inlärare. I: Hyltenstam, Kenneth & Inger Lindberg (red.), s. 169–195.
- Eklund Heinonen, Maria 2009. *Processbarhet på prov: Bedömning av muntlig språkfärdighet hos vuxna andraspråksinlärare*. Uppsala: Institutionen för Nordiska språk, Uppsala universitet.
- Eriksson Gunnar & Janne Lundberg 2004. CrossCheck-korpusen – en elektronisk svensk inlärankorpus.<www.csc.kth.se/tcs/projects/xcheck/rapporter/lindbergeriksson04.doc> Hämtat 2016-05-14.
- Gardner & Davies 2014. A New Academic Vocabulary List. In: *Applied Linguistics* 35 (3): 305–327.
- Garne, Birgitta (red.), 2002. *Språk på väg. Om elevers språk och skolans möjligheter*. Uppsala: Hallgren & Fallgren
- Grundin, Hans 1975. *Läs- och skrivförmågans utveckling genom skolåren*. Stockholm: Liber läromedel.
- Hallencreutz, Katharina 2002. Särskrivningar och andra skrivningar – nu och då. I: Garne, Birgitta (red.), s. 101–112.
- Hammarberg, Björn 2013. Andraspråksforskning med ASU-korpusen. I: *Nordand – Nordisk Tidskrift för andrespråksforskning* 2013, 8 (1): 7–34.
- Hellspång & Ledin 1997. *Vägar genom texten. Handbok i brukstextanalys*. Lund: Studentlitteratur.
- Henriksen, Birgit 1999. Three dimensions of Vocabulary Development. In: *Studies in Second Language Acquisition* 1999, 21:303–317.
- Holmegaard, Margareta 2007. Långa ord – en svårighet för flerspråkiga studerande? I: *OrdiL – En korpusbaserad kartläggning av ordförrådet i läromedel för grundskolans senare år*. (ROSA 8) s. 13–60. Institutet för svenska som andraspråk, Institutionen för svenska språket, Göteborgs universitet.
- Hultman, Tor.G. 2008. *Svenska Akademiens språklära*. Stockholm: Norstedts.
- Hultman, Tor.G. & Margareta Westman 1977. *Gymnasistsvenska*. Lund: Liber läromedel.

- Hyltenstam, Kenneth & Inger Lindberg (red.), 2013. *Svenska som andraspråk – i forskning, undervisning och samhälle*. Lund: Studentlitteratur.
- Jiménez Catalán, Rosa María, Fitzpatrick, Tess 2014. Frequency Profiles of EFL Learners' Lexical Availability. In: *Lexical Availability in English and Spanish as a Second Language*, s. 83–100.
- Johansson Kokkinakis, Sofie & Ulrika Magnusson 2011. Computer Based Quantitative Measures Applied to First and Second Language Student Writing. In: Källström, Roger & Inger Lindberg (eds.), s. 105–124.
- Josefsson, Gunlög 2005. *Ord*. Lund: Studentlitteratur.
- Källström, Roger 2012. *Svenska i kontrast*. Lund: Studentlitteratur.
- Källström, Roger & Inger Lindberg (red.). 2011. *Young Urban Swedish Variation and Change in Multilingual Settings*. Göteborg: Institutionen för svenska språket.
- Lagerholm, Per 2008. *Stilistik*. Lund: Studentlitteratur.
- Laufer, Batia 1997. What's in a word that makes it hard or easy: some intralexical factors that affect the learning of words. In: *Vocabulary Description, Acquisition and Pedagogy*. Cambridge: Cambridge University Press. s. 140–155.
- Laufer, Batia & Paul Nation 1995. Vocabulary size and use: lexical richness in L2 written Production. In: *Applied Linguistics*, 16 (3): 307–322.
- Laufer, Batia & Tina Waldman 2011. Verb-noun collocations in second language writing: A corpus analysis of learners' English. In: *Language Learning*, 61 (2): 647–672.
- Lim Falk, Maria & Per Holmberg 2016. Paths to Academic Writing in a Globalized World: A Longitudinal Study of Content and Language Integrated Learning in Upper Secondary School in Sweden. In: *Recherches en écritures: regards pluriels* / (eds.) Sylvie Plane et al., Crem: Université de Lorraine , s. 187–211.
- Mackey, Alison & Gass, Susan, M. 2012. *Research Methods in Second Language Acquisition: A Practical Guide*. Malden, MA: Blackwell Publishing, Ltd.
- Magnusson, Ulrika 2011. *Skolspråk i utveckling – En- och flerspråkiga elevers bruk av grammatiska metaforer i senare skolår*. (Göteborgsstudier i nordisk språkvetenskap 17.) Göteborg: Institutionen för svenska språket, Göteborgs universitet.

- Milton J. 2009. *Measuring Second Language Vocabulary Acquisition*. Bristol: Multilingual Matters.
- Nation, Paul 2013. *Learning Vocabulary in another Language*. Cambridge: Cambridge University Press.
- Nordenfors, Mikael 2011. *Skriftspråsutveckling under högstadiet*. (Göteborgsstudier i nordisk språkvetenskap 16.) Göteborg: Institutionen för svenska språket, Göteborgs universitet.
- Norrby, Catrin & Gisela Håkansson 2007. The Interaction of Complexity and Grammatical Processability – The Case of Swedish as a Foreign Language. *IRAL* 45: 45–68.
- Pienemann, Manfred, 1998. *Language processing and second language development: Processability theory*. Amsterdam: Benjamins.
- Pienemann, Manfred & Gisela Håkansson 1999. A unified approach towards the development of Swedish as L2: A processability account. In: *Studies in Second Language Acquisition* 21: 383–420.
- Prentice, Julia 2010. *Käppen i hjulen. Behärskning av svenska konventionaliserade uttryck bland gymnasieelever med varierande språklig bakgrund*. (ROSA 12.) Göteborg: Institutet för svenska som andraspråk, Institutionen för svenska språket, Göteborgs universitet.
- Prentice, Julia & Emma Sköldberg 2013. *Flerordsenheter – ur ett andraspråks-perspektiv*. I: Hyltenstam, Kenneth & Inger Lindberg (red.), s. 197–220.
- Reichenberg, Monica 2008. *Vägar till läsförståelse*. Stockholm: Natur & Kultur.
- Ribeck, Judy, Håkan Jansson & Emma Sköldberg 2013. Från aspekt till övergripande – en ordlista över svensk akademisk vokabulär. I: *Nordiska studier i lexikografi* 12: 370–384
- Rydenvald, Marie 2014. "eftersom jag har två språk". *Språkbruk bland svensktalande ungdomar i Europa*. (ROSA 19.) Göteborg. Institutet för svenska som andraspråk, Institutionen för svenska språket, Göteborgs universitet.
- svenska som andraspråk, Institutionen för svenska språket, Göteborgs universitet.
- SAG = Teleman, Ulf, Staffan Hellberg & Erik Andersson 1999. *Svenska Akademiens grammatik* 1999. Stockholm: Norstedts ordbok.

- Selinker, Larry 1972. Interlanguage. In: *International Review of Applied Linguistics*, 10: 209–231.
- Skolverket 2017. *Bygga svenska*. <<https://www.skolverket.se/bedomning/bedomning/bedomningsstod/nyan-elevs-sprakutveckling/bygga-svenska-ett-bedomningsstod-for-nyanlanda-elevs-sprakutveckling-1.258477>>. Hämtat 2017-05-15.
- SO = *Svensk ordbok utgiven av Svenska Akademien* 2009. Utarbetad vid Redaktionen för Svenska Akademiens samtidsordböcker, Lexikaliska institutet, Institutionen för svenska språket, Göteborgs universitet. Stockholm: Norstedts.
- Språkbanken*, Göteborgs universitet: <<https://spraakbanken.gu.se/>>. Hämtat 2016-12-12.
- Stockholms universitet. Institutionen för svenska och flerspråkighet. Kursbeskrivning NS0333, Behörighetsgivande kurs, ht16. Stockholms universitet. <http://www.su.se/polopoly_fs/1.289191.1472457561!/menu/standard/file/>. Hämtat 2017-03-15.
- Svenskt språkbruk. Ordbok över konstruktioner och fraser* 2003. Stockholm: Norstedts ordbok.
- Tisus, Institutionen för svenska och flerspråkighet, Stockholms universitet: <<http://www.su.se/svefler/tisus/om-tisus-kontakt>>. Hämtat 2016-12-15.
- Tisus utland, Institutionen för svenska och flerspråkighet, Stockholms universitet: <<http://www.su.se/svefler/tisus/provorter/provorter-tisus-utland/provorter-tisus-utland-1.30042>>. Hämtat 2017-03-05.
- Westman, Margareta 1974. *Bruksprosa*. Lund: Liber läromedel.
- Vetenskapsrådet 2002. Forskningsetiska principer inom humanistisk-samhällsvetenskaplig forskning. <<http://www.codex.vr.se/texts/HSFR.pdf>>. Hämtat 2016-12-31.
- Volodina, Elena, Lena Granstedt, Beata Megyesi, Julia Prentice, Monica Reichenberg, Gunlög Sundberg, & Mats Wirén. 2016. A Friend in Need? Research Agenda for Electronic Second Language Infrastructure. <http://www8.cs.umu.se/~johanna/slrc2016/abstracts/SLTC_2016_paper_11.pdf> Hämtat 2017-05-15.
- Östlund-Stjärnegårdh, Eva 2002. Hur lärare bedömer. I: Garne, Birgitta (red.), s. 52–69.

Tisus

Test i svenska för universitets- och högskolestudier

Bilaga 1

Skriftlig färdighet

bedömningsblad

Bilagan fanns med i uppsatsen vid opponeringen, men efter det belades den med sekretess av prefekt Gunlög Sundberg, Institutionen för svenska och flerspråkighet vid Stockholms universitet.

Tisus

Test i svenska för universitets- och högskolestudier

Skriftlig färdighet
Bedömningsblad

Betygsskala 1-15 (9=godkänd)

| | | | |
|-------------|-----------|--------|--------|
| Testtagare: | Bedömare: | Datum: | Betyg: |
|-------------|-----------|--------|--------|

Kommentarer:

| | | | | | |
|---------------------|---|--------------------------|--------------------------|--------------------------|--------------------------|
| Innehåll: | har instruktionerna följts självständighet ämnesbehandling | nej | ? | ja | + |
| | | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |
| Genre: | rätt genre mottagaranpassad | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |
| Disposition: | styckeindelning sammanhang textbindning | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |
| Lexikon: | ordförråd konstruktioner, variation idiomatik, exakthet | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |
| Syntax: | ordföljd, meningsbyggnad komplexitet, bestämdhet tempus, passiv | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |
| Form: | morfologi, stavning interpunktion | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |
| Betyg: | | <input type="text"/> | | | |

Skriv gärna fler
kommentarer på baksidan!

Namnteckning (bedömare):

Bilagan fanns med i uppsatsen vid opponeringen, men efter det belades den med sekretess av prefekt Gunlög Sundberg, Institutionen för svenska och flerspråkighet vid Stockholms universitet.

Nominalkvot (n/v)

Bilaga 4

| G (3) | Tokens subst | % av alla tokens i texten | Tokens verb | % verb | Totalt antal tokens | Nominalkvot (n/v) |
|------------|--------------|---------------------------|-------------|--------|---------------------|-------------------|
| Text 1 | 186 | 24 | 166 | 21 | 776 | 1,12 |
| Text 2 | 98 | 24 | 76 | 19 | 404 | 1,28 |
| Text 3 | 95 | 18 | 102 | 20 | 520 | 0,93 |
| Text 5 | 144 | 19 | 170 | 23 | 744 | 0,85 |
| Text 6 | 129 | 26 | 99 | 20 | 499 | 1,30 |
| Text 9 | 89 | 19 | 99 | 21 | 467 | 0,90 |
| Text 10 | 127 | 23 | 98 | 17 | 561 | 1,30 |
| Text 13 | 118 | 26 | 76 | 17 | 458 | 1,55 |
| Medelvärde | 123 | 22 | 111 | 20 | 554 | 1,15 |

| G (5) | Tokens subst | % av alla tokens i texten | Tokens verb | % verb | Totalt antal ord | Nominalkvot (n/v) |
|------------|--------------|---------------------------|-------------|--------|------------------|-------------------|
| Text 23 | 120 | 25 | 99 | 20 | 484 | 1,21 |
| Text 48 | 131 | 22 | 124 | 20 | 608 | 1,06 |
| Text 68 | 95 | 22 | 86 | 20 | 431 | 1,10 |
| Text 72 | 117 | 18 | 128 | 20 | 646 | 0,91 |
| Text 84 | 146 | 20 | 151 | 21 | 720 | 0,97 |
| Text 87 | 127 | 20 | 136 | 22 | 627 | 0,93 |
| Text 95 | 103 | 23 | 91 | 20 | 444 | 1,13 |
| Text 102 | 99 | 20 | 94 | 19 | 502 | 1,05 |
| Medelvärde | 117 | 21 | 114 | 20 | 558 | 1,05 |

| U (2) | Tokens subst | % av alla tokens i texten | Tokens verb | % verb | Totalt antal ord | Nominalkvot (n/v) |
|------------|--------------|---------------------------|-------------|--------|------------------|-------------------|
| Text 16 | 96 | 19 | 108 | 21 | 506 | 0,89 |
| Text 17 | 80 | 17 | 104 | 22 | 469 | 0,77 |
| Text 18 | 125 | 23 | 124 | 23 | 548 | 1,00 |
| Text 19 | 54 | 23 | 50 | 21 | 237 | 1,08 |
| Text 29 | 98 | 16 | 141 | 24 | 600 | 0,69 |
| Text 30 | 71 | 16 | 94 | 21 | 452 | 0,76 |
| Text 31 | 87 | 16 | 113 | 21 | 534 | 0,77 |
| Text 33 | 87 | 18 | 102 | 22 | 471 | 0,85 |
| Text 35 | 59 | 18 | 66 | 20 | 322 | 0,89 |
| Text 36 | 103 | 23 | 107 | 24 | 454 | 0,96 |
| Text 38 | 102 | 22 | 104 | 23 | 459 | 0,98 |
| Text 45 | 93 | 15 | 139 | 23 | 608 | 0,67 |
| Text 49 | 86 | 16 | 119 | 22 | 535 | 0,72 |
| Text 52 | 110 | 16 | 168 | 25 | 671 | 0,65 |
| Text 53 | 88 | 23 | 74 | 19 | 389 | 1,19 |
| Text 55 | 84 | 19 | 99 | 23 | 436 | 0,85 |
| Medelvärde | 89 | 19 | 107 | 22 | 481 | 0,86 |

Substantiv i de 32 Tisustexterna som återfinns i AO

Bilaga 5 (1/3)

(Siffror efter ordet betyder antal förekomster (tokens), ingen siffra betyder en förekomst. Siffrorna t.ex. G3.1 14/19 betyder G3 text nummer 1, 14 typer, 19 tokens av textens subst. som återfinns i AO.)

| | | | |
|---------------|--------------|-------------|--------------|
| G5.23 | 11/18 | förväntning | |
| brist | | krav | 4 |
| forskning | | lösning | 4 |
| grund | | mening | |
| konsekvens | 2 | minne | 2 |
| kontroll | 4 | praktik | |
| lösning | 2 | situation | |
| minne | | | |
| orsak | 2 | G5.87 | 7/14 |
| situation | | antal | |
| verk | | funktion | |
| åsikt | 2 | grund | |
| | | innehåll | |
| G5.48 | 18/21 | kompetens | |
| antal | | lösning | |
| brist | | omvärld | |
| fenomen | | | |
| forskning | | G5.95 | 13/16 |
| fördel | 2 | anledning | |
| företeelse | | antal | 3 |
| förklaring | | faktum | 2 |
| grund | | följd | |
| gräns | | individ | |
| krav | 2 | kollega | |
| lösning | | krav | |
| minne | | lösning | |
| möjlighet | 2 | minne | |
| omständighet | | press | |
| ställning | | process | |
| tillgång | | skillnad | |
| utveckling | | studie | |
| verklighet | | | |
| | | G5.102 | 18/22 |
| G5.68 | 15/19 | allvar | |
| begrepp | | alternativ | |
| bemärkelse | | antal | |
| brist | | brist | 2 |
| form | | effekt | |
| förutsättning | | egenskap | |
| grund | | faktor | |
| inslag | | fall | |
| konsekvens | | fenomen | 2 |
| orsak | 2 | kollega | |
| perspektiv | | krav | |
| situation | 3 | känsla | |
| skillnad | | lösning | |
| status | | minne | |
| utveckling | | press | 2 |
| värde | 2 | situation | 2 |
| | | teori | |
| G5.72 | 12/13 | värdering | |
| bedömning | | G3.1 | 14/19 |
| brist | | aktivitet | |
| forskning | | faktor | 3 |
| grund | | form | |
| gräns | 2 | grund | 4 |
| kollega | | hot | |
| känsla | | individ | |
| minne | | lösning | |
| omgivning | | minne | |
| perspektiv | | orsak | |
| ram | | roll | |
| uppgift | | studie | |
| | | typ | 2 |
| G5.84 | 11/18 | uppgift | |
| ansats | | utrymme | |
| artikel | | | |
| erfarenhet | | G3.2 | 12/16 |
| fall | | behov | 2 |

Bilaga 5 (2/3)

| | | | |
|---------------|--------------|---------------|--------------|
| forskning | 3 | forskning | 3 |
| förmåga | | förhållande | |
| förväntning | | hot | |
| kompetens | | individ | 2 |
| kraft | | inflytande | |
| krav | | läsare | |
| led | | mening | |
| möjlighet | 2 | motsats | |
| orsak | | möjlighet | |
| system | | nivå | 2 |
| utgångspunkt | | position | |
| | | press | |
| G3.3 | 8/10 | sanning | |
| forskning | | situation | 2 |
| grund | 3 | skäl | |
| hot | | uppfattning | |
| känsla | | uppmärksamhet | |
| lösning | | verklighet | |
| roll | | åsikt | |
| situation | | | |
| ålder | | G3.13 | 16/19 |
| G3.5 | 15/19 | allmänhet | |
| aktivitet | | antal | |
| avsnitt | | effekt | 2 |
| fall | | fall | |
| försök | | fenomen | |
| grad | | individ | |
| gräns | 2 | kollega | |
| konsekvens | 2 | konflikt | |
| krav | | konsekvens | |
| minne | | krav | 2 |
| omständighet | | minne | |
| orsak | | omständighet | |
| slutsats | | skillnad | 2 |
| studie | 2 | ställning | |
| typ | 2 | syfte | |
| utveckling | | uttalande | |
| | | | |
| G3.6 | 20/25 | U2.16 | 9/13 |
| ambition | | avsikt | |
| antal | | grund | 3 |
| begrepp | | hot | |
| diskussion | 2 | krav | |
| faktor | | minne | |
| gräns | | modell | |
| lösning | 2 | roll | 3 |
| metod | | skillnad | |
| minne | 2 | typ | |
| mängd | | | |
| möjlighet | | U2.17 | 5/5 |
| påverkan | | forskning | |
| resurs | | grund | |
| roll | | minne | |
| situation | | orsak | |
| status | | universitet | |
| uppmärksamhet | | | |
| utveckling | 2 | U2.18 | 10/13 |
| ålder | | bakgrund | |
| åsikt | | behov | |
| | | gräns | 2 |
| G3.9 | 7/8 | kön | |
| förmåga | | orsak | 3 |
| hot | | samtal | |
| krav | 2 | situation | |
| minne | | status | |
| system | | verklighet | |
| uppmärksamhet | | åsikt | |
| | | | |
| G3.10 | 26/32 | U2.19 | 10/12 |
| allmänhet | | aktivitet | |
| brist | | fall | |
| definition | 2 | förmåga | 2 |
| diskussion | | helhet | |
| engagemang | | lösning | |
| fall | | omständighet | |
| fenomen | | | |

Bilaga 5 (3/3)

| | | | | |
|-------------|-------------|---------------|--------------|-------------|
| osäkerhet | | krav | | |
| sammanhang | | orsak | | |
| situation | 2 | osäkerhet | | |
| svårighet | | system | | |
| | | utveckling | | |
| U2.29 | 8/10 | | | |
| faktum | | U2.45 | 13/18 | |
| fall | | betydelse | | |
| forskning | 3 | faktor | | |
| metod | | forskning | | |
| minne | | förklaring | | |
| möjlighet | | förmåga | 3 | |
| orsak | | historia | | |
| skillnad | | hot | | |
| U2.30 | 7/10 | lösning | | |
| förhållande | | möjlighet | 3 | |
| grund | 3 | situation | | |
| kollega | | uppmärksamhet | | |
| lösning | | ålder | 2 | |
| minne | | åsikt | | |
| orsak | | U2.49 | 9/14 | |
| åsikt | 2 | allvar | | |
| U2.31 | 7/8 | erfarenhet | | |
| forskning | | fördel | | |
| företeelse | | förändring | | |
| hot | | kollega | | |
| minne | | krav | 5 | |
| roll | | möjlighet | | |
| typ | | period | | |
| åsikt | 2 | urval | | |
| U2.33 | 7/9 | U2.52 | 8/9 | |
| erfarenhet | | alternativ | | |
| grund | | behov | | |
| hot | | intresse | | |
| intresse | 2 | minne | | |
| lösning | | möjlighet | | |
| möjlighet | 2 | reaktion | | |
| roll | | situation | 2 | |
| U2.35 | 7/11 | tal | | |
| forskning | 4 | U2.53 | 11/12 | |
| förklaring | 2 | allmänhet | | |
| historia | | fördel | | |
| krav | | gräns | | |
| orsak | | | | kompetens 2 |
| utveckling | | | | |
| U2.36 | 7/10 | konflikt | | |
| förklaring | | krav | | |
| förmåga | | minne | | |
| grund | 4 | möjlighet | | |
| minne | | projekt | | |
| praktik | | situation | | |
| princip | | villkor | | |
| situation | | U2.55 | 5/5 | |
| U2.38 | 9/12 | förmåga | | |
| alternativ | | individ | | |
| erfarenhet | 3 | krav | | |
| grund | | minne | | |
| konsekvens | 2 | nivå | | |

| G (3) | Types subst. | antal subst. i AO | % subst. i AO |
|------------|--------------|-------------------|---------------|
| Text 1 | 102 | 15 | 14,7 |
| Text 2 | 58 | 13 | 22,4 |
| Text 3 | 52 | 10 | 19,2 |
| Text 5 | 91 | 16 | 17,6 |
| Text 6 | 86 | 20 | 23,3 |
| Text 9 | 60 | 7 | 11,7 |
| Text 10 | 74 | 29 | 39,2 |
| Text 13 | 73 | 17 | 23,3 |
| | | | |
| Medelvärde | 75 | 16 | 21,4 |
| G (5) | Types subst. | Antal subst. i AO | % subst i AO |
| Text 23 | 58 | 13 | 22,4 |
| Text 48 | 94 | 18 | 19,1 |
| Text 68 | 60 | 15 | 25 |
| Text 72 | 89 | 13 | 14,6 |
| Text 84 | 97 | 12 | 12,4 |
| Text 87 | 77 | 9 | 11,7 |
| Text 95 | 79 | 15 | 19 |
| Text 102 | 79 | 20 | 25,3 |
| | | | |
| Medelvärde | 79 | 14 | 18,7 |

| U (2) | Types subst. | Antal subst. i AO | % subst. i AO |
|------------|--------------|-------------------|---------------|
| Text 16 | 62 | 10 | 16,1 |
| Text 17 | 45 | 5 | 11,1 |
| Text 18 | 62 | 11 | 17,7 |
| Text 19 | 39 | 10 | 25,6 |
| Text 29 | 71 | 8 | 11,3 |
| Text 30 | 42 | 8 | 19 |
| Text 31 | 49 | 9 | 18,4 |
| Text 33 | 66 | 9 | 13,6 |
| Text 35 | 39 | 7 | 17,9 |
| Text 36 | 56 | 9 | 16,1 |
| Text 38 | 60 | 9 | 15 |
| Text 45 | 66 | 14 | 21,2 |
| Text 49 | 54 | 13 | 24,1 |
| Text 52 | 69 | 8 | 11,6 |
| Text 53 | 63 | 11 | 17,5 |
| Text 55 | 42 | 5 | 11,9 |
| Totalt | | 146 | |
| Medelvärde | 55 | 9 | 16,8 |

Bilaga 7

| Tisustexter G (3) | Totalt antal ord | Types | Antal substantiviska sammansättningar | % substantiviska sammansättningar |
|-------------------|------------------|-------|---------------------------------------|-----------------------------------|
| Text 1 | 776 | 341 | 32* | 4,1 |
| Text 2 | 404 | 230 | 20* | 5,0 |
| Text 3 | 520 | 234 | 5 | 1,0 |
| Text 5 | 744 | 341 | 23* | 3,0 |
| Text 6 | 499 | 286 | 21* | 4,2 |
| Text 9 | 467 | 241 | 11* | 2,4 |
| Text 10 | 561 | 309 | 12* | 2,1 |
| Text 13 | 458 | 248 | 10 | 2,2 |
| Summa | 4429 | 2230 | 134 | |
| Medelvärde | 554 | 279 | 16 | 3 |

(* upprepning av samma sammansättning har medräknats)

| Tisustexter G (5) | Totalt Antal ord | Types | Antal substantiviska sammansättningar | % substantiviska sammansättningar |
|-------------------|------------------|-------|---------------------------------------|-----------------------------------|
| Text 23 | 484 | 218 | 10* | 2,1 |
| Text 48 | 608 | 335 | 27* | 4,4 |
| Text 68 | 431 | 231 | 11 | 2,6 |
| Text 72 | 646 | 341 | 19* | 2,9 |
| Text 84 | 720 | 344 | 21* | 2,9 |
| Text 87 | 627 | 298 | 36* | 5,7 |
| Text 95 | 444 | 265 | 15 | 3,4 |
| Text 102 | 502 | 276 | 16 | 3,2 |
| Summa | 4462 | 2308 | 155 | |
| Medelvärde | 558 | 289 | 19 | 3,4 |

G3 + G5 % sammansatta ord i genomsnitt = 3,2
 Antal sammansatta ord (tokens) i genomsnitt = 18
 Totalt antal ord (tokens) i genomsnitt = 556

| Tisustexter U (2) | Totalt antal ord | Types | Antal substantiviska sammansättningar | % substantiviska sammansättningar |
|-------------------|------------------|-------|---------------------------------------|-----------------------------------|
| Text 16 | 506 | 238 | 3 | 0,6 |
| Text 17 | 469 | 215 | 3 | 0,6 |
| Text 18 | 548 | 244 | 8* | 1,5 |
| Text 19 | 237 | 151 | 11* | 4,6 |
| Text 29 | 600 | 274 | 11 | 1,8 |
| Text 30 | 452 | 193 | 6* | 1,3 |
| Text 31 | 534 | 223 | 10* | 1,9 |
| Text 33 | 471 | 257 | 10* | 2,1 |
| Text 35 | 322 | 175 | 8* | 3,1 |
| Text 36 | 454 | 212 | 18* | 4 |
| Text 38 | 459 | 234 | 11* | 2,4 |
| Text 45 | 608 | 267 | 8* | 1,3 |
| Text 49 | 535 | 233 | 14* | 2,6 |
| Text 52 | 671 | 315 | 11* | 1,6 |
| Text 53 | 389 | 206 | 13* | 3,3 |
| Text 55 | 436 | 194 | 6* | 1,4 |
| Summa | 7691 | 3631 | 151 | |
| Medelvärde | 481 | 227 | 10 | 2,1 |

Tre led

G3.1: arbetslivserfarenhet, fritidsaktivitet
 G3.2: tillhandahållande, utgångspunkt, övernattningsrum
 G3.6: korttidsstress, långtidsminne, långtidsstress, överlevnadssituation
 G5.48: fritidsaktivitet, fritidsintresse
 G5.84: deltidjobb, kom-ihåg-listor
 G5.87: friskvårdsförmån, ”stressbehandlingstjänst”
 G5.95: friskvårdspersonal
 U2.19: livsbakgrund
 U2.29: minnesträningsuppsats, minnesträningskurs
 U2.45: nykomlingsproblem
 U2.49: sysselsättningspolitik
 U2.52: arbetsmiljölag
 U2.53: daglig-jobb-konkurrens, sjukledighetsdag

Fyra led

G3.9: fritidssysselsättning

Dubbel syftning

G3.3: alkohol- och drogmissbruk
 G3.5: arbets- och privatliv, del- eller heltidsjobb
 G5.87: flex- eller övertid

Särskrivna sammansättningar

G3.1: arbets uppgift (2), diskussions tema. fritids aktivitet, gratis biljett, rast tid
 G3.5: extra jobb, föräldrar möte, industri samhälle
 G3.6: företags ledare, stress faktorer, överlevnads situation
 G3.13: aerobics klass, friskvård grupper
 G5.72: flock djur
 G5.87: projekt anställningar, sjukskrivnings period
 U2:17: huvud orsaken
 U2:18: mötes tid, vardag liv
 U2.29: minnestränings kursen, rättighets problem, varje dags livet
 U2.33: fika rasten
 U2.35: utveckling länderna
 U2.36: livs situation, privat livet
 U2.45: nykomling problem, stress faktor, över levnad
 U2.49: fri tid, invandrar politiken, mellan mjölkets, prestations krav, prov anställningar,
 stress nivå (2), utvecklings möjligheter, välfärd system
 U2.52: orolighets moment, privat liv, stress symptom
 U2.53: blod trycket

| Antal | U2 | G3 | G5 | Exempel |
|---------------------------------|-----------|-----------|-----------|-----------------------|
| tre led | 8 | 9 | 7 | fritidsintresse |
| fyra led | - | 1 | - | fritidssysselsättning |
| dubbel syftning drogmissbruk | - | 3 | 1 | alkohol- och |
| särskrivningar | 26 | 14 | 3 | privat livet |

Antal stavelser per substantivisk sammansättning / tokens
(särskrivningar medräknade)

Bilaga 9

| | 2 | 3 | 4 | 5 | 6 | 7 | |
|---------------|----------|----------|----------|----------|----------|----------|-----------|
| G3.1 | 3 | 10 | 8 | 9 | 1 | 1 | 32 |
| G3.2 | 3 | 11 | 1 | 4 | 1 | | 20 |
| G3.3 | 1 | 1 | 1 | 2 | | | 5 |
| G3.5 | 6 | 9 | 3 | 4 | 1 | | 23 |
| G3.6 | 6 | 7 | 6 | | 1 | 1 | 21 |
| G3.9 | 4 | 1 | 2 | 3 | 1 | | 11 |
| G3.10 | 1 | 3 | 3 | 4 | 1 | | 12 |
| G3.13 | <u>2</u> | <u>2</u> | <u>5</u> | <u>1</u> | | | <u>10</u> |
| Totalt antal: | 26 | 44 | 29 | 27 | 6 | 2 | 134 |

| | 2 | 3 | 4 | 5 | 6 | 7 | 8 | |
|---------------|----------|----------|----------|----------|----------|----------|----------|-----------|
| G5.23 | 2 | 5 | 1 | 2 | | | | 10 |
| G5.48 | 6 | 11 | 5 | 2 | 3 | | | 27 |
| G5.68 | 1 | 3 | 3 | 2 | 1 | | 1 | 11 |
| G5.72 | 4 | 5 | 4 | 3 | 3 | | | 19 |
| G5.84 | 8 | 5 | 3 | 4 | 1 | | | 21 |
| G5.87 | 9 | 13 | 3 | 8 | 3 | | | 36 |
| G5.95 | 2 | 7 | 2 | 2 | 2 | | | 15 |
| G5.102 | <u>3</u> | <u>5</u> | <u>4</u> | <u>4</u> | | | | <u>16</u> |
| Totalt antal: | 35 | 54 | 25 | 27 | 13 | 0 | 1 | 155 |

G3+G5 61 98 54 54 19 2 1 289

| | | | | | | | | |
|---------------|----------|----------|----------|----|---|---|--|----------|
| U2.16 | 1 | 1 | 1 | | | | | 3 |
| U2.17 | 1 | 1 | | 1 | | | | 3 |
| U2.18 | 2 | 6 | | | | | | 8 |
| U2.19 | 3 | 5 | 3 | | | | | 11 |
| U2.29 | 3 | 2 | 3 | 1 | 2 | | | 11 |
| U2.30 | 5 | 1 | | | | | | 6 |
| U2.31 | 1 | 4 | 3 | 2 | | | | 10 |
| U2.33 | 5 | 4 | | | 1 | | | 10 |
| U2.35 | | 5 | 1 | 1 | 1 | | | 8 |
| U2.36 | 1 | 3 | 10 | 4 | | | | 18 |
| U2.38 | 2 | 1 | 4 | 4 | | | | 11 |
| U2.45 | | 3 | 4 | 1 | | | | 8 |
| U2.49 | 3 | 3 | 4 | 1 | 1 | 2 | | 14 |
| U2.52 | 1 | 7 | 1 | 1 | 1 | | | 11 |
| U2.53 | 2 | 4 | 3 | 2 | 2 | | | 13 |
| U2.55 | <u>4</u> | <u>1</u> | <u>1</u> | | | | | <u>6</u> |
| Totalt antal: | 34 | 51 | 38 | 18 | 8 | 2 | | 151 |

Bilaga 10 (1/2)

Jämförelse mellan substantiv i Tisustexterna och de frekvensbaserade ordlistorna 1000-5000 forts.

| G3 | subst 1000 % | subst 2000 % | subst 3000 % | subst 4000 % | subst 5000 % | subst totalt % 1- 5000 | Subst % inte 1- 5000 | TC % hela texten 1-5000 |
|----|--------------------|--------------------|--------------------|--------------------|--------------------|------------------------------------|-------------------------------|----------------------------------|
| 1 | 25,5 | 13,7 | 10,8 | 7,8 | 6,9 | 64,7 | 35,3 | 85,3 |
| 2 | 19,0 | 10,3 | 12,1 | 6,9 | 10,3 | 58,6 | 41,4 | 82,4 |
| 3 | 28,8 | 26,9 | 17,3 | 1,9 | 7,7 | 82,6 | 17,3 | 90,3 |
| 5 | 23,1 | 12,1 | 9,9 | 3,3 | 9,9 | 58,3 | 41,8 | 82,7 |
| 6 | 24,4 | 17,4 | 15,1 | 2,3 | 9,3 | 68,5 | 31,4 | 84,3 |
| 9 | 30 | 16,7 | 13,3 | 6,7 | 5,0 | 71,7 | 28,3 | 89,7 |
| 10 | 20,5 | 19,2 | 16,4 | 8,2 | 6,8 | 71,1 | 28,8 | 82 |
| 13 | 27,4 | 17,8 | 16,4 | 2,7 | 4,1 | 68,4 | 31,5 | 83,8 |

| G5 | subst 1000 % | subst 2000 % | subst 3000 % | subst 4000 % | subst 5000 % | subst totalt % 1- 5000 | Subst WT % inte 1- 5000 | TC % hela texten 1-5000 |
|-----|--------------------|--------------------|--------------------|--------------------|--------------------|------------------------------------|-------------------------------------|----------------------------------|
| 23 | 29,3 | 15,5 | 15,5 | 5,2 | 8,6 | 74,1 | 25,9 | 85,9 |
| 48 | 19,1 | 12,8 | 22,3 | 3,2 | 8,5 | 65,9 | 34 | 81,6 |
| 68 | 28,3 | 15 | 10 | 3,3 | 3,3 | 59,9 | 40 | 84,5 |
| 72 | 24,7 | 15,7 | 14,6 | 4,5 | 2,2 | 61,7 | 38,2 | 84 |
| 84 | 29,6 | 15,3 | 8,2 | 4,1 | 4,1 | 61,3 | 38,8 | 87,2 |
| 87 | 24,4 | 9,8 | 11 | 4,9 | 6,1 | 56,2 | 43,9 | 86,1 |
| 95 | 18,8 | 10,6 | 7,6 | 2,9 | 9,4 | 50,6 | 49,3 | 81 |
| 102 | 22,8 | 13,9 | 8,9 | 5,1 | 15,2 | 65,9 | 34,2 | 82,5 |

Bilaga 10 (2/2)

Jämförelse mellan substantiv i Tisustexterna och de frekvensbaserade ordlistorna 1000-5000 forts.

| U2 | subst 1000 % | subst 2000 % | subst 3000 % | subst 4000 % | subst 5000 % | subst totalt % 1- 5000 | Subst % inte 1- 5000 | TC % hela texten 1-5000 |
|----|--------------------|--------------------|--------------------|--------------------|--------------------|------------------------------------|-------------------------------|----------------------------------|
| 16 | 35,5 | 22,6 | 9,7 | 4,8 | 8,1 | 80,7 | 19,4 | 85,4 |
| 17 | 35,6 | 24,4 | 13,3 | 2,2 | 6,7 | 82,2 | 17,8 | 91,3 |
| 18 | 33,9 | 17,7 | 9,7 | 8,1 | 8,1 | 77,5 | 22,6 | 91 |
| 19 | 33,3 | 12,8 | 17,9 | 7,7 | 0 | 71,7 | 28,2 | 88,2 |
| 29 | 36,6 | 16,9 | 5,6 | 2,8 | 8,5 | 70,4 | 29,6 | 91,5 |
| 30 | 26,2 | 11,9 | 16,7 | 2,4 | 11,9 | 69,1 | 31 | 85,4 |
| 31 | 26,5 | 14,3 | 16,3 | 4,1 | 10,2 | 71,4 | 28,6 | 89,7 |
| 33 | 30,3 | 16,7 | 15,2 | 4,5 | 7,6 | 74,3 | 25,8 | 87,2 |
| 35 | 23,1 | 17,9 | 15,4 | 7,7 | 7,7 | 71,8 | 28,2 | 79,2 |
| 36 | 19 | 19 | 12,1 | 6,9 | 5,2 | 62,2 | 37,9 | 80,4 |
| 38 | 30 | 18,3 | 8,3 | 5 | 6,7 | 68,3 | 31,7 | 84,6 |
| 45 | 28,8 | 15,2 | 19,7 | 1,5 | 6,1 | 71,3 | 28,8 | 88,6 |
| 49 | 24,1 | 20,4 | 14,8 | 0 | 11,1 | 70,4 | 29,6 | 87,2 |
| 52 | 29 | 24,6 | 8,7 | 5,8 | 5,8 | 73,9 | 26,1 | 90,2 |
| 53 | 23,8 | 19 | 11,1 | 3,2 | 7,9 | 65 | 34,9 | 83,2 |
| 55 | 25 | 20 | 12,5 | 2,5 | 7,5 | 67,5 | 32,5 | 85,7 |

| Medelvärden | subst 1000 % | subst 2000 % | subst 3000 % | subst 4000 % | subst 5000 % | subst totalt % 1- 5000 | Subst % inte 1- 5000 | TC % hela texten 1- 5000 |
|--------------------|--------------------|--------------------|--------------------|--------------------|--------------------|------------------------------------|-------------------------------|--------------------------------------|
| G5 | 24,6 | 13,6 | 12,3 | 4,1 | 7,2 | 62 | 38 | 84,1 |
| G3 | 24,8 | 16,7 | 13,9 | 5 | 7,5 | 68 | 32 | 85 |
| G3+G5 | 24,7 | 15,2 | 13 | 4,6 | 7,3 | 65 | 33,2 | 84,6 |
| U2 | 28,8 | 18,2 | 12,9 | 4,3 | 7,4 | 71,7 | 28,3 | 86,8 |