

FORORD

Det er flere personer som fortjener en stor takk i forbindelse med denne masteroppgaven. Først og fremst vil jeg takke den fantastiske veilederen min, Kristin Melum Eide, som tipset meg om temaet til denne masteroppgaven. Tusen takk for store mengder inspirasjon, kyndig veiledning og fyldige tilbakemeldinger. Takk for godt humør som alltid smitter, engasjement for faget ditt, og for god humor. Takk for gode råd, og for at du får meg til å tro på meg selv og det jeg gjør.

Jeg vil også rette en spesiell takk til Kari Tenfjord, som var så snill at hun inviterte meg hjem til henne på ASK-kurs (ASK-kurset jeg egentlig skulle på ble avlyst, men jeg hadde allerede bestilt flybilletter til Bergen). Dette var et veldig godt, informativt og hyggelig møte som hjalp meg veldig. I tillegg har Kari hjulpet meg gjennom flere runder på e-post, i arbeidet med å finne de riktige søkeuttrykkene i ASK. Dette har jeg satt umåtelig stor pris på. Olav Husby fortjener også en takk for å ha hjulpet meg (indirekte gjennom Kristin) med vietnamesiske eksempelfraser.

Lars har vært en helt uerstattelig støtte. Tusen takk, Lars, for at du er så tålmodig, snill, avslappet, og for at du setter alt i perspektiv. Takk til Torunn for ekstra doser motivasjon når jeg har trengt det, og til pappa, Hans Petter, for hjelp i sammenheng med tyske eksempelfraser.

Trondheim, mars 2016

Solveig Gjerdset

Innhold

1 INNLEDNING	1
1.1 Disposisjon	3
2 TEORETISK GRUNNLAG	5
2.1 Dobbel bestemthet	5
2.2 Andrespråk, andrespråksinnlærer og kritisk periode	6
2.3 Mellomspråk	8
2.4 Prinsipp- og parameterteorien	9
2.5 Transfer	12
2.6 Kompetanse og performanse	14
3 METODE	17
3.1 Korpuslingvistikk	17
3.1.1 Kritikken mot korpuslingvistikk	17
3.1.2 Moderne korpus	18
3.1.3 Fordeler og problemer med moderne korpus	20
3.2 ASK – norsk andrespråskorpus	21
3.2.1 Designkriterer i ASK	22
3.2.2 Søk i ASK	23
3.3 Jarvis’ metode for påvisning av tverrspråklig innflytelse	25
3.4 Signifikansnivå	27
4 UNDERSØKELSEN	29
4.1 Valg av språk	29
4.1.1 Tysk	30
4.1.2 Engelsk	31
4.1.3 Vietnamesisk	31
4.1.4 Oversikt over språklige trekk	32
4.2 Empiri	33
4.2.1 Søk i ASK	33
4.2.2 Kategorisering	35
4.2.3 Engelskkunnskaper som variabel	36
4.3 Resultater og analyse	38
4.3.1 Frekvens	39

4.3.2 Avvik.....	42
4.3.3 Oppsummering av resultater.....	50
4.3.4 Avsluttende oppsummering.....	51
LITTERATUR.....	55
VEDLEGG 1: SØK – NORSK KONTROLLGRUPPE.....	59
VEDLEGG 2: SØK – TYSK FØRSTESPRÅKSGRUPPE.....	63
VEDLEGG 3: SØK – ENGELSK FØRSTESPRÅKSGRUPPE.....	67
VEDLEGG 4: SØK – VIETNAMESISK FØRSTESPRÅKSGRUPPE.....	71
VEDLEGG 5: MASTERARBEIDETS RELEVANS.....	75
VEDLEGG 6: SAMMENDRAG.....	77
VEDLEGG 7: ABSTRACT.....	79

1 INNLEDNING

Denne masteroppgaven er skrevet innenfor fagfeltet norsk som andrespråk, med bakgrunn i chomskyansk generativ grammatikk. Forskning på andrespråklæring er en relativt ny fagdisiplin som har vokst fram særlig i løpet av de siste tiårene (Berggreen & Tenfjord, 2011: 15). Med større språkkontakt og migrering over hele verden, er det større behov enn noen gang for kunnskap om andrespråk og prosessene som er involvert i læring av andrespråk. Dette er også en av årsakene til at jeg synes dette fagfeltet er så spennende. Som kommende lektor er dette noe jeg utvilsomt vil erfare i eget klasserom, og som det vil være verdifullt å ha kompetanse i. Emnet for denne masteroppgaven er nominale fraser, nærmere bestemt *dobbel bestemthets*-konstruksjonen. Kort forklart er dobbel bestemthet en konstruksjon som både har et fritt foranstilt determinativ og en etterhengt artikkel på nomenet, som begge uttrykker bestemthet – som i frasen ”den grønne bilen”. Disse markørene må også kongruere i både numerus og genus. Fenomenet vil forklares nærmere under 2.1.

Jeg har valgt akkurat dette emnet av flere årsaker. For det første er dobbel bestemthet et svært sjeldent og eksotisk fenomen i språk på verdensbasis – konstruksjonen er svensk, norsk og færøysk (Haider, Olsen, & Vikner, 1995: 25). Ettersom det er en såpass sjelden konstruksjon, finnes det derfor foreløpig lite forskning på emnet. For det andre er konstruksjonen redundant – informasjonen som meddeles er overflødig¹, og man kan fint gjøre seg forstått uten den. Redundante trekk, som ulike bøyingskategorier og kongruens, er noe som andrespråksinnlærere ofte har en del problemer med (Berggreen & Tenfjord, 2011: 32). Ettersom dobbel bestemthet både er et sjeldent og redundant fenomen, mener jeg det vil være svært interessant å foreta en komparativ undersøkelse av konstruksjonen. Dette vil jeg gjøre med utgangspunkt i informantgrupper med ulike førstespråk. De ulike førstespråksgruppene er valgt ut fra hvilke egenskaper førstespråket deres har når det gjelder trekk som opptrer i nominale fraser. Ingen av språkene har imidlertid dobbel bestemthetskonstruksjonen som i norsk. Målet med undersøkelsen er å finne svar på følgende problemstilling:

1) Hva karakteriserer bruken av dobbel bestemthets-konstruksjoner i norsk mellomspråk hos voksne andrespråksinnlærere med tysk, engelsk og vietnamesisk som førstespråk, med tanke på det rent grammatiske og når det gjelder relativ frekvens?

¹ Det finnes også situasjoner der informasjonen som meddeles ikke er overflødig: Julien (2002) forklarer at det finnes en arbeidsfordeling mellom fraser som ”den hvite mann” og ”den hvite mannen” (denne informasjonen er hentet fra Roehrs 2009: 58). Dette er avansert teori som jeg ikke vil gå nærmere inn i her, og som jeg velger å se bort fra i denne masteroppgaven.

2) *Varyerer disse karakteristikaene ut fra hvilket førstespråk man har, og kan man eventuelt påvise tverrspråklig innflytelse ut fra resultatene?*

De ulike morsmålsgruppene jeg skal undersøke er altså tysk, engelsk og vietnamesisk. Språkene er som nevnt valgt på bakgrunn av deres ulike egenskaper i sammenheng med nominalfraser. Norsk har morfologisk uttrykte markører for numerus, genus og definitthet, og krever synlige kongruenstrekk både på adjektivet og determinativet for disse tre grammatiske kategoriene (Jin & Eide, 2015: 295). Tysk har de samme trekkene som er nevnt over, men har ikke den etterhengte artikkelen på nomenet som er med på å skape dobbel bestemthets-konstruksjonen i norsk. Engelsk, på sin side, har morfologisk uttrykte markører for numerus og definitthet, og det kreves kun synlige kongruenstrekk for numerus på determinativet (Jin & Eide, 2015: 302-303). Til sammenligning har vietnamesisk verken numerus, genus eller definitthet (og derfor heller ingen kongruens når det gjelder disse grammatiske kategoriene). Ettersom språkene er såpass ulike, mener jeg at en komparativ undersøkelse vil være svært interessant. Språkene vil presenteres i nærmere detalj under 4.1.

Når man skal utføre komparative undersøkelser av denne arten, må man helst ha tilgang til store mengder data. Datagrunnlaget til denne masteroppgaven er hentet fra norsk andrespråskorpus (ASK), et søkbart parallellkorpus med data hentet fra Norsk Språktests arkiv over innvandrere som har tatt *Språkprøven i norsk* og *Test i norsk – høyere nivå* (Universitetet i Bergen, 2014). I ASK-korpuset er det hundre besvarelser per morsmålsgruppe, noe som også gjør det mulig å finne statistisk signifikante resultater basert på tekstene i korpuset (Tenfjord, Hagen & Johansen, 2009: 5). Undersøkelsen jeg foretar vil i hovedsak være kvantitativ, og basere seg på besvarelser fra andrespråksinnlærere som har tatt *Språkprøven i norsk*. Informanttekstene jeg benytter meg av vil derfor ha bakgrunn i den samme testsituasjonen, noe som gir datagrunnlaget mitt en større reliabilitet. Jeg benytter meg av Jarvis' (2010) metode for påvisning av tverrspråklig innflytelse i undersøkelsen, som vil diskuteres nærmere under 3.3.

Min utgangshypotese for denne undersøkelsen er at alle informantgruppene vil ha noen problemer knyttet til dobbel bestemthets-konstruksjonen, men at den tyske førstespråksgruppen vil ha minst vanskeligheter ettersom det tyske nominalsystemet er mest likt det norske i denne undersøkelsen. Vietnamesisk har et nominalsystem som skiller seg fra de andre, og som er klart minst likt norsk. Derfor vil det være rimelig å anta at den vietnamesiske førstespråksgruppen vil ha størst problemer. Dette kan gi utslag i avvik, eller

eventuelt i en underproduksjon eller unnvikende produksjon av konstruksjonen (mer om disse fenomenene i kapittel 2.3).

Resultatene som kommer fram av denne masteroppgaven, vil være et bidrag til andrespråksforskningen, særlig ettersom det finnes såpass lite forskning på akkurat dette emnet fra før. I tillegg mener jeg også at masteroppgaven kan ha nytteverdi i sammenheng med andrespråksundervisning, som blir mer og mer relevant for enhver norsklærer.

1.1 Disposisjon

Denne masteroppgaven er delt inn i tre hoveddeler, samt en innledning. Den første hoveddelen danner et teoretisk bakteppe for undersøkelsen min, og tar opp og forklarer de viktigste begrepene og teoriene som brukes i teksten. Andre hoveddel er et metodekapittel. Her vil jeg diskutere korpussøk som metode, korpus som metodologi, hvordan ASK er satt sammen og hvordan søkefunksjonen i korpuset fungerer, og hvordan man kan påvise tverrspråklig innflytelse (eller transfer) med *Jarvis' metode*. Den siste hoveddelen presenterer selve undersøkelsen. Her vil jeg først begrunne valg av førstespråksgrupper i denne undersøkelsen, forklare empirien for undersøkelsen min, og kontrollere for engelskkunnskaper som variabel. Til slutt vil resultatene presenteres og analyseres.

2 TEORETISK GRUNNLAG

I denne delen av oppgaven vil jeg gi masteroppgaven min en teoretisk forankring, og jeg vil forklare de viktigste begrepene som blir brukt. Først vil jeg gå nærmere inn på hva dobbel bestemthet er i 2.1, og hvorfor det etter min mening er et så spennende fenomen å undersøke i andrespråkssammenheng. Deretter vil jeg gjøre rede for begrepene *andrespråk*, *andrespråksinnlærer* og *kritisk periode* under 2.2, før jeg under 2.3 vil forklare termen *mellomspråk*. Under 2.4 vil jeg framstille *Prinsipp- og parameterteorien* og en del begreper som hører til under denne, og forklare hvorfor teorien relevant for andrespråksforskning og for min undersøkelse. Jeg vil så diskutere begrepet *transfer* og potensiell problematikk rundt dette under 2.5. Helt til slutt problematiserer jeg undersøkelsen min gjennom diskusjon av begrepene *kompetanse* og *performanse* under 2.6.

2.1 Dobbelt bestemthet

Dobbelt bestemthet (eller dobbelt definitthet, som det også kalles) er en konstruksjon der man både har et fritt foranstilt determinativ og en etterhengt artikkel på nomenet, som begge uttrykker bestemthet (definitthet). Disse markørene må også kongruere i både numerus og genus. Determinativet brukes og oppfattes riktignok vanligvis bare som en bestemt artikkel når ordene står sammen med adjektiv (Golden, Mac Donald, & Ryen, 2014: 46). Determinativet fungerer altså ikke som demonstrativ i denne situasjonen, med mindre man legger trykk på det (merk at asterisken * betyr at konstruksjonen er ugrammatikalsk):

1 (a)

Norsk: **Det** røde huset

*Røde huset

*Det røde hus

(b)

Dansk: **Det** røde hus

*Røde hus

*Det røde huset

I tillegg til dette må adjektivet i frasen stå i bestemt form (det må med andre ord ha svak bøyning), og det må kongruere med substantivet. Når adjektivet står i bestemt form har de fleste adjektivene riktignok den samme e-endingen:

2 (a)

kaldt – *kalde* (nøyt.)

(b)

kald – *kalde* (mask. og fem.)

(c)

kalde – *kalde* (plu.)

Adjektivet kan også gradbøyes innenfor konstruksjonen (Golden, et al., 2014 : s. 20). Jeg undersøker altså ikke bare fraser av typen "*det røde huset*", men også av typen "*det rødeste*

huset". I tillegg må også konstruksjoner med ordenstall også regnes med, som for eksempel "*det første huset*", "*det andre huset*" og så videre.

Dobbel bestemthet en ren formell konstruksjon, der noe av informasjonen som meddeles er semantisk overflødig. Man "trenger" strengt tatt ikke en dobbel markør i frasen for å formidle bestemthet, og kongruensen i frasen er også redundant. Her er det altså flere rent formelle krav. Dermed er det store mengder informasjon som ikke er nødvendig kommunikasjonsmessig: Man forstår svært godt hva som menes om noen bare sier "rød hus". Dette er noe av det som gjør dobbel bestemthet veldig interessant i et andrespråksperspektiv. Redundante språktrekk utvikles som regel sent i andrespråksinnlærerens språk, og enkelte trekk læres kanskje aldri (Berggreen & Tenfjord, 2011: 32). Det er dessuten påvist at kongruenstrekk er et område der mange voksne andrespråksinnlærere har vedvarende problemer (Jin & Eide, 2015: 287). Av disse årsakene mener jeg dobbel bestemthet er særlig interessant: Det er et sjeldent fenomen som innlærerne ikke har innarbeidet i sitt eget morsmål, og flere forhold tyder på at det er komplisert å lære denne typen konstruksjoner. Spørsmålet er om innlæringen varierer på tvers av ulike førstespråksbakgrunn, og eventuelt hvorfor. Dette kommer jeg særlig tilbake til i 2.4, der jeg tar for meg prinsipp- og parameterteorien og hvordan den er relevant for denne undersøkelsen.

2.2 Andrespråk, andrespråksinnlærer og kritisk periode

Termen *andrespråk* ser ved første øyekast ut til å være nokså selvforklarende. Lærer man et andrespråk, lærer man et språk i tillegg til *førstespråket* sitt (altså morsmålet). Begrepet dekker ikke bare ett, men alle språk som blir lært etter førstespråket – både andre, tredje, fjerde etc. lærte språk (Berggreen & Tenfjord, 2011: 16). En person som lærer et andrespråk har ulike navn i andrespråkslitteraturen. I denne masteroppgaven vil personer som lærer andrespråk omtales som *andrespråksinnlærere* (eventuelt bare *innlærere*). Årsaken til dette er at en andrespråkslærer fort kan oppfattes som noen som underviser et andrespråk, framfor noen som lærer det.

Det er i praksis to måter andrespråkslæring kan skje på. Den første måten skjer hovedsakelig i en klasseromssituasjon, der man lærer et andrespråk i sitt eget førstespråksmiljø, med liten eller ingen tilgang på faktiske førstespråksbrukere av det språket som læres. Dette blir gjerne kalt *fremmedspråksundervisning*, og er ikke den typen andrespråksinnlæring som er interessant for de fleste andrespråksforskere (Åfarli, 2015: 60). Den andre typen, som forskerne er mest interesserte i, er den som skjer naturalistisk. Det vil si en andrespråkslæring som skjer når innlæreren har stor påvirkning fra andrespråket i

dagliglivet. Dette er en situasjon som innvandrere typisk befinner seg i (Åfarli, 2015: 61). Det er den naturalistiske måten å tilegne seg andrespråk på som jeg skal undersøke i denne masteroppgaven. Når jeg bruker begrepet andrespråk, er det altså kun denne varianten jeg omtaler, og ikke fremmedspråk. Informantene mine har riktignok fått norskopplæring, men ettersom dette har skjedd innenfor en naturalistisk kontekst, med eksponering for norsk språk i dagliglivet, vil jeg likevel regne dette som andrespråksinnlæring og ikke fremmedspråklæring.

En rekke framstående forskere innenfor andrespråksfeltet poengterer også at det er av stor betydning *når* andrespråket blir lært. Det å si at et andrespråk blir lært etter førstespråket, er altså en noe vag formulering. Berggreen og Tenfjord presiserer at det er vanlig å sette et skille ved treårsalderen: Lærer man to språk før man har fylt tre år, regnes man vanligvis som tospråklig (2011: 16). Man vil altså ha to førstespråk om man begynner å lære begge språkene før man har fylt tre år. Årsaken til at det er fokus på dette, er at alderen ser ut til å ha svært mye å si for hvordan man lærer språk. Allerede på 60-tallet formulerte psykologen Lenneberg en hypotese som sa at alle har en medfødt evne til språklæring, som bare er intakt fram til individet når puberteten (Berggreen & Tenfjord, 2011: 54). Dette ble omtalt som den *kritiske perioden* for språkinnlæring. Dette begrepet har bestått, selv om innholdet har forandret seg noe. Det har vært en del uenigheter om hvor lang denne perioden er, og om den kan deles inn i kritiske perioder for ulike språklige moduler (Eide & Busterud, 2015: 39). Det de aller fleste andrespråksforskerne imidlertid er enige om, er at det *finnes* en eller flere kritiske perioder for språklæring.

Det vanligste er å anta at en startalder på seks-sju år som regel fører til en aksentfri og innfødtlik målspråkskompetanse, og at grensen for tilegnelse av grammatikk, morfologi og syntaks går ved puberteten (Eide & Busterud, 2015: 40). Når man er voksen, går man gjennom en mye raskere morfologisk og syntaktisk utvikling enn det barn gjør. Men de som starter språklæringen som barn, oppnår høyere sluttnivå enn de som starter som voksne i det lange løp. Dette omtales ofte som ”Older is faster, but younger is better” i andrespråksmiljøet (Berggreen & Tenfjord, 2011: 280). I denne masteroppgaven undersøker jeg kun språket til voksne andrespråksinnlærere, det vil si andrespråksinnlærere som alle har passert den kritiske perioden for språklæring. Informantene mine vil derfor med stor sannsynlighet ha større vanskeligheter med å tilegne seg norsk enn andrespråksinnlærere som begynner å lære norsk i barndommen. Dermed har mine informanter stort sett det samme utgangspunktet, noe som gir datagrunnlaget for denne undersøkelsen større reliabilitet.

2.3 Mellomspråk

Når man arbeider med andrespråklæring, er begrepet *mellomspråk* av stor betydning. Om noen lærer seg et andrespråk, bygger vedkommende opp en egen internalisert grammatikk og sitt eget sett med regler for hvordan språket skal brukes (Husby, 2015: 420). Dette systemet blir mer og mer komplisert med tiden, og det vil ligne mer på målspråket etter hvert som innlæreren tilegner seg mer kompetanse. Grammatikksystemet som bygges opp av hver enkelt innlærer kalles et *mellomspråk*. Dette var en term som oppsto som et alternativ til å beskrive innlærerens språk basert på avvik fra målspråket, eller det man tradisjonelt har kalt *feil*. Begrepet går også et skritt videre ved å forsøke å beskrive andrespråksinnlærerens grammatikk som en helhet (Husby, 2015: 420). Slik ser man ikke lenger bare på avvik fra normen, men på alle sider av innlærerens språk. Når man bruker begrepet *mellomspråk*, er altså hele språkssystemet til andrespråksinnlæreren interessant. I denne masteroppgaven er det nettopp *mellomspråket* til ulike andrespråksinnlærere jeg undersøker. Når jeg skriver om avvik betyr dette avvik fra målspråksnormen, og har ikke noe med ”feil” innad i *mellomspråket* å gjøre.

Selv om *mellomspråk* er ulike fra person til person, er det bred enighet blant andrespråksforskere om at det finnes det en del generelle trekk som karakteriserer disse grammatikksystemene. For det første er *mellomspråkene* strukturelt enkle språk i tidlige faser (Berggreen & Tenfjord, 2011: 29). Dette betyr at redundante språktrekk, som for eksempel bøyningkategorier, bare er til stede i liten grad. Slike trekk kan falle på plass senere i språklæringen, men *mellomspråket* kan også *fossiliseres* før dette skjer. Det vil si at språkutviklingen stopper opp og at språket ”stivner”, slik at de formene som avviker fra målspråket blir permanente deler av innlærerens språk (Berggreen & Tenfjord, 2011: 370). Det vil dermed være en del individuell variasjon mellom språkbrukere når det gjelder hvor enkle eller komplekse *mellomspråkene* deres ender opp med å bli. Likevel vil altså redundante trekk, som jeg undersøker i denne masteroppgaven, generelt være vanskelige for innlærerne, og vil typisk være noe av det siste som faller på plass i *mellomspråket* (om det i det hele tatt gjør det).

Mellomspråket følger også gjerne forutsigbare utviklingssteg på veien mot beherskelse av målspråket. Det finnes en viss regelmessighet i hva slags fenomener som blir lært tidlig, og i hvilke som blir lært senere. Dette skjer også uten at læringskonteksten eller førstespråksbakgrunnen har noen særlig påvirkningskraft (Berggreen & Tenfjord, 2011: 40). Det er for eksempel en kjent sak at læring av negasjon har en viss utviklingsløype i *mellomspråket*: Først bruker innlæreren som regel innlærte formularer i form av hele ytringer

som ”Jeg vet’ke”. Andrespråksinnlæreren kan imidlertid ikke skille mellom de ulike komponentene i ytringen, og har kun lært dette som en helhet. I samme fase produserer innlæreren ytringer med ekstern nekting, som ”ikke venn”, eller med preverbal nekting, som i ”Jeg ikke spørre”. På neste stadium kommer postverbal nekting av typen ”jeg huske ikke”, eller ”jeg fortelle ikke” (Berggreen & Tenfjord, 2011: 42). Dette har det blitt forsket mye på, og mønsteret ser altså ut til å være det samme i alle språk, uansett hva slags språklig bakgrunn man har fra før. Jeg har allerede vært inne på det som er mest relevant for denne undersøkelsen i sammenheng med forutsigbare utviklingstrekk; at redundante og rent formelle trekk ofte viser seg sent i læringsrekkefølgen, eller aldri. Det som blir interessant i min undersøkelse, er å se om det er noen merkbar forskjell mellom gruppene med ulik førstespråksbakgrunn i sammenheng med dette.

I mellomspråkene kan man også ofte observere at noen typer konstruksjoner eller ytringer som er vanlige i målspråket, ikke blir brukt i like stor grad av innlæreren. Dette forklares ofte ved hjelp av begrepene *underproduksjon* eller *unnnvikende produksjon*. *Underproduksjon* av en bestemt type konstruksjon (for eksempel relativsetninger), er når andrespråksinnlæreren ikke bruker konstruksjonen fordi vedkommende rett og slett ikke vet at den eksisterer. Dette må ikke forveksles med *unnnvikende produksjon*, som er en bevisst eller ubevisst unngåelse av det relevante fenomenet på grunn vanskeligheter (Abrahamsson, 2009: 237). Både underproduksjon og unnnvikende produksjon kan også føre til en *overproduksjon* – en mer frekvent bruk av enkelte andre konstruksjoner enn det som er vanlig i målspråket (Abrahamsson, 2009: 238). I tillegg til egenskapene som allerede er nevnt, er også mellomspråk ustabile og variable språk, og de viser dessuten tegn til såkalt tverrspråklig innflytelse eller *transfer*. Dette vil jeg komme nærmere inn på under 2.5, der jeg forklarer hva transfer egentlig er, og under 2.6, der jeg forklarer begrepene kompetanse og performanse.

2.4 Prinsipp- og parameterteorien

Denne masteroppgaven er som nevnt skrevet med bakgrunn i chomskyansk generativ grammatikk, innenfor fagfeltet norsk som andrespråk. Prinsipp- og parameterteorien, som er viktig for disse to feltene i kombinasjon, er derfor relevant for andrespråksforskning innenfor generativ grammatikk, og derfor også for denne masteroppgaven. For å forklare denne teorien må vi aller først se nærmere på begrepene generativ grammatikk og universalgrammatikk.

En *generativ grammatikk* kan i korte trekk beskrives som en grammatikk som nøyaktig spesifiserer hva som er grammatikalske eller ugrammatikalske setninger i et gitt språk. Man kan på mange måter si at den generative grammatikken er en modell av

menneskets språklige intuisjon, som genererer grammatikalske setninger for språkbrukeren uten noen anstrengelse fra språkbrukerens side. Den generative grammatikken må altså forstås som en kognitiv grammatikk, og dessuten som en del av menneskets medfødte og universelle språkevne. Denne medfødte språkevnen er gjerne kjent under navnet *universalgrammatikken*, og refereres ofte til som forkortelsen *UG* (Berggreen & Tenfjord, 2011: 29).

Man regner med at alle verdens språk genereres med utgangspunkt i den samme universalgrammatikken, og at dette er en biologisk medfødt utrustning på lik linje med for eksempel syn eller hørsel (Åfarli & Eide, 2008: 17). Men ettersom alle individer på jorden er født med den samme språkevnen, skulle ikke alle da egentlig ha snakket det samme språket? Dette er selvfølgelig ikke tilfellet. Mennesker har ikke ett enkelt medfødt språk– UG svarer heller til selve språkevnen som mennesket har. Mange har også omtalt denne evnen som et *instinkt*. Darwin påpekte allerede i *The Decent of Man* at mennesket har en instinktiv evne til å snakke, noe vi ser i bablingen til små barn (mens ingen barn har noen instinktiv tendens til å bake, sykle, eller skrive) (Boeckx, 2010: 23). Derfor er UG helt essensiell både innenfor førstespråklæring og andrespråklæring.

Forklaringen på at ikke alle ender opp med å ha nøyaktig det samme språket, er at UG inneholder to ulike typer prinsipper: Den ene typen er såkalte absolutte prinsipper, som gjerne bare kalles *prinsipper*. Disse er felles for alle språk, og er en del av den faste kjernen til universalgrammatikken. Den andre typen er åpne prinsipper, som gjerne kalles *parametre*. Disse kan ha ulike verdier i ulike språk (Åfarli, 2015: 64). Man tilegner seg ulike parametriske verdier ut fra hva slags språk man blir eksponert for – altså hva slags språklig stimuli man får gjennom barndommen. Eide og Busterud (2015: 33) forklarer parametrene med utgangspunkt i en bryterpanelmetafor som opprinnelig er hentet fra Chomsky:

”Vi kan tenke på parametre som brytere på et bryterpanel [...]. Bryterpanelet og det apparatet det er tilkoblet, er språkevna (universalgrammatikken), og bryterne er dermed medfødt. Men bryternes stilling er ikke bestemt ved fødselen.”

Hvis man velger å se på universalgrammatikken som et bryterpanel, kan man tenke seg at et nyfødt barn har alle bryterne ”nullstilt”. Det er først når barnet får språklig stimuli fra omverdenen at bryterne stiller seg inn. Om barnet får språklig input i form av norsk, vil bryterne stille seg inn på et ”norsk” system, og om barnet vokser opp med vietnamesisk stimuli, vil det få et ”vietnamesisk” system. Dette systemet vil deretter generere setninger etter hva slags språklig stimuli man har blitt eksponert for (Åfarli & Eide, 2008: 18). Et spørsmål som er særlig interessant i sammenheng med dette, er hvor rigide disse bryterne er

når de først har stilt seg inn, og om de kan nullstilles igjen ved læring av nye språk. Dette spørsmålet er viktig for akkurat det jeg skal undersøke i denne masteroppgaven.

Det er uenighet om hvor stor tilgang man har til sin egen universalgrammatikk i sammenheng med andrespråklæring. Noen har argumentert for at man har full tilgang til UG hele livet, altså at bryterpanelet kan nullstilles fullstendig ved når man begynner å lære et andrespråk (Åfarli, 2015: 69). Dette samsvarer imidlertid dårlig med de erfaringene vi har med andrespråklæring. Førstespråklæring skjer uten hard anstrengelse fra barnet som lærer å snakke, og språkstrukturen blir sjelden noe annet enn perfekt. Hvis andrespråklæring kun baserer seg på full aksess til UG, ville man kunne forvente at resultatet hadde blitt det samme som ved førstespråklæring (Åfarli, 2015: 70). Det er riktignok svært stor spredning i hvor godt noen klarer å lære et andrespråk. De aller fleste som begynner å lære et andrespråk, særlig om dette skjer etter den kritiske perioden, oppnår aldri en fullstendig språkkompetanse av den typen førstespråksbrukerne har. Dette indikerer at det ikke kan være UG alene som styrer hvordan andrespråksinnlæringen foregår (Eide & Busterud, 2015: 34). Andrespråklæring er altså en kompleks prosess, og de aller fleste innenfor andrespråksfeltet mener det er flere komponenter som bør tas med i regnestykket. Tverrspråklig innflytelse eller *transfer*, altså en overføring av trekk fra førstespråket til andrespråket, er et av elementene som må være med for å forklare innlæringsprosessen. Transfer er et omdiskutert emne innenfor andrespråksforskningen, men det er fremdeles bred enighet om at det alltid vil være en eller annen form for transfer involvert i andrespråksinnlæring (Berggreen & Tenfjord, 2011: 201). Dette vil jeg diskutere i nærmere detalj under 2.5.

I moderne komparative språkstudier prøver man ofte å finne ut hvilke strukturelle trekk som finnes under samme parametriske verdi (Berggreen & Tenfjord, 2011: 50). I denne masteroppgaven undersøker jeg tre innlærergrupper med ulik førstespråksbakgrunn komparativt, og det interessante er at ingen av dem har dobbel bestemthet i førstespråket slik som norsk har. I sammenheng med dette er det flere viktige spørsmål som kan stilles for denne undersøkelsen: Er bryterne som er involvert i tilegnelse av dobbel bestemthet nullstilte fordi det ikke har vært ”bruk for” dem før? Eller har de allerede har blitt brukt i tilegnelse av andre grammatiske fenomener? Og hvis bryterne allerede har stilt seg inn – er det i en retning som legger til rette for eller vanskeliggjør dobbel bestemthet? I denne undersøkelsen vil det være spennende å se hvordan de parametriske verdiene for nominale fraser som tysk har, som er veldig likt norsk, vil påvirke bruk av dobbel bestemthet i norsk mellomspråk. Til tross for likhetene har jo ikke tysk dobbel bestemthet. Det blir også spennende å se hvordan dette skjer i den engelske førstespråksgruppen: Engelsk har vesentlige forskjeller fra norsk når det

gjelder nominalfraser, og derfor sannsynligvis færre parametriske verdier til felles med norsk i sammenheng med dette. På samme måte blir det interessant å se hvordan de parametriske verdiene for nominale fraser som vietnamesisk har, vil påvirke tilegnelse og bruk av dobbel bestemthet i norsk mellomspråk. Vietnamesiske nominale fraser er bygd opp på en helt annen måte enn norske, og derfor har sannsynligvis også vietnamesisk færre parametriske verdier til felles med norsk på dette feltet – sannsynligvis også færre til felles enn både tysk og engelsk har.

2.5 Transfer

Under 2.4 argumenterte jeg for at UG alene ikke kan forklare hvordan andrespråklæring foregår, særlig hvis det er snakk om språklæring som skjer etter den kritiske perioden (som er tilfellet for informantene i denne masteroppgaven). UG er altså bare en del av puslespillet. Som jeg allerede har vært inne på, er en annen sentral del av andrespråklæring påvirkning fra førstespråket. Dette har jeg valgt å hovedsakelig omtale som *tverrspråklig innflytelse* i denne masteroppgaven (begrunnelse for dette ser man nedenfor), men er enda mer kjent under navnet *transfer*. Transfer er et omdiskutert og kontroversielt tema i andrespråkforskningen, som det finnes mange ulike syn på og teorier om. Det er likevel få teorier som ser bort fra førstespråkets sentrale rolle i andrespråklæringen (Berggreen & Tenfjord, 2011: 201). De fleste andrespråkforskerne mener altså at den språkkompetansen man allerede har fra førstespråket, vil påvirke hvordan man lærer et andrespråk. Dette er noe som kan gi synlige og i noen tilfeller mange utslag på mellomspråket til andrespråksinnlærerne.

Transfer kan defineres på mange måter. I denne masteroppgaven ønsker jeg å bruke Jarvis' definisjon (Jarvis, 2000: 252, her sitert fra Johansen 2010: 161):

”L1 influence refers to any instance of learner data where a statistically significant correlation or (probability-based relation) is shown to exist between some feature of learners' IL performance and their L1 background”.²

Jarvis legger mer vekt på statistisk signifikans og evidens i sin definisjon enn mange andre, og definisjonen er dessuten teorinøytral. En av de grunnleggende antakelsene til Jarvis er at transferens essens foreløpig er utenfor forskerens rekkevidde, men at påvirkning fra førstespråket kan komme til syne som såkalte *transfer-effekter* (Johansen, 2010: 161). Og det

² L1 er en forkortelse for ”language 1”, eller førstespråk på norsk. IL er en forkortelse for ”internal language”, og kalles ofte ”I-language”, eller ”I-språk” på norsk. I-språket er kompetansen til den enkelte andrespråksinnlæreren (Johannessen, 2003: 147). Når Jarvis skriver ”learners' IL performance” i sitatet over, betyr dette altså performansen til hver enkelt andrespråksinnlærer. Dette er en noe uvanlig måte å bruke termen I-språk på, fordi performansen vanligvis er knyttet til E-språk (”external language”).

er nettopp disse effektene man kan studere nærmere. Dette kommer jeg tilbake til i metodekapittel 3.3 om Jarvis' metode for påvisning av tverrspråklig innflytelse.

I tillegg til at man kan observere transfereffekter fra førstespråket, mener også mange at påvirkningen kan komme fra et hvilket som helst språk innlæreren tidligere har lært – derfor er termen *tverrspråklig innflytelse* på mange måter mer presis (Berggreen & Tenfjord, 2011: 204). Både førstespråk og eventuelle andrespråk kan altså ha en viss innflytelse på innlærernes mellomspråk, noe jeg har forsøkt å ta høyde for i denne undersøkelsen (se spesielt kapittel 4.3.2 om engelskkunnskaper som variabel hos de vietnamesiske og tyske andrespråksinnlærerne).

Transfer gir seg ofte synlige utslag i andrespråksinnlærerens mellomspråk, slik som for eksempel i disse setningene som er produsert av en andrespråksinnlærer av norsk med engelsk som førstespråk (eksemplene er hentet fra Berggreen & Tenfjord 2011: 43):

3(a) ”Som jeg er student kan jeg få en billigere billett på flyreiser”.

(b) ”Som jeg har allerede sett, jeg kan ikke koke godt”.

I 3(a) kan vi se at den engelske andrespråksinnlæreren har brukt ordet ”som” akkurat på samme måte som man ville brukt subjunksjonen ”as” på i engelsk. Innlæreren har antageligvis lært at ”som” har samme betydning som ”as” i noen setninger, som for eksempel i ”Arne er like snill som deg”. Videre har innlæreren mest sannsynlig overgeneralisert, og tenkt at ”som” kan brukes som ”as” på alle vis, og slik endt opp med setningen i 3(a). Vi kan forøvrig også se et mulig tilfelle av transfer i ordet ”billiger”, som er bøydd nøyaktig på samme måte som man ville bøydd det i engelsk (cheap-er, billig-er). I 3(b) kan vi også se flere indikasjoner på transfer. Et av tegnene på innflytelse fra engelsk kan vi se i leddstillingen, som ikke er målspråkskonform. I tillegg ser vi at innlæreren har brukt ordet ”koke” på samme måte som man ville brukt ordet ”cook” i engelsk.

Det finnes i hovedsak to typer transfer. Den ene typen er overføringer av strukturer eller trekk fra førstespråket til andrespråket som avviker fra målspråksnormen. Denne typen kalles ofte *interferens* eller *negativ transfer*, og er den typen transfer vi foreløpig har sett på i eksemplene i 3(a) og (b). Når overføringene på den andre siden skaper mellomspråksstrukturer som er målspråkskonforme, kalles dette ofte *positiv transfer*. De trekkene eller strukturene man overfører finnes i disse tilfellene både i førstespråket og andrespråket, og derfor samsvarer den formen man overfører fra førstespråket med målspråksnormen (Berggreen & Tenfjord, 2011: 204). Den positive transferen blir ofte glemt

eller oversett, nettopp fordi formen ikke ser annerledes ut enn i målspråket – det er ikke noe man trenger å ”rette opp”. Dette betyr imidlertid at man har et syn på mellomspråk som er fiksert på avvik fra målspråksnormen, og at man overser rollen mellomspråket har som et eget språkssystem. Den positive transferen er altså like viktig som interferensen i andrespråksforskningen, og begge typene transfer kan fortelle oss noe om hvordan språklæring foregår. Hvis vi ser på eksemplene 3(a) og (b), ser vi forskjellen mellom interferens og positiv transfer. I begge eksemplene brukes ”som” parallelt med ”as” i engelsk. I (a) avviker bruken fra målspråket, mens det i (b) samsvarer med målspråket.

Et av de store metodiske problemene med transfer ligger i hvordan man kan vite at det man observerer faktisk *er* transfer. I eksemplene jeg viste i 3(a) og (b) er det mye som tyder på transfer fra engelsk, men i noen tilfeller kan det være vanskelig å bedømme dette på en vitenskapelig måte. Hvordan kan man for eksempel isolere faktoren transfer fra andre faktorer som har betydning i andrespråklæringen? Dette og flere andre metodiske spørsmål kommer jeg tilbake i kapittel 3.3, der jeg presenterer Jarvis’ metode for påvisning av tverspråklig innflytelse, som jeg benytter meg av i denne undersøkelsen.

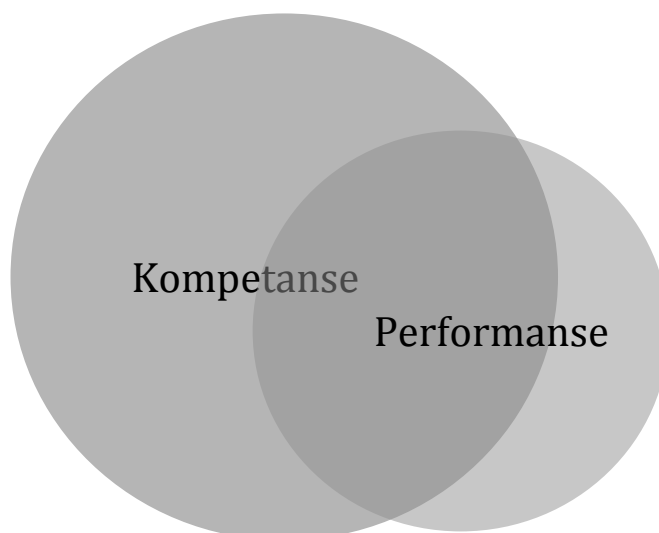
2.6 Kompetanse og performanse

Som jeg så vidt var inne på under 2.3, er mellomspråk både ustabile og variable språk. En naturlig del av å lære et språk er at man stadig utvikler seg, og derfor endres også grammatikksystemet over tid. Det er imidlertid også slik at innlærerne ofte bruker ulike former om hverandre innenfor et kort tidsrom, til og med innenfor samme tekst (Åfarli, 2015: 71). Det kan være mange årsaker til en slik språklig variasjon hos én enkelt innlærer. Ofte har dette med ytre forhold å gjøre – som for eksempel om man skal snakke eller skrive, hvor mye tid man har til rådighet, hvordan man behersker oppgaveformen, eller om man for eksempel er nervøs, stresset eller distraheret. Ofte kan man intellektuelt vite hvordan et språklig fenomen skal se ut, men i det øyeblikket man skal produsere ytringen, kan man likevel ende opp med en form som avviker fra målspråket. Denne uttalelsen av en gutt i 9. klasse som holder på å lære seg fransk, beskriver godt hvordan dette ofte oppleves (Swain 1982: 94, sitert her fra Eide og Busterud 2015: 21):

”I understand everything anyone says to me, and I hear in my head how I should sound when I talk, but it never comes out that way.”

Ettersom ytre faktorer kan påvirke den språklige produksjonen til en andrespråksinnlærer, finnes det et viktig skille mellom en innlærers *kompetanse* og *performanse*. Kompetansen er selve kunnskapen som utgjør den internaliserte grammatikken til en språkbruker. Performansen svarer på den andre siden til den faktiske produksjonen av ytringer – altså til det som blir skrevet eller sagt (Hawkins, 2001: 23). For gutten i sitatet ovenfor kan man ane en tydelig forskjell mellom disse to fenomenene. Kompetansen som gutten har i fransk er større enn det som kommer fram i de språklige ytringene hans, altså i performansen. Innenfor generativ andrespråksforskning er det kompetansen som er av interesse, fordi det er den som forteller mest om språkevnen til mennesker (Eide & Busterud, 2015: 50). Alle, inkludert førstespråksbrukere, kan dessuten forsnakke seg og bli påvirket av situasjonen man er i når man skal ytre seg. Derfor er kompetansen det viktigste studieobjektet for generative andrespråksforskere.

Chomsky påpekte i sine tidligere dager at lingvistisk teori primært handler om idealiserte talere i et fullstendig homogent språksamfunn, og at forhold som minnebegrensninger, distraksjoner og så videre ikke kan fortelle oss noe om grammatikken i seg selv og hvordan den fungerer (Chomsky, 1965: 31). Performansen kan på grunn av forsnakkelser og andre ytre faktorer inneholde strukturer og trekk som ikke kan finnes i kompetansen i det hele tatt, og derfor kan den gi et feil bilde på hva slags internalisert grammatikk språkbrukeren har. Eide og Busterud viser hvordan forholdet mellom kompetanse og performanse ser ut i en figur som jeg har satt inn nedenfor (2015: 50):



Interessen for kompetansen til en innlærer gjør forskning og undersøkelser innenfor det generative andrespråksparadigmet komplisert. Det vil være helt umulig å få fullstendig tilgang

til en andrespråksinnlærers kompetanse, nettopp fordi de språklige ytringene man studerer alltid til en viss grad vil være påvirket av ytre omstendigheter. Noen metoder har likevel blitt utviklet for å gi et større innblikk i språkbrukernes kompetanse, og disse er ofte kjent som grammatikalitetstester eller akseptabilitetsbedømminger. Dette er tester der informantene først får presentert en eller flere setninger, og så skal de bedømme om setningen er en mulig ytring, altså om den er grammatikalsk (Schütze & Sprouse, 2013: 28). I teorien undersøker man dermed ikke språkbrukerens ytringer i seg selv, men heller den språklige intuisjonen til vedkommende³.

Den undersøkelsen jeg foretar i denne masteroppgaven, er ikke basert på grammatikalitetstester. Jeg undersøker tekster som er produsert av andrespråksinnlærere i en bestemt testsituasjon, og det er derfor kun performansen til disse innlærerne jeg får innblikk i. Dette kan sies å være en svakhet med denne undersøkelsen, fordi det bare er noen deler av informantenes internaliserte grammatikk (eller kompetanse) som avdekkes. Jeg mener likevel at det er stor verdi i undersøkelser av denne typen også, noe jeg nå vil begrunne nærmere i kapitlet om metode.

³ Schütze & Sprouse (2013: 29) diskuterer også om akseptabilitetsvurderinger faktisk viser kompetansen eller performansen til en innlærer. Jeg vil imidlertid ikke gå dypere inn i denne diskusjonen i denne masteroppgaven.

3. METODE

I denne delen av teksten tar jeg for meg metoden for denne undersøkelsen. Aller først presenterer jeg korpuslingvistikk under 3.1, og her diskuterer jeg blant annet kritikken mot korpus, og hvorfor jeg likevel har valgt å benytte meg av korpussøk som metode. Jeg vil også definere nærmere hva et moderne korpus er, og viktige begreper knyttet til dette. Så introduserer jeg *norsk andrespråskorpus* (ASK) under 3.2, som jeg har samlet inn datagrunnlaget til denne undersøkelsen fra. Her tar jeg opp muligheter og begrensninger som korpuset har, samt designkriterier, representativitet og hva slags søkemuligheter man har. Deretter vil jeg under 3.3 forklare *Jarvis' metode for påvisning av tverrspråklig innflytelse*, som jeg har valgt å bruke i min undersøkelse. Til slutt forklarer jeg under 3.4 begrepet *statistisk signifikans*, som er viktig i sammenheng med tolkningen av resultatene fra undersøkelsen.

3.1 Korpuslingvistikk

Datagrunnlaget for undersøkelsen jeg foretar i denne masteroppgaven er som nevnt hentet fra ASK, og derfor er min metode i hovedsak korpussøk. Innledningsvis vil jeg også nevne at jeg, i tråd med McEnery, Xiao og Tono (2006), anser korpuslingvistikk som en metodologi, og ikke en teori: "Corpus linguistics should be considered as a methodology with a wide range of applications across many areas and theories of linguistics" (McEnery, Xiao, & Tono, 2006: 8). Korpuslingvistikken kan altså anses som en metodologisk tilnærming som kan brukes på mange måter, på tvers av mange lingvistiske teorier. Dermed må man ikke anse korpuslingvistikken som en teori eller et eget forskningsfelt, og heller ikke en metode i seg selv. Selve korpusundersøkelsen jeg gjør er metoden i denne masteroppgaven. Det er også viktig å understreke at det å benytte seg av korpus ikke er det samme som å bruke kvantitative metoder (Johannessen, 2003: 149). Et korpus kan brukes både kvalitativt og kvantitativt.

3.1.1 Kritikken mot korpuslingvistikk

Chomsky var tidlig ute med å kritisere korpus som empirisk grunnlag for forskning innenfor lingvistikken. Dette sitatet fra Chomsky (1962), gjengitt av McEnery & Wilson (2001: 10), forteller oss om noe av grunnlaget for kritikken mot korpuslingvistikken:

"Any natural corpus will be skewed. Some sentences won't occur because they are obvious, others because they are false, still others because they are impolite. The corpus, if natural, will be so wildly skewed that the description [based upon it] would be no more than a mere list."

Chomsky påpekte altså at et korpus alltid kun vil bestå av enkelte deler av et språk. Korpuset kan umulig gjengi nøyaktig hvordan språket egentlig er, både fordi situasjonen rundt tekstproduksjonen til informantene kan påvirke språket, for eksempel ved at man unngår å skrive uhøflige eller stygge setninger, og blant annet fordi informantenes sinnsstemning kan påvirke hvordan den språklige produksjonen foregår. Er informanten stresset, forvirret, distraheret og så videre, vil det naturligvis påvirke hvordan vedkommende ytrer seg. Her ble det skapt et skille mellom begrepene kompetanse og performanse, som ble forklart i nærmere detalj under 2.6.

Det må riktignok framheves at Chomskys kritikk mot korpus på 60-tallet gjaldt helt andre korpus enn de som finnes i dag. På 60-tallet var korpusene små, manuelle samlinger av tekster, som ikke kunne representere noe språk på en god måte. Her var det imidlertid en enorm framgang når datamaskinene for alvor begynte å utvikle seg. Etter hvert kunne store mengder data samles i elektroniske korpus, og dette kunne gjøres svært oversiktlig, brukervennlig og gi muligheter for mange ulike typer søk (Ragnhildstveit, 2009: 49).

Svært mye har altså skjedd innen utvikling av korpus siden 60-tallet, både når det gjelder mengde data i hvert korpus, designkriterier, lagring av metadata og systematisering (for å nevne noe). Man kan altså ikke la den gamle kritikken til Chomsky avskrive korpuslingvistikken. Man må heller, som McEnery og Wilson påpeker, bruke kritikken for å utforme mest mulig representative korpus (McEnery & Wilson, 2001: 30). Det har det blitt arbeidet mye med de siste årene, og korpus danner i dag det som anses som et godt og sikkert empirisk grunnlag for svært mange undersøkelser og studier.

3.1.2 Moderne korpus

I prinsippet kan alle samlinger som inneholder mer enn én tekst kalles et korpus. I en moderne lingvistisk kontekst er imidlertid denne definisjonen for enkel. Derfor kan et korpus i denne sammenhengen heller defineres som en samling med maskinleselige autentiske tekster, enten muntlige eller skriftlige, som kan annoteres (se nedenfor) med ulike former for lingvistisk informasjon (McEnery, et al., 2006: 4). Det er en stor fordel å ha maskinleselige tekster i et korpus, framfor den originale formen (for eksempel lydopptak eller håndskrevet tekst). Årsaken til dette er hovedsakelig at man kan søke på måter som ikke ville ha vært mulig ellers, noe som effektiviserer arbeidet med datainnsamlingen enormt. Med tidligere korpus måtte man lese alt manuelt og notere informasjon underveis, noe som kunne ta dagevis eller ukevis (McEnery & Wilson, 2001: 31). I et moderne søkbart korpus bruker man kun minutter på dette, ved å bruke en konkordansefunksjon eller ganske enkelt ved å ta et ordsøk. Dessuten

kan man legge inn mange former for *annotering* i et korpus, som er ulike former for lingvistisk informasjon lagret i morfem, ord eller på setningsnivå (McEnery & Wilson, 2001: 32). Dette kan bidra til å gjøre søkene enda enklere, og dessuten mye mer spesifikke. Et annotert korpus sparer forskeren for svært mye arbeid. Det er imidlertid mennesker som må annotere korpus, og derfor kan det selvfølgelig også oppstå menneskelig svikt og feiltagging. Et godt korpus har derfor egne retningslinjer for annotering, som skal sikre en mest mulig homogen praksis blant ulike kodere (Tenfjord, et al., 2009: 60). Retningslinjene til ASK i sammenheng med dette vil diskuteres i kapittel 3.2.2.

Et annet stikkord som er svært viktig i sammenheng med korpus er representativitet. Et lite korpus vil ikke kunne være representativt for et helt språk, fordi det ikke kan inneholde alle mulige setninger i språket. Det vil også være helt tilfeldig om sjeldne konstruksjoner har blitt ytret av noen av informantene, og om noen har benyttet seg av uttrykk som ikke er typiske for språket i det hele tatt. Når størrelsen på korpusene øker, i form av mer og mer språklig data, vil imidlertid gjerne representativiteten bli bedre. Sjansen er større for at frekvensen på typiske og atypiske uttrykk er mer i tråd med det virkelige språket i et stort korpus med mange informanttekster, enn i et lite korpus med få informanttekster. Et stort og representativt korpus gjør kvantitative undersøkelser mulig, og man kan derfor bruke funnene til å generalisere ulike trekk for det språket man undersøker (Ragnhildstveit, 2009: 50). Størrelsen har altså mye å si for representativiteten til korpuset, og derfor har altså moderne elektroniske korpus et klart fortrinn.

Det er imidlertid ikke *bare* størrelsen på korpuset som avgjør om det er representativt. Et korpus må også være velbalansert, noe som avhenger av hva slags tekster det inneholder, og hvordan disse er samlet inn (Johannessen, 2003: 148). Dette varierer naturligvis ut fra korpusets bruksområde. Derfor blir det laget designkriterier til hvert enkelt korpus, ut fra hva som er hensikten med korpuset, og hva slags studier og undersøkelser det er beregnet på. For eksempel krever synkrone studier innsamling av data fra samme tid, og skal korpuset brukes til dette, må man designe korpuset deretter (Ragnhildstveit, 2009: 51). Under 3.2 vil jeg forklare representativitet og designkriterer i ASK-korpuset nærmere.

McEnery et. al definerer også korpuset som ”a body of naturally occurring language” (McEnery, et al., 2006: 4). I tillegg til å være maskinleselige, må altså tekstene i moderne korpus også være autentiske. Autentiske tekster fra et korpus kan gi oversikt over bredden som finnes i et språk. Ved introspeksjon finner man ut hva som er grammatikalsk for akkurat det individet som utfører introspeksjonen. Ettersom alle har en individuell internalisert grammatikk, kan svært mye variere mellom ulike språkbrukere av for eksempel språket norsk.

Selv om jeg bruker et uttrykk svært frekvent, trenger ikke det bety at dette er et kjennetegn ved norsk generelt. Derfor kan korpusstudier og –undersøkelser være nyttige når man ønsker å få tak i ny kunnskap om hele enkeltspråk, og på tvers av språk. Det er imidlertid ikke alle forskningsspørsmål som det egner seg å besvare ved hjelp av empiri fra korpus. For at korpusbruk kan forsvares, er det som Johannessen skriver ”nødvendig med en god porasjon kritisk sans og helst tilleggsundersøkelser med bruk av informanter og introspeksjon” (2003: 148). Derfor kan man si at grammatikalitetsvurderinger og korpusundersøkelser ikke er to metoder som utelukker hverandre, men at de heller komplementerer hverandre.

3.1.3 Fordeler og problemer med bruk av moderne korpus

Som jeg har vært inne på, har utviklingen av det moderne elektroniske korpuset, til tross for noen svakheter i sammenheng med kompetanse- og performanseskillet, gitt store muligheter innenfor forskning. Muligheten for å arbeide med store mengder data blir i stor grad fremhevet som en positiv side ved moderne korpus (Johannessen, 2003: 149). Der intuisjonstestene kan gi innsikt i kompetansen til den enkelte informanten, det som tilsvarer *I-språket* (”internal language”), kan korpusstudier blant annet være nyttige i komparative undersøkelser (Johannessen, 2003: 147). Hvis man studerer *E-språket* til en gruppe mennesker (”external language” – en term som svarer til performansen), som jeg gjør i denne masteroppgaven, vil bruk av korpus både være nyttig og effektivt. For at man skal kunne si noe om parametriske forskjeller mellom store grupper som snakker samme språk eller dialekt, og ikke forskjeller mellom enkeltindivider, er det en enorm fordel å ha tilgang til store mengder språkdata. Dette gjør generaliseringer av nasjonalspråk og dialekter mulig.

Et av problemene med korpusbruk er imidlertid at det kan være vanskelig å formulere spørsmål om negative data. Det kan som nevnt være mange årsaker til at en konstruksjon ikke finnes i et korpus – det kan være at konstruksjonen er typisk muntlig, eller kanskje den er så merkelig pragmatisk sett at ingen ville finne på å bruke den (Johannessen, 2003: 147). I et innlærerkorpus kan det også være en underproduksjon eller unnvikende produksjon av visse fenomener (som nevnt i kapittel 2.3). Ettersom forskeren samler inn data rett fra korpuset, har ikke vedkommende muligheter til å stille oppfølgingsspørsmål for å gå nærmere inn i slik problematikk (Johannessen, 2003: 149). Dette faktum kan også skape vansker hvis relevant informasjon om informantens bakgrunn ikke er å finne i korpuset. Mange korpus har imidlertid store mengder metadata tilgjengelig, noe som kan være med på å løse deler av dette problemet. I tillegg vil en ikke-kontakt med informantene også gi en stor

fordel: Man unngår at selve datainnsamlingen farges av forskerens hypoteser (Johannessen, 2003: 149).

3.2 ASK – norsk andrespråskorpus

Undersøkelsen jeg utfører i denne masteroppgaven, er basert på data som er hentet fra norsk andrespråskorpus (ASK). ASK er altså et *andrespråskorpus*, noe som gjør at det skiller seg fra de fleste andre innlærerkorpus, som ofte er *fremmedspråskorpus* (Johansen, 2010: 164). Disse korpusene består av tekster skrevet av innlærere som stort sett har sin språklige input i en klasseromssituasjon. ASK består på den andre siden av datamateriale samlet inn fra innvandrere som har lært norsk i en naturalistisk sammenheng (Johansen, 2010: 164).

Når man bygger et andrespråskorpus, må man ta hensyn til at det er stor variasjon i andrespråksbruk, og at det er mange faktorer som påvirker både språklæring og språkbruk (Tenfjord, et al., 2009: 55). Derfor er gode designkriterier og mange andre tekniske løsninger av stor betydning for hvordan et andrespråskorpus konstrueres. ASK har fått anerkjennelse for flere av sine tekniske løsninger når det gjelder både lagring, annotering og søk (Johansen, 2010: 164). Senere i dette kapitlet vil jeg i nærmere detalj gå inn på akkurat hvordan ASK er oppbygd. I første omgang kan vi se hvordan Hilde Johansen oppsummerer oppbygningen til andrespråskorpuset (2010: 163, 164).

”ASK (...) kan kort oppsummeres som et korpus av tekster skrevet av innlærere med ulike morsmålsbakgrunner, tilhørende persondata, annotering med feilkoder og grammatiske tagger, og et søkegrensesnitt hvor man kan kombinere ulike språklige trekk med personvariabler.”

Foreløpig består ASK av over 1700 tekster, hver på omtrent 200-400 ord, som er skrevet av andrespråksinnlærere som har tatt *Språkprøven i norsk for voksne innvandrere* og *Test i norsk – høyere nivå*. I den undersøkelsen som jeg foretar i denne masteroppgaven bruker jeg tekster fra *Språkprøven i norsk*, som tilsvarer nivå B1 (Threshold level) i henhold til definisjonene i *The Common European Framework for Reference for Languages* (CEFR) (Tenfjord, et al., 2009: 55). Dermed har alle informantene mine bakgrunn i samme testsituasjon, noe som øker reliabiliteten til denne undersøkelsen. Ved å bruke testbesvarelser som data, sikrer man at tekstene er nokså homogene både når det gjelder sjanger og kontekst for tekstproduksjonen (Tenfjord, et al., 2009: 55). Tekstene som ligger i ASK-korpuset møter også kravet om autentisitet, fordi tekstene er innlærernes egenproduserte tekster (Ragnhildstveit, 2009: 59).

I tillegg til å skape et likt utgangspunkt i situasjon, oppgaveform, tidsbegrensning og så videre, er det også lagt inn metadata i ASK. Her kan man finne opplysninger om flere variabler som kan ha påvirkning på språklæringen (Tenfjord, et al., 2009: 55). Derfor blir det også mulig for meg å kryssjekke andre variabler som er kjent for å påvirke språklæring i denne undersøkelsen, for eksempel hva slags engelskkunnskaper de tyske og vietnamesiske informantene eventuelt har fra før. Hver informanttekst er dessuten markert med en egen tekst-ID, som gjør det enkelt å lenke personopplysninger til den enkelte informant (Tenfjord, et al., 2009: 56). Dette er en funksjon jeg har vært helt avhengig av i min undersøkelse, fordi den blant annet kunne vise meg hvor mange forekomster av dobbel bestemthet det var per tekst-ID, altså per informant – ettersom hver informant har én tekst.

I tillegg til hovedkorpuset består også ASK av et kontrollkorpuser. Disse tekstene er av samme type som i hovedkorpuset, men de er skrevet av personer som har norsk som førstespråk. Likevel må det nevnes at kontrollkorpuset skiller seg fra hovedkorpuset, nettopp fordi tekstene ikke er skrevet i noen testsituasjon. Dette er noe som kan påvirke performansen til informantgruppene på ulike måter, men som også naturligvis er uunngåelig, ettersom en oppgave man får på B1-nivå i norsk ikke vil være noen utfordring for voksne med norsk som førstespråk (Tenfjord, et al., 2009: 55).

3.2.1 Designkriterier i ASK

Som tidligere nevnt (i kapittel 3.1.2) avhenger representativiteten i et korpus blant annet av hvordan innsamlingen av dataene er gjort. Utvalget av informanttekster i ASK er samlet inn med utgangspunkt i tre designkriterier. Det første av disse er *morsmålskriteriet*, som sier at informantene ”skulle representere en rimelig typologisk variasjon av førstespråk” (Tenfjord, et al., 2009). Dette kriteriet var altså teoretisk motivert, fordi utvalget muliggjorde forskning på hvordan et førstespråk eventuelt påvirker læring av et andrespråk. Derfor var også en viss språktypologisk variasjon mellom førstespråkene ønskelig.

I tillegg kommer også *bestått-kriteriet*, som sier at hver ”informant skulle ha en definert kommunikativ minimumskompetanse i norsk, noe som er sikret ved at de har bestått testen de har avlagt” (Tenfjord, et al., 2009: 55). Dette vil derfor bety at tekstene i ASK i prinsippet er på eller over et visst funksjonelt nivå, noe som (i tillegg til det som ble nevnt under 3.2) bidrar til å gi en viss homogenitet. Dette kravet er helt essensielt for den metodologiske validiteten til ASK-korpuset (Tenfjord, et al., 2009: 57).

Det siste designkriteriet i ASK er *info-kriteriet*, som kan gjengis slik: ”For hver informant skulle det foreligge personopplysninger som kunne tenkes å være relevante for

språkutviklingen, og som skulle kunne lenkes til kandidatens besvarelse” (Tenfjord, et al., 2009: 56). Dette kriteriet gir forskeren muligheter til å finne opplysninger om for eksempel alder, kjønn, oppholdstid i Norge, antall timer med språkundervisning og så videre, som tidligere nevnt under 3.2. Denne typen informasjon om alle informantene gjør det mulig å kontrollere for en rekke ulikheter både i innlæringskontekst og innlærerbakgrunn, noe som kan være av stor betydning (Tenfjord, et al., 2009: 57).

3.2.2 Søk i ASK

Informanttekstene i ASK har blitt annotert med lingvistisk informasjon som man søker fram gjennom ulike søkekommandoer. Det er vanlig å bruke feiltagging som del av dette i innlærerkorpus, og slik er det også i ASK (Ragnhildstveit, 2009: 56). Feilkodingen i ASK har blitt gjort manuelt, ved hjelp av to prinsipper som skal sikre en mer eller mindre homogen praksis blant ulike kodere: *Det pragmatiske probabilitetsprinsippet* skal sikre at koderen velger den tolkningen av innholdet som er mest sannsynlig ut fra et pragmatisk perspektiv når man tar teksten som helhet i betraktning. *Det minimale modifikasjonsprinsippet* skal sikre at koderen også velger den rekonstruksjonen av teksten som utgjør minst mulig endring av originalen (Tenfjord, et al., 2009: 60). Feilkodingen er ikke gjort ut fra en mellomspråksanalyse, men ut fra en helhetlig tolkning og forståelse av informantteksten. Som Tenfjord, et al. (2009: 59) skriver, var et svært viktig poeng her at ”systematisk mellomspråksanalyse *ikke* skulle være nødvendig for en vellykket tolkning.” En slik analyse har faktisk vært eksplisitt forbudt i retningslinjene for feilkodingen (Tenfjord, et al., 2009: 59).

Det er imidlertid mange søkemuligheter i ASK som ikke nødvendigvis involverer søk på feil. Datamaterialet i korpuset kan lett brukes uten å utføre søk på feilkoder – for eksempel kan man søke etter forekomster av bestemte ordklasser, enkeltord eller lemma, eller strenger av ord (Tenfjord, et al., 2009: 71). Alle søkene kan enten gjøres på et eget søkespråk, eller ved hjelp av søk i menyer. Hvis man bruker menyer, gir dette enkle og oversiktlige muligheter for ulike søk, ved at man legger til ønskede attributter (som for eksempel at alle resultater skal være fra *Språkprøven*) og egenskaper gjennom undermenyer, som vist på bildene nedenfor. I mine søk, som vil gjennomgå i kapittel 4.2.1, kombinerte jeg blant annet lemma og bestemte ordklasser med visse attributter for å finne fram til dobbel bestemthetskonstruksjoner gjennom søkemenyene. Da bygger også søkespråket seg automatisk opp i tekstboksen, noe som gjør det enkelt å lagre søkeuttrykket til senere.

Enkelt søk | Avansert søk

Bruk denne tekstboksen for å skrive et søk i korpus-søkespråket

Søk | Tilbakestill | Bygg et grafisk søk

Her kan du lage et søk ved å bruke menyer.

Velg et **delkorpus**:

Morsmål != "norsk"

legg til:

attributt: -

Velg **posisjonelle betingelser**. Ignorer strukturelle posisjoner

target

Ord = %c

gjentagelse: 1

legg til:

attributt: -

struct: -

Søk

Søkesidene i ASK.

Når man har gjennomført søket, kommer man inn på konkordanssidene, der man får oversikt over alle treffene som søket har fått. Her kan man blant annet velge å vise KWIC-konkordansen, altså konteksten for treffet på hver side, som vist i bildet nedenfor. Man kan også trykke seg inn på de enkelte tekstene for å se hele konteksten. I tillegg kan man velge å sortere treffene ut fra tekst-ID (som vist til venstre i bildet), eller ut fra persondata som for eksempel engelskkunnskaper, førstespråk, kjønn, alder og så videre.

Enkelt søk | Avansert søk

`[[lemma = "det|den|de" %c] [pos = "adj"] ([pos = "subst" & morph = ("be" "fl")]] | [pos = "subst" & morph = ("be" "ent")]) \ <> \\ \ :: language = "norsk" &`

Søk | Lagrede søk | Save Query as

Treff 1 - 40 av 110 | neste 40 treff | Gå til: | Last ned (Excel-modus) | Type: kwic | Vis annotasjon | Vis linjefilter | Att

Tekst-id	treff
s2010	ngene og belaster et hardt presset helsebudsjett. For ikke å snakke om alle de psykiske problemene folk har. " Det si
s2003	meg selv. Men jeg må da finne ut og sette ord på hva som har fanget meg. Det andre begrepet - ansvar - betyr
s2005	viktig her i livet. Men hva er viktigst her i livet? Kristne vil si JESUS. I følge det <sic> Bibelske </sic> synet , spiller det over

Konkordanssidene i ASK.

Korpuset tilbyr også en distribusjonsfunksjon, som kan være nyttig særlig i frekvensspørsmål. Her kan man sortere alle treffene man har fått etter ulike faktorer, som for eksempel etter tekst-ID. Slik kan man få god oversikt over alle treffene man har fått. Jeg brukte denne funksjonen blant annet for å finne ut hvor mange informanter (basert på tekst-IDene) som

produserte dobbel bestemthets-konstruksjonen, og hvor mange forekomster det var per informant.

Fractions sum up to 1.0 in each row. Fractions in blue are unweighted means of group fractions. Fractions in green

Vis distribusjon av Ord %c, Δ: 0 filter: relative to Tekst-id
 type: absolutt | counts only | include structures

Side 1/1 av 2x7. Previous Next | Last ned

(sum)	< de psykiske problemene	Det andre begrepet	det <sic> Bibelske </sic> synet	den viktigste verdien	Den siste tiden	de anc sid
(sum)	110 (100,0)	1 (0,9)	1 (0,9)	2 (1,8)	2 (1,8)	
	110 (100,0)	1 (0,8)	1 (1,7)	2 (1,7)	2 (1,4)	
s2003	1 (100,0)		1 (100,0)			
s2005	2 (100,0)			1 (50,0)	1 (50,0)	
s2008	3 (100,0)					1 (33,3)
s2010	2 (100,0)	1 (50,0)				

Distribusjonssidene i ASK.

3.3 Jarvis' metode for påvisning av tverrspråklig innflytelse

I denne masteroppgaven har jeg valgt å ta utgangspunkt i Jarvis' metode for påvisning av tverrspråklig innflytelse. Denne metoden ble først presentert i artikkelen "Methodological Rigor in the study of transfer: Identifying L1 influence in the interlanguage lexicon" fra 2000, der Jarvis tar et oppgjør med metodikken som tidligere har blitt brukt i sammenheng med transferstudier (Johansen, 2010: 160). Jarvis savnet et enhetlig og teorinøytralt rammeverk for forskning på transfereffekter, og presenterte derfor et eget forslag.

I følge Jarvis blir tverrspråklig innflytelse ofte behandlet som et "you-know-it-when-you-see-it"-fenomen, noe som kan være problematisk. Et av problemene med denne holdningen er at transfer kan være så subtilt eller skjult av andre faktorer, at det ikke kan oppdages kun ved å se på dataene. Andre sentrale problemer er at det som ligner på tverrspråklig innflytelse ikke nødvendigvis er det, og at det er ikke alltid er så enkelt å isolere transfer fra andre faktorer (Jarvis & Pavlenko, 2010: 27). Ved å introdusere et nytt og enhetlig rammeverk for påvisning av tverrspråklig innflytelse, mente Jarvis at han kunne forhindre unødvendig forvirring og motstridende påstander både med tanke på når, hvor, i hvilken form og i hvilken grad transfer vil manifestere seg i innlærernes bruk av eller kunnskap om et andrespråk (Johansen, 2010: 160). Som en del av det nye rammeverket sitt, introduserte Jarvis tre typer evidens man må ha for å kunne påvise tverrspråklig innflytelse (her forklart i Jarvis & Pavlenko 2010: 35):

1. Intragroup homogeneity (intragruppe-homogenitet): "Evidence that the behavior in question is not an isolated incident, but instead a common tendency of individuals who know the same combination of languages".
2. Intergroup heterogeneity (intergruppe-heterogenitet): "Evidence that the behavior in question is not something that all language users do regardless of the combinations of L1s and L2s that they know".
3. Crosslinguistic performance congruity (tverrspråklig performansesamsvar): "Evidence that a language user's behavior in one language really is motivated by her use (i.e., the way she demonstrates her knowledge) of another language".

For å kunne påvise tverrspråklig innflytelse, må man undersøke alle disse tre punktene. I de fleste tilfellene vil påvisning av minst to av disse være tilstrekkelig evidens for transfer (Johansen, 2010: 162). For at jeg skal kunne påvise tverrspråklig innflytelse i min undersøkelse, må jeg derfor møte minst to av kravene under:

1. Intragruppe-homogenitet: Det må framstilles evidens for at bruken av dobbel bestemthets-konstruksjonen viser en felles tendens innenfor hver førstespråksgruppe.
2. Intergruppe-heterogenitet: Det må framstilles evidens for at bruken av dobbel bestemthets-konstruksjonen er ulik mellom de ulike førstespråksgruppene.
3. Tverrspråklig performansesamsvar: Det må framstilles evidens for at det er samsvar mellom førstespråket og mellomspråket til innlærerne når det gjelder bruken av dobbel bestemthets-konstruksjonen.

I tillegg til dette påpeker Jarvis at man også bør kontrollere for relevante variabler, som for eksempel alder, utdanning, sosial og kulturell bakgrunn, målspråksferdigheter, og mengde og type av målspråksinnputt. Kontroll på variabler og påvisning av de tre typene transfereffekter skaper sikrere påstander om tverrspråklig innflytelse, men han understreker også at "en slik rigorøsitet kun er et ideal som en bør gjøre sitt beste for å komme i møte innenfor de rammene en har" (Johansen, 2010: 162). Å kunne kontrollere for alle variabler vil være for omfattende for denne masteroppgaven. Likevel har jeg valgt å kontrollere for noen utvalgte variabler som jeg mener er viktige, for å gi en større reliabilitet til undersøkelsen min.

Alder er allerede kontrollert for i denne undersøkelsen (i alle fall til en viss grad), fordi alle informantene er voksne (fra 18 og oppover). Målspråksferdighetene til alle informantene er også kontrollert for, ettersom jeg kun bruker tekster fra *Språkprøven i norsk*,

som tilsvarer nivå B1. I tillegg har jeg undersøkt engelskkunnskapene hos de tyske og vietnamesiske innlærerne (se 4.2.3). Engelskkunnskapene er selvrapportert, og det kan være en mulig feilkilde i dette.⁴

3.4 Signifikansnivå

For å kunne påvise tverrspråklig innflytelse, må man også ta utgangspunkt i signifikansnivå når man skal analysere dataene. Et statistisk signifikant resultat kan betegnes på denne måten (Kunnskapsbasert praksis, 2012):

”Et resultat som det er usannsynlig er fremkommet ved tilfeldigheter. Den sedvanlige grense for denne vurderingen er at resultatet, eller mer ekstreme resultater, ville forekomme med en sannsynlighet mindre enn 5 % hvis nullhypotesen var sann. Statistiske tester gir en p-verdi som brukes for å vurdere dette”.

Nullhypotesen for denne undersøkelsen sier at det ikke er noen tverrspråklig innflytelse i noen av førstespråksgruppene. Min utgangshypotese sier på den andre siden at jeg forventer en viss grad av tverrspråklig innflytelse i alle førstespråksgruppene.

Hvis signifikansnivået ligger på $p > .05$, betyr dette at det er en sjanse på 5 % (eller 1 av 20) for at forskjellene som ble funnet *ikke* var på grunn av utgangshypotesen, men på grunn av andre årsaker (Salkind, 2007: 203). Selv om forskerens jobb innebærer å ta høyde for variabler som kan påvirke resultatene, er det umulig å kontrollere for absolutt alle faktorer. Det er heller ikke mulig å være 100% sikker på at informantene man forsker på er et helt representativt utvalg (Salkind, 2007: 204). Derfor er signifikansnivået en pekepinn på hvor sannsynlig det er at resultatene har framkommet ved en tilfeldighet (Salkind, 2007: 203). Med utgangspunkt i dette vil jeg regne alle forskjeller i resultatene som er 5% eller over som statistisk signifikante – det vil si at jeg anser det som usannsynlig at disse resultatene har framkommet ved en tilfeldighet.

⁴ Det finnes imidlertid studier som tyder på at utdannede, voksne andrespråksinnlærere greier å rapportere sine egne språklige nivå med en stor grad av nøyaktighet. Det faktiske nivået til informantene ble målt gjennom tester, og resultatene korrelerte med informantenes selvrappotering (Wilson, 1999: 26).

4 UNDERSØKELSEN

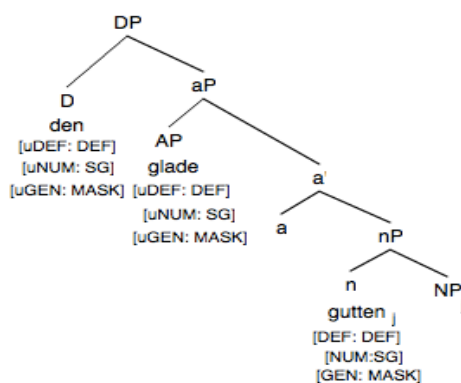
I denne delen av teksten presenterer jeg selve undersøkelsen jeg har gjort. Under 4.1 vil jeg først begrunne valg av førstespråksgrupper til denne undersøkelsen, og forklare hvordan de relevante språkene skiller seg fra hverandre med tanke på fraser som i funksjon tilsvarer dobbel bestemthet. I 4.2 vil jeg forklare undersøkelsens empiri, med forklaring av hva slags fremgangsmåter jeg brukte i innsamling av data, hva slags søkeuttrykk jeg brukte, hvordan jeg kategoriserte og systematiserte dataene, og hvordan jeg kontrollerte for engelskkunnskaper som variabel hos de tyske og vietnamesiske innlærerne. Til slutt vil jeg presentere resultatene fra undersøkelsen og analysere dem under 4.3.

4.1 Valg av språk

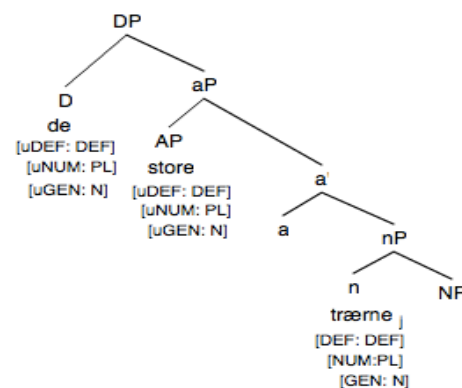
Det var på ingen måte tilfeldig hva slags språk jeg valgte til denne undersøkelsen. De tre språkene tysk, engelsk og vietnamesisk har svært ulike systemer for nominale fraser, både når det gjelder ulike grammatiske kategorier og ulike typer kongruens. Derfor vil kontrastive analyser av dobbel bestemthet i teorien være mulig. I dette underkapitlet vil jeg vise hvordan de ulike språkssystemene skiller seg fra hverandre, og fra norsk.

Norsk har morfologisk uttrykte markører for numerus, genus og definitthet, og krever synlige kongruenstrekk på både adjektivet og determinativet for disse tre grammatiske kategoriene (Jin & Eide, 2015: 295). Norsk er det eneste av språkene i undersøkelsen som har dobbel bestemthets-konstruksjonen. Jeg bruker en forenklet versjon av trestruktursystemet fra Julien (2005) for å vise fram eksempler på norske DP-er (determinativsfraser). På hvert ledd i frasen er de relevante trekkene satt inn, med en *u* foran på de trekkene som er rent formelle. Dette er en vanlig notasjonsmåte i generativ grammatikk, og *u*-en står opprinnelig for ”uninterpretable” – noe som i norsk gjerne kalles *utolkbare trekk* (Jin, Eide & Busterud, 2015: 252). *U*-en signaliserer at trekket ikke har noen semantisk effekt på tolkningen til frasen:

4 (a) Den glade gutten



(b) De store trærne



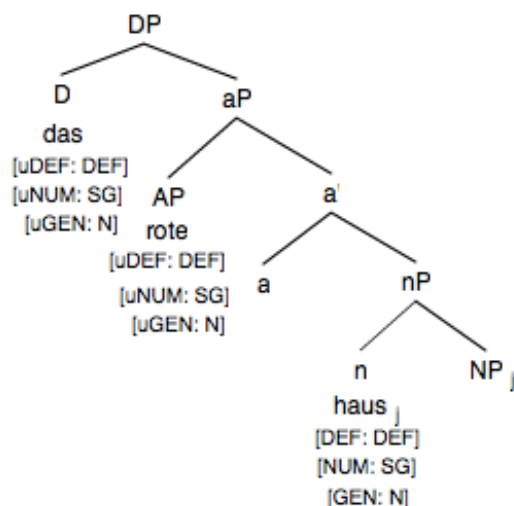
Adjektivet kan også gradbøyes innenfor konstruksjonen (Golden, et al., 2014 : 20). Dermed undersøker jeg ikke bare fraser som ”de store trærne”, men også ”de største trærne” (der adjektivet står i superlativ). Noen adjektiv kan også stå i dobbel bestemthets-konstruksjonen i komparativ form, som i ”de eldre brukerne”, eller ”den yngre garde”. I tillegg har man konstruksjoner med ordenstall, som i ”det første huset”, ”det andre huset” og så videre, og man kan man dessuten linke flere adjektiv sammen, som i frasen ”den store målbare verdien”.

4.1.1 Tysk

Tysk er det språket i undersøkelsen som ligner mest på norsk med tanke på nominalfraser. Språket har morfologisk uttrykte markører for numerus, genus og definitthet, og det kreves synlige kongruenstrekk på adjektivet og determinativet for alle disse markørene (Reiten, 1995: 25, 31, 55). Dette er dermed helt likt norsk. Tysk har imidlertid ikke den samme doble bestemtheten som norske nominalfraser har – det er ingen etterhengt artikkel på nomenet som uttrykker bestemthet:

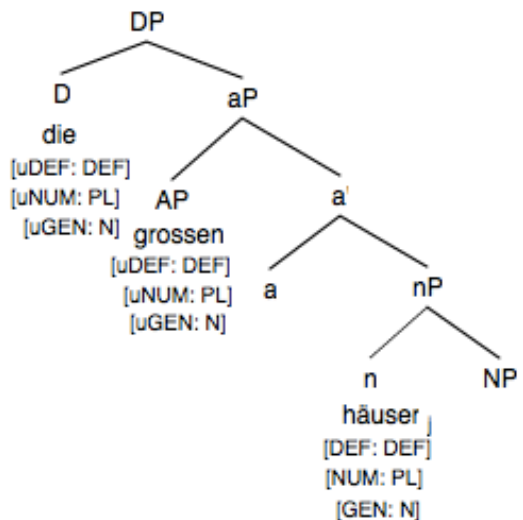
5 (a) Das rote haus

Det røde huset



(b) Die grossen häuser

De store husene



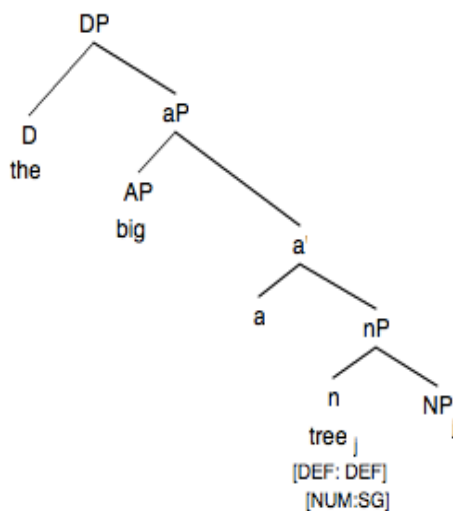
Tysk skiller seg også fra norsk ved at det har synlig kasusbøying, mens norsk kun har rester (hovedsakelig i personlige pronomen). Adjektivet kongruerer også med substantivets kasus (Reiten, 1995: 54). Derfor er også tyske nominalfraser på flere måter mer komplekse enn norske. I entallsformer av svake adjektiv som står i nominativ, har tysk imidlertid e-ending i alle genuskategoriene, noe som tilsvarer det vi har i norsk (Reiten, 1995: 55). I tillegg har adjektivet gradbøying, som i norsk.

4.1.2 Engelsk

Engelsk har morfologisk uttrykte markører for numerus og definitthet, og det kreves kun synlige kongruenstrekk for numerus på determinativet (Jin & Eide, 2015: 302-303). Kongruenstrekkene på determinativet er imidlertid ikke i bruk i fraser som tilsvarer dobbel bestemthet i norsk. Det er kun på demonstrativer som *this/these* at det er kongruens – ikke på den bestemte artikkelen *the* (Jin & Eide, 2015: 290). Det er heller ingen kongruenstrekk på de engelske adjektivene, men man gradbøyer adjektivene i likhet med det norske systemet. Forskjellene mellom det norske og det engelske nominalsystemet er dermed nokså store: I de engelske frasene er det ingen kongruens i det hele tatt (i fraser som tilsvarer dobbel bestemthet), og kun morfologisk uttrykte markører for numerus og definitthet:

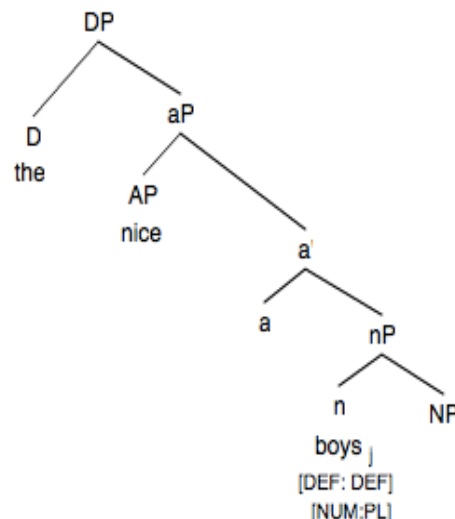
6 (a) The big tree

Det store treet



(b) The nice boys

De snille guttene



4.1.3 Vietnamesisk

Vietnamesisk har verken morfologiske uttrykte markører for numerus, genus eller definitthet, og derfor heller ingen kongruens når det gjelder disse grammatiske kategoriene (Husby, 1991: s. 93-94). For å uttrykke definitthet eller numerus må man sette til bestemte ord. Dessuten blir svært mye uttrykt gjennom klassifikatorer. Ved å bruke klassifikatorer kan man omtale de enkelte enhetene i hver gruppe, og i tillegg kan enkeltenheter klassifiseres etter ulike kvaliteter – for eksempel om noe er levende eller ikke-levende, om det er menneskelig, om det har en bestemt form og så videre (Husby, 1991: s. 93). Vietnamesisk har dessuten heller ingen ordklasse som tilsvarer det vi på norsk kan kalle adjektiv. De ordene som tilsvarer

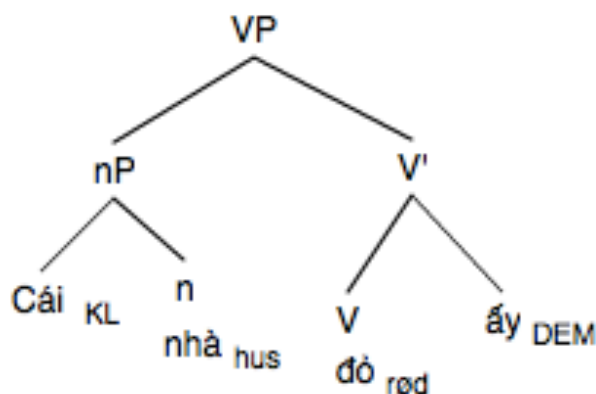
adjektiv er en type verb på vietnamesisk. Ordet ”nhanh”, som på norsk ville blitt ”rask”, må derfor heller oversettes til ”å være rask” (Husby, 1991: s. 123-124). Derfor får eksemplet under, ”cái nhà đỏ ấy”, betydningen ”det huset som er rødt” i norsk:

7 (a) Cái nhà đỏ ấy⁵

KL⁶ hus rød DEM⁷

Det huset som er rødt

(Det røde huset)



4.1.4 Oversikt over språklige trekk

For å skape en viss oversikt, har jeg plassert de språklige trekkene inn i to tabeller. 8 (a) viser hva slags grammatiske kategorier og kongruenstrekk som finnes i nominale fraser i de relevante språkene, og 8 (b) viser hva slags kongruenstrekk som finnes på determinativ og adjektiv. Som nevnt er en vanlig notasjonsmåte i generativ grammatikk er å markere trekk som rent formelle ved å skrive en *u* foran. U-en står opprinnelig for ”uninterpretable”, og kalles i norsk gjerne *utolkbare trekk* (Jin, Eide, & Busterud, 2015: 252). Utolkbare trekk er rent formelle, og gir ingen semantisk effekt på tolkningen til setningen (Jin, et al., 2015: 243). Denne notasjonsmåten vil jeg også bruke her:

⁵ Det vietnamesiske eksemplet (cái nhà đỏ ấy) har jeg fått gjennom personlig kommunikasjon med Olaf Husby, og trestrukturen er også basert på dette. Strukturen er løst omtolket til systemet fra Julien (2005).

⁶ KL: klassifikator.

⁷ DEM: demonstrativ.

8 (a) Grammatiske kategorier og kongruens

Grammatisk kategori	Norsk	Tysk	Engelsk	Vietnamesisk
Numerus	X	X	X	O
Numeruskongruens	X	X	X*	O
Genus	X	X	O	O
Genuskongruens	X	X	O	O
Definitthet	X	X	X	O
Definittkongruens	X	X	O	O
Dobbel bestemthet	X	O	O	O

X: Språket har trekket O: Språket har ikke trekket *Engelsk har kun numeruskongruens på demonstrativer

(b) Kongruenstrekk på determinativ og adjektiv⁸

Språk	Determinativ			Adjektiv		
	[uNUM]	[uGEN]	[uDEF]	[uNUM]	[uGEN]	[uDEF]
Norsk	X	X	X	X	X	X
Tysk	X	X	X	X	X	X
Engelsk	X*	O	O	O	O	O
Vietnamesisk	O	O	O	O	O	O

X: Språket har trekket O: Språket har ikke trekket *Engelsk har kun numeruskongruens på demonstrativer

4.2 Empiri

Arbeidet med datainnsamling til denne masteroppgaven har vært sentrert rundt søk i ASK.

I denne delen av teksten vil jeg gjøre rede for hvordan jeg har søkt opp, registrert og kategorisert dobbel bestemthets-konstruksjonen fra korpuset, og hvordan jeg har kvalitetssikret det arbeidet jeg har gjort.

4.2.1 Søk i ASK

Arbeidet med søkeuttrykk i ASK var en prosess som tok forholdsvis lang tid, og som involverte en del prøving og feiling. Det søkeuttrykket som ble stående igjen til slutt var dette:

```
[lemma = "det|den|de" %c] [pos = "adj"] ([pos = "subst" & morph = ("be" "fl")] | [pos = "subst" & morph = ("be" "ent")]) \ < \ \ <del> :: language = "norsk" & testlevel = "Språkprøven"
```

⁸ Disse tabellene er direkte inspirert av Jin og Eide (2015: 302-303), og informasjonen om norsk og engelsk er også hentet derfra. Informasjonen om tysk har jeg hentet fra Reiten (1995), og opplysningene om vietnamesisk er samlet gjennom personlig kommunikasjon med Olav Husby og fra Husby (1991).

Basic search | **Advanced search**

[lemma = "det|den|de" %c] [pos = "adj"] ([pos = "subst" & morph = ("be" "fl")] | [pos = "subst" & morph = ("be" "ent")]) \ <> \ \ :: language =

Run Query | Saved queries | Save Query as

Hit 1 - 40 of 110 | [next 40 hits](#) | Go to: | Download (Excel mode) | Type: | Show annotation | Show line filter | Attrib

pid	match	
s2003	meg selv. Men jeg må da finne ut og sette ord på hva som har fanget meg.	Det andre begrepet - ansvar - betyr for meg
s2005	n har festes sin litt til Jesus. Jesus som døde for alle menneskers synder er	den viktigste verdien /personen. Det er stort fo
s2005	viktig her i livet. Men hva er viktigst her i livet? Kristne vil gi JESUS. I følge	det <sic> Bibelske </sic> synet , spiller det over ikke så s
s2008	lse og tretthet t, plager som i stor grad kan knyttes til brød og melk. Me På	den andre siden ser man at en ukritisk tilb
s2008	kosthold som er sunt og allsidig, men hvor det kan være rom for å kutte ut	de sunne matvarene du selv kjenner at du rea
s2008	HVEM BESTEMMER HVA SOM ER SUNT?	Den siste tiden har det gått en diskusjon,
s2010	søte ting synes umettelig. Det snakkes om en epedemi som sprer seg i alle	de vestlige landene , og at vi er i ferd med å f
s2010	rgene og belaster et hardt presset helsebudsjett. For ikke å snakke om alle	de psykiske problemene folk har. " Det skal god ry
s2014	an. Menneskene jakter jo også på lykken. En spør seg: Ligger den snarere i	de enkle tingene som ofte er helt gratis?

Dette søket har først et lemma, som består av determinativene *det*, *den* eller *de*, som er de obligatoriske første delene i en dobbel bestemthets-konstruksjon. Det neste leddet i søket er ”adj”, eller et adjektiv, som er mellomleddet i dobbel bestemthets-konstruksjonen. Til slutt kommer ”subst”, altså substantivet. Egenskapene som er satt for substantivet er ”be”, altså bestemt form, enten i flertall, ”fl”, eller entall, ”ent”. Til slutt kommer spesifikasjonene for språk og for testnivå, som for denne undersøkelsen er satt til *Språkprøven*.

I tillegg til dette foretok jeg også søk med doble adjektiv, som for eksempel ”de best betalte jobbene”, eller ”den store målbare verdien”. Søkeuttrykket var ellers helt likt det som er beskrevet ovenfor:

Basic search | **Advanced search**

[lemma = "det|den|de" %c] [pos = "adj"] [pos = "adj"] ([pos = "subst" & morph = ("be" "fl")] | [pos = "subst" & morph = ("be" "ent")]) \ <> \ \ ::

Run Query | Saved queries | Save Query as

Hit 1 - 8 of 8 | Download (Excel mode) | Type: | Show annotation | Show line filter | Attributes ... |

pid	match	
s2014	å penger for å ha mest mulig å rutte med til livets opphold. Det gjelder å få	de best betalte jobbene , selv om e
s2014	jobbene, selv om ens egne interesser ikke er helt med. I iver etter å oppnå	de helt store summene er det man
s2045	trappa istedenfor heisen Vi kan gjøre noe selv for å holde oss i form. Det er	de små daglige innsatsene som teller;
s2066	jen om det å ikke være så materialistiske i velferdstaten Norge. Sett pris på	de små ubetalelige tingene . Venninne
s2090	en påkjønning utsetter man helsen sin for? For å få til en virkelig ending av	den stadig økende overvekten , må man e
s2097	Penger og andre verdier. I vårt samfunn er penger	den store målbare verdien . Har du m
s2099	ykket, være fysisk velsignet og vedlikeholdt, samt skilte med friskt mot og	de rette moralske verdiene . I våre dag
s2100	den minste form for utfordring til regime. Samtidig så Saddam muligheter i	den stadig økende eurokursen og fallende

Hit 1 - 8 of 8

Jeg brukte nøyaktig de samme søkene på alle førstespråksgruppene, men jeg forandret naturligvis språket som var involvert i søkeuttrykket, slik at det sto enten ”tysk”, ”engelsk” eller ”vietnamesisk” der ”norsk” nå står. Jeg valgte å ha på KWIC-konkordansen i søket, slik at konteksten for treffene skulle komme godt frem. I tillegg inkluderte jeg en del attributter som viste informantenes metadata, som alder, kjønn og hva slags land vedkommende kom fra. På de to gruppene med tysk og vietnamesisk som førstespråk, la jeg også til engelskkunnskaper som synlig attributt, for å kunne kontrollere for denne variabelen.

Engelsknivået ble klassifisert etter seks ulike grupper: *ikke noe*, *begynnernivå*, *grunnivå*, *mellomnivå*, *høyere nivå*, og *ukjent* (mer om dette i kapittel 4.2.3).

På grunn av en usikkerhet rundt om alle konstruksjonene virkelig dukket opp i søkene jeg gjorde, foretok jeg meg en kvalitetssikring. Dette gjorde jeg ved å grundig lese gjennom minst 15 ulike tekster fra hver førstespråksgruppe, og ved å manuelt markere alle tilfeller av dobbel bestemthet i disse tekstene. Denne metoden avdekte en del unøyaktigheter i søkeuttrykkene. Da kontaktet jeg professor Kari Tenfjord ved UiB, som gav meg veldig god hjelp ved å foreslå endringer på søkeuttrykkene mine.

Etter at jeg hadde brukt søkeuttrykket på hver førstespråksgruppe, lagret jeg dem i Microsoft Excel. Alle søkene ligger vedlagt i Excel-format (som vedlegg 1-4), fordi dette er den mest systematiske måten å vise søkene i sin helhet. I dette formatet gikk jeg over alle treffene jeg hadde fått, og huket vekk resultater som hadde kommet med til tross for at de ikke var dobbel bestemthets-konstruksjoner (disse markerte jeg med en X, noe som vil være synlig i vedleggene). Dette var nokså få av disse.

4.2.2 Kategorisering

I kategoriseringen av dataene sorterte jeg antall forekomster av dobbel bestemthets-konstruksjonen under distribusjonsfunksjonen i korpuset. Jeg sorterte dem ut fra tekst-ID, slik at jeg kunne se hvor mange forekomster av dobbel bestemthet det var per informanttekst. De resultatene jeg huket vekk tidligere, trakk jeg fra manuelt ettersom antallet var så lite, og det gjorde jeg også med utgangspunkt i tekst-IDen.

Basic search | **Advanced search**

[lemma = "det|den|de" %c] [pos = "adj"] ([pos = "subst" & morph = ("be" "fi")] | [pos = "subst" & morph = ("be" "ent")]) \ <> \\ :: language =

Run Query | Saved queries | Save Query

Fractions sum up to 1.0 in each row. Fractions in blue are unweighted means of group fractions. Fractions in green are distr

Show distribution of word %c, Δ: 0 filter: relative to pid %c, Δ: type: absolute | counts only | include structures

Page 1/1 of 2x7. Previous Next | Download

	(sum)	de psykiske problemene	Det andre begrepet	det <sic> Bibelske </sic> synet	den viktigste verdien	Den siste tiden	den andre siden
(sum)	110 (100,0)	1 (0,9)	1 (0,9)	1 (0,9)	2 (1,8)	2 (1,8)	3 (2,7)
	110 (100,0)	1 (0,8)	1 (1,7)	1 (0,8)	2 (1,7)	2 (1,4)	3 (3,1)
s2003	1 (100,0)		1 (100,0)				
s2005	2 (100,0)			1 (50,0)	1 (50,0)		
s2008	3 (100,0)					1 (33,3)	1 (33,3)
s2010	2 (100,0)	1 (50,0)					

Det samme gjorde jeg med det doble adjektivsøket. Resultatene fra begge søkene la jeg sammen manuelt, også her med utgangspunkt i tekst-ID for å holde styr på hvor mange forekomster det var per informanttekst (denne oversikten er også vedlagt). Dette gav meg

oversikt over hvor mange informanter som produserte dobbel bestemthet i hver førstespråksgruppe (og i den norske kontrollgruppen), og hvor mange forekomster av dobbel bestemthet det samlet sett var per gruppe. Distribusjonsfunksjonen ble også brukt til å finne spredningen av engelskkunnskaper hos de tyske og vietnamesiske førstespråksgruppene. Her oppsto riktignok problemer med at distribusjonen kun viste antall forekomster per nivågruppe (*ikke noe, ukjent, begynnernivå, grunnivå, mellomnivå, høyere nivå*) i engelsk – jeg fikk ikke funksjonen til å sortere etter antall tekst-IDer. En informant som hadde produsert 4 dobbel bestemthets-konstruksjoner gav derfor 4 utslag på det nivået vedkommende lå på, og kunne få dette nivået til å se kunstig høyt ut. Derfor gikk jeg også over tekst-IDene manuelt, og passet på at kun én av konstruksjonene ble medregnet per informanttekst. Oversikt over engelskkunnskaper for de tyske og vietnamesiske førstespråksgruppene er lagt til i underkapittel 4.2.3.

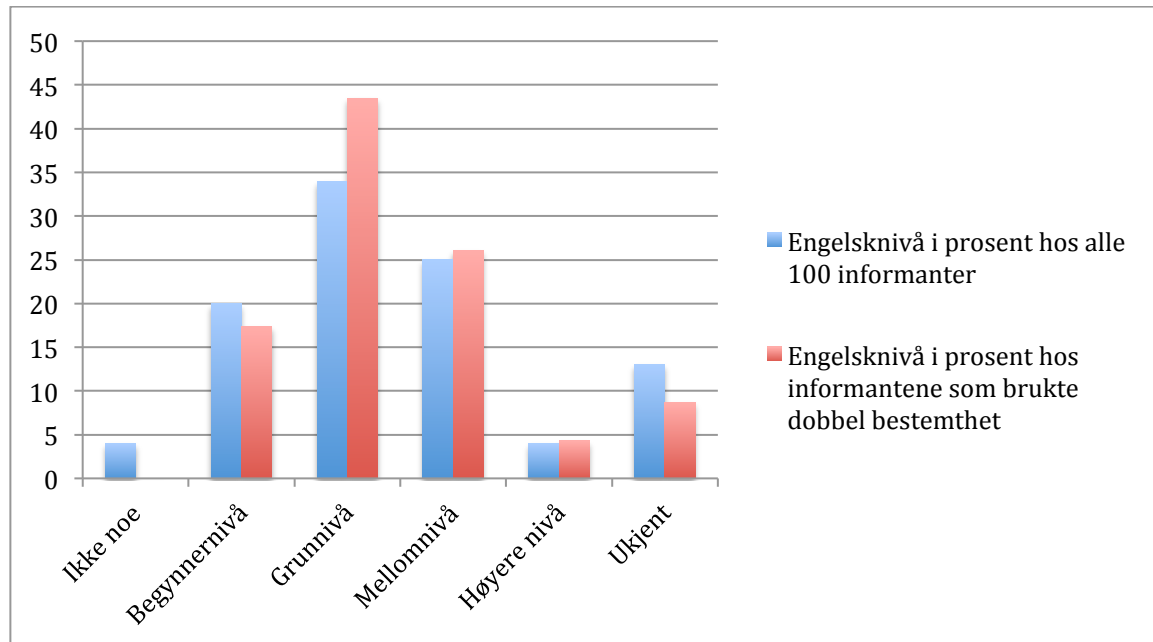
For å få en oversikt over de avvikende konstruksjonene som informantene hadde produsert, gikk jeg over Excel-dokumentene manuelt, og sorterte avvikene inn i ulike grupper: En gruppe som viste absolutt alle avvik i antall, en som viste antall avvik på kongruensen til adjektivet, en som viste antall avvik på genuskongruensen til determinativet, og en som viste antall avvik på numeruskongruensen til determinativet. Rene skrivefeil, samsvarende bruk av feil genuskategori (som ”det nyeste modellet”), og kontekstuelle avvik som for eksempel ”den store søsteren” (der man på norsk vanligvis ville valgt å si ”storesøster”) blir ikke regnet som avvik i denne masteroppgaven. Det er kun det rent grammatiske som blir undersøkt her.

4.2.3 Engelskkunnskaper som variabel

Som nevnt i teorikapittel 2.5, kan transfereffekter komme fra et hvilket som helst språk innlæreren tidligere har lært. Derfor valgte jeg å undersøke faktoren *engelskkunnskaper* både for den tyske og den vietnamesiske førstespråksgruppen. Dette gjorde jeg ved å først finne engelsknivået til alle de 100 informantene i hver førstespråksgruppe, og senere finne engelsknivået hos de informantene som brukte dobbel bestemthets-konstruksjonen. Engelsknivået er klassifisert ut fra 6 ulike grupper: *ikke noe, begynnernivå, grunnivå, mellomnivå, høyere nivå, og ukjent*. Når jeg visste hvor mange informanter som lå på hvilket nivå, regnet jeg ut prosenten slik at nivået hos alle 100 informanter kunne sammenlignes direkte med nivået hos informantene som brukte dobbel bestemthet (oversikt over prosentene finner man nedenfor). For å gi et oversiktlig bilde over denne variabelen, plottet jeg

prosentene inn i søylediagrammer. Her kan man se spredningen av de ulike engelsknivåene i den vietnamesiske førstespråksgruppen:

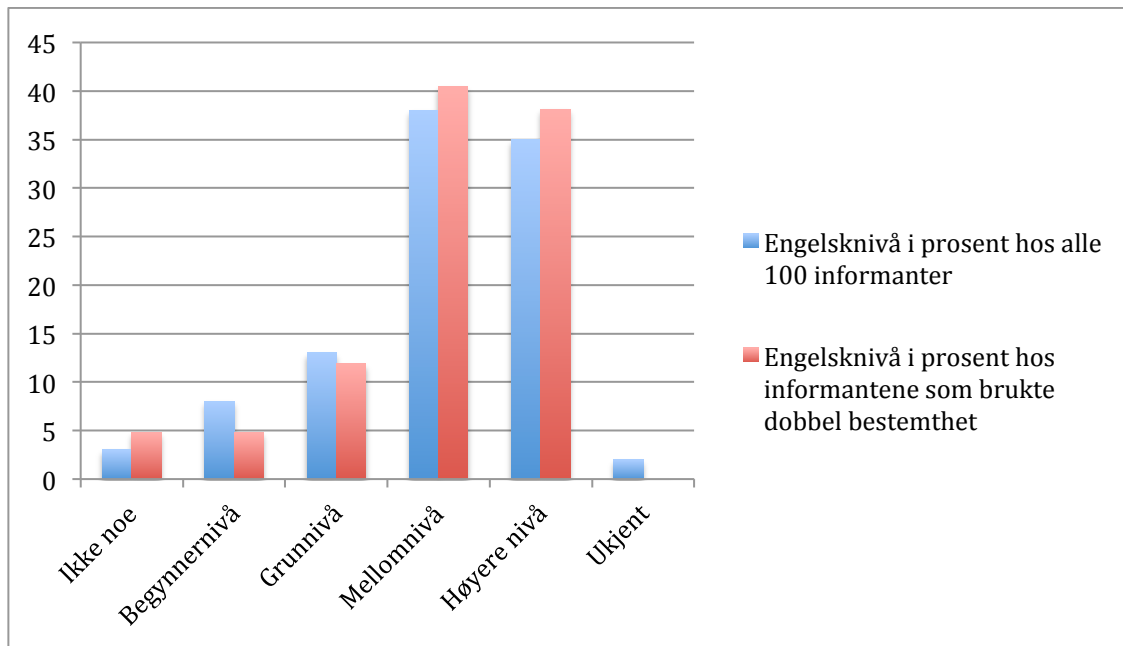
Engelsknivå i vietnamesisk førstespråksgruppe



Dette diagrammet viser fordelingen av informanter på ulike engelsknivå hos den vietnamesiske førstespråksgruppen. De blå søylene viser nivåene til alle 100 informanter i prosent, med 4% på ikke noe, 20% på begynnernivå, 34% på grunnivå, 25% på mellomnivå, 4% på høyere nivå og 13% på ukjent. De røde søylene viser nivåene til de 23 informantene som brukte dobbel bestemthet i prosent, med 0% på ikke noe, 17,4% på begynnernivå, 43,5% på grunnivå, 26,1% på mellomnivå, 4,3% på høyere nivå og 8,7% på ukjent.

Det er ingen statistisk signifikante forskjeller (se 3.4) i disse prosentene, foruten i grunnivågruppen. Der er det en forskjell på 9,5%. Hvis man legger sammen alle de faktiske nivåene av engelskkunnskaper (man fjerner med andre ord ikke-gruppen og ukjent-gruppen), ser man det samme. 83% av alle informantene har engelskkunnskaper i en eller annen grad, mens dette er på 91,3 % i dobbel bestemthets-gruppen – en forskjell på 8,3%. Dette kan indikere at engelskkunnskapene hos den vietnamesiske førstespråksgruppen har innvirkning på om dobbel bestemthets-konstruksjonen blir produsert overhodet: Det kan altså se ut til at generelle engelskkunnskaper øker sjansen for at konstruksjonen blir produsert. Man må likevel ta høyde for at en viss prosent er ukjent – 13% hos alle informantene, og 8,7% hos dobbel bestemthets-gruppen. I tillegg må man ta høyde for at selvrapporing er en mulig feilkilde i denne sammenheng. Derfor er dette resultatet noe usikkert.

Engelsknivå i tysk førstespråksgruppe



Dette diagrammet viser fordelingen av informanter på ulike engelsknivå hos den tyske førstespråksgruppen. De blå søylene viser nivåene til alle 100 informanter i prosent, med 3% på ikke noe, 8% på begynnernivå, 13% på grunnivå, 38% på mellomnivå, 35% på høyere nivå og 2% på ukjent. De røde søylene viser nivåene til de 42 som brukte dobbel bestemthet i prosent, med 4,8% på ikke noe, 4,8% på begynnernivå, 11,9% på grunnivå, 40,5% på mellomnivå, 38,1% på høyere nivå og 0% på ukjent. Her er det ingen statistisk signifikante forskjeller mellom alle informantene og informantene som brukte dobbel bestemthet.

4.3 Resultater og analyse

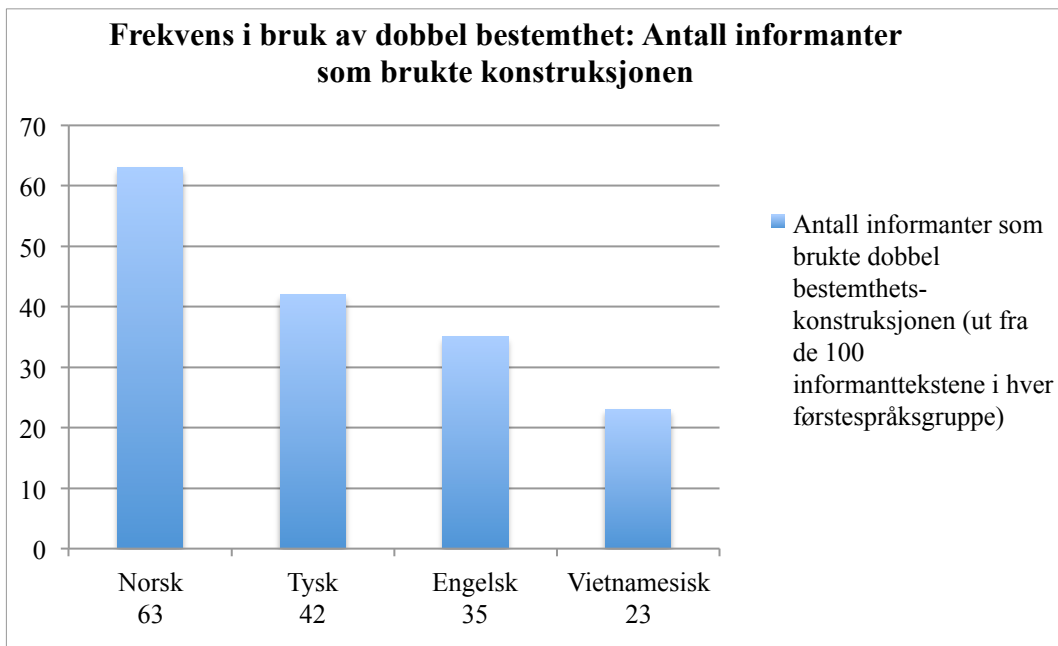
I denne delen av masteroppgaven skal jeg presentere resultatene fra undersøkelsen min, og gjøre en analyse av disse. Først vil jeg gå inn på resultater og analyse knyttet til frekvens i 4.3.1, og i 4.3.2 vil resultater knyttet til avvik diskuteres og analyseres. Under 4.3.3 vil det være en kort oppsummering av alle resultatene. Til slutt kommer en avsluttende oppsummering under 4.3.4. Før jeg presenterer resultatene for undersøkelsen jeg har gjort, kommer en repetisjon av problemstillingen for denne masteroppgaven:

- 1) *Hva karakteriserer bruken av dobbel bestemthets-konstruksjoner i norsk mellomspråk hos voksne andrespråksinnlærere med tysk, engelsk og vietnamesisk som førstespråk, med tanke på det rent grammatiske og når det gjelder relativ frekvens?*

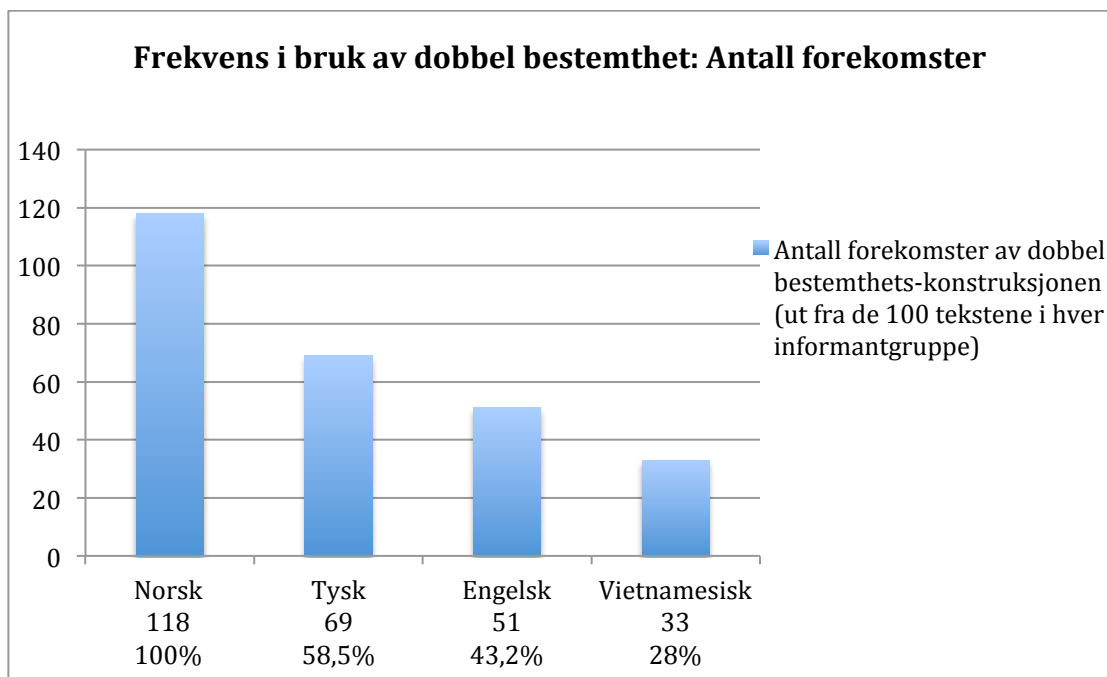
2) *Varyerer disse karakteristikaene ut fra hvilket førstespråk man har, og kan man eventuelt påvise tverrspråklig innflytelse ut fra resultatene?*

4.3.1 Frekvens

Frekvensen som dobbel bestemthets-konstruksjonene forekommer med, er i stor grad med på å karakterisere hvordan konstruksjonen blir brukt i de ulike førstespråksgruppene. Funnene i denne undersøkelsen tyder på at det er stor variasjon mellom de tre førstespråksgruppene som ble undersøkt opp mot den norske kontrollgruppen, både når det gjelder antall informanter som brukte dobbel bestemthet, og når det gjelder det samlede antallet forekomster: Den norske kontrollgruppen hadde 118 forekomster av dobbel bestemthet, fordelt på 63 informanttekster. Det var altså 63 av de 100 norske informanttekstene som i det hele tatt inneholdt dobbel bestemthet. I den tyske førstespråksgruppen var det 69 forekomster av konstruksjonen fordelt på 42 informanttekster. Det var altså vesentlig færre forekomster samlet, og også et lavere antall informanttekster som inneholdt konstruksjonen. I den engelske førstespråksgruppen var det 51 forekomster, fordelt på 34 informanttekster. Til sammenligning hadde den vietnamesiske førstespråksgruppen bare 33 forekomster fordelt på 23 tekster.



Dette diagrammet viser også prosenten, fordi det er 100 informanttekster per førstespråksgruppe.



I dette diagrammet ser man antall forekomster av dobbel bestemthet i den norske kontrollgruppen, og i de tre førstespråksgruppene. Dette er riktignok ikke i prosent. Prosenten står under, og den er regnet ut med utgangspunkt i at de norske 118 forekomstene tilsvarer 100%. Det betyr at jeg indikerer at 118 forekomster per 100 informanter er normalen i norsk på innfødnivå. En mulig feilkilde i dette tilfellet er enkeltinformanter som skiller seg ut. To av informantene i den norske kontrollgruppen hadde så mye som seks forekomster av konstruksjonen hver, og dette utgjør store utslag på prosentene. Derfor er dette kun en indikasjon.

Tysk, som er det språket som ligner mest på norsk, hadde det høyeste antallet informanter som brukte konstruksjonen, og det høyeste antallet forekomster samlet av de tre førstespråksgruppene. Likevel er det store forskjeller mellom den tyske førstespråksgruppen og den norske kontrollgruppen – det var 21% færre informanter brukte dobbel bestemthet, og hele 41,5% lavere forekomst samlet, noe som viser en statistisk signifikant intergruppeheterogenitet mellom den tyske førstespråksgruppen og den norske kontrollgruppen. Den tyske DP-en er lik den norske på de fleste områder, men tysk har ikke dobbel bestemthet. En forklaring på disse resultatene er at det kan være en underproduksjon av dobbel bestemthetskonstruksjonen i den tyske førstespråksgruppen, eller eventuelt en unnvikende produksjon (se diskusjon av disse begrepene i kapittel 2.3) på grunn av ulikhetene mellom språkene. Hvilket av disse fenomenene vi har med å gjøre, er derimot vanskeligere å si. En unnvikende produksjon betyr at innlæreren har kompetansen, at vedkommende vet hvordan dobbel bestemthetskonstruksjonen skal produseres, men unngår det på grunn av mangel på

selsikkerhet eller automatisering. Underproduksjonen skyldes i motsetning til dette manglende kompetanse – at innlæreren ikke vet hva dobbel bestemthet er (Abrahamsson, 2009: 237). I en korpusundersøkelse er det vanskelig å finne ut noe om intensjoner eller tanker rundt valg av ord og setninger, eller om informanten vet at en konstruksjon eksisterer eller ikke. Man har ikke muligheter til å stille oppfølgingsspørsmål, og av denne årsaken er det ikke enkelt å kunne si noe sikkert angående forekomstene i denne undersøkelsen. Man kan imidlertid foreslå disse to fenomenene som mulige forklaringer på den lavere relative frekvensen i den tyske førstespråksgruppen, uten at man egentlig kan vite hvilken av dem som er mest sannsynlig. I tillegg er det et viktig poeng at man ikke kan snakke om noen unnvikende produksjon eller underproduksjon i den norske kontrollgruppen, selv om det var 37% som *ikke* produserte dobbel bestemthet der også. Det var en høyere prosent som ikke produserte konstruksjonen i den tyske gruppen, men hele denne prosenten kan altså ikke forklare som unnvikende produksjon eller underproduksjon.

Resultatene for den engelske førstespråksgruppen var også interessante. Det engelske nominalsystemet skiller seg fra det norske på en god del områder, noe som kan være med på å forklare resultatene: Det er ingen morfologisk uttrykte markører for genus i engelsk, ingen kongruenstrekk på verken adjektivet eller determinativet (hvis man ser bort fra demonstrativer), og ingen dobbel bestemthet. Engelske DP-er skiller seg altså mer fra norsk enn de tyske DP-ene. Dette kan ha gjort utslag på frekvensen som dobbel bestemthet forekommer med i den engelske førstespråksgruppen: Det var 28% færre informanter som brukte dobbel bestemthet i den engelske førstespråksgruppen sammenlignet med den norske kontrollgruppen, og hele 56,8% færre forekomster samlet. Dette viser en statistisk signifikant intergruppe-heterogenitet mellom den engelske førstespråksgruppen og den norske kontrollgruppen – men også mellom den engelske og den tyske: Det var 7% færre informanter som brukte dobbel bestemthets-konstruksjonen i den engelske enn i den tyske førstespråksgruppen, og 15,3% færre forekomster samlet. Ettersom engelsk har færre likheter med norsk enn det tysk har, kan dette være med på å forklare et lavere antall informanter som brukte dobbel bestemthet, og færre forekomster samlet. Her kan også en mulig forklaring ligge i under- eller unnvikende produksjon, som diskutert ovenfor. Hvis man bruker denne forklaringen, kan man også antyde at forskjellene mellom nominalsystemene i norsk og engelsk gir en enda større grad av underproduksjon eller unnvikende produksjon enn det gjør i den tyske førstespråksgruppen.

Vietnamesisk ligger klart lengst unna det norske nominalsystemet av språkene i denne undersøkelsen – det har verken morfologisk uttrykte markører for numerus, genus eller

definitthet, og derfor ingen kongruens når det gjelder disse grammatiske kategoriene, og ingen dobbel bestemthet. Språket har heller ingen ordklasse som tilsvarer det vi på norsk kan kalle adjektiv. Dette kan se ut til å ha hatt innvirkning på frekvensen som dobbel bestemthet forekom med i den vietnamesiske førstespråksgruppen: Det var 40% færre informanter som produserte dobbel bestemthet enn i den norske kontrollgruppen, og det samlede antallet forekomster var hele 72% lavere. Her var det kanskje en viss påvirkning fra engelsk, ved at generelle engelskkunnskaper så ut til å øke sjansen for at konstruksjonen ble produsert (se 4.2.3). Hvis dette stemmer, kan man også si at resultatene er enda mer slående. Tallene viser en statistisk signifikant intergruppe-heterogenitet mellom den vietnamesiske førstespråksgruppen og kontrollgruppen. Dette kan vi også se mellom den vietnamesiske førstespråksgruppen og de andre førstespråksgruppene: Det var 19% færre informanter som produserte dobbel bestemthet enn i den tyske førstespråksgruppen, og 12% færre enn i den engelske. Det var også 30,5% færre forekomster samlet i den vietnamesiske førstespråksgruppen enn i den tyske, og 15,5% færre enn i den engelske.

Resultatene gir samlet sett et inntrykk av jo mer ulikt nominalsyste­met i førstespråkene er med målspråket norsk, jo lavere frekvens forekommer dobbel bestemthetskonstruksjonen med. En mulig forklaring på dette er at det er en underproduksjon eller en unnvikende produksjon i alle førstespråksgruppene, og at denne blir større jo større forskjellen er mellom det relevante nominalsyste­met og nominalsyste­met i norsk. Hvis man følger Jarvis' metode for påvisning av tverrspråklig innflytelse, viser frekvensforskjellene en statistisk signifikant intergruppe-heterogenitet og relativt klart tverrspråklig performansesamsvar i alle førstespråksgruppene. Om man kan finne evidens for to av de tre punktene for påvisning av transfer (intergruppe-heterogenitet, intragruppe-homogenitet og tverrspråklig performansesamsvar), er dette i følge Jarvis' tilstrekkelig med evidens for å kunne påvise tverrspråklig innflytelse. Med utgangspunkt i dette, mener jeg det er sannsynlig at det er snakk om tverrspråklig innflytelse i sammenheng med relativ frekvens i alle førstespråksgruppene.

4.3.2 Avvik

For å finne ut hva som karakteriserer bruken av dobbel bestemthet hos de ulike førstespråksgruppene, er avvikene viktige. Her vil jeg først undersøke om det i det hele tatt finnes avvik i de ulike gruppene. Hvis det gjør det, vil jeg undersøke hva slags avvik det er snakk om. Som nevnt, blir rene skrivefeil, og kontekstuelle avvik som for eksempel "den store søsteren" (der man på norsk vanligvis ville valgt å si "storesøster") ikke regnet som

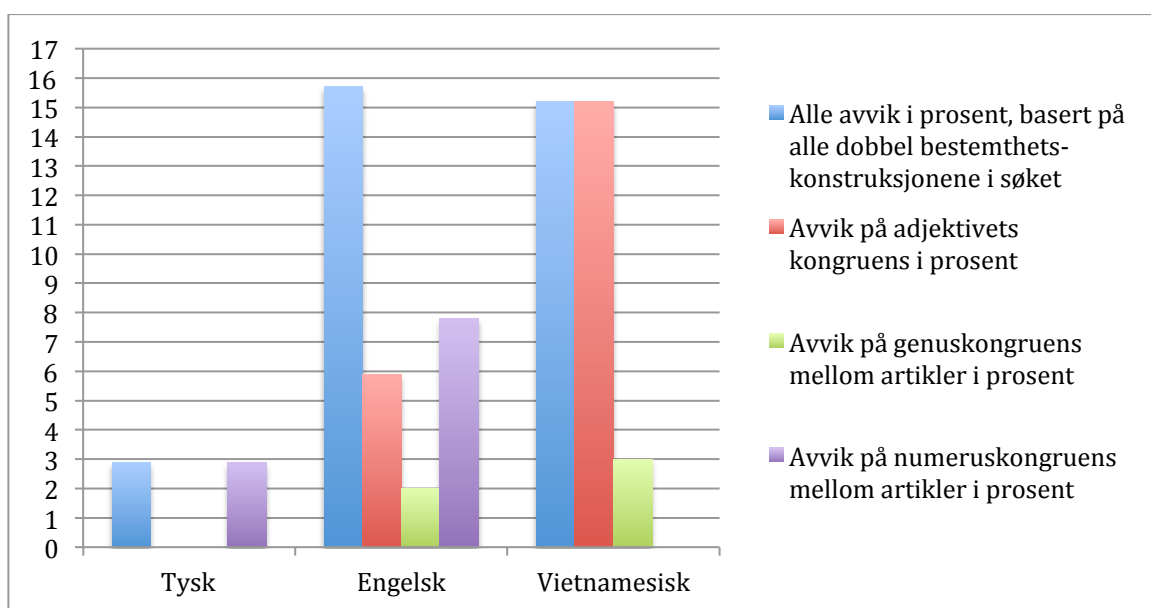
avvik i denne masteroppgaven. Valg av avvikende genuskategori med samsvarende kongruens, som i ”det nyeste modellet” blir heller ikke regnet som avvik – i motsetning til ”det største grunnen”, som mangler samsvarende genus. Det er kun det rent grammatiske som blir undersøkt i denne masteroppgaven.

Jeg har delt inn avvikene i ulike grupper: En gruppe som viser *alle avvik* i antall, en som viser *avvik på kongruensen til adjektivet*, en som viser *avvik på genuskongruensen til determinativet*, og en som viser *avvik på numeruskongruensen til determinativet*. I tabellen under ser man en oversikt over avvikene som ble funnet i denne undersøkelsen, og hva slags avvik dette var. Her kan man se det samlede antallet avvik, samt antall avvik i hver kategori. Der det under for eksempel tysk står ”2 av 69” betyr det at 2 av de 69 forekomstene av dobbel bestemthets-konstruksjonen var avvikende i den relevante gruppen. ”(2/42)” betyr at avvikene forekom i 2 av de 42 informanttekstene. Dette er gjort for å vise om individuelle informanter eventuelt har flere avvik. Alle avvikene er også markert med tekst-ID i parentes bak, slik at man får oversikt over hvilke avvik hver informant har gjort:

Språk	Alle avvik	Avvik på adjektivets kongruens	Avvik på genuskongruensen til determinativet	Avvik på numeruskongruensen til determinativet
Norsk	0 av 118 (0/63)	0 av 118 (0/63)	0 av 118 (0/63)	0 av 118 (0/63)
Tysk	2 av 69 (2/42) ”den gamle dagene”(0124), ”den første frostnettene” (0453),	0 av 69 (0/42)	0 av 69 (0/42)	2 av 69 (2/42) ”den gamle dagene”(0124), ”den første frostnettene” (0453),
Engelsk	8 av 51 (7/35) ”den gode tidene” (0316), ”den dårlige tidene” (0316), ”den negative punktene” (0329), ”den siste årene” (0342), ”den grusom lyden” (0346), ”det største grunnen” (0738), ”det fast	3 av 51 (3/35) ”den grusom lyden”(0346), ”det fast telefonsystemet” (0194), ”det mest viktig sake” (0194)	1 av 51 (1/35) ”det største grunnen” (0738)	4 av 51 (3/35) ”den gode tidene” (0316), ”den dårlige tidene” (0316), ”den negative punktene” (0329), ”den siste årene” (0342)

	telefonsystemet” (0751), ”det mest viktig sake” (0194),			
Vietnamesisk	5 av 33 (5/23) ”den interessant sporten” (0583), ”den størst søsteren” (0625), ”det morsomt programet” (0638), ”den ordentlig livet” (0683), ”det nødvendig spørsmålet” (0885)	5 av 33 (5/23) ”den interessant sporten” (0583), ”den størst søsteren” (0625), ”det morsomt programet” (0638), ”den ordentlig livet” (0683), ”det nødvendig spørsmålet” (0885)	1 av 33 (1/23) ”den ordentlig livet”(0683)	0 av 33 (0/23)

Jeg har også omregnet antallet avvik i tabellen over til prosent, som vist i søylediagrammet under. Tallene i diagrammet er basert på alle dobbel bestemthets-konstruksjonene i søkene jeg har gjort, og om en konstruksjon var avvikende på flere måter, ble den plassert i alle passende kategorier. Eksemplet ”den ordentlig livet” fra en informant i den vietnamesiske informantgruppa, passer både inn i avvik på genuskongruensen og i avvik på adjektivets kongruens. Derfor overlapper prosentene delvis hverandre:

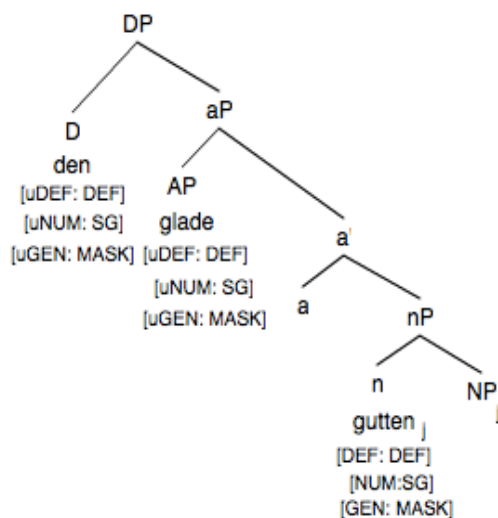


I den norske kontrollgruppen var det ingen avvik, og derfor vil jeg også regne dette som normalen i målpråket. For å ha en helt innfødtlik kompetanse i norsk, mener jeg altså at man normalt sett ikke har noen avvik på adjektivets kongruens, eller på genus- og numeruskongruensen på determinativet. Jeg vil nå gå gjennom hver førstespråksgruppe og gi en komparativ analyse av avvikene jeg fant i denne undersøkelsen.

Tysk førstespråksgruppe

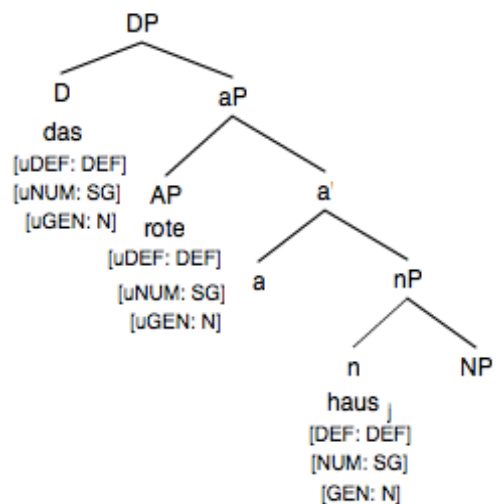
Den tyske førstespråksgruppen hadde færrest avvik av de tre førstespråksgruppene i denne undersøkelsen – kun 2, som tilsvarer en prosent på 2,9. Dette er ikke et statistisk signifikant tall. Jeg mener likevel det kan være interessant å se nærmere på de to avvikene. De er begge avvik på numeruskongruensen på determinativet: ”den gamle dagene” og ”den første frostnettene”. Tysk har som nevnt et nominalsystem som er nokså likt norsk:

9 (a) Den glade gutten



(b) Das rote haus

Det røde huset



Studerer man på disse DP-ene komparativt, kan man se at den eneste forskjellen (sett bort fra kasus) er at norsk har dobbel bestemthet, mens tysk ikke har det. Numeruskongruensen er imidlertid helt lik. Derfor gir ikke de to avvikene ”den gamle dagene” og ”den første frostnettene” indikasjoner på transfer ut fra en komparativ tolkning av norsk og tysk. Ved nærmere gjennomsyn kan man imidlertid se at de to avvikene er produsert av to informanter med gode engelskkunnskaper, der den ene lå på et mellomnivå, og den andre på et høyere

nivå. Dette kan ha hatt en innvirkning, men fordi resultatet ikke er statistisk signifikant, kan dette også være en ren tilfeldighet.

Ettersom det bare er to avvik i tysk, som ikke tilsvarer en statistisk signifikans, kan det ved første øyekast se ut til at det ikke er noen tverrspråklig innflytelse i dette tilfellet. Jeg vil imidlertid argumentere for at dette ikke er helt riktig, fordi man da setter et likhetstegn mellom tverrspråklig innflytelse og avvik. Ettersom det tyske nominalsystemet er så likt det norske, mener jeg det kan være snakk om en positiv transfer her. Likheten mellom nominalsystemene kan ha ført til at det har vært enklere for de tyske informantene å produsere dobbel bestemthets-konstruksjoner uten avvik. Den tyske førstespråksgruppen viser en intragruppe-homogenitet i avvikene – det er en felles tendens i gruppen at det ikke er noen avvik. Man kan også argumentere for at man ser tverrspråklig performansesamsvar i denne gruppen, i form av positiv transfer. Intergruppe-heterogeniteten er ikke statistisk signifikant mellom den tyske førstespråksgruppen og den norske kontrollgruppen, men det er en intergruppe-heterogenitet mellom tysk og de andre førstespråksgruppene. Om man kan finne evidens for to av de tre punktene i Jarvis' metode, skal dette være tilstrekkelig evidens for at det er snakk om tverrspråklig innflytelse. Derfor mener jeg at det mest sannsynlig snakk om en tverrspråklig innflytelse her, hvis man tar utgangspunkt i at det er tverrspråklig performansesamsvar i form av positiv transfer.

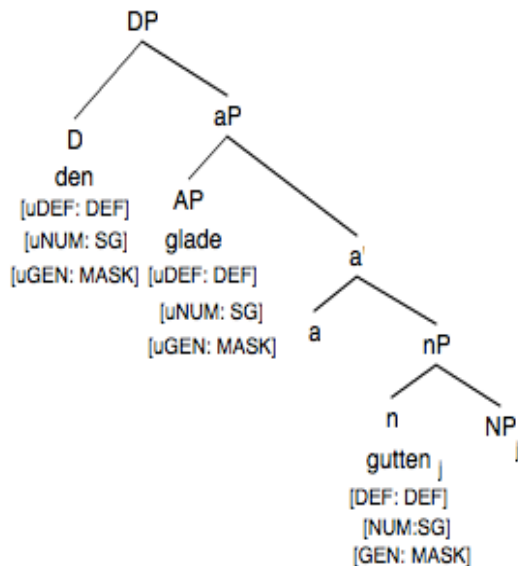
Engelsk førstespråksgruppe

I den engelske førstespråksgruppen fant jeg den største samlede andelen avvik, som lå på 15,7%. 5,9% av dette var avvik på kongruensen til adjektivet, 2% var avvik på genuskongruensen, og 7,8% var avvik på numeruskongruensen. Avvikene på genuskongruensen var altså ikke statistisk signifikante, og derfor kan dette være rent tilfeldig. De andre to resultatene er imidlertid svært interessante: Avvikene på adjektivkongruensen er statistisk signifikante, og det er også avvikene på numeruskongruensen. Det må imidlertid poengteres at det er få avviksføremster med i dette bildet (se avvikstabell ovenfor), og at disse avvikene utgjør forholdsvis store utslag på prosentene. I tillegg produserte én av informantene 2 av de 4 avvikene på numeruskongruensen, så i realiteten bør man antagelig regne med 3 avvik her – og derfor 5,9% i stedet for 7,8%. Prosenten er fremdeles statistisk signifikant for både adjektivkongruensen og numeruskongruensen, men det lave antallet avviksføremster kan altså være en mulig feilkilde.

Hvis vi likevel tar utgangspunkt i de avviksprosentene som er regnet ut, kan man se at avvikene på numeruskongruensen og adjektivkongruensen kan være knyttet til hvordan

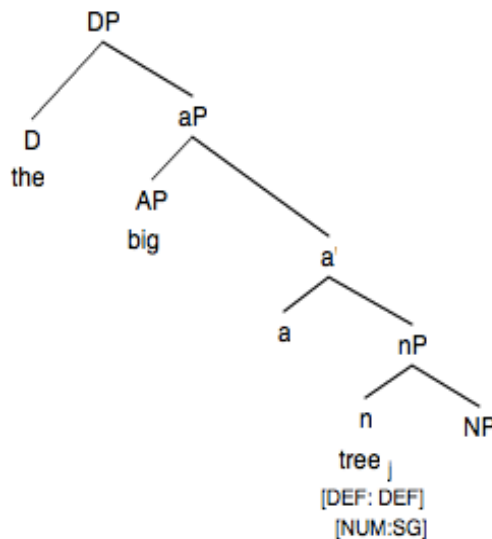
engelske nominalfraser er bygd opp. Engelsk har som nevnt et nominalsistem som skiller seg fra norsk på flere måter, og forklaringene på avvikene kan ligge her:

10 (a) Den glade gutten



(b) The big tree

Det store treet



Som trestrukturene over viser, har ikke engelsk kongruens på adjektivet i det hele tatt, i motsetning til norsk. Dette kan være en mulig forklaring på hvorfor det er en statistisk signifikant avviksprosent akkurat på adjektivets kongruens, som i eksemplene ”den grusom lyden” og ”det mest viktig sake”. I tillegg har ikke engelsk numeruskongruens i fraser som tilsvarer dobbel bestemthet, som framstilt i trestrukturen over. Dette kan på samme måte være en mulig forklaring på avviksprosenten for numeruskongruensen, og dessuten på hva som karakteriserer disse avvikene: Alle avvik på numeruskongruensen består av determinativ i entall, men substantiv som står i flertall, som i ”den siste årene” – ikke motsatt.

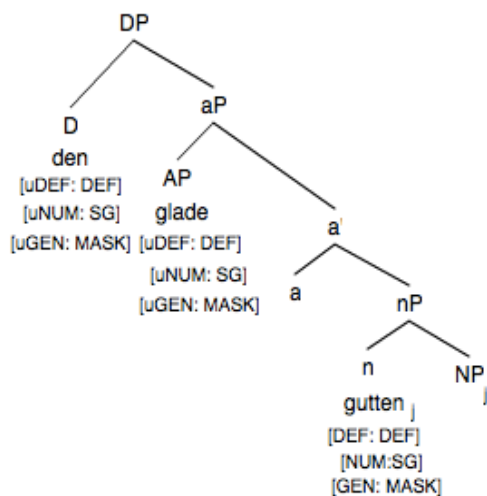
Det ser altså ut til å være tverrspråklig performansesamsvar mellom engelsk og mellomspråket til informantene i den engelske førstespråksgruppen, både når det gjelder numeruskongruensen og adjektivets kongruens. Ettersom det er en felles tendens når det gjelder avvik innad i gruppen, kan man også si at det er en intragruppe-homogenitet her. I tillegg er det en statistisk signifikant intergruppe-heterogenitet mellom den engelske førstespråksgruppen og den norske kontrollgruppen (samt de andre førstespråksgruppene). Dermed er alle de tre punktene for evidens i Jarvis’ metode for påvisning av tverrspråklig innflytelse funnet. Det er selvfølgelig, som nevnt overfor, en viss usikkerhet på grunn av et såpass lavt antall avviksføremønstre som gir store utslag på prosentene. Jeg mener likevel det

er sannsynlig at det er snakk om tverrspråklig innflytelse i den engelske førstespråksgruppen når det gjelder avvik på numeruskongruensen og på adjektivets kongruens.

Vietnamesisk førstespråksgruppe

I den vietnamesiske førstespråksgruppen lå det samlede antallet avvik på 15,2%. Det som var slående i denne gruppen, var at alle avvikene kunne defineres som avvik på adjektivets kongruens. I tillegg var ett av avvikene, ”den ordentlig livet”, også avvikende på genuskongruensen. Dette gav 15,2% avvik på adjektivkongruensen, og 3% avvik på genuskongruensen. Dermed var det kun avvikene på adjektivkongruensen som var statistisk signifikante. En mulig forklaring på disse avviksprosentene kan ligge i forskjellene mellom det norske og det vietnamesiske nominalsystemet:

10 (a) Den glade gutten

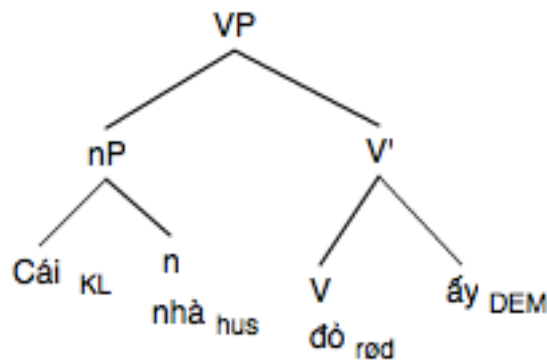


(b) Cái nhà đỏ ấy

KL hus rød DEM

Det huset som er rødt

(Det røde huset)



Hvis vi ser komparativt på norsk og vietnamesisk, vil vi finne mange forskjeller. En av de virkelig fundamentale forskjellene mellom systemene er adjektivene: I vietnamesisk er det som på norsk ville tilsvart et adjektiv, egentlig en type verb (som forklart i 4.1.3). De tre kategoriene numerus, definitthet og genus er heller ikke til stede i vietnamesisk, og det er derfor heller ingen kongruens der. Samlet sett kan disse forskjellene mellom språkene forklare hvorfor det var 15,2% avvik på adjektivets kongruens, av typen ”den interessant sporten”. Man kan altså si at det er tverrspråklig performansesamsvar mellom avvikene i informantenes

mellomspråk og førstespråket vietnamesisk. Fordi dette er en felles tendens i gruppen, kan man også si at det er en intragruppe-homogenitet her. I tillegg er det en statistisk signifikant intergruppe-heterogenitet mellom den vietnamesiske førstespråksgruppen og den norske kontrollgruppen. I den vietnamesiske førstespråksgruppen kan man altså krysse av på alle de tre punktene for evidens i Jarvis' metode for påvisning av tverrspråklig innflytelse. Det er imidlertid få avvik samlet i denne gruppen (se avvikstabellen ovenfor), og prosenten er regnet ut med utgangspunkt i det samlede antallet forekomster av dobbel bestemthet i gruppen – som også var lav. De 5 avvikene (av 33 forekomster) som ble funnet gir altså et veldig stort utslag på prosentene. Derfor må jeg også her, i likhet med i den engelske førstespråksgruppen, anta at dette kan være en mulig feilkilde. Til tross for dette mener jeg likevel at det er sannsynlig at det er snakk om en tverrspråklig innflytelse i den vietnamesiske førstespråksgruppen når det gjelder avvik på adjektivkongruensen.

Det er også en interessant informant jeg gjerne vil trekke fram spesielt i den vietnamesiske førstespråksgruppen. Et av avvikene, ”den størst søsteren”, er ett av tre eksempler på dobbel bestemthet hos denne informanten. De andre to eksemplene er begge varianter av ”den store søsteren”. I det ene tilfellet har informanten tilføyd ”den” etterpå (når dette skjer, markeres det med taggen <add> der tillegget begynner, og </add> der det slutter, som nedenfor), og i det andre tilfellet har frasen blitt skrevet rett fram (<sic> er en feiltagg, og brukes i ”den store søsteren” på grunn av kontekstuellet avvik):

s0625	av morden. Den morden skulle gifte seg med	den <sic> størst </sic> søsteren	snart. H
s0625	om drept søsteren sin. Det var så trist for at	den </add> <sic> store søsteren	ble drept
s0625	de to jentene. Denne mannen snakket med	den </sic> <sic> store søsteren	og faren

Informanten er dermed inkonsekvent i sin måte å skrive frasen på. Dette er ofte typisk hos innlærere i tidlige faser av språkinnlæringen. Denne informanten har imidlertid en oppholdstid i Norge på over fem år (akkurat hvor lenge er det ingen informasjon om), og ligger også på eller over nivå B1 (*Språkprøven* er bestått). Dette viser at selv erfarne språkbrukere kan ha problemer med fraser som har mye formell informasjon i form av kongruens. Vietnamesisk og norsk har grunnleggende forskjellige nominalfraser, og starter man innlæringen som voksen, er det svært sjelden å oppnå en kompetanse som er helt målpråkskonform. Språket fossiliseres gjerne et eller annet sted i prosessen, når man har nådd et nivå som er tilstrekkelig for kommunikative hensyn (som nevnt i kapittel 2.3). Det kan også hende at innlæreren har kompetansen, men at dette ikke nødvendigvis kommer

konsekvent fram i performansen. Informanten kan altså ha kunnskaper om hvordan konstruksjonen egentlig skal være, men likevel ende opp med å produsere noe som ikke er konsekvent i sammenheng med dette (av ulike årsaker, som er diskutert under 2.6).

4.3.3 Oppsummering av resultater

Jeg har argumentert for at alle førstespråksgruppene viste tegn til tverrspråklig innflytelse i denne undersøkelsen. Frekvensundersøkelsen gav samlet sett et inntrykk av at jo mer ulikt nominalsyste­met i førstespråkene er sammenlignet med målspråket norsk, jo lavere frekvens forekommer dobbel bestemthets-konstruksjonen med. Det så ut til at det var en underproduksjon eller en unnvikende produksjon i alle førstespråksgruppene, men at denne ble større jo større forskjellen var mellom det relevante nominalsyste­met og nominalsyste­met i norsk. Jeg konkluderte med at det var vanskelig å si noe sikkert om hvilket av disse fenomenene det var snakk om i noen av tilfellene. Frekvensforskjellene viste en statistisk signifikant intergruppe-heterogenitet og tverrspråklig performansesamsvar i alle førstespråksgruppene. Ut fra Jarvis' metode for tverrspråklig innflytelse, er to av de tre punktene for evidens nok for å påvise transfer. Derfor argumenterte jeg for at det sannsynligvis er snakk om tverrspråklig innflytelse når det gjelder relativ frekvens i alle førstespråksgruppene.

Jeg argumenterte også for at det var tverrspråklig innflytelse i alle førstespråksgruppene ut fra avvikene som ble presentert. I den tyske førstespråksgruppen var det ikke noen statistisk signifikante resultater i avvikene som ble funnet i denne undersøkelsen, men jeg argumenterte likevel for at det kunne være en tverrspråklig innflytelse i form av positiv transfer. I den engelske førstespråksgruppen så det ut til å være tverrspråklig performansesamsvar mellom det engelske nominalsyste­met og avvikene, en felles tendens i gruppen, altså en intragruppe-homogenitet, og en statistisk signifikant intergruppe-heterogenitet. På grunn av få avviksforekomster som gav store prosentutslag, samt én innlærer som produserte to avvik på numeruskongruensen, kan man riktignok derfor også si at det er en viss usikkerhet i disse resultatene. Til tross for dette argumenterte jeg for at det likevel sannsynligvis var snakk om tverrspråklig innflytelse (i sammenheng med avvik på numeruskongruensen og adjektivets kongruens) i den engelske førstespråksgruppen.

I den vietnamesiske førstespråksgruppen fant jeg avvik på adjektivets kongruens, noe som viste tverrspråklig performansesamsvar, fordi vietnamesisk ikke har det som på norsk tilsvarer adjektiv (der brukes en type verb). Her var det også en intragruppe-homogenitet, samt en statistisk signifikant intergruppe-heterogenitet. Det var riktignok få

avviksforekomster også i den vietnamesiske førstespråksgruppen, samt få forekomster av dobbel bestemthet samlet sett. Derfor kunne dette gi høye utslag i prosentene, og potensielt utgjøre en mulig feilkilde. Jeg argumenterte likevel for at det sannsynligvis var snakk om tverrspråklig innflytelse i den vietnamesiske førstespråksgruppen i sammenheng med avvik på adjektivkongruensen.

4.3.4 Avsluttende oppsummering

I denne masteroppgaven har jeg foretatt en komparativ korpusundersøkelse av dobbel bestemthet i norsk mellomspråk. Dobbelt bestemthet er en konstruksjon som både har et fritt foranstilt determinativ og en etterhengt artikkel på nomenet, som begge uttrykker bestemthet – som i ”*det røde huset*”. Problemstillingen for masteroppgaven gikk ut på at jeg skulle gjøre en komparativ undersøkelse av mellomspråket til tre førstespråksgrupper, med henholdsvis tysk, engelsk og vietnamesisk som førstespråk (samt en norsk kontrollgruppe). Målet var i hovedsak å finne ut om det var tegn til tverrspråklig innflytelse med tanke på relativ frekvens og det rent grammatiske. Min utgangshypotese var at alle førstespråksgruppene ville ha visse problemer, men at den vietnamesiske førstespråksgruppen antagelig ville ha størst vanskeligheter, som enten kunne gi utslag i lavere relativ frekvens eller i avvik fra målspråksnormen.

Masteroppgaven ble gitt en teoretisk forankring i eget kapittel. Dette gjorde jeg gjennom å forklare og diskutere viktige begreper fra andrespråksteori og generativ grammatikk som var relevante for undersøkelsen min, som *andrespråk*, *andrespråksinnlærer*, *kritisk periode*, *mellomspråk*, *underproduksjon*, *unnvikende produksjon*, *transfer* og *kompetanse og performanse*. I tillegg framstilte jeg sentrale punkter fra *Prinsipp- og parameterteorien*, som er viktig og relevant for fagfeltene generativ grammatikk og norsk som andrespråk i kombinasjon.

Metoden for denne masteroppgaven var korpussøk, som ble gjort i *norsk andrespråkskorpus* (ASK). Et metodologisk krav var at alle informantene i undersøkelsen skulle ligge på eller over samme språklige nivå, ved å ha tatt *Språkprøven i norsk for voksne innvandrere*. Dette og andre valg ble diskutert i et eget metodekapittel. Her diskuterte jeg korpus som metodologi, kritikken mot korpuslingvistikken, og fordeler og ulemper ved å bruke korpussøk som metode. ASK ble også presentert – både hvordan korpuset er oppbygd, hva slags designkriterer som er lagt til grunn, og hvordan man utfører søk i korpuset. I denne masteroppgaven har jeg brukt *Jarvis’ metode for påvisning av tverrspråklig innflytelse* i analysen av dataene som ble samlet inn.

Den siste hoveddelen av denne masteroppgaven presenterte selve undersøkelsen som ble gjort. Her så vi først en presentasjon av dobbel bestemthets-konstruksjonen i norsk i detalj, og en oversikt over hvordan tilsvarende fraser ser ut i tysk, engelsk og vietnamesisk. Denne oversikten viste at tysk er klart mest likt norsk, men at språket ikke har noen etterhengt artikkel på nomenet som uttrykker bestemthet (som i norsk). Vi så også at engelsk har vesentlige forskjeller fra norske nominalfraser, ved at det ikke har genus, og ingen kongruens på adjektivet eller determinativet i fraser som tilsvarer dobbel bestemthet i norsk. Vietnamesisk er det språket i undersøkelsen som er mest ulikt norsk, blant annet fordi det verken har numerus, genus eller definitthet, og fordi det som tilsvarer adjektiv i norsk er en type verb på vietnamesisk. I undersøkelsesdelen av masteroppgaven forklarte jeg også empirien for undersøkelsen min, med en oversikt over hvordan jeg gikk fram i søkene jeg gjorde i ASK, og hvordan jeg kategoriserte dataene jeg samlet inn. Det ble også kontrollert for engelskkunnskaper som variabel hos de tyske og vietnamesiske innlærerne. Til slutt ble resultatene fra undersøkelsen presentert og analysert.

Etter analyse av dataene med utgangspunkt i *Jarvis' metode for påvisning av tverrspråklig innflytelse*, konkluderte jeg med at det sannsynligvis var tverrspråklig innflytelse i alle førstespråksgruppene, både med tanke på relativ frekvens og ut fra avvikene som ble presentert. Jeg la vekt på statistisk signifikans i tolkningen av dataene. Frekvensundersøkelsen gav samlet sett et inntrykk av at jo mer ulikt nominalsystemet i førstespråkene er sammenlignet med målspråket norsk, jo lavere frekvens forekommer dobbel bestemthets-konstruksjonen med. Når det gjelder avvik, argumenterte jeg for at det var positiv transfer i den tyske førstespråksgruppen, og tverrspråklig innflytelse i den engelske førstespråksgruppen i sammenheng med avvik på numeruskongruensen og adjektivets kongruens. I tillegg argumenterte jeg for at det var en tverrspråklig innflytelse i den vietnamesiske førstespråksgruppen i sammenheng med avvik på adjektivets kongruens.

Det har vært utrolig spennende å få arbeide med denne masteroppgaven. En av årsakene til at jeg valgte å skrive masteroppgaven min innenfor *norsk som andrespråk*, var nettopp at jeg har en stor interesse for feltet. I en verden der språkkontakten øker voldsomt, er det også større behov enn noen gang for kunnskap om andrespråk og prosessene som er involvert i læring av andrespråk. Arbeidet med denne masteroppgaven har vært utrolig interessant, og har gitt meg en dypere kompetanse i *norsk som andrespråk* – noe som vil komme godt med i virket som lektor.

LITTERATUR

- Abrahamsson, N. (2009). *Modermålets inflytande: dess manifestasjoner, begrænsninger og orsaker Andrespråksinl ring*. Lund: Studentlitteratur.
- Berggreen, H. & Tenfjord, K. (2011). *Andrespr ksl ring*. Oslo: Gyldendal Norsk Forlag.
- Boeckx, C. (2010). *Language in Cognition*. Oxford: Wiley-Blackwell.
- Chomsky, N. (1965). *Aspects of the Theory of Syntax*. Cambridge: MIT Press.
- Eide, K. M. (red.) (2015). *Norsk andrespr kssyntaks*. Oslo: Novus AS.
- Eide, K. M. & Busterud, G. (2015). Kompetanse og performanse: kunnskap og produksjon. I Eide (red.) *Norsk andrespr kssyntaks*. Oslo: Novus Forlag.
- Golden, A., Mac Donald, K. & Ryen, E. (2014). *Norsk som fremmedspr k*. Oslo: Universitetsforlaget.
- Haider, H., Olsen, S. & Vikner, S. (1995). *Studies in Comparative Germanic Syntax*. London/Boston: Kluwer Academic Publishers.
- Hawkins, R. (2001). *Second Language Syntax: A Generative Introduction*. Malden: Blackwell Publishing.
- Husby, O. (1991). *Vietnamesisk grammatikk : en innf ring i vietnamesisk grammatikk sett med norske  yne*. Oslo: Friundervisningens forlag.
- Husby, O. (2015). Andrespr kstilegnelse: anvendt og teoretisk perspektiv. I Eide (red.) *Norsk andrespr kssyntaks*. Oslo: Novus Forlag.
- Jarvis, S. (2000). Methodological Rigor in the Study of Transfer: Identifying L1 Influence in the Interlanguage Lexicon. *Language Learning*, 50(2).
- Jarvis, S. & Pavlenko, A. (2010). *Crosslinguistic Influence in Language and Cognition*. London & New York: Routledge.
- Jin, F. & Eide, K. M. (2015). Tilegnelse av nominale kategorier. I Eide (red.) *Norsk andrespr kssyntaks*. Oslo: Novus Forlag.
- Jin, F., Eide, K. M. & Busterud, G. (2015). Nyere andrespr ksteorier I Eide (red.) *Norsk andrespr kssyntaks*. Oslo: Novus forlag.
- Johannessen, J. B. (2003). *P  spr kjakt*. Oslo: Unipub forlag.
- Johansen, H. (2010). Kontroll, Analyse, Reliabilitet, Innl rerkorpus og Transfer - Slik Egner seg Korpus til Studier av Transfer hos Innl rere. I H. Johansen, A. Golden & J. E. Hagen (red.), *Systematisk, variert, men ikke tilfeldig: Antologi om norsk som andrespr k i anledning Kari Tenfjords 60- rsdag*: Novus Forlag.

- Julien, M. (2005). *Nominal Phrases from a Scandinavian Perspective*. Amsterdam/Philadelphia: John Benjamins Publishing Company.
- Kunnskapsbasert praksis. (2012). *Statistisk signifikant*. Lokalisert 6. februar 2016, på kunnskapsbasertpraksis.no/ordliste/statistisk-signifikant/
- McEnery, T. & Wilson, A. (2001). *Corpus Linguistics: An Introduction*. Edinburgh: Edinburgh University Press.
- McEnery, T., Xiao, R. & Tono, Y. (2006). *Corpus-Based Language Studies: An advanced resource book*. New York: Routledge.
- Ragnhildstveit, S. (2009). *Genustildeling og morsmålstransfer i norsk mellomspråk: En korpusbasert studie*. Masteroppgave. Universitetet i Bergen, Bergen.
- Reiten, H. (1995). *Tysk grammatikk*. Otta: Aschehoug.
- Roehrs, D. (2009). *Demonstratives and Definite Articles as Nominal Auxiliaries*. Amsterdam/Philadelphia: John Benjamins Publishing Company.
- Salkind, N. J. (2007). *Statistics for People Who (Think They) Hate Statistics*. Los Angeles: SAGE Publications.
- Schütze, C. T. & Sprouse, J. (2013). Judgment data. I R. J. Podesva & D. Sharma (Red.), *Research Methods in Linguistics*. Cambridge: Cambridge University Press.
- Statistisk sentralbyrå. (2015). *Nøkkeltall for innvandring og innvandrere*. Lokalisert 22.02 2016, på <https://http://www.ssb.no/innvandring-og-innvandrere/nokkeltall/innvandring-og-innvandrere>
- Swain, M. (1982). Immersion Education: Applicability for Nonvernacular Teaching to Vernacular Speakers. I B. Hartford, A. Valdman & C. R. Foster (Red.), *Issues in International Biligual Education: The Role of the Vernacular*. New York: Springer.
- Tenfjord, K., Hagen, J. E. & Johansen, H. (2009). Norsk andrespråskorpus (ASK) - design og metodiske forutsetninger. *NOA norsk som andrespråk*, 25(1). Universitetet i Bergen. (2014). Norsk andrespråskorpus. ask.uib.no
- Wilson, K. M. (1999). *Validity of Global Self-Ratings of ESL Speaking Proficiency Based on an FSI/ILR-Referenced Scale*. Lokalisert 16.03 2016, på <https://www.ets.org/Media/Research/pdf/RR-99-13.pdf>
- Åfarli, T. A. (2015). UG-basert andrespråksteori og syntaktisk struktur. I Eide (red.) *Norsk andrespråkssyntaks*. Oslo: Novus Forlag.
- Åfarli, T. A. & Eide, K. M. (2008). *Norsk generativ syntaks*. Oslo: Novus forlag.

NORSK KONTROLLGRUPPE - SØK MED DOBLE ADJEKTIV

KWIC Concordance.
 # Format: tab-separated, UTF-8. Corpus engine: Corpuscle, Uni Computing, Bergen

Corpus: ASK

Query: [lemma = "det|den|de" %c] [pos = "adj"] [pos = "adj"] [pos = "subst" & morph = ("be" "f")] [pos = "subst" & morph = ("be" "ent")] \<> \ :: language = "norsk" & testlevel = "Språkprøven"

Matches: 8

Excluded: 0

Execution time: 4.921818 sec. (4.967971 CPU sec.)

left-context

match

right-context

years

english

gender

age

language

country

testlevel

mest mulig å rutte med til livets opphold. Det gjelder å få
 m ens egne interesser ikke er helt med. I iver etter å oppnå
 heisen Vi kan gjøre noe selv for å holde oss i form. Det er
 e være så materialistiske i velferdstaten Norge. Sett pris på
 etter man helsten sin for? For å få til en virkelig endring av
 Penger og andre verdier. I vårt samfunn er penger
 isk velignet og vedlikeholdt, samt skulle med friskt mot og
 m for utfordring til regime. Samtidig så Sadding muligheter i

de best betalte jobbene
 de helt store summene
 de små daglige innsatsene
 de små ubetalte tingene
 den stadig økende overvekten
 den store målbare verdien
 de rette moralske verdiene
 den stadig økende eurokursen

selv om ens egne interesser ikke er helt med. I iver etter
 er det mange som er med i spill om penger. Innsatsen kan of
 som teller, en liten spaseretur kveldstur, aketur med ungene
 - Venninnen min måtte gå. Det begynte å bli seint, og vi sku
 , må man endre folks holdninger, og dermed skape varige resu
 - Har du mye eller lite penger får det store konsekvenser fo
 - I våre dager kan man sitte igjen med inntrykket av at mang
 og fallende dollarkursen. Han bestemte seg for å begynne å

66 norsk Norge Språkprøven
 66 norsk Norge Språkprøven
 55 norsk Norge Språkprøven
 20 norsk Norge Språkprøven
 21 norsk Norge Språkprøven
 38 norsk Norge Språkprøven
 26 norsk Norge Språkprøven
 26 norsk Norge Språkprøven

NORSK KONTROLLGRUPPE: DISTRIBUTJON

Tekst-ID	Antall forekomster	Forekomster i adjektivsøket	Av
s2003	1		100
s2005	2		100
s2008	3		100
s2010	2		100
s2014	1	2	100
s2015	2		100
s2017	2		100
s2019	2		100
s2020	1		100
s2023	1		100
s2024	1		100
s2026	3		100
s2027	2		100
s2028	1		100
s2032	2		100
s2033	1		100
s2034	4		100
s2036	2		100
s2038	1		100
s2039	1		100
s2040	1		100
s2042	1		100
s2044	1		100
s2045	1	1	100
s2046	1		100
s2047	1		100
s2048	2		100
s2049	2		100
s2050	1		100
s2053	6		100
s2054	1		100
s2055	2		100
s2056	2		100
s2058	1		100
s2060	1		100
s2061	1		100
s2062	2		100
s2066		1	
s2067	1		100
s2068	2		100
s2070	4		100
s2072	1		100
s2073	4		100
s2074	1		100
s2076	4		100
s2077	2		100
s2078	1		100
s2081	1		100
s2082	2		100
s2083	1		100
s2084	1		100
s2085	3		100
s2086	1		100
s2087	3		100
s2088	1		100
s2089	1		100
s2090	5	1	100
s2092	2		100
s2093	3		100
s2096	1		100
s2097		1	
s2099		1	
s2100	3	1	100
(63 tekst-IDer)			
(Unweighted r	110	8	100
(Total)		118	100


```

TYSK FØRSTESPRÅKSGRUPPE - SBK MED DOBLE ADJEKTIV
# KWIC Concordance.
# Format: tab-separated, UTF-8. Corpus engine: Corpuscle, Uni Computing, Bergen
#
# Corpus: ASK
# Query: [lemma = det] (den | de "%") [pos = "adj"] [pos = "ad"] ([pos = "subst" & morph = {"be" "fl"}] | [pos = "subst" & morph = {"be" "ent"}]) <|> \<del> :: language = tysk & testlevel = "Språksprøven"
# Matches: 1
# Excluded: 0
# Execution time: 4.237527 sec. (4.228852 CPU sec.)
#
# left-context      match      right-context
# (1) Fernsynsprogram for barn. Hver gang begynner det med   den lille <|> <|> orange </|> hovedpersonen som kommer inn på bildet. Litt senere kommer den andre hove s0977 0-1 høyere nivå kvinne age language country testlevel
#                                     Tykkland Språksprøven

```

TYSK FØRSTESPRÅKSGRUPPE: DISTRIBUTJON

Tekst_ID	Antall forekomster	Forekomster i adjektivsøket	Av
s0026	1		100
s0027	1		100
s0029	1		100
s0063	2		100
s0121	1		100
s0124	7		100
s0127	1		100
s0430	1		100
s0433	1		100
s0438	1		100
s0444	2		100
s0447	2		100
s0448	1		100
s0449	2		100
s0450	3		100
s0453	1		100
s0456	1		100
s0458	2		100
s0459	1		100
s0460	1		100
s0461	1		100
s0462	1		100
s0464	1		100
s0466	1		100
s0468	1		100
s0469	3		100
s0470	1		100
s0471	1		100
s0911	2		100
s0915	3		100
s0923	1		100
s0928	1		100
s0934	3		100
s0947	1		100
s0966	4		100
s0973	2		100
s0977	3	1	100
s0989	1		100
s0990	1		100
s0991	1		100
s0995	1		100
s1000	1		100
(42 tekst-IDer)			
(Unweighted)	68		1 100
(Total)			69 100

ENGELSK FØRSTESPRÅKSGRUPPE - SØK MED DOBLE ADJEKTIV

KWIC Concordance.
 # Format: tab-separated, UTF-8. Corpus engine: Corpuscle, Uni Computing, Bergen
 # Corpus: ASK
 # Query: [lemma = "det|den|de" %c] [pos = "adj"] [pos = "subst" & morph = {"be" "ent"}] [pos = "subst" & morph = {"be" "ent"}] \<>\<del = language = "engelsk" & testlevel = "Språkprøven"

Matches: 3
 # Excluded: 0
 # Execution time: 4.125424 sec. (3.886811 CPU sec.)

#	left-context	match	right-context	pid	years	english	gender	age	language	country	testlevel
#	GAVE 2 BOMLUGØ Første og fremst, jeg synes at ro og fred er	det </sic> mest <sic> viktig </sic> <sic> sake	for meg og familie min. Vi flyttet til Norge for ca et år s	s0194	01.feb	høyere nivå	mann		43 engelsk	Storbritannia	Språkprøven
#	ing og fysiske egenskaper, jeg vil jeg først ha nærheten til	den fine norske naturen	for ski- og sykkeltur muligheter. Fine hager med gress og b	s0298	01.feb	høyere nivå	kvinne		21 engelsk	Australia	Språkprøven
#	av og benytte tiden alene. " Jeg kommer sent " må være en av	de <sic> mette </sic> <sic> vangelige </sic> grunnene	for å bruke en mobiltelefon. Og det er også en fordel. Dess	s0357	01.feb	høyere nivå	kvinne		37 engelsk	Storbritannia	Språkprøven

ENGELSK FØRSTESPRÅKSGRUPPE: DISTRIBUTJON

Tekst-ID	Alle forekomster	Forekomster i adjektivsøket	Av
s0011	1		100
s0014	1		100
s0019	1		100
s0020	1		100
s0194		1	
s0218	1		100
s0228	1		100
s0231	1		100
s0289	2		100
s0291	2		100
s0293	1		100
s0296	1		100
s0297	1		100
s0298		1	
s0299	2		100
s0301	2		100
s0304	1		100
s0306	2		100
s0307	1		100
s0310	3		100
s0311	1		100
s0316	2		100
s0329	2		100
s0331	1		100
s0332	3		100
s0333	1		100
s0342	2		100
s0344	2		100
s0346	1		100
s0350	1		100
s0353	1		100
s0354	1		100
s0355	1		100
s0357	1	1	100
s0738	2		100
s0751	1		100
(35 tekst-IDer)			
Sum	48	3	100
(Total)		51	100

VEDLEGG 4: SØK – VIETNAMESISK FØRSTESPRÅKSGRUPPE

	match	right-context	pid	years	english	gender	age	language	country	testlevel
VIETNAMESISK FØRSTESPRÅKSGRUPPE - HOVEDSØK										
■ KWIC Concordance.										
■ Format: tab-separated, UTF-8. Corpus engine: Corpusio, Uni Computing, Bergen										
■ Corpus: ASK										
■ Query: [lemma = den den de *%] [pos = "adj"] ([pos = "sub"] [pos = "subit"] & morph = [{"be": "nt"}]) <> \ \ :: language = "vietnamesisk" & testlevel = "Språkopøven"										
■ Matches: 33										
■ Excluded: 0										
■ Execution time: 4.310757 sec. (3.457499 CPU sec.)										
■ Left-context										
nok mat, klar, å gå på skole å ha en moderne bil osv., En av	dem </sio> viktigere behovene	er å ha et koselig, pent sted å bo. Men hva som gjør et ste	s0016	01. feb grunnnivå	01. feb grunnivå	kvinne	19 vietnamesisk	Vietnam	Språkopøven	
mest i verden. I Vietnam er det få kvinner som kan rykke r.	De viktigste årsakene	ti at folk reiser så mye i våre dager, er at man har det b	s0093	04. mar grunnivå	04. mar grunnivå	kvinne	32 vietnamesisk	Vietnam	Språkopøven	
Opp 1:	De fleste menneskene	som jobber i bar kan rykke. Vietnamere tenker at det ik	s0180	02. mar grunnivå	02. mar grunnivå	kvinne	39 vietnamesisk	Vietnam	Språkopøven	
lig når alle familiemedlemene sitter sammen for å både spise	den vietnamesiske maten	som bor i verden, har ofte vaere og tradisjoner. Folk flest	s0181	5 (over) grunnivå	5 (over) grunnivå	kvinne	44 vietnamesisk	Vietnam	Språkopøven	
konkurrer med hverandre for å få et godt sted, hvis jeg får	den vietnamesiske maten	og diskutere om hva forfædre gjorde i sitt liv. Vi mener	s0181	5 (over) grunnivå	5 (over) grunnivå	kvinne	44 vietnamesisk	Vietnam	Språkopøven	
karakter systeme i første klasse. De må lære hardt for å få	de beste karakterene	. har jeg mange penger - jeg blir kjent og har mange patient	s0184	01. feb mellomnivå	01. feb mellomnivå	kvinne	35 vietnamesisk	Vietnam	Språkopøven	
eksempel må vi være høflig og respekt når vi ta kontakt med	de </sio> gamle </sio> </sio> menneskerne	. Vi må høre fra dem, og gjøre hva de vil noen ganger. Tidsp	s0184	01. feb mellomnivå	01. feb mellomnivå	kvinne	35 vietnamesisk	Vietnam	Språkopøven	
tygge livet i framtiden. Jeg er en student fra utlandet så	den nære framtiden	ønsker jeg meg å ta så mye kunnskap som mulig, først av all	s0557	01. feb begynnernivå	01. feb begynnernivå	mann	21 vietnamesisk	Vietnam	Språkopøven	
eren (jeg har en jobb nå). Jeg går på en teologiske skole og	den </sio> fjære </sio> framtiden	vi jeg gjerne en ungdomspartner når slutten er ferdig. Jeg er	s0557	01. feb begynnernivå	01. feb begynnernivå	mann	21 vietnamesisk	Vietnam	Språkopøven	
skole, er det avslappende å se på fotball-kampen. Jeg liker	den </sio> </sio> interessant </sio> sporten	veldig godt, og det er veldig morsomt å spille fotball med	s0583	01. feb mellomnivå	01. feb mellomnivå	kvinne	25 vietnamesisk	Vietnam	Språkopøven	
støren ble drept av mordren. Den mordren skulle gifte seg med	den </sio> størst </sio> søsteren	snart. Helevis fast en mann ut han som drepte de to jentene	s0625	5 (over) grunnivå	5 (over) grunnivå	kvinne	45 vietnamesisk	Vietnam	Språkopøven	
nne ut, hvem som drept søsteren sin. Det var så trist for at	den </sio> </sio> store søsteren	ble drept av mordren. Den mordren skulle gifte seg med den s	s0625	5 (over) grunnivå	5 (over) grunnivå	kvinne	45 vietnamesisk	Vietnam	Språkopøven	
nin ut han som drept de to jentene. Denne mannen snakket med	den </sio> </sio> store søsteren	og faren jentene. Til slutt, tilstod han at han hadde drept	s0635	5 (over) grunnivå	5 (over) grunnivå	kvinne	45 vietnamesisk	Vietnam	Språkopøven	
paren spille kort, drikk vin, rykke osv. hun ble borteomt.	Den gamle damen	ble trist og vart, for seg, fordi hun ga mye penger for dem	s0626	01. feb grunnivå	01. feb grunnivå	mann	22 vietnamesisk	Vietnam	Språkopøven	
stort hus. Om kvelden kom han hjem, men han kunne ikke finne	den gamle damen	og noboen hennes. Det var en gang i Thailand i en by som he	s0626	01. feb grunnivå	01. feb grunnivå	mann	22 vietnamesisk	Vietnam	Språkopøven	
nsdag @dato i Norge Da så jeg og mannen min på TV kl.10, var	det gamle huset	. Der stod det nytt, stort hus der. Men kona hans var ikke f	s0631	01. feb begynnernivå	01. feb begynnernivå	kvinne	31 vietnamesisk	Vietnam	Språkopøven	
mal er den viktigste dagen i Norge. Folket feirer dagen med	den største </sio> gløden	som hete " baat for baat ". Jeg feirte vont i maggen men jeg	s0638	01. feb mellomnivå	01. feb mellomnivå	kvinne	33 vietnamesisk	Vietnam	Språkopøven	
r for 17. mai. De har fest hele landet. Jeg synes 17. mai er	den viktigste dagen	og er veldig stolte av landet ditt.	s0641	03. mar grunnivå	03. mar grunnivå	kvinne	29 vietnamesisk	Vietnam	Språkopøven	
jærighet når de blir triste eller glade så det er henge av	det </sio> mest foreldrene	i Norge. Folket feirer dagen med den største gløden og er v	s0641	02. mar grunnivå	02. mar grunnivå	kvinne	29 vietnamesisk	Vietnam	Språkopøven	
Men fleste av dem trenger for å prate eller snakke sammen om	den lange sykepleierutanningen	. Når de blir myndige, bestemmer de seg selv, men det er ikk	s0648	03. mar	03. mar	kvinne	31 vietnamesisk	Vietnam	Språkopøven	
mange land for å studere. Hans bok var det som inspirerte av	de </sio> </sio> forskjellige </sio> lingene	min fordi arbeidsgiven stiller ganske høye krav til språk o	s0666	02. mar	02. mar	kvinne	23 vietnamesisk	Vietnam	Språkopøven	
fra hjemlandet mitt. Han heter Nguyen Du. Nguyen Du er en av	den </sio> </sio> orientant </sio> livet	. Det er de fattige, de rike og de forskjeller, osv. Han har	s0671	5 (over) grunnivå	5 (over) grunnivå	kvinne	25 vietnamesisk	Vietnam	Språkopøven	
medier: for eksempel en ny sofa eller bord, stoler. Det er	de store forfattere	i 18 århundre i Viet nam. Han peiler å skriver om kvinne. N	s0683	0-1 mellomnivå	0-1 mellomnivå	kvinne	15 vietnamesisk	Vietnam	Språkopøven	
eg skal ha undervisning i første året og jobbe som lærer i	de </sio> små </sio> mobilene	og vi kan gjøre rest. Likevel er det også noe ulemper. Når	s0693	01. feb begynnernivå	01. feb begynnernivå	kvinne	26 vietnamesisk	Vietnam	Språkopøven	
eg på voksen opplæring for å lære meg hvordan jeg kan hjelpe	den viktigste grunnen	. Tiden går så fort og jeg håper at det er ikke lenge til je	s0854	02. mar grunnivå	02. mar grunnivå	kvinne	35 vietnamesisk	Vietnam	Språkopøven	
sport og de som arrangere sportsarrangementer, for eksempel	de </sio> gamle </sio> menneskerne	til det. De er så små og søte. Hver gang ser på dem leker,	s0855	02. mar grunnivå	02. mar grunnivå	kvinne	21 vietnamesisk	Vietnam	Språkopøven	
e? Svaret ligger bak fridagsaktiviteter. Sport er stort sett	det </sio> engelske </sio> fotballaget	. Jeg synes samfunnet i Norge er så bra. De tenker alt for d	s0857	01. feb grunnivå	01. feb grunnivå	kvinne	22 vietnamesisk	Vietnam	Språkopøven	
få ha uavnlige i oppveksten. Problemer er ofte lungesykdomer.	den største typen	. Liverpool, Caribøeg investerer millioner av kroner i lag	s0876	01. feb høyere nivå	01. feb høyere nivå	kvinne	28 vietnamesisk	Australia	Språkopøven	
	det </sio> </sio> medvendt </sio> spørsmåljet	av fridagsaktiviteter. En annen grunn til at sports har fått	s0876	01. feb høyere nivå	01. feb høyere nivå	kvinne	28 vietnamesisk	Australia	Språkopøven	
	De fleste </sio> rykene	har fått blir fått lungekreft, statistikken viser at hvert s	s0885	0-1 grunnivå	0-1 grunnivå	kvinne	19 vietnamesisk	Vietnam	Språkopøven	

```

VIETNAMEISK FØRSTESPRÅKSGRUPPE - SØK MED DOBLE ADJEKTIV
# KWIC Concordance.
# Format: tab-separated, UTF-8. Corpus engine: Corpuscle, Uni Computing, Bergen
#
# Corpus:      ASK
# Query:      [lemma = "djet|den|de" %c] [pos = "adj"] [pos = "adj"] ([pos = "subst" & morph = ("be" "ft")] | [pos = "subst" & morph = ("be" "ent")]) \<> \<del> :: language = "vietnamesisk" & testlevel = "Språkprøven"
# Matches:    1 (ingen relevante)
# Excluded:   0
# Execution time: 4.2441 sec. (3.973383 CPU sec.)
# left-context      match      right-context
# light. Men bak min leilighet har masse trærn. I sommeren er   X skinner gjennom, derfor får jeg kjølig om sol</sic> solen   X
#                   det veldig vanskelig</sic> </sic> solen   X
#                   X skinner gjennom, derfor får jeg kjølig om sol</sic> solen   X
#                   29 vietnamesisk Vietnam
#                   age
#                   gender
#                   english
#                   years
#                   pid
#                   5 (over)
#                   begynerniv:kvinne
#                   language
#                   country
#                   testlevel
#                   Språkprøven

```

VIETNAMESISK FØRSTESPRÅKSGRUPPE: DISTRIBUTJON

Tekst-ID	Antall forekomster	Antall forekomster i adjektivsøket	Av
s0016	1		100
s0093	1		100
s0180	1		100
s0181	2		100
s0184	2		100
s0185	1		100
s0557	2		100
s0583	1		100
s0625	3		100
s0626	2		100
s0631	1		100
s0638	1		100
s0641	2		100
s0648	1		100
s0666	1		100
s0671	1		100
s0683	1		100
s0693	1		100
s0854	1		100
s0855	2		100
s0857	1		100
s0876	2		100
s0885	2		100
(23 tekst-IDer)			
Sum:	33	0	100
(Total)		33	100

VEDLEGG 5: MASTERARBEIDETS RELEVANS

I dette vedlegget vil jeg gjøre rede for masterarbeidets relevans for mitt virke som lektor: Som jeg nevnte i innledningen av denne masteroppgaven, er det større behov enn noen gang for kunnskap om andrespråk og prosessene som er involvert i læring av andrespråk. Dette skyldes større språkkontakt, både på grunn av massemedia og migrering over hele verden. Innvandringen til Norge har også økt kraftig de siste tjue årene. Det bor innvandrere i alle landets kommuner, med bakgrunn i 222 forskjellige land og selvstyrte regioner (Statistisk sentralbyrå, 2015). Antallet innvandrere og norskfødte med innvandrerforeldre har steget fra 364 981 i 1995, til 805 964 i 2015. 15,6% av befolkningen i Norge i dag er altså innvandrere eller norskfødte med innvandrerforeldre – *uten* at nordiske statsborgere regnes med (Statistisk sentralbyrå, 2015). Denne voksende innvandringen skaper utfordringer i den norske skolen, der *norsk som andrespråk* er relativt nytt. Derfor er det et stort behov for lærere med denne formen for kompetanse.

Som kommende lektor vil kunnskapene jeg har tilegnet meg i arbeidet med denne masteroppgaven (og tidligere arbeid med andrespråksteori) være av stor nytteverdi. Jeg vil utvilsomt komme i situasjoner der det vil være en fordel for både meg og eleven at jeg har denne kompetansen. Min interesse og mitt engasjement for dette fagfeltet vil gi meg stor glede i møte med tverrspråklige utfordringer, og forhåpentligvis vil noe av denne gleden kunne smitte over på og inspirere eleven i sitt arbeid.

VEDLEGG 6: SAMMENDRAG

Denne masteroppgaven er skrevet innenfor fagfeltet *norsk som andrespråk*, med bakgrunn i *chomskyansk generativ grammatikk*. Jeg har foretatt en komparativ korpusundersøkelse av dobbel bestemthet i norsk mellomspråk. Dobbelt bestemthet er en konstruksjon som både har et fritt foranstilt determinativ og en etterhengt artikkel på nomenet, som begge uttrykker bestemthet – som i ”*det røde huset*”. Målet var å finne svar på følgende problemstilling:

- 1) *Hva karakteriserer bruken av dobbelt bestemthets-konstruksjoner i norsk mellomspråk hos voksne andrespråksinnlærere med tysk, engelsk og vietnamesisk som førstespråk, med tanke på det rent grammatiske og når det gjelder relativ frekvens?*
- 2) *Varierer disse karakteristikaene ut fra hvilket førstespråk man har, og kan man eventuelt påvise tverrspråklig innflytelse ut fra resultatene?*

De ulike førstespråksgruppene ble valgt på grunnlag av førstespråkets variasjon i nominalfraser. Ingen av dem har imidlertid dobbelt bestemthet. Tysk er klart mest likt norsk i denne undersøkelsen, men språket har ingen etterhengt artikkel på nomenet som uttrykker bestemthet. Engelsk har vesentlige forskjeller fra norsk – det har ingen kongruens i fraser som tilsvarende dobbelt bestemthet. Vietnamesisk er mest ulikt norsk: Det har ikke genus, numerus eller definitthet, og det som tilsvarende adjektiv i norsk er en type verb på vietnamesisk.

Metoden for denne undersøkelsen har vært korpussøk, som har blitt gjort i *norsk andrespråkskorpus* (ASK). Et metodologisk krav var at alle informantene skulle ligge på eller over samme språklige nivå, ved å ha tatt *Språkprøven*. Jeg brukte *Jarvis' metode for påvisning av tverrspråklig innflytelse* i analysen av dataene, for å finne ut om det var tegn til transfer i noen av gruppene. Min utgangshypotese var at alle gruppene ville vise varierende grad av transfer, men at den vietnamesiske gruppen ville ha størst problemer .

Etter analyse av dataene med *Jarvis' metode*, konkluderte jeg med at det sannsynligvis var transfer i alle førstespråksgruppene, både med tanke på relativ frekvens og ut fra avvikene som ble presentert. Frekvensundersøkelsen gav inntrykk av at jo mer ulikt nominalsysteget i førstespråkene er sammenlignet med norsk, jo lavere frekvens forekommer dobbelt bestemthet med. I avvikanalysen argumenterte jeg for at det var en positiv transfer i den tyske førstespråksgruppen, og anslo det som sannsynlig at det var transfer i den engelske førstespråksgruppen i sammenheng med numeruskongruensen og adjektivets kongruens. I tillegg mente jeg det var sannsynlig at det var transfer i den vietnamesiske førstespråksgruppen i sammenheng med avvik på adjektivets kongruens.

VEDLEGG 7: ABSTRACT

This thesis is written within the discipline *second language acquisition*, with a background in *Chomskyan generative grammar*. I have made a comparative corpus study of double definiteness of L2-users of Norwegian. Double definiteness is a phenomenon in Norwegian where both the determinative and a suffixed article on the noun express definiteness: “*det røde huset*” (“*the red house-the*”). My goal was to answer the following issues:

- 1) *What characterizes the use of double definiteness in Norwegian IL performance in adult second language learners with German, English and Vietnamese as first languages, considering purely grammatical issues and relative frequency?*
- 2) *Do these characteristics vary according to the learner’s first language, and is it possible to prove crosslinguistic influence based on the results?*

The different first language-groups were chosen based on variation in nominal phrases. None of the L1’s, however, have double definiteness like Norwegian has. German is most like Norwegian in this study, but there is no suffixed article on the noun to express definiteness. English displays significant differences – it has no agreement in phrases that correspond to double definiteness. Vietnamese doesn’t have gender, number or definiteness, and what corresponds to adjectives in Norwegian is a type of verb in Vietnamese.

I have used corpus searches as a method in this study, and this has been done with *ASK*, which is a corpus with data collected from second language learners of Norwegian (immigrants). A methodological requirement for this study has been that all L2-learners had to be on or above a threshold level by taking the same test for adult immigrants: *Språkprøven*. I used *Jarvis’ method for detecting crosslinguistic influence* in the analysis of the data I collected. My hypothesis was that all groups would show L1-influence, but that the Vietnamese group would have the most problems.

After analysis of the data with *Jarvis’ method*, I concluded that it was probably L1-influence in all groups, both concerning relative frequency and in the deviations presented. The frequencies implied that the more different the nominal phrases in the L1-languages were from Norwegian, the lower the frequency of double definiteness occurs. After analysing the deviations, I argued that it was positive transfer in the German L1-group, and I also deemed it probable that there was L1-influence in the English group on the number- and adjective agreement. In addition to this, I see it as probable that it is L1-influence in the Vietnamese L1-group on the adjective agreement as well.