

Interskandinavisk ordforståelse: En internetbaseret undersøgelsesmetode

Sebastian Kürschner & Charlotte Gooskens

Skandinavisk Afdeling, Groningen Universitet, Holland

1. Baggrund

Ved Skandinavisk Afdeling på Groningen Universitet i Holland udføres i løbet af 2007 en stor internetundersøgelse af forståelsen af nærtbeslægtede sprog i det germanske sprogområde. Formålet er at undersøge, hvilke lingvistiske faktorer, der spiller en rolle for ordforståelsen. Syv germanske sprog (dansk, norsk, svensk, tysk, plattysk, hollandsk og frisisk) bliver testet blandt gymnasieelever med seks forskellige modersmål (dansk, norsk, svensk, hollandsk, frisisk og tysk). Vi vil her beskrive undersøgelsen samt præsentere de planlagte analyser af ordforståelsen mellem dansk, norsk og svensk. Yderligere information og resultater fra hele undersøgelsen findes via <http://www.let.rug.nl/micrela>.

Der er hidtil blevet udført flere forskellige undersøgelser af den interskandinaviske sprogforståelse (Maurud 1976, Bö 1978, Börestam 1987, Delsing & Åkesson 2005). Formålet med disse undersøgelser var først og fremmest at få et generelt indtryk af, hvor godt skandinaver forstår hinanden. Resultaterne viser, at den svensk-danske talesprogsforståelse ofte er problematisk. Især svenskerne har svært ved at forstå dansk. Nordmændene er bedst til at forstå nabosprogene, mens svenskerne og danskerne har sværere ved at forstå norsk. Desuden viser resultaterne, at det er lettere at læse nabosprogene end at forstå dem, når de tales. De fleste af de nævnte undersøgelser er tekstforståelseksperimenter. Sådanne undersøgelser giver et alment indtryk af sprogforståelsen, men det er ikke muligt at undersøge, hvorfor det ene sprog er lettere at forstå end det andet, og hvor problemerne præcist ligger. Vi håber med vores undersøgelse at kunne bidrage med en mere detaljeret analyse af de lingvistiske fænomener, som kan forklare forskellene i den indbyrdes sprogforståelse. Vi håber bl.a. at kunne give svar på følgende forskningsspørgsmål:

- Hvor godt forstår skandinaver hinandens ordforråd?

- Findes der asymmetri i ordforståelsen?
- Hvilke forskelle er der på forståelsen af talte og skrevne ord?
- Hvilken rolle spiller forskellige lingvistiske faktorer (fx afstande på forskellige lingvistiske niveauer, ordlængde og 'lumske venner') mht. ordforståelsen?
- Er der en forskel på forståelsen af arveord og låneord af forskellig oprindelse?
- Er der en forskel på forståelsen af ord brugt i formel og uformel kontekst?

For at kunne besvare disse spørgsmål har vi udviklet en ordforståelsestest, som gymnasieelever kan deltage i via internettet. Formålet med denne artikel er i første omgang at beskrive materialet (paragraf 2) og metoden (paragraf 3) samt at forklare, hvordan vi kan bruge resultaterne til at få nærmere kendskab til, hvilke lingvistiske faktorer der spiller en rolle for ordforståelsen (paragraf 4). Vi er stadig i gang med at udføre eksperimenter på skoler i de deltagende lande. Derfor vil vi her kun præsentere de mulige kvantitative og kvalitative analysemetoder samt hypoteser om resultaterne. Vi planlægger en detaljeret fremstilling og analyse af resultaterne, når indsamlingen af data er afrundet. Vi slutter af med en metodisk diskussion (paragraf 5).

2. Materiale

For at ordforståelseksperimentet ikke skulle blive for langt og antallet af testpersoner for stort, var vi nødt til at begrænse antallet af testord. For at modellere sprogkontakt-situationen så godt som muligt, valgte vi at basere udvalget af testordene på en database med frekvente ord fra talesproget. Vi vil her beskrive denne database og derefter forklare, hvordan vi valgte testordene fra databasen.

Vi antager, at man bruger flere 'internationale' ord og nye låneord i formel tale end i uformel tale. Sådanne ord har andre, formentlig mere transparente, fonetiske og ortografiske ligheder med de tilsvarende ord på modersmålet end uformelle ord, som ofte er arveord eller låneord, der har været en del af sproget i en længere periode og derfor har haft tid til at tilpasse sig sproget. Eftersom man kan gå ud fra, at de fleste mennesker konfronteres med begge slags sprog, og det endelige mål med databasen er at kunne

undersøge sprogforståelsesprocessen, mente vi, at det er vigtigt at databasen indeholder frekvente ord fra både formel og uformel tale.

Vi har ekstraheret ord fra situationer, som er typiske for uformel tale, ved at bruge det hollandske talesprogs-korpus Corpus Gesproken Nederlands (CGN). Det drejer sig om samtaler mellem familie og venner i hjemlige omgivelser. Som repræsentant for ord i en formel kontekst valgte vi et korpus baseret på hollandske og svenske monologer, der er blevet holdt som taler under møder i Europaparlamentet (Europarl-korpusset). Vi valgte de 1500 mest frekvente ord (lemmaer) fra de to korpusser. Disse to frekvensordslister fjede vi derefter sammen. En del ord fandtes i begge korpusser, så efter sammenlægningen var der i alt 2575 forskellige ord i vores database. Alle disse ord oversatte vi derefter til følgende germanske sprog: hollandsk, vestfrisisk, højtysk, plattysk, dansk, svensk og norsk. De 2575 ord på syv germanske sprog danner grundlaget for vores database.

I databasen fjede vi derefter følgende information til hvert ord på hvert sprog:

- ordklasse
- fonetisk transskription
- oprindelse (arveord eller låneord)
- hvis det er et låneord: a) sprog som ordet er lånt direkte fra; b) det oprindelige sprog som ordet kommer fra; c) evt. årstal hvor ordet er kommet ind i sproget (hvis vi havde oplysninger om det)
- stil (CGN=uformel, Europarl=formel, CGN+Europarl=neutral)
- kognat (historisk beslægtet med de tilsvarende ord på de andre sprog) eller ikke-kognat (ikke beslægtet)

Databasen kan bruges til forskellige formål, og det er muligt at udvide den med flere sprog. I undersøgelsen, som beskrives her, brugte vi den til at udvælge testord til internetundersøgelsen, som beskrives i paragraf 3. Det er vigtigt, at testen kan udføres i løbet af en skoletime. Det er derfor ikke realistisk at teste alle 2575 ord i vores database. Vi besluttede kun at teste substantiver, eftersom de ofte er mindre abstrakte og mere centrale for

sprogforståelsen end andre ordklasser. I alt var der 815 substantiver. Det er muligt, at gymnasieelever ikke kender betydningen af alle disse ord. Derfor gennemførte vi en test for at finde ud af, hvilke ord der var forståelige for denne aldersgruppe. Vi forelagde alle de hollandske substantiver for 24 hollandske gymnasieelever (15 til 17 år) og bad dem sætte kryds ved ord, som de ikke forstod. I alt blev der sat kryds ved 82 ord af en eller flere elever. Disse ord blev fjernet fra listen med substantiver. Derefter fjernede vi yderligere 349 ord efter tilfældighedens princip.

De resterende 384 ord lod vi indtale på de syv testsprog af voksne, mandlige talere, som har de pågældende standardsprog som modersmål. Da der ikke findes et officielt talt norsk standardsprog, valgte vi at bruge en taler fra Oslo-området. Tilsvarende bad vi en medarbejder fra 'Institut für niederdeutsche Sprache', som også er nyhedslæser ved Radio Bremen, om at indtale ordene på plattysk. Ordene blev optaget i et optagelsesstudio, hvorefter de blev klippet og bevaret digitalt ved hjælp af programmet CoolEdit. De 384 indtalte ord og deres nedskrevne form (efter det pågældende sprogs normerede retskrivning) udgør vores testmateriale. Norsk repræsenteres af bokmål. Til de nedertyske oversættelser brugte vi de retskrivningsregler, som er blevet udviklet af Johannes Sass (1957).

3. Eksperiment

Vi har indrettet en speciel internetportal til undersøgelserne, det såkaldte Language Research System. Portalen er offentlig tilgængelig på <http://www.let.rug.nl/lrs>. Den er egnet til forskellige slags internetbaserede eksperimenter og kan udvides til andre projekter. Prøvedeltagelse i det nuværende eksperiment er mulig ved at logge sig ind med brugernavnet 'germanic' og adgangskoden 'guest'.

3.1 Testpersoner

Vi tester elever i alderen 16-19 år på gymnasier og andre uddannelser, som giver direkte adgang til universitetet. Når vi har valgt at teste netop denne gruppe, skyldes det, at det er forholdsvis let at finde en homogen gruppe, hvad angår alders- og uddannelsesniveau (på den måde kan sociale faktorer holdes nogenlunde konstante), og desuden er det en fordel, at

de fleste gymnasier råder over faciliteter, som gør det muligt at teste en hel gruppe på en gang. I alt skal mere end tusind elever testes i det germanske sprogområde hvoraf to tredjedele i Skandinavien. For at få så mange elever til at deltage, søgte vi kontakt med skoleledere og lærere på gymnasier og bad dem om at bruge en skoletime på eksperimentet. Udover gymnasier i Danmark, Norge og Sverige er der også planlagt tests på skoler i Holland (inkl. Frisland) og Tyskland. Følgende rapport handler udelukkende om den skandinaviske del af undersøgelsen.

Vi tester hele klasser, hvilket betyder, at nogle af testpersonerne ikke har den ønskede baggrund. Disse testpersoners resultater bliver udelukket fra den videre analyse af resultaterne. Udelukkelsen baseres på oplysninger, som testpersonerne udfylder i et spørgeskema inden testens begyndelse (se paragraf 3.2). De analyserede testpersoners modersmål er det officielle sprog, som tales i landet, de bor i. De er alle monolingvale modersmålstalere, dvs. at de taler det samme sprog med begge forældre. På den måde udelukker vi testpersoner, som evt. ikke behersker landets officielle sprog i tilstrækkelig grad. Desuden inddeler vi deltagerne mht. a) geografi og b) fremmedsproglig baggrund.

ad a) Vi skelner mellem deltagere fra grænseområderne til et af de testede sprog og deltagere fra andre områder. For Danmarks vedkommende betyder det fx, at deltagere fra hovedstadsområdet danner en speciel gruppe i dataanalysen, fordi de bor i nærheden af den svenske grænse. På samme måde kategoriseres svenske deltagere fra Skåne samt hele grænseområdet mellem Sverige og Norge som specialgrupper etc. Målet med denne inddeling er at kunne teste, i hvor høj grad regional nærhed spiller en rolle for evnen til at forstå nabosprogene.

ad b) Informationerne om de indlærte fremmedsprog bruges til at kunne inddele eleverne i grupper med baggrundsviden om fremmedord. En nærliggende hypotese er fx, at elever, som har lært latin, har en bedre forståelse af latinske fremmedord end andre elever.

3.2 Gennemførelse af testen

For at deltage i testen skal hver testperson først logge ind på systemet. Deltageren får automatisk tildelt et systeminternt ID-nummer, som gør det muligt at koble testpersonerne

til responserne i den senere analyse. Efter at have logget ind bliver deltagerne bedt om at angive, hvad deres modersmål er. Deltagerne får nu en kort præsentation af projektets mål. For at udfordre deltagere til at yde deres bedste får de at vide, at de kan sammenligne deres eget resultat med de andre deltageres resultater efter selve testen. Desuden udloves priser til de bedste deltagere i eksperimentet. Teknikken tjekkes med en lydtest, hvorefter den individuelle del af testen begynder.

På næste skærmbillede skal deltagerne udfylde et skema med personlige informationer: køn, alder, boligområde, dialektkundskaber, hvilket sprog de taler med deres forældre, deres aktuelle uddannelsessted og -niveau samt deres kendskab til forskellige fremmedsprog (hvor mange år de har haft engelsk, tysk, fransk, spansk og latin i skolen). Disse informationer bruges til at udelukke deltagere, som ikke hører til målgruppen (jf. afsnit 3.1), fra analysen og til at inddele de godkendte deltagere i forskellige grupper.

Når testpersonerne har udfyldt skemaet med personlige oplysninger, starter selve eksperimentet. Hver person bliver kun testet i et af testsprogene. Dette sprog tildeles automatisk af systemet, når testpersonen logger ind. Det er for omfattende at teste alle 384 ord både skriftligt og mundtligt. Derfor har vi delt ordene i fire grupper à 96 ord efter tilfældighedens princip. Hver testperson lytter til 96 talte ord og læser 96 skrevne ord fra to forskellige ordgrupper. Halvdelen af testpersonerne starter med lyttedelen og den anden halvdel starter med læsedelen. På den måde vil vi udelukke, at præstationsniveauet i en af de to dele bliver påvirket af træthed pga. testdelenes rækkefølge.

I læsedelen præsenteres ordene via computerskærmen og i lyttedelen via hovedtelefoner. Deltagernes opgave består i at prøve at oversætte ordene til deres eget modersmål. Et nyt skærmbillede åbnes for hvert ord, og deltagerne bliver bedt om at skrive ordet på deres modersmål i det dertil beregnede felt. De får maksimalt 10 sekunder til at skrive oversættelsen. Hvis de er hurtigere end det, kan de gå videre til næste ord med det samme.

Efter at have gennemført eksperimentet får deltagerne at vide, hvor stor en procentdel af ordene de har oversat korrekt. De kan vælge at sammenligne deres resultat med gennemsnittet af de andre deltageres resultater. Desuden kan de komme med kommentarer eller spørgsmål til testen.

3.3 Databehandling

Alle resultater inkl. baggrundsinformationerne bliver ført direkte ind i en relationel database, som ligger til grund for vores analysearbejde. I denne database findes desuden den lingvistiske information om testordene (se paragraf 2). Systemet sørger for, at de forskellige varianter af testene (se paragraf 3.2) fordeles ligeligt blandt testpersonerne. Når en hel klasse testes på en gang, bliver eleverne derfor testet i forskellige sprog og i forskellige udvalg og rækkefølger af de 384 ord. Fordelen er, at eleverne ikke kan hjælpe hinanden med at finde de rigtige svar, og at træthedseffekterne er ens for alle data.

Systemet matcher elevernes svar med de korrekte svar i vores database og afgør om svaret er korrekt eller ej. Dette betyder, at stavefejl også regnes som forkerte svar (se paragraf 5). I nogle tilfælde er der flere korrekte svarmuligheder. Fx kan det svenske ord *rum* oversættes som både *rum* og *værelse* på dansk. Det er muligt at tilføje disse alternative svarmuligheder, således at begge svar regnes som korrekte. Alternativerne kan tilføjes til enhver tid, hvorefter alle tidligere resultater genberegnes.

Det er muligt at udvælge en del af testresultaterne eller testordene til videre analyse ved hjælp af forskellige udvælgelseskriterier. Man kan fx vælge at sammenligne, hvor godt københavnere og andre danskere forstår svensk eller finde ud af, om kvinder har klaret testen bedre end mænd. Man kan også se på den gennemsnitlige forståelse af hvert enkelt ord eller sammenligne testpersonernes forståelse af tyske og engelske låneord osv.

4. Analyse af resultaterne

Ved hjælp af dataene håber vi bl.a. at kunne give svar på forskningsspørgsmålene formuleret i paragraf 1. Da dataindsamlingen stadigvæk er i gang, og vi i denne artikel har valgt at fokusere på metodiske spørgsmål, vil vi i det følgende kun præsentere, hvilke resultater der kan forventes, og hvordan vi kan analysere dem ved hjælp af databasen. Resultaterne kan vi derefter sammenligne med dem fra hidtidige undersøgelser af inter-skandinavisk forståelse, hvoraf Delsing/Lundin-Åkesson (2005) er det nyeste og omfangsrigeste studie. I de hidtidige studier blev den generelle tekstforståelse testet, mens vi i vores undersøgelse tester forståelsen af isolerede ord. Ud over en mere præcis analyse

af de lingvistiske fænomener, som spiller en rolle i sprogforståelsen, kan vi ved sammenligningen med tidligere undersøgelser derfor også se, hvordan ordforståelsen afviger fra tekstforståelsen. Vi begynder med at præsentere diverse kvantitative analysemuligheder og diskuterer derefter, hvordan resultaterne kan analyseres kvalitativt.

4.1 Kvantitative analysemuligheder

I første omgang er vi interesseret i at se på, hvor godt forskellige grupper af testpersoner forstår nabosprogene. Derefter håber vi at kunne forklare forskellene mellem forskellige gruppers præstationer ved hjælp af et antal faktorer.

Et af de spørgsmål, som vi ønsker at få besvaret, er, om ordforståelsen mellem de forskellige sprog er asymmetrisk ligesom i de tidligere tekstforståelseseksperimenter. Specielt for dansk-svensk og svensk-norsk viser Delsing/Lundin-Åkesson (2005: 77), at danskerne forstår svensk noget bedre, end svenskerne forstår dansk, og at svenskerne er meget dårligere til at forstå norsk, end nordmændene er til at forstå svensk. En af forklaringerne på den dansk-svenske asymmetri er, at dansk talesprog på grund af den lydige struktur er meget svært at segmentere, dvs. at det især for en ikke-modermålstaler er svært at høre, hvor et ord ender, og hvor et nyt begynder. Dette problem er formodentlig i meget mindre grad til stede, når ordene præsenteres isoleret. Det er derfor muligt, at vi ikke finder asymmetriske dansk-svenske resultater.

Vi vil også undersøge forskellene mellem forståelsen af talte og skrevne ord. Vores hypotese er, at skrevne ord er lettere at forstå end talte ord, fordi talesprogene har udviklet sig mere end skriftsprogene, ligesom der er mindre sproglig variation i de skrevne standardsprog end i de talte standardsprog. Specielt talt dansk har inden for det seneste århundrede gennemgået en massiv udvikling, som har resulteret i en stor lingvistisk afstand til svensk og norsk talesprog. Skriftsproget er derimod mere konservativt og derfor mindre afvigende fra de andre sprogs skriftsprog. I deres undersøgelse af tekstforståelsen er Delsing/Lundin Åkesson (2005: 76-79) kommet frem til, at læseforståelsen er bedre end høreforståelsen, men at dette ikke gælder i lige høj grad for de forskellige sprogekombinationer. Mens nordmænd og svenskere er næsten lige gode til at forstå hinandens talte og skrevne sprog,

forstår de skrevet dansk meget bedre end talt dansk. I vores eksperiment vil vi undersøge, om vi får lignende resultater, når vi tester forståelsen af isolerede ord.

Vi håber specielt at kunne identificere lingvistiske faktorer, som spiller en rolle for ordforståelsen ved at måle afstande på forskellige lingvistiske niveauer, fx det fonetiske og det leksikale niveau. Til at måle fonetiske og ortografiske afstande er det fx muligt at bruge den såkaldte Levenshtein-algoritme. Denne algoritme udtrykker andelen af ændringer, som er nødvendige for at ændre ordet på et sprog til det tilsvarende ord på et andet sprog. Ved at korrelere de fonetiske og ortografiske afstande med resultaterne af sprogforståelsestestene er det muligt at få nærmere indblik i disse afstandes rolle i ordforståelsen. I Heeringa (2004) beskrives metoden i detaljer. Et af målene med vores forskning er at tilpasse Levenshtein-algoritmen så den kan modellere sprogforståelse så optimalt som muligt. Fx er det sandsynligt, at lyde i begyndelsen af ord er vigtigere for forståelsen end lyde i andre dele af ordet. Vi kan også undersøge, om lange ord er lettere at forstå end korte ord, fordi der er mere fonetisk/ortografisk substans i lange ord. Hvis vi finder evidens for disse hypoteser i forståelsestesten, bør det indbygges i algoritmen.

Hvad angår det leksikale niveau, vil vi i første omgang sammenligne forståelsen af kognater og ikke-kognater. I princippet er ikke-kognater uforståelige, med mindre man har kendskab til sproget eller til et andet sprog, som har de tilsvarende ikke-kognater. En faktor, som er speciel for ordniveauet og derfor ikke kunne isoleres i de tidligere undersøgelser af tekstforståelsen, er betydningen af ordenes historiske oprindelse for sprogforståelsen. Da alle ord i databasen er annoteret med informationer om deres etymologi, kan vi inddrage ordenes oprindelse i analysen. Det kan forventes, at arveord har været udsat for længere og mere omfattende fonologiske udviklinger i enkeltsprogene end (nyere) låneord. Vi får muligheden for at sammenligne, hvor godt arveord og fremmedord forstås, og vi vil desuden være i stand til at sammenligne forskellige grupper af fremmedord. Det kan fx forventes, at latinske og engelske internationalismer er forholdsvis lette at forstå. Det gælder sandsynligvis ikke kun i sammenligning med arveordene, men også med de nedertyske låneord, som i højere grad er blevet tilpasset de skandinaviske sprogs fonologisk-grammatiske systemer og har deltaget i de fleste

fonologiske udviklinger siden låneperioden i middelalderen. I denne forbindelse bliver det også spændende at se, i hvor høj grad kategoriseringen i formelle vs. uformelle ord spiller en rolle for forståelsen. Det kan forventes, at formelt sprog er lettere at forstå end uformelt sprog, fordi der bruges mange internationale låneord.

Et særligt problem for ordforståelsen udgøres af de såkaldte 'falske venner', fx dansk *rolig* vs. svensk *rolig* 'sjov'. Disse ord kan være direkte hindrende for sprogforståelsen, og derfor er det vigtigt at inddrage disse forhold i analysen. Det samme gælder falske venner på andre sprog, fx dansk *bed* – engelsk *bed* 'seng'. Et tilsvarende problem udgør forståelsen af ord, som har lignede kognater indenfor det samme betydningsfelt, fx dansk *pige* – svensk *pige* 'tjenestepige'. Her er det ikke på forhånd tydeligt, om kognaterne er til hjælp for sprogforståelsen. Det er muligt, at en svensker vil oversætte det danske ord *pige* forkert til *pige*, men på den anden side kan det også være til en vis hjælp, at ordet *pige* har en semantisk overlapning med det svenske ord *pige*. Ordforståelsestesterne kan gøre det muligt at drage mere sikre konklusioner om dette.

4.2 Kvalitative analysemuligheder

Vores data giver anledning til en del kvalitative spørgsmål. Disse spørgsmål drejer sig først og fremmest om den morfologiske ordstruktur. Fx opstår en del 'fejl' i testpersonernes oversættelse af testordene ved at derivationsaffikser, som både i modersmålet og i testsproget er præget af høj produktivitet, bruges i tilfælde, hvor de pga. leksikalisering egentlig ikke er korrekte, dvs. at de bliver blokkeret. Det svenske derivat *omgivning* svarer til bokmål *omgivelse*. Det er altså dannet med de to forskellige suffikser - *ning* og -*else* i de respektive sprog. En nordmand, som blev konfronteret med det svenske ord, dannede et tilsvarende ord på norsk med -*ning*. Eftersom suffikset -*ning* har høj produktivitet i norsk, kan dette også forstås som en fejlagtig produktiv brug af -*ning*. Lignende 'nydannelser' kan også iagttages ved andre sprogpar, jf. en dansk deltagers oversættelse af svensk *överenskommelse* til *overenskommelse* og af *undersökning* til *undersøgning* (de korrekte danske ord er *overenskomst* og *undersøgelse*). Dette fænomen giver anledning til at formode, at morfologisk blokkering af bestemte affikser pga. leksikaliserede former ikke nødvendigvis

virker under tidspres, når en anden orddannelsesmodel står til rådighed. Således kan vores data også bruges som materiale til kvalitative morfologiske undersøgelser.

5. Metodisk diskussion

Vi har valgt et eksperimentelt design for at undersøge ordforståelse. I det følgende vil vi gå ind på to diskutabile punkter i forbindelse med design og databehandling og diskutere, hvorfor vi har valgt at udføre vores analyser som beskrevet ovenfor. Det drejer sig om a) isoleret ordpræsentation og b) forskellige slags fejl.

ad a) Ordene præsenteres isoleret i vores eksperiment. Det kan kritiseres, at vi ikke har undersøgt, om forståelsesgraden af ord i isolation ligner forståelsen af ord, som bliver ytret i kontekst. Det er muligt, at forståelsen af ord i et fremmedsprog øges, jo mere kontekst man får, fordi ukorrekte tolkninger kan sorteres fra ved hjælp af information herfra. Alligevel mener vi, at vores metode har den fordel, at rent strukturelle faktorer, som lydstruktur og ordenes morfologiske struktur, isoleres i forståelsestesten. Det kan betragtes som en ulempe, at testen derved ikke simulerer naturlige sproglige situationer særlig godt, men til gengæld er det lettere at isolere lingvistiske faktorer end i en test med sætnings- eller samtalekontekster. Vi går ud fra, at ulempen ved ikke at have kontekst er lige stor for alle grupper. Ved hjælp af vores resultater er det også muligt at sammenligne isoleret ordforståelse med tekstforståelse, som den er blevet testet i forskellige tidligere undersøgelser. På den måde forventer vi desuden at få større indsigt i to forskellige slags sprogperception, nemlig 'global listening' vs. 'analytical listening'. I første tilfælde er forståelsen baseret på tekst- og kontekstinformationer, mens betydningstolkningen i det andet tilfælde kun baseres på lyd- og morfemstrukturen.

ad b) Oversættelsesfejl kan være af forskellig art. Mens det i de fleste tilfælde drejer sig om forkerte oversættelser eller om tilfælde, hvor deltageren ikke skrev noget som helst pga. manglende forståelse, bliver også stavefejl ved korrekte oversættelser eller orddannelsesfejl af den slags beskrevet i afsnit 4.2 kategoriseret som forkerte oversættelser. Selvom vi er klar over, at der findes en kvalitativ forskel mellem fejltyper, mener vi ikke, at det er metodisk korrekt at skelne mellem disse slags fejl. Vi er først og fremmest

interesseret i kvantitative data, og vi regner med en vis andel af retskrivnings- og orddannelsesfejl, som er ens fordelt over vores testgrupper, så vi alligevel får fuldstændigt fortolkelige og pålidelige tal. Korrektioner fra vores sider ville derimod føre til, at vi ville tolke testpersonernes svar, og dette ville indebære en subjektiv vurdering af fejlene, da det ikke altid er tydeligt, om der er tale om en stavefejl eller orddannelsesfejl eller ej. Det ville være problematisk, hvis det gennemsnitlige antal stavefejl pr. ord pr. testgruppe var forskelligt, fordi testsprogenes ortografi ikke var lige svære for eleverne, eller fordi eleverne ikke har fået lige meget stavetræning i de tre lande. Desværre har vi ikke viden om disse forhold, og vi kender ikke til sammenlignende studier inden for dette område.

Sammenfattende forventer vi at få svar på en hel række spørgsmål om den interskandinaviske forståelse vha. vores eksperiment. Sammenlignet med tidligere forståelsesundersøgelser er vi med vores testmetoder i højere grad i stand til at inddrage lingvistiske faktorer i forklaringen af sprogforståelsen. Derudover er ordniveauet aldrig tidligere blevet testet så præcist, at en resultatsammenligning kan tjene til at undersøge forskellen mellem global kontekstorienteret tekstforståelse og lokal lydkædebaseret ordforståelse.

Litteratur

- Bø, Inge (1978). *Ungdom og naboland. En undersøgelse av skolens og fjernsynets betydning for nabospråksforståelsen*. Stavanger: Rogalandsforskningen.
- Börestam, Ulla (1987). *Dansk-svensk språkgemenskap på undantag. Nordisk språkförståelse i nutidshistoriskt och regionalt perspektiv belyst av svenska gymnasieungdomars förståelse av äldre och nutida talad danska*. FUMS rapport, 137. Uppsala: Uppsala Universitet, Institutionen för nordiska språk.
- Delsing, Lars-Olof & Katarina Lundin-Åkesson (2005). *Håller språket ihop Norden? En forskningsrapport om ungdomars förståelse av danska, svenska och norska*. TemaNord 2005: 573. Köpenhamn: Nordiska ministerrådet.
- Heeringa, Wilbert (2004). *Measuring Dialect Pronunciation Differences using Levenshtein Distance*. Grodil 46. Groningen: Dissertation. <http://irs.ub.rug.nl/ppn/258438452>
- Maurud, Øivind (1976). Reciprocal comprehension of neighbour languages in Scandinavia. An investigation of how well people in Denmark, Norway and Sweden understand each other's written and spoken languages. *Scandinavian journal of educational research* 20: 49-72.
- Sass, Johannes (1957). *Kleines plattdeutsches Wörterbuch, nebst Regeln für die plattdeutsche Rechtschreibung*. Hamburg: Verlag der Fehrs-Gilde.