

*Peter Widell og Ulf Dalvad Berthelsen (udg.):
11. Møde om Udforskningen af Dansk Sprog
Århus 2006*

Internordisk sprogforståelse i et dialektperspektiv.

af Charlotte Gooskens (Skandinavisk afdeling, Groningen Universitet, Holland)

1. Introduktion

De tre skandinaviske sprog er så nært beslægtede, at nordmænd, svenskere og danskere kan tale med hinanden på deres respektive sprog. Den interskandinaviske kommunikation foregår dog ikke altid lige gnidningsløst. Der er hidtil blevet foretaget flere undersøgelser med det formål at få et indblik i, hvor godt skandinaver forstår hinanden (f.eks. Maurud 1976, Delsing & Åkesson 2005). Det er kendetegnende for disse projekter, at de kun tester forståelsen af standardvarianterne dansk, norsk og svensk. Dette kan især ses som en mangel, når det gælder forståelsen af norsk, eftersom de fleste nordmænd taler dialekt, og det derfor ikke kun er en standardvariant, man konfronteres med, når man hører norsk. Der er desuden et sparsomt kendskab til danskernes forståelse af nationale dialekter.

Den interskandinaviske sprogforståelse kan forventes at afhænge af tre faktorer: a) holdning over for landets sprog og dets indbyggere (attitude), b) erfaring med sproget gennem f.eks. medierne eller personlig kontakt, c) lingvistisk afstand mellem sprogene. De fleste undersøgelser af emnet, har først og fremmest søgt at belyse og forklare betydningen af de første to faktorer. Det har dog været svært at finde en direkte sammenhæng mellem sprogforståelse og de ekstra-lingvistiske faktorer, holdning og kontakt. Der findes mange kontrastive beskrivelser af de skandinaviske sprog, men hidtil er der kun gjort få forsøg på at undersøge det direkte forhold mellem nabosprogsforståelse og lingvistiske afstande. Gooskens (2006, under publicering a, b) fandt en høj korrelation mellem fonetiske afstande og målingerne af sprogforståelse i Delsing og Åkessons undersøgelse (2005). Bezooijen & Gooskens (under publicering) udviklede en metode til at måle afstande på forskellige lingvistiske niveauer og brugte denne metode til at forklare hollænderes forståelse af frisisk og afrikaans, som er nært beslægtede med nederlandsk.

Undersøgelsen, som her skal beskrives, havde til formål at kortlægge forholdet mellem danskeres forståelse af forskellige nordiske dialekter¹ og lingvistiske afstande til disse dialekter. Forskningsspørgsmålene kan formuleres som følger:

- (1) Hvor godt kan danskere forstå forskellige nordiske dialekter?
- (2) Hvor store er de lingvistiske afstande mellem disse nordiske dialekter?
- (3) Hvilken sammenhæng er der mellem sprogforståelsen og lingvistiske afstande?



Figur 1: De 18 dialekters placering i Norden.

2. Materiale

Vi testede forståelsen af i alt 18 forskellige dialekter. En del af materialet, som blev brugt til lytteeksperimentet, var allerede indsamlet. Det drejer sig om optagelser af fabelen *Nordenvinden og solen* på mere end 50 forskellige norske dialekter.² Jeg valgte otte af disse dialekter til min undersøgelse. Som det fremgår af Figur 1, er dialekterne godt fordelt over Norge, ligesom de er

¹ I virkeligheden drejede det sig om både dialekter og standardsprog. For nemheds skyld henviser jeg til alle 18 sprogvarianter med termen 'dialekt' i resten af artiklen.

² Optagelserne blev lavet af Jørn Almberg og Kristian Skarbø fra NTNU. De er tilgængelige via <http://www.ling.ntnu.no> sammen med de fonetiske transskriptioner samt oplysninger om informanternes baggrund. Jeg vil gerne takke for tilladelsen til at bruge materialet.

en rimelig repræsentation af den dialektologiske inddeling af landet. Der er to nordnorske (Tromsø og Rana), to vestlandske (Gaular og Fyresdal), to trønderske (Bjugn og Oppdal) og to østlandske (Trysil og Oslo) dialekter. Selvom der ikke findes et standardsprog i Norge, betragtes dialekten i Oslo i denne undersøgelse som en repræsentant for norsk standardsprog.

For at kunne sammenligne genkendelsen af de norske dialekter med genkendelsen af andre dialekter i Norden lavede vi ekstra optagelser af færøsk (Thorshavn), rigssvensk (taleren er fra Stockholm) og fire svenske dialekter, som repræsenterer de største svenske dialektgrupper, nemlig Gryttinge (sydsvensk), Lidköping (götamål), Storliden (nordlandsk) og Helsinki (østsvensk), samt rigsdansk (taleren er fra København) og tre danske dialekter, nemlig Høgsted (vendelbomål), Katrad (vestjysk) og Hjordkær (sønderjysk).

Talerne fik forelagt teksten om *Nordenvinden og solen* på standardsproget og fik derefter tid til at tilpasse den til deres egen dialekt. Den danske, svenske og færøske tekst var oversættelser af den norske tekst.

Optagelserne af de 18 dialekter blev brugt til lytteeksperimenterne, som skulle give indsigt i danskernes forståelse af forskellige nordiske dialekter (se Sektion 3.). Derudover blev der lavet fonetiske transskriptioner af alle optagelserne. Disse transskriptioner blev brugt til at udregne de fonetiske afstande (se Sektion 4).

3. Sprogforståelse

Metode

Hver lytteprøve bestod af alle seks sætninger fra fabelen *Nordenvinden og solen*. Hver sætning blev præsenteret på en anden dialekt. Eftersom vi ville teste 18 dialekter, og fabelen kun har seks sætninger, blev dialekterne opdelt i tre grupper. Hver gruppe havde seks forskellige rækkefølger af dialekter og sætninger, således at hver sætning på hver dialekt kunne blive testet. Der fandtes altså i alt 18 forskellige versioner af lytteprøven. Inden for hver version blev sætningerne og dialekterne præsenteret i vilkårlig rækkefølge. Hver sætning blev præsenteret to gange. Først fik testpersonerne hele sætningen at høre. Derefter blev den præsenteret i mindre stykker. På den måde havde testpersonerne tid nok til at skrive oversættelsen ned, og det var ikke nødvendigt, at de skulle huske hele sætninger.

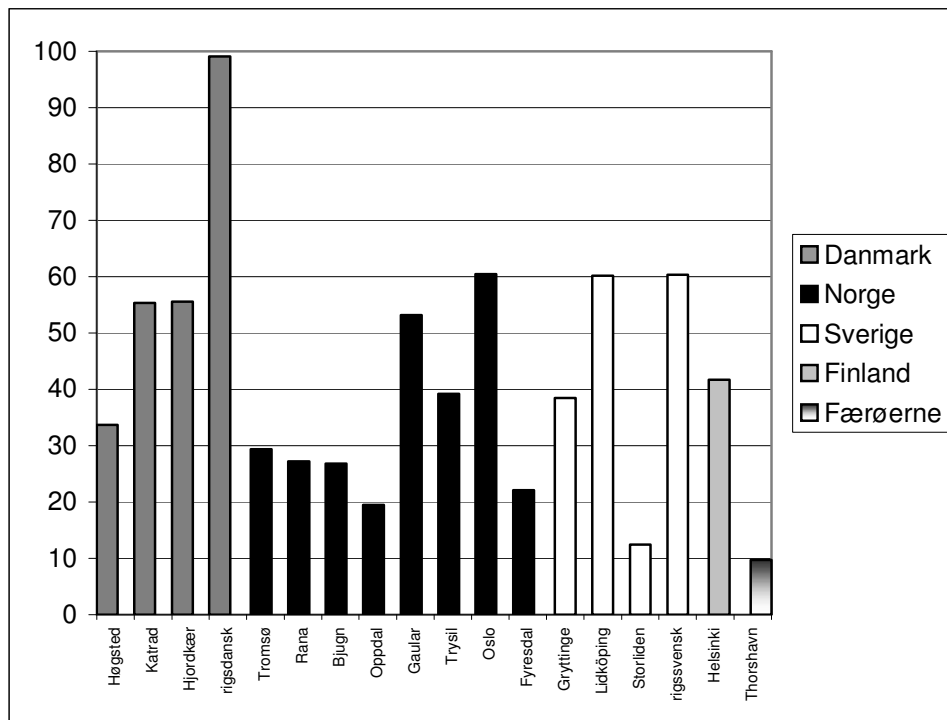
Testpersonerne var alle elever fra københavnske gymnasier. I alt deltog 351 elever i undersøgelsen. Jeg var især interesseret i forholdet mellem lingvistiske afstande og

sprogforståelse. De lingvistiske afstande blev målt mellem testpersonernes sprog og hver af de 18 dialekter. Eftersom alle testpersonerne boede i København og havde danske forældre, gik jeg ud fra at de talte en rigssprogsvariant eller i hvert fald havde grundigt kendskab til rigsdansk.

Testpersonerne blev bedt om at oversætte de seks sætninger ord for ord.

Resultater

For hvert ord i hver dialekt blev der tildelt et point for en korrekt oversættelse. Delvist korrekt oversatte ord, f.eks. hvis kun den sidste halvdel af ordet *nordenvinden* var korrekt oversat, fik et halvt point. I Figur 2 gengives procentdelen af korrekt oversatte ord pr. dialekt, dvs. gennemsnittet af korrekte oversættelser i alle seks sætninger.



Figur 2: Gennemsnitlig procentdel korrekt oversatte ord pr. dialekt.

Som det fremgår af Figur 2, er der stor forskel på, hvor godt testpersonerne forstår de forskellige dialekter. Rigsdansk oversættes som forventet næsten helt korrekt. Dette viser, at opgaven med at oversætte sætningerne i sig selv ikke var for svær. De danske dialekter har testpersonerne derimod svært ved at forstå. Dette gælder især for den nordjyske dialekt fra Høgsted, hvor kun

34,2% af ordene blev oversat korrekt, men også de to andre jyske dialekter har testpersonerne svært ved at forstå (Katrad 55,3% og Hjortkær 55,5%).

To af de norske dialekter forstår testpersonerne relativt godt, nemlig dialekterne fra Oslo (60,5%) og Gaular (53,2%). De øvrige norske dialekter har de svært ved at oversætte med under 30% korrekte oversættelser for alle dialekter bortset fra dialekten i Trysil (39,2%).

I Sverige er det rigssvensk (60,3%) og dialekten fra Lidköping (60,2%), som testpersonerne har lettest ved at forstå. Storliden-dialekten er så godt som uforståelig (12,5%) og Gryttinge-dialekten er også svær (38,5%). Observationen at '[d]et är en myt att danskar skulle ha lättare att förstå skånska än rikssvenska' (Nordisk Råd 2002) kan altså bekræftes. Af den finlandssvenske dialekt fra Helsinki blev 41,2% oversat korrekt.

Den allersværeste dialekt er den færøske dialekt fra Thorshavn (9,6%). Dette kommer ikke som en overraskelse, eftersom færøsk tilhører en anden gren inden for den nordiske sproggruppe, men der er altså norske og svenske dialekter, som forstås næsten lige så dårligt.

Som nævnt i indledningen er der flere forskellige faktorer, som kan være afgørende for, hvor godt de forskellige dialekter forstås. I denne artikel vil jeg dog koncentrere mig om forholdet mellem sprogforståelse og lingvistiske afstande. I følgende afsnit regnes de lingvistiske afstande mellem rigsdansk og de 18 dialekter ud, og i afsnit 5 korreleres de lingvistiske afstande med resultaterne af sprogforståelsestesten.

4. Lingvistiske afstande

Fonetiske afstande

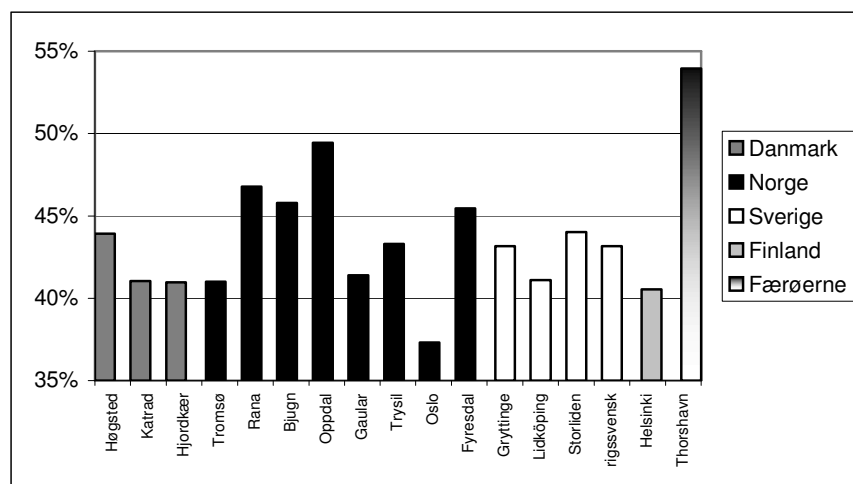
Fonetiske transskriptioner af de 18 dialekter danner grundlaget for de lingvistiske afstandsmålinger. Kun historisk beslægtede ord (*cognates*) blev inkluderet i målingerne, eftersom det ikke er meningsfyldt at måle afstandene mellem ord, som ikke er beslægtede (*non-cognates*). De ikke-beslægtede ord vil blive brugt til at måle de leksikale afstande (se nedenfor). Levenshtein-afstande mellem rigsdansk og en af de 18 dialekter er omkostningerne ved at ændre et ord på rigsdansk til det samme ord på den pågældende dialekt, jo flere omkostninger jo større er afstanden. Basisomkostningerne er tilføjelser, borttagelser og erstatninger af fonetiske symboler. Der findes flere versioner af metoden (se Heeringa 2004), men i min undersøgelse blev omkostningerne beregnet som følger:

operation	point
- tilføjelser og borttagelser	1
- erstatninger af identiske symboler	0
- erstatninger af en vokal med en vokal eller af en konsonant med en konsonant	0,5
- erstatninger af en vokal med en konsonant eller omvendt	1
- diacritiske tegn (regnes sammen med det foregående symbol)	0,25

Afstanden mellem f.eks. [a] og [a:] var altså 0,25, mellem [a] og [o] 0,5 og mellem [o] og [a:] 0,75. Der blev kompenseret for ordenes længde ved at dele summen af omkostninger med antallet af symboler. Her gives et eksempel på, hvordan man kan regne Levenshtein-afstanden ud mellem det danske ord *gade* og det tilsvarende svenske ord *gata*.

	1	2	3	4
dansk	g	ε:	ð	o
svensk	g	a:	t	a
omkostninger	0	0,5	0,5	0,5

De totale omkostninger ($0,5 + 0,5 + 0,5 = 1,5$) divideres med antallet af symboler (4). Dette giver en afstand på 37,5%. Afstanden mellem de to sprogvarieteter er gennemsnittet af afstandene mellem alle ordparene i en tekst. Afstandene udtrykkes i procent. Resultaterne af afstands-målingerne præsenteres i Figur 3. Afstanden til rigsdansk er i sagens natur nul og vises ikke.

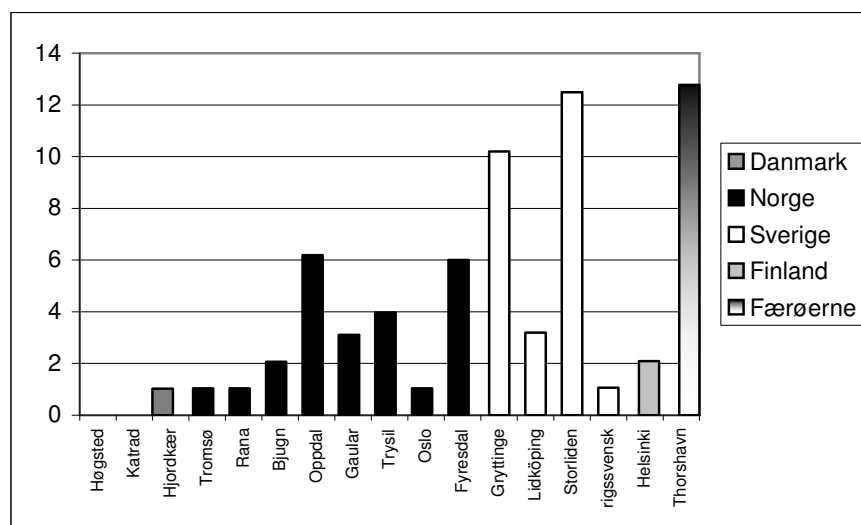


Figur 3: Levenshtein-afstanden fra rigsdansk til de øvrige 17 dialekter i undersøgelsen.

Som forventet er afstanden til den færøske dialekt størst (54,0%), men der er også et par norske dialekter som er meget afvigende. Fire af de norske dialekter har en afstand til rigsdansk på over 45%. På den anden side er Oslo-dialekten den dialekt i undersøgelsen, som ligner rigsdansk mest (37,2%). De svenske og de danske dialekter ligger mellem 40 og 45%.

Leksikale afstande

Ofte er dialekter afvigende på flere lingvistiske niveauer. I denne undersøgelse har jeg koncentreret mig om de to niveauer, som jeg går ud fra har størst betydning for forståelsen af de nordiske dialekter, nemlig det fonetiske og det leksikale niveau. Ofte vil dialekter, som er fonetisk meget afvigende fra dansk, også have mange ord, der ikke er genetisk beslægtede med de tilsvarende ord på dansk, og som derfor er uforståelige for en dansker. I sådanne tilfælde vil de to lingvistiske niveauer forstærke hinanden og gøre det endnu sværere at forstå dialekterne, end hvis afvigelserne kun findes på det ene niveau. Der er dog ikke nødvendigvis en sammenhæng mellem fonetiske afstande og leksikale afstande. Det er muligt, at dialekter fonetisk set ligner rigsdansk, men at der alligevel er mange ikke-beslægtede ord i dialekten. Det omvendte kan også være tilfældet. Eftersom et enkelt uforståeligt ord kan være meget forstyrrende for forståelsen af en hel sætning, er det vigtigt også at måle leksikale afstande. Disse afstande måles ved at regne ud, hvor stor del af ordene i dialektteksten, der ikke findes i en genetisk beslægtet form på rigsdansk (Figur 4).



Figur 4: Leksikale afstande mellem rigsdansk og de øvrige 17 dialekter udtrykt som procentdelen af ord pr. dialekt, som ikke findes i en genetisk beslægtet form på rigsdansk.

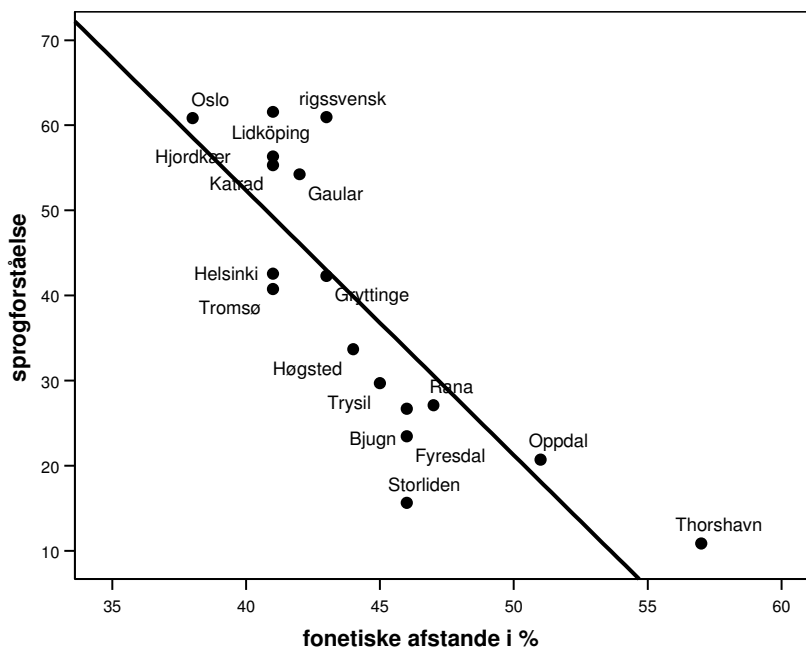
Som det fremgår af Figur 4, er den leksikalske afstand til færøsk stor (12,8% ikke-beslægtede ord). I modsætning til de fonetiske afstande er det to svenske dialekter (Storliden 12,5% og Gryttinge 10,2%), som har de største leksikale afstande til rigsdansk, mens de norske dialekter har en betydelig mindre leksikal afstand. De øvrige tre svenske dialekter samt alle de danske dialekter har også en meget lille leksikalsk afstand til dansk. Korrelationen mellem leksikalske og

fonetiske afstande er signifikant ($r=.55$, $p=.02$), hvilket viser, at der er en vis sammenhæng mellem de to niveauer, men at de i en del tilfælde er uafhængige af hinanden.

5. Sammenhængen mellem sprogforståelse og lingvistiske afstande

Fonetiske afstande

Sammenhængen mellem sprogforståelse og fonetiske afstande kan man undersøge ved at regne korelationskoefficienten ud mellem den gennemsnitlige procentdel korrekt oversatte ord pr. dialekt (Figur 2) og de tilsvarende fonetiske afstande (Figur 3). Korrelationen er høj og signifikant på 1% niveauet ($r=0,82$, $p=.00$ og $r=0,80$, $p=.000$ hvis rigsdansk udelades).



Figur 5: Punktdiagram som viser forholdet mellem sprogforståelsesresultaterne fra Figur 2 og de fonetiske afstande fra Figur 3 ($r=.80$). Rigsdansk er udeladt.

I Figur 5 vises sprogforståelsesresultaterne og de fonetiske afstande i et punktdiagram. Som det fremgår af heraf, kan forståelsen af de fleste dialekter i høj grad forudsiges ud fra de fonetiske afstande. F.eks. kan man se, at forklaringen på, at de to jyske dialekter fra Hjordkær og Katrad forstås bedre end dialekten fra Høgsted, efter al sandsynlighed er, at den vendsysselske udtale er mere afvigende fra rigsdansk end den vestjyske og den sønderjyske. Storliden ligger lidt længere under regressionslinjen end de andre dialekter. Man kunne ud fra de fonetiske afstande altså forvente, at testpersonerne havde været bedre til at forstå denne dialekt, end de i virkeligheden

var. På den anden side blev rigssvensk samt dialekten fra Lidköping forstået bedre end forventet. Jeg kommer ind på dette nedenfor.

Leksikale afstande

Korrelation mellem sprogforståelse og leksikale afstande var lavere end med de fonetiske afstande og ikke signifikant ($r = .582, p = 0,07$). Hvis rigsdansk udelades, er korrelationen lavere ($r = 0,47, p = 0,11$).

Det er altså ikke i så høj grad muligt at forudsige sprogforståelsen på basis af leksikale afstande. Sandsynligvis hænger dette sammen med, at det er meget varierende, hvor vigtige forskellige ord er for sprogforståelsen. Hvis et ord ikke har et særligt vigtigt semantisk indhold, har det ikke så store konsekvenser, hvis det er uforståeligt. I andre tilfælde kan et enkelt uforståeligt ord bevirke, at resten af en sætning bliver uforståelig, hvis det drejer sig om et centralt begreb. De leksikale afstande kan dog være med til at give indsigt i, hvorfor nogle af dialekterne blev forstået dårligere, end man skulle forvente ud fra de fonetiske afstande, mens andre blev forstået bedre.

Især Storliden-dialekten blev noget dårligere forstået end forventet ud fra de fonetiske afstande. I Figur 4 kan vi se, at det også er den dialekt, som, bortset fra Thorshavn-dialekten, indeholder de fleste ikke-beslægtede ord. Gryttinge har en næsten ligeså stor leksikal afstand, men forstås dog efter forventningerne. Det er muligt at en ikke-lingvistisk faktor spiller en rolle her. Gryttinge ligger nemlig tæt på København, så det er større chance for at testpersonerne har erfaring med denne dialekt end med den nordlige dialekt fra Storliden. De andre to dialekter i Sverige indeholder meget få ikke-beslægtede ord og forstås også bedre end forventet. Det er altså muligt at de leksikale afstande kan forklare en del af de svenske resultater. På den anden side ser der ikke ud til at være nogen tydelig sammenhæng mellem de leksikale afstande og sprogforståelse ved de øvrige dialekter.

6. Konklusion

Resultaterne af undersøgelsen viste, at der er stor forskel på, hvor godt de forskellige dialekter i Norden forstås af unge danskere fra København. Det er påfaldende, at testpersonerne har næsten ligeså svært ved at forstå jyske dialekter som nogle af de svenske og norske dialekter. Generelt set forstås standardvarianterne bedst. Når det drejer sig om interskandinavisk kommunikation, er

det derfor måske værd at følge rådet i Nordisk Råds rapport (2002) om at undgå udtalevarianter, som kun findes i dialekter. Forklaringen på at standardvarianterne forstås bedst, ser ud til bl.a. at have en lingvistisk basis, men kan også hænge sammen med at det er disse sprogvarianter, som testpersonerne oftest hører, hvis de konfronteres med nabosprogene. De nordligste dialekter i Norge og Sverige forstås generelt set dårligst, hvilket også kunne betyde, at kontakt er en vigtig faktor, eftersom det er rimeligt at forvente at testpersonerne har haft størst chance for at komme i kontakt med de sydligste dialekter.

Sprogforståelsen kan i høj grad forudsiges ud fra fonetiske afstande til de forskellige dialekter. Selvom korrelationen med de leksikale afstande er lav, kan undtagelser ofte forklares ved hjælp af procentdelen af ord som ikke er genetisk beslægtede. Resultaterne viser altså en tydelig sammenhæng mellem sprogforståelse og fonetiske afstande, men det er en vigtig fremtidig opgave at analysere de fonetiske forskelle i nærmere detaljer, så man kan få et dybere indblik i, hvilke fonetiske faktorer, der især virker hindrende på sprogforståelsen.

Litteraturliste

- Bezooijen, Renée van & Gooskens, Charlotte (under publicering) Interlingual text comprehension: linguistic and extralinguistic determinants In: Jan D. ten Thije & Ludger Zeevaert (eds.). *Receptive Multilingualism and intercultural communication*. Hamburger Studies in Multilingualism. Amsterdam: Benjamins.
- Delsing, Lars-Olof & Lundin Åkesson, Katarina (2005) *Håller språket ihop Norden? En forskningsrapport om ungdomars förståelse av danska, svenska och norska*. Köbenhavn: Nordiska ministerråd.
- Gooskens, Charlotte (2006) Linguistic and extra-linguistic predictors of Inter-Scandinavian intelligibility. I Jeroen van de Weijer & Bettelou Los (red.). *Linguistics in the Netherlands* 23. Amsterdam: John Benjamins, 101-113.
- Gooskens, Charlotte (under publicering a) Contact, attitude and phonetic distance as predictors of inter-Scandinavian communication. *Proceedings Collateral languages 2, Limerick June 2005*. Paris: L'Harmattan.
- Gooskens, Charlotte (under publicering b) The contribution of linguistic factors to the intelligibility of closely related languages. *Journal of Multilingual and multicultural development*.
- Heeringa, Wilbert (2004) *Measuring Dialect Pronunciation Differences using Levenshtein Distances*. Groningen: Groningen dissertations in linguistics (Grodil).
- Maurud, Øyvind (1976) *Nabospråksförståelse i Skandinavia: en undersøkelse om gjensidig förståelse av tale- og skriftspråk i Danmark, Norge og Sverige*. Stockh: Skandinavisk råd.
- Nordisk råd (2002) *Att förstå varandra i Norden: språkråd till nordbor i nordiskt samarbete: en handledning*. Stockholm: Nordiska Rådet. http://www.nordisk-sprakrad.no/folder_att_forstaa/att_forstaa.doc.