

## Skriftsprogets rolle i den dansk-svenske talesprogsforståelse

Af GERARD DOETJES og CHARLOTTE GOOSKENS

---

### Abstract

*Doetjes, Gerard*, gerard.doetjes@fremmedspraaksenteret.no, Research Fellow, Norwegian Centre for Foreign Languages in Education, Østfold University College, Halden, Norway; *Gooskens, Charlotte*, c.s.gooskens@rug.nl, Associate Professor, Dept. of Scandinavian Linguistics and Literatures, University of Groningen, The Netherlands: "The role of orthography in Danish-Swedish intelligibility". *Språk och stil* 19, 2009, pp. 105–123.

The fact that mutual intelligibility of Danish-Swedish spoken language is asymmetric can be explained by various linguistic and extra-linguistic factors. The present paper focuses on the role of orthography. Both orthographies are rather conservative, but Danish pronunciation has changed more rapidly during the last century than Swedish pronunciation has. As a consequence of the conservative spelling, Danes can be expected to have more support from the orthography of their mother tongue than Swedes when listening to the neighbouring language. This hypothesis is tested, first by measuring phonetic distances between the spoken languages' adaption and taking into consideration the help that the listeners can receive from the orthography when listening to the neighbouring language. The distance corrected for orthography turns out to be smaller for the Danes than for the Swedes. This indicates that Danes have more potential help from the orthography than Swedes do. Second, the distance is correlated with the results of a word intelligibility experiment among Danes. The distance corrected for the influence of orthography shows higher correlations with the intelligibility scores than pure phonetic distances. From this, the authors conclude that Danish listeners indeed seem to make use of the additional information that the orthography can provide.

*Keywords:* Swedish, Danish, mutual intelligibility, orthography, linguistic distance, Levenshtein.

---

## 1. Indledning

På grundlag af diverse empiriske undersøgelser af den interskandinaviske sprogforståelse (Maurud 1976 s. 184–185,<sup>1</sup> Bø 1978 s. 91 og Delsing & Lundin

---

Vi takker Sebastian Kürschner, Kirsten Lundgaard Kolstrup samt tre anonyme referenter for værdifulde kommentarer til denne artikel.

<sup>1</sup> I Maurud 1976 præsenteres resultaterne som medianer. Vi har beregnet gennemsnittet pr. deltager på grundlag af informationerne i appendikset.

Åkesson 2005 s. 73 og 75<sup>2</sup>) kan man drage de samme generelle konklusioner:<sup>3</sup> talt norsk forstås temmelig godt af både danskere og svenskere (se tabel 1); nordmænd er forholdsvis gode til at forstå nabosprogene; hvorimod danskere og svenskere har svært ved at forstå hinandens talesprog. Den dansk-svenske sprogforståelse har desuden vist sig at være asymmetrisk, således at danskere er bedre til at forstå svensk, end svenskere er til at forstå dansk. Denne asymmetri findes dog kun ved talesprogsforståelsen, mens svenskere og danskere har lige svært ved at forstå hinandens skriftsprog (se tabel 2). Det er generelt lettere for skandinaver at læse hinandens sprog end at forstå dem, når de tales.

*Tabel 1.* Empiriske undersøgelser af den indbyrdes forståelse af dansk, norsk og svensk talesprog i Maurud 1976, Bø 1978 og Delsing & Lundin Åkesson 2005.

	dansk i Norge	dansk i Sverige	norsk i Danmark	norsk i Sverige	svensk i Danmark	svensk i Norge
Maurud (1976)	72 %	21 %	65 %	41 %	40 %	81 %
Bø (1978)	62 %	32 %	63 %	74 %	49 %	80 %
Delsing & Lundin Åkesson (2005)	41 %	19 %	35 %	56 %	27 %	69 %

*Tabel 2.* Empiriske undersøgelser af den indbyrdes forståelse af dansk, norsk og svensk skriftsprog i Maurud 1976, Bø 1978 og Delsing & Lundin Åkesson 2005.

	dansk i Norge	dansk i Sverige	norsk i Danmark	norsk i Sverige	svensk i Danmark	svensk i Norge
Maurud (1976)	94 %	68 %	81 %	76 %	60 %	87 %
Bø (1978)	75 %	41 %	48 %	56 %	44 %	76 %
Delsing & Lundin Åkesson (2005)	77 %	55 %	73 %	55 %	63 %	74 %

### 1.1 Den asymmetriske dansk-svenske talesprogsforståelse

Det er altså omtrent lige let for svenskere og danskere at læse hinandens sprog, mens der findes en asymmetri, når det gælder talesprogsforståelse. Svenskere

<sup>2</sup> I Delsing & Lundin Åkesson 2005 blev der udført både en audiotest og en videotest. Her præsenteres kun resultaterne af audiotesten, eftersom videotestene var forskellige for hvert sprog og derfor ikke sammenlignelige. Resultaterne er omregnet til procent.

<sup>3</sup> De samme konklusioner er man kommet frem til i tre undersøgelser, hvor sprogforståelsen ikke blev testet direkte, men hvor deltagerne i undersøgelsen blev spurgt om, hvor godt de kunne forstå indbyggerne i nabolandene (Haugen 1953, Nordiska rådet 1973, Börestam Uhlmann 1991).

har temmelig svært ved at forstå talt dansk, mens det for danskere er lettere at forstå talt svensk. Der findes forskellige lingvistiske og ekstralingvistiske forklaringer på, at den svensk-danske sprogforståelse er asymmetrisk. Hidtil har de ekstralingvistiske forklaringer fået den største opmærksomhed. I de fleste undersøgelser går man ud fra, at især attitude spiller en vigtig rolle. Braunmüller (1999) ser den individuelle vilje som et grundlæggende kriterium for, at den interskandinaviske kommunikation lykkes (s. 309). I Skandinavien har Sverige længe været det økonomisk dominerende land. Dette kan have medført, at lillebroderen Danmark har haft større interesse i at forstå svensk end omvendt. Disse holdninger genspejles også i de bedømmelser af sprogene, som testpersonerne har givet i tidligere studier (se referencer i afsnit 1), hvor man inkluderede spørgsmål om holdninger over for nabosprogene. Der så ud til at være en vis sammenhæng mellem holdninger og sprogforståelse. Både i de ældre undersøgelser og i den nyeste undersøgelse af Delsing & Lundin Åkesson (2005) var de dansk-svenske attituder asymmetriske: danskerne synes bedre om det svenske sprog end omvendt. Korrelationerne mellem attitudeskalaerne og sprogforståelsen er dog lave, og det er svært at bevise, at der er en direkte sammenhæng (se Gooskens 2006).

En anden ekstralingvistisk faktor er den kontakt med sproget, som lytterne har haft i forvejen via personlige relationer og forskellige medier. Man går ud fra, at kontakt med sproget resulterer i, at man lærer sproget og derfor er bedre i stand til at forstå det. Det lader til, at danskere ser mere svensk fjernsyn, læser flere svenske aviser og oftere besøger Sverige end omvendt. I Delsing & Lundin Åkesson 2005 er der en vis sammenhæng mellem sprogforståelsen og et kontaktindeks, som blev beregnet ud fra svarene på forskellige spørgsmål om kontakt med nabolandet (s. 101).

Vi vil her se nærmere på en faktor, som også kan være afgørende for den asymmetriske talesprogsforståelse mellem svenskere og danskere, nemlig indflydelse fra skriftsproget. Det danske talesprog har udviklet sig meget hurtigt i det sidste århundrede (Brink & Lund 1975, Grønnum 1998), hvilket har resulteret i en stor afstand mellem skriftsproget og talesproget. Afstanden er mindre i svensk, hvor talesproget har fjernet sig fra skriftsproget i et langsommere tempo. I nogle tilfælde har talesproget oven i købet nærmet sig skriftsproget (Wessén 1965 s. 152, Birch-Jensen 2007). De afvigende forhold mellem skrift og tale i de to lande har resulteret i forskellige udgangspunkter, når det gælder forståelsen af nabosproget. Når danskere hører et svensk ord, har de formodentlig ofte støtte fra deres kendskab til stavningen af det tilsvarende ord på dansk, fordi den danske ortografi ligger tæt ved den svenske udtale. Teleman (1980) opstiller følgende hypotese:

Det är oftare så att man från ett *dansk* uttal inte kan härleda motsvarande ortografiska uttryck med hjälp av *svenskans* regler för ljud-skrift-relation, än det är att man från ett *svenskt* uttal inte kan härleda motsvarande ortografiska uttryck med det en *dansk* vet om relationen ljud/skrift (s. 273).

Delsing & Lundin Åkesson (2005 s. 137) förklarar också resultaten af den dansk-svenske talesprogsförståelse med de forskellige forhold mellem skrift og tale i Danmark, Norge og Sverige:

Danskarnas lilla övertag ska nog i första rummet tillskrivas det faktum att tal och skrift i danskan ligger så pass långt ifrån varandra, medan svenskt och norskt uttal ligger närmare skriftspråket. De tre skriftspråken liknar varandra ganska mycket, och danskarna har den vägen god hjälp att avkoda t.ex. svenska ord som *bägare*, [bɛ:garə], med hjälp av den danska skriftformen *bæger*, medan svenskarna har svårare att förstå att det köpenhamnska uttalet [bæ:ɔ] ska förknippas med svenska ordet *bägare* (s. 137).

Dette kan også illustreres med ordparret da. *hånd* – sv. *hand*. Udtalen af ordet er forskellig på de to sprog: da. [hɔnʔ] – sv. [hand]. Bl.a. udtales slutkonsonanten ikke på dansk, men eftersom den findes i den skrevne form, kan danskere afkode det svenske ord via stavemåden. Omvendt har svenskere mindre hjælp fra den svenske stavemåde, når de hører den danske udtale, fordi d'et ikke udtales. Eksempler, som peger i modsat retning, findes også. Det danske ord *kylling* (udtalt med [k] i begyndelsen af ordet) er formodentligt lettere for svenskere at forstå, end det tilsvarende svenske ord *kyckling* (udtalt med [ç]) er for danskere. En dansker ved formodentligt ikke, at et *k* kan udtales som [ç], mens det for en svensker ikke er ualmindeligt, at skriftsprogets *k* svarer til udtalen [k], jf. fx ordet *katt* [kat]. Det er dog sandsynligt, at fordelene er større for danskere end for svenskere, fordi det især er det danske talesprog, som har ændret sig gennem det seneste århundrede. Det vil sige, at afstanden mellem dansk udtale og ortografi er større end afstanden mellem svensk udtale og ortografi.

I denne artikel vil vi måle afstande mellem sprogene og vise, at det i sådan en beregning er muligt at tage højde for den hjælp, som danskere har fra ortografien, når de hører talt svensk. Spørgsmålet, som vi ønsker at besvare, er som følger:

I hvor høj grad bidrager den danske ortografi til danskeres forståelse af svensk talesprog?

Vi har valgt at basere vores beregninger på den såkaldte Levenshtein-algoritme (Heeringa 2004). I afsnit 2 viser vi først, hvordan Levenshtein-algoritmen kan bruges til at måle ortografiske og fonetiske afstande mellem svensk og dansk.

Derefter udvikler vi en ny variant af Levenshtein-algoritmen, som modellerer skriftsprogets indflydelse på den dansk-svenske sprogforståelse. I afsnit 3 tester vi denne model ved hjælp af et lytteeksperiment. I modsætning til de fleste tidligere undersøgelser begrænser vi vores materiale til isolerede ord. Fordelen ved denne fremgangsmåde er, at det i højere grad er muligt at undersøge, hvilke sproglige fænomener der har indflydelse på forståelsen, end hvis man undersøger ord i en kontekst. Det er nemlig svært at kontrollere konteksten på en sådan måde, at den påvirker ordforståelsen i lige høj grad i alle tilfælde. Ulemper ved at teste isolerede ord er, at testsituationen i mindre grad modellerer virkeligheden, eftersom man kun sjældent kommer ud for at skulle forstå ord i isolation. Desuden kan man i tilfælde af isolerede ord ikke teste virkningen af fænomener, som kun forekommer på frase- eller sætningsniveau såsom assimilationer henover ordgrænser samt intonation. I afsnit 4 undersøger vi andre sproglige faktorer, som også kan spille en rolle for sprogforståelsen, og i afsnit 5 drager vi nogle konklusioner.

## 2. Måling af sproglige afstande

I dette afsnit vil vi først beskrive det materiale, som vi baserede vores afstandsmålinger på (afsnit 2.1). Derefter vil vi vise, hvordan vi beregnede afstandene mellem dansk og svensk skriftsprog (afsnit 2.2) samt mellem dansk og svensk talesprog (afsnit 2.3).

### 2.1 Materiale

Afstandsberegningerne er baseret på en liste med 96 ord, som oprindeligt blev indsamlet til diverse hollandske undersøgelser (se fx Gooskens & Van Bezooijen 2008). Ordene refererer til dagligdags genstande og koncepter inden for elleve forskellige domæner: dele af huset, genstande i dagligstuen, køkkenet og soveværelset, fugle og insekter, dyr i og omkring huset og i naturen, mad og drikke, kropsdele, tøj, blomster, planter og træer, vejr og landskab. 25 mænd og kvinder mellem 22 og 61 år (gennemsnitsalder 40 år) med forskellig social baggrund blev bedt om at nævne det ord, som de først kom til at tænke på inden for hvert domæne. De 87 ord (40 % af alle ordene), der blev nævnt af mindst ti personer, blev optaget på listen. Desuden blev de føl-

gende 9 ord føjet til listen: *hus, fugl, dyr, grøft, eng, skov, blomst, plante og træ*. Disse ord blev ikke nævnt af informanterne, eftersom de udgør titlerne på de domæner, som blev præsenteret for dem. I alt bestod listen således af 96 ord.

Vi oversatte den hollandske liste til dansk og til svensk og noterede, hvilke ord der var kognater, dvs. ordpar som er historisk beslægtede i de to sprog. Eksempler på dansk-svenske kognater er ordparrene da. *hus* sv. *hus* eller da. *kylling* sv. *kyckling*. Ti af de danske ord var ikke beslægtede med de tilsvarende svenske ord (ikke-kognater), fx ordparrene da. *kartoffel* sv. *potatis* og da. *frø* sv. *groda*. Disse ord blev udeladt fra videre analyse, da det i sagens natur ikke er relevant at måle ortografiske eller fonetiske afstande mellem ikke-kognater. Desuden er det i princippet ikke muligt at forstå nabosprogets ikke-kognater uden forkundskab til sproget. Ikke-kognater bør naturligvis medtages i en undersøgelse, hvis formål det er at undersøge, hvor godt danskere og svenskere forstår hinandens ordforråd, men her er formålet et andet, nemlig at undersøge hvor stor rolle skriftsproget spiller for den indbyrdes talesprogsforståelse. Kognater har forskellige grader af lighed og ulighed (se fx Rekdal 1981 s. 191–194, Barðdal et al. 1997 kap. 3), hvilket vi tager højde for ved at måle fonetiske og ortografiske afstande mellem ordparrene (se afsnit 2.2 og 2.3). En liste over de 86 kognater findes i appendikset. Vi transkriberede disse ord fonetisk ved hjælp af en dansk udtaleordbog (Brink et al. 1991) og en svensk udtaleordbog (Hedelin & Elert 1997). Begge ordbøger har rigssproget som udgangspunkt.

## 2.2 Måling af ortografiske afstande mellem svensk og dansk

Vi brugte Levenshtein-algoritmen til at måle de sproglige afstande. Med denne algoritme er det muligt at måle afstande mellem dialekter. Den blev for første gang brugt af Kessler (1995) til at lave dialektkort over gæliske dialekter i Irland, og har siden været brugt til at klassificere dialekterne i flere forskellige sprogområder, fx det hollandske (Heeringa 2004), tyske (Nerbonne & Siedle 2005) og norske (Gooskens & Heeringa 2004). Algoritmen defineres som »a numerical value of the cost of the least expensive set of insertions, deletions or substitutions that would be needed to transform one string into another« (Kruskal 1999 s. 18–23). Det drejer sig altså om en algoritme, som beregner omkostningerne ved at ændre en »string« (i vores tilfælde en række af bogstaver eller fonetiske symboler) til en anden med det fær-

rest mulige antal operationer. Operationer kan være tilføjelser, udeladelser og erstatninger. Jo flere operationer man har brug for, desto større er afstanden. Princippet illustreres med eksempler nedenfor. Sprogforståelsesundersøgelser har vist, at der er en tydelig sammenhæng mellem målinger af sproglige afstande udført med denne algoritme og forståelsen af forskellige sprogvarianter i Skandinavien (Gooskens 2007a og 2007b). Selvom metoden hidtil har været brugt til at måle fonetiske afstande, er den også velegnet til at måle ortografiske afstande.

Når man vil beregne afstanden mellem to sprogvarianter ved hjælp af Levenshtein-algoritmen, måler man først afstanden mellem alle ordpar i parallelle ordlister fra de to sprogvarianter. I vores tilfælde var materialet de to ordlister med de samme 86 kognater (beslægtede ord) på svensk og dansk (se afsnit 2.1). Hvis vi fx ser på ordparret da. *pande* – sv. *panna*, kan vi fastslå, at to ud af fem bogstaver er forskellige, jvf. eksempel (1). Man tager højde for ordlængde ved at dividere antallet af forskelle med antallet af bogstaver. Forskellen mellem de skrevne former af de to ord er altså 0,40 eller 40 % (jo lavere tallet er, desto mindre er afstanden). Bogstaverne matches på en sådan måde, at den mindst mulige afstand opnås.

(1)

Dansk	p	a	n	d	e
Svensk	p	a	n	n	a
	0	0	0	1	1

Afstand            2/5 eller 40 %

Levenshtein-metoden tager også hensyn til, at bogstaver sommetider falder bort eller tilføjes som fx i ordparret da. *mælk* – sv. *mjölk*. Disse forskelle beregnes på samme måde som i de tilfælde, hvor bogstaverne er forskellige, se eksempel (2).

(2)

Dansk	m	-	æ	l	k
Svensk	m	j	ö	l	k
	0	1	1	0	0

Afstand            2/5 eller 40 %

Vi ved ikke, i hvor høj grad det virker hindrende for svenskere og danskere, der skal læse nabosproget, at svensk *ä* og dansk *æ* samt svensk *ö* og dansk *ø* ser

forskellige ud. Vi har valgt at betragte tegnene som ens, fx i ordparret da. *øje* – sv. *öga*.

(3)				
Dansk	ø	j	e	
Svensk	ö	g	a	
	0	1	1	

Afstand                    2/3 eller 67 %

Da vi udregnede de ortografiske afstande mellem alle 86 ordpar, viste det sig, at 29 af ordparrene blev skrevet helt ens på dansk og svensk og derfor havde en afstand på 0 %. De mest forskellige ordpar havde en afstand på 67 %, mens gennemsnittet af alle 86 afstande var 24 %.

### 2.3 Måling af fonetiske afstande mellem svensk og dansk

På baggrund af de fonetiske transskriptioner af de 86 kognater i ordlisten målte vi også afstanden mellem dansk og svensk talesprog. I de fonetiske transskriptioner repræsenteres hver lyd af et forskelligt fonetisk tegn. Stød og længdeforskelle (korte og lange vokaler/konsonanter) talte vi ikke med i beregningerne.<sup>4</sup> Beregningerne af Levenshtein-afstandene udføres efter samme principper som for afstandene mellem skriftsprogene (afsnit 2.2). For hvert ordpar blev afstanden mellem de fonetiske transskriptioner beregnet, hvorefter gennemsnittet af alle afstandene blev udregnet. Nedenstående eksempler (4) og (5) med ordparrene da. *pande* – sv. *panna* og da. *mælk* – sv. *mjölk* illustrerer, hvordan vi målte de fonetiske afstande. De ortografiske afstandsregninger gentages som reference. Hvis vi sammenligner de fonetiske og de ortografiske afstande, ser vi, at de fonetiske afstande for begge ord er større (50 % og 60 %) end de ortografiske afstande (begge ord har en ortografisk afstand på 40 %).

(4)	<i>Ortografisk afstand</i>	<i>Fonetisk afstand</i>
Dansk	p a n d e	p æ n ə
Svensk	p a n n a	p a n a
	0 0 0 1 1	0 1 0 1

Afstand                    2/5 eller 40 %

2/4 eller 50 %

<sup>4</sup> Det drejer sig her om et første forsøg på at kvantificere ortografiens rolle for sprogforståelsen. I fremtidig forskning vil vi inddrage mere detaljerede forskelle i algoritmen. Der findes fx varianter af Levenshtein-algoritmen, som tager hensyn til, at nogle lyde ligner hinanden mere end andre, se Heeringa 2004.



(5)	<i>Ortografisk afstand</i>	<i>Fonetisk afstand</i>
Dansk	m – æ l k	m – ε l g
Svensk	<u>m j ö l k</u>	<u>m j ø l k</u>
	0 1 1 0 0	0 1 1 0 1
Afstand	2/5 eller 40 %	3/5 eller 60 %

Ifølge resultaterne af de dansk-svenske sprogforståelsestests (se afsnit 1) kan man forvente, at afstanden mellem talesprogene er større end mellem skriftsprogene. Vi forventer altså, at afstandene i de fleste tilfælde viser samme tendens som ved ordene *pande* og *mælk*. Gennemsnittet af afstandene baseret på hele listen bekræfter hypotesen, idet afstanden mellem talesprogene er 53 %, altså 29 % højere end mellem skriftsprogene (24 %, se afsnit 2.2). Dette høje tal kan være en forklaring på, at det som regel er sværere for svenskere og danskere at forstå hinandens talte sprog end at læse dem. Påfaldende er det lave antal ord, som udtales ens på de to sprog, nemlig 7, i forhold til 27 som skrives helt ens. Der var to ord med en maksimal udtaleafstand på 100 %: *dyr/djur* (*r* vokaliseres på dansk) og det tidligere nævnte eksempel *øjelöga*, se (3). Tre af de 86 ordpar havde en større ortografisk end fonetisk afstand (*væg/vägg*, *seng/säng* og *eng/läng*), elleve ordpar havde den samme afstand, mens 72 ordpar havde en større fonetisk end ortografisk afstand.

#### 2.4 Fonetiske afstande korrigeret for stavning

I dette afsnit vil vi vise, hvordan man kan beregne afstandene mellem danske og svenske ord under hensyntagen til den hjælp, som man kan gå ud fra, at lytteren har fra sit eget modersmåls skriftsprog. Disse afstande kalder vi her *fonetiske afstande korrigeret for stavning*, dvs. at vi i afstandene har indberegnet den hjælp, som en dansker eller svensker kan få fra modersmålets skriftsprog, når vedkommende hører nabosproget. Eftersom udgangspunktet ikke er det samme for en svensker og en dansker, kan man gå ud fra, at afstandene kan være asymmetriske, altså at afstanden fra dansk til svensk ikke nødvendigvis er den samme som afstanden fra svensk til dansk. Dette kan igen illustreres med eksemplet da. *pande* – sv. *panna*, se (6). Hvis en dansker hører dette ord udtalt af en svensker, vil han kunne genkende a'et (udtalt som [a]) som en af de to muligheder, man har for at udtale skriftsprogets *a* (den anden mulighed er [æ]). Så selvom [a] og [æ] fonetisk set er forskellige, kan en dansker formodentlig let genkende begge lyde som repræsentanter for skriftsprogets *a*. For dette ord-

par betyder den tilpassede beregning altså, at den fonetiske afstand set fra en danskers perspektiv falder fra 50 % (rene fonetiske afstande) til 25 % (fonetiske afstande korrigeret for stavning).

(6)	<i>Ortografisk afstand</i>	<i>Fonetisk afstand</i>	<i>Fonetisk afstand korrigeret for dansk stavning</i>
Dansk	p a n d e	p æ n ə	p a/æ n ə
Svensk	<u>p a n n a</u>	<u>p a n a</u>	<u>p a n a</u>
	0 0 0 1 1	0 1 0 1	0 0 0 1
Afstand	2/5 eller 40 %	2/4 eller 50 %	1/4 eller 25 %

Omvendt har den nye beregningsmetodik ingen følger for afstandsberegningen for dette ordpar, når man tager situationen for en svensk lytter som udgangspunkt, se (7). Svensk har kun en måde at udtale a'et på, og derfor kan en svensker højst sandsynligt ikke genkende det danske [æ] som en mulig udtale af skriftsprogets *a*. Den korrigerede afstand er altså større for en svensker (50 %) end for en dansker (25 %), hvilket peger på, at det for en svensker sandsynligvis er sværere at forstå nabosprogets ord, end det er for en dansker.

(7)	<i>Ortografisk afstand</i>	<i>Fonetisk afstand</i>	<i>Fonetisk afstand korrigeret for svensk stavning</i>
Dansk	p a n d e	p æ n ə	p æ n ə
Svensk	<u>p a n n a</u>	<u>p a n a</u>	<u>p a n a</u>
	0 0 0 1 1	0 1 0 1	0 1 0 1
Afstand	2/5 eller 40 %	2/4 eller 50 %	2/4 eller 50 %

Den danske og den svenske ortografi ligner hinanden mere end udtalen (se afstandsmålingerne i afsnit 2.2 og 2.3), og man kan gå ud fra, at det især er den danske udtale, som har ændret sig, mens ortografien er mere konservativ. Dette indebærer, at en dansker, som hører et svensk ord, sandsynligvis oftere har støtte fra sit eget sprogs ortografi end en svensker, når han skal finde det tilsvarende ord på sit modersmål. Dette ser man fx i ordparet da. *pande* – sv. *panna*, men effekten er ganske udbredt og gælder de fleste ordpar på vores liste (se næste afsnit). Den rolle, som kendskab til ens eget skriftsprog har for sprogforståelsen, er altså ikke den samme for svenskere og danskere. Det, der ellers kan betragtes som en stor fordel for svensksprogede, nemlig den store lighed mel-

lem svensk i skrift og tale, bliver således en ulempe, når det gælder forståelsen af dansk.

Hvis vi vil bruge fonetiske afstandsmålinger til at modellere ortografiens rolle for den svensk-danske talesprogsforståelse, bør vi tage med i beregningerne, at lytterne kan gøre brug af deres eget skriftsprog, når de skal forstå nabosproget. Vi tager udgangspunkt i fonetiske afstande mellem ordparrene og reducerer afstandene i de tilfælde, hvor ortografien kan gøre det lettere at finde det tilsvarende ord på modersmålet. Det kan forventes, at dette resulterer i forskellige afstande mellem sprogene, afhængigt af om udgangspunktet er en svensk taler og en dansk lytter eller en dansk taler og en svensk lytter, hvilket også viser sig at være tilfældet. Afstanden, som modellerer en svenskers forståelse af dansk (de fonetiske afstande korrigeret for stavning), er højere (gennemsnitligt 46 % over alle 86 kognater) end afstanden, som modellerer en danskers forståelse af svensk (30 %). De nye afstande er altså lavere i begge retninger end den fonetiske afstand, som vi beregnede i afsnit 2.3 (53 %). Af de 86 ordpar er der 61 (71 %), som får en lavere afstand, når det gælder om at modellere danskernes svenske ordforståelse. Omvendt er der kun 22 ord (26 %), der får en lavere afstand, når svenskernes danske ordforståelse bliver modelleret. Dette viser, at danskere i højere grad end svenskerne kan gøre brug af modersmålets ortografi i forståelsen af nabosproget.

### 3. Lytteeksperiment

Resultaterne af afstandsmålingerne i afsnit 2 ser ud til at bekræfte vores hypotese om, at danskere har større nytte af deres eget sprogs ortografi end svenskerne, når de skal forstå det talte nabosprog. For at undersøge, om der virkelig findes en sammenhæng mellem sprogforståelsen og afstandsberegningerne, og om denne sammenhæng er stærkere, når afstandsberegningerne tager hensyn til ortografiens rolle, end når ortografien ikke tages i betragtning, udførte vi et ordforståelseseksperiment. Foreløbigt har vi kun testet danskeres forståelse af de 86 svenske ord. Det er her man kan forvente at finde den største hjælp fra ortografien (se afsnit 2.4), eftersom der er større forskel på de fonetiske afstande og de fonetiske afstande korrigeret for ortografi for danskerne end for svenskerne (se resultaterne af målingerne sidst i afsnit 2.4).

### 3.1 Metode

Undersøgelsens testmateriale bestod af en optagelse af alle ord i den oprindelige ordliste, som også omfattede 10 ikke-kognate ordpar. Taleren kom fra Uppsala og hans sprog kunne karakteriseres som rigssvensk. For at undgå eventuelle negative eller positive effekter af listens rækkefølge lavede vi tre forskellige testversioner, som omfattede hver 64 ord. Mellem hvert ord lavede vi en pause på 5 sekunder (noget mere efter meget lange ord), hvilket resulterede i en lydfil med en længde på ca. 7 minutter og 30 sekunder. Testen blev gennemført med i alt 54 elever mellem 16 og 19 år i tre 2. g-klasser på et gymnasium i Aabenraa i foråret 2006.<sup>5</sup> Aabenraa ligger ikke langt fra den tyskdanske grænse, og man må forvente, at eleverne ikke havde noget eller i hvert fald meget lidt kendskab til svensk.<sup>6</sup> Deltagerne fik til opgave at lytte til de svenske ord på cd'en og bruge pauserne til at skrive egne oversættelser af ordene ned på dansk. De blev bedt om altid at prøve at skrive et svar ned, også når de var i tvivl om, hvorvidt deres bud var (helt) rigtigt.

Vi gav ét point for hvert korrekt oversat ord. Fejlagtige eller ikke oversatte ord blev tildelt nul point. Ikke-kognatene på listen blev ikke taget i betragtning. Et så enkelt pointsystem medfører at vi måtte foretage et vist skøn, når det gjaldt besvarelser, som ikke var helt rigtige. For at få et point for korrekt forståelse vurderede vi, om det var tydeligt, at eleven havde forstået, hvilket koncept det drejede sig om. Vi lagde ikke vægt på, om eleven var i stand til at udtrykke sig helt præcist og på (grammatisk) korrekt dansk. Her skal to eksempler gives: Testordet *socka* blev af mange oversat til *sok* (rigtigt tal) eller *sokker* (forkert tal men ellers rigtig betydning), som begge blev godkendt som korrekte svar. Imidlertid var der også nogen, som havde oversat ordet med *sukker*, hvilket må betragtes som et forkert svar, eftersom det drejer sig om et ord, som semantisk betragtet er helt forskelligt. Et andet eksempel er ordet *linne*, der blev korrekt oversat til *linned* og *sengetøj* af nogle testpersoner, mens oversættelser til *linde* betragtedes som forkert.

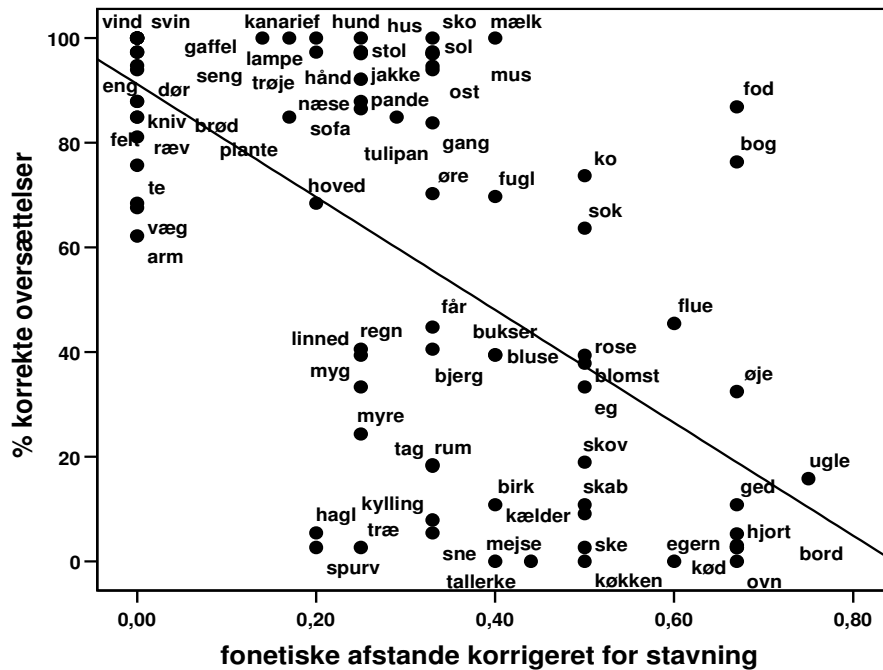
### 3.2 Resultater og korrelationer

Den gennemsnitlige procentdel af korrekte svar pr. ordpar korrelerede vi med afstandsmålingerne for de samme ordpar. Det viste sig, at korrelationen med de

<sup>5</sup> Denne test udgjorde en del af en større test, som skulle undersøge effekten af tilpasset sprogbrug i dansk-svensk sprogforståelse (se Doetjes 2007 for præliminære resultater fra denne undersøgelse).

<sup>6</sup> Den korte afstand til den tyske grænse medførte imidlertid, at forholdsvis mange elever havde gode tyskkundskaber, noget som kan have bidraget til forståelsen af enkelte ord.

fonetiske afstande ikke er særlig høj ( $r = 0,54$ ), men den forøges, når der korreleres med de fonetiske afstande med staveafstande som beskrevet i afsnit 4 ( $r = 0,63$ ). Dette viser, at der er en sammenhæng mellem fonetiske afstande og ordforståelse: jo lavere fonetisk afstand desto bedre forståelse, samt at det er muligt at modellere disse afstande på en måde, som viser, at danskere til en vis grad gør brug af den danske ortografi, når de skal afkode svenske ord. Figur 1 viser et punktdiagram med procentdelen af korrekte svar og de fonetiske afstande korrigeret for stavning pr. ordpar. Ord med en stor fonetisk afstand til det tilsvarende ord på modersmålet forstås generelt dårligt (fx ordene *hjort*, *egern* og *bord*), mens ord som ligner det tilsvarende ord på testpersonernes eget sprog er lettere at forstå (fx *vind*, *svin* og *eng*). Eftersom korrelationen mellem procentdelen af korrekte svar og de fonetiske afstande er forholdsvis lav, er det tydeligt, at vores afstandsberegninger i mange tilfælde ikke er i stand til at forudsige, hvor godt testpersonerne forstår ordene. Fx. forstås ordene *fod* og *bog* bedre, end man skulle forvente ud fra de fonetiske afstande, mens man ud fra de fonetiske afstande ville forvente en bedre forståelse af ordene *hagl* og *spurv*.



Figur 1. Punktdiagram som viser forholdet mellem sprogforståelsesresultaterne i procent på y-aksen og de fonetiske afstande korrigeret for stavning på x-aksen ( $r = 0,63$ ).

#### 4. Forslag til forbedringer af afstandsmålingerne

Formålet med vores undersøgelse var at finde ud af i hvor høj grad ortografien bidrager til danskeres forståelse af svenske ord. Resultaterne som præsenteret i figur 1 viser, at forståelsen af de forskellige ord kun i store linjer kan forklares med de forholdsvis simple beregninger af fonetiske afstande korrigeret for testpersonernes kendskab til modersmålets skriftsprog. Der er mange andre faktorer, som kan gøre nogle ord sværere at forstå, end man skulle forvente ud fra disse afstande (ordene under regressionslinjen i figur 1). På den anden side er der også faktorer, som kan gøre det lettere for en testperson at finde frem til den korrekte oversættelse af et testord (ordene over linjen). Her vil vi nævne nogle af disse faktorer og give et indtryk af, i hvor høj grad nogle af disse faktorer bidrager til forståelsen. Vi har valgt at inddrage fire faktorer, som i den psycholinguvistiske litteratur har vist sig at være afgørende for sprogforståelsen. Ud fra de undersøgelser, som vi kender, kan vi ikke sige noget om, hvor stor effekten af disse faktorer vil være i praksis. Det er derfor svært at afgøre, hvor meget de forskellige faktorer skal vægtes i afstandsmålingerne. De fonetiske afstande ligger mellem 0,0 og 1,0. Vi har valgt at lade hver af de fire faktorer tælle mellem  $-0,3$  og  $0,3$ , alt efter om faktoren gør det lettere ( $-0,3$ ) eller sværere ( $+0,3$ ) at forstå ordet. Fortegnet er positivt eller negativt alt efter om faktoren forventes at bidrage positivt eller negativt til forståelsen.<sup>7</sup> De fire faktorer er:

##### 1. *Anderledes begyndelseslyd* (+0,3)

Det er rimeligt at forvente, at ordenes begyndelse er vigtigere for ordforståelsen end slutningen, og at lyde, som ikke forekommer i det danske lydinventar, især gør ord svære at forstå, hvis de forekommer i begyndelsen af ordet. Fx er der stor sandsynlighed for, at det svenske ord *kött* er svært at forstå for danskerne, fordi begyndelsesbogstavet ikke udtales som det danske [k] men som [ç].

##### 2. *Andet stavelsesantal* (+0,3)

Forskelle i antallet af stavelser mellem svensk og dansk (fx sv. *hagel* med to stavelser vs. da. *hagl* med en stavelse) udgør sandsynligvis også en hindring i genkendelsen af ord.

<sup>7</sup> Da vores endelige mål er at udarbejde en model for ordforståelsen i nært beslægtede sprog, arbejder vi i øjeblikket på en undersøgelse, hvor vi inddrager et større antal faktorer og ser nærmere på deres mere præcise bidrag til forståelsen.

3. *Naboord* (+0,3)

En faktor, der uden tvivl også har en stor indflydelse på ordforståelsen, er det, vi her vil referere til som *naboord* (Luce & Pisoni 1998). Hermed mener vi ord, som har lige så stor eller større lighed med testordet end den korrekte oversættelse. Fx er der 30 ud af 39 testpersoner, som oversatte det svenske *träd* med *træt* i stedet for det korrekte *træ*, formodentlig fordi *træt* ligner det svenske ord mere end *træ*. Derimod er der ikke nogen ord, der ligner *soffa* mere end det korrekte *sofa*, og derfor er der sandsynligvis ingen testpersoner, der har forvekslet *soffa* med et andet ord. Hvis mere end halvdelen af testpersonerne har oversat et ord med et andet mere lignende ord som i tilfældet med *træt* i stedet for *træ*, tæller det hos os som et ord med *naboord*.

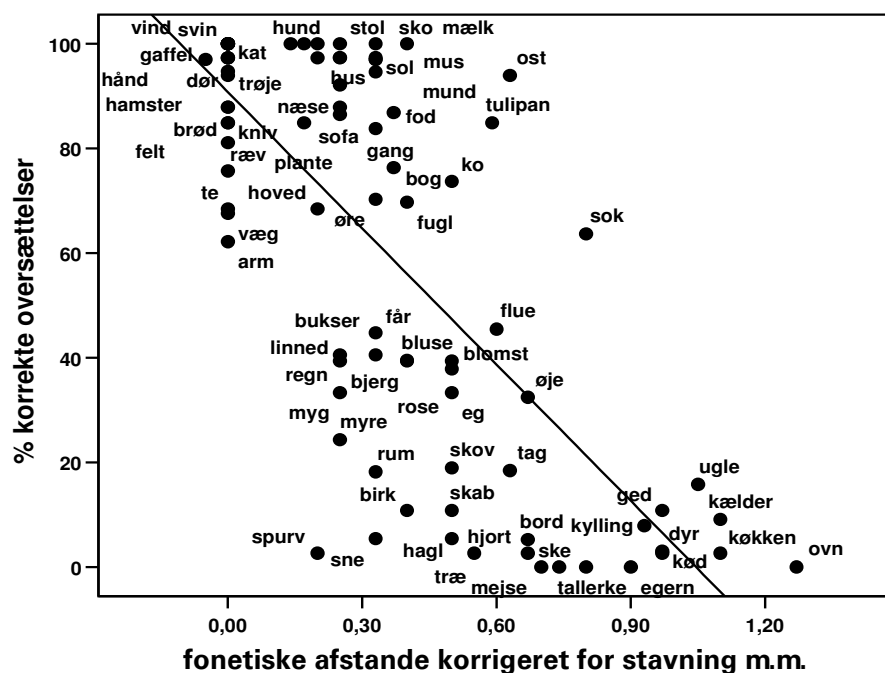
4. *Hjælp fra andre sprog* (-0,3)

Hvis ordet findes på engelsk eller tysk i en form, som ligner testordet mere end den danske, er det muligt, at testpersonerne har nytte af deres kendskab til disse fremmedsprog. Eksempler er sv. *foot* 'fod', der udtales [fø:t], som ligner det engelske ord *foot* [fot] mere end det danske *fod* [fo:ʔð] og sv. *fågel* [fögəl] 'fugl', som ligner det tyske ord *Vogel* [fo:gəl] mere end det danske *fugl* [fu?l].

For hvert ord har vi tilføjet ekstra afstande der, hvor de fire faktorer gør sig gældende. Teoretisk set kan et ords afstand altså øges med maksimalt 0,9 point eller formindskes med højst 0,3 point. Resultaterne af de nye beregninger vises i figur 2. Korrelationen er højere ( $r = 0,74$ ) end uden korrektionerne ( $r = 0,63$ ). Det ser altså ud til at de nævnte faktorer er relevante, når det drejer sig om at forklare sprogforståelsen på ordniveau, selvom korrelationen langt fra er perfekt.

## 5. Konklusion og diskussion

I denne artikel har vi vist, hvordan man kan måle afstande mellem sprog – både på udtaleniveau og på skriftsprogsniveau – med den såkaldte Levenshtein-algoritme. Vores målinger af afstandene mellem svensk og dansk viste, at afstanden mellem skriftsprogene er mindre end mellem talesprogene. Dette stemmer overens med den kendsgerning, at danskere og svenskere i forskellige



Figur 2. Punktdiagram som viser forholdet mellem sprogforståelsesresultaterne og de fonetiske afstande korrigeret for stavning samt fire andre faktorer, se tekst ( $r = 0,74$ ).

undersøgelser har vist sig at have lettere ved at forstå hinandens sprog, når de læser det, end når de hører det. De samme undersøgelser har også vist, at den dansk-svenske talesprogsforståelse er asymmetrisk. Danskere forstår bedre svensk end omvendt. Det er sandsynligt, at denne asymmetri bl.a. kan forklares med, at danskere i højere grad har støtte fra deres eget skriftsprog end svenskere, når de skal forstå det talte nabosprog. Det danske talesprog har nemlig udviklet sig i et hurtigere tempo end det svenske, mens begge skriftsprog er mere konservative. I denne artikel udviklede vi en variant af Levenshtein-algoritmen, som kan modellere dette forhold. Disse nye afstandsmålinger bekræftede, at danskerne i hvert fald teoretisk set har større hjælp fra deres eget skriftsprog end svenskerne, når de skal forstå nabosproget. Et ordforståelseseksperiment viste, at modellen til en vis grad stemmer overens med virkeligheden. Det er dog tydeligt, at der også er andre faktorer, der spiller en vigtig rolle for sprogforståelsen, og som derfor bør inkorporeres i algoritmen. Vi opnåede en højere korrelation med sprogforståelsesresultaterne, når vi korrigerede afstandene for de følgende fire faktorer: anderledes begyndelseslyd, andet stavelsesantal, naboord og hjælp fra andre sprog.



Ideelt set burde sådanne faktorer altså inkorporeres i algoritmen, som beregner afstandene. Det vil dog kræve en del eksperimentel forskning samt afstands-målinger baseret på et større antal ord at finde frem til, hvordan disse faktorer skal vejes i forhold til hinanden. Desuden er det heller ikke tydeligt, hvilke faktorer det drejer sig om. Ud over de fire nævnte faktorer er der flere forhold, som kan forventes at spille en vigtig rolle. Fx ser det ud til, at konsonanter er vigtigere for sprogforståelsen end vokaler. Beijering (2007) testede danskeres forståelse af forskellige nordiske dialekter og fandt højere korrelationer mellem sprogforståelsen og konsonantafstande end mellem sprogforståelsen og vokalafstande. Konsonantforskelle bør derfor givetvis vægtes tungere end vokaler.

Ligeledes vil algoritmen sandsynligvis forbedres, hvis gradvise forskelle mellem de forskellige lyde medregnes. Nærmere undersøgelse af figur 2 giver fx anledning til at formode, at danskere ikke har svært ved at interpretare ord, som udtales med [u] på svensk men med [o] på dansk. Dette kan sandsynligvis forklares med, at de to lyde ligner hinanden så meget, at danskerne tror, de hører et /o/, når en svensker udtaler et /u/. Disner (1983 s. 59) viste, at der er stor fonetisk overlap mellem dansk /o/ og svensk /u/. Således vil en algoritme, som tager hensyn til, at nogle lyde ligner hinanden mere end andre, formodentlig være bedre i stand til at forudsige ordforståelsen.

## Litteratur

- Barödal, Jóhanna, Jörgensen, Nils, Larsen, Gorm & Martinussen, Bente, 1997: Nordiska. Våra språk förr och nu. Lund.
- Beijering, Karin, 2007: The role of phonetic and lexical distances in the intelligibility and perception of Scandinavian language varieties for speakers of standard Danish. MA-thesis, Rijksuniversiteit Groningen.
- Birch-Jensen, Jörgen, 2007: Från rista till chatta. Svenska språkets historia. Göteborg.
- Braunmüller, Kurt, 1999: Die skandinavischen Sprachen im Überblick. Tübingen.
- Brink, Lars & Lund, Jørn, 1975: Dansk rigsmål: lydudviklingen siden 1840 med særligt henblik på sociolekterne i København. København.
- Brink, Lars, Lund, Jørn, Heger, Steffen & Jörgensen, J. Normann, 1991: Den store danske udtaleordbog. København.
- Bø, Inge, 1978: Ungdom og naboland. (Rogalandsforskning rapport 4.) Stavanger.
- Börestam Uhlmann, Ulla, 1991: Språkmöten och mötesspråk i Norden. Nordisk språksekretariat, Oslo.
- Delsing, Lars-Olof & Lundin Åkesson, Katarina, 2005: Håller språket ihop Norden? En forskningsrapport om ungdomars förståelse av danska, svenska och norska. Nordisk ministerråd, København.
- Disner, Sandra Ferrari, 1983: Vowel quality: The relation between universal and language-specific factors. University of California, Los Angeles.

- Doetjes, Gerard, 2007: Understanding differences in inter-Scandinavian language understanding. I: Receptive multilingualism, ed. by Jan ten Thije & Ludger Zeeväert. Hamburg. S. 217–230.
- Gooskens, Charlotte, 2006: Linguistic and extra-linguistic predictors of Inter-Scandinavian intelligibility. I: Linguistics in the Netherlands 23, ed. by Jeroen van de Weijer & Bettelou Los. Amsterdam. S. 101–113.
- Gooskens, Charlotte, 2007a: Internordisk sprogforståelse i et dialektperspektiv. I: 11. Møde om Udforskningen af Dansk Sprog, red. af Peter Widell & Ulf Dalvad Berthelsen. Århus. S. 170–180.
- Gooskens, Charlotte, 2007b: The contribution of linguistic factors to the intelligibility of closely related languages. I: Journal of multilingual and multicultural development 28 (6). S. 445–467.
- Gooskens, Charlotte & van Bezooijen, Renée, 2008: Linguistic distances between Flemish and Norwegian around 1900: subjective views and objective measurements. I: The Beloved Mother tongue: Ethnolinguistic Nationalism in Small Nations. Inventories and Reflections, ed. by Hans Vandevoorde, Petra Broomans & Goffe Jensma. (Groningen Studies in Cultural Change.) Leuven.
- Gooskens, Charlotte & Heeringa, Wilbert, 2004: Perceptive evaluation of Levenshtein dialect distance measurements using Norwegian dialect data. I: Language Variation and Change 16 (3). S. 189–207.
- Grønnum, Nina, 1998: Fonetik og fonologi. Almen og dansk. København.
- Haugen, Einar, 1953: Nordiske språkproblemer – en opinionsundersøkelse. I: Nordisk tidskrift for vetenskap, konst och industri 29. S. 225–249.
- Hedelin, Per & Elert, Claes-Christian, 1997: Norstedts svenska uttalslexikon. Stockholm.
- Heeringa, Wilbert, 2004: Measuring dialect pronunciation differences using Levenshtein distances. (Grodil = Groningen dissertations in linguistics.) Groningen.
- Kessler, Brett, 1995: Computational dialectology in Irish Gaelic. I: Proceedings of the 7th Conference of the European Chapter of the Association for Computational Linguistics. EACL, Dublin. S. 60–67.
- Kruskal, Joseph B., 1999: An overview of sequence comparison. I: Sanko, David & Kruskal, Joseph B. (eds.), Time warps, string edits, and macromolecules. The Theory and Practice of Sequence Comparison. CSLI, Stanford. S. 1–44.
- Luce, Paul A. & Pisoni, David B., 1998: Recognizing spoken words: The neighborhood activation model. I: Ear and Hearing 19 (1). S. 1–36.
- Maurud, Øyvind, 1976: Nabospråksforståelse i Skandinavia: en undersøkelse om gjensidig forståelse av tale- og skriftspråk i Danmark, Norge og Sverige. Stockholm.
- Nerbonne, John & Siedle, Christine, 2005: Dialektklassifikation auf der Grundlage aggregierter Ausspracheunterschiede. I: Zeitschrift für Dialektologie und Linguistik 72 (2). S. 129–147.
- Nordiska rådet, 1973: Den nordiska allmänheten och det nordiska samarbetet 1973. Attitydundersökning. (Nordisk utredningsserie, 4/73.) Stockholm.
- Rekdal, Olaug, 1981: Frekvensordlister i nabospråksundervisningen. I: Internordisk språkförståelse, red. av Claes-Christian Elert. Umeå universitet. S. 179–203
- Teleman, Ulf, 1980: Om den dansk/svenska hörförståelsens betingelser. I: Nysvenska studier 59/60. S. 268–280.
- Wessén, Elias, 1965: Svensk språkhistoria. Ljudlära och ordböjningslära. Stockholm.

Appendiks: oversigt over de 86 kognater som blev brugt i analysen

<b>dansk</b>	<b>svensk</b>	<b>dansk</b>	<b>svensk</b>
køkken	kök	kanariefugl	kanariefågel
hus	hus	plante	planta
bog	bok	brød	bröd
kat	katt	dyr	djur
ovn	ugn	tulipan	tulpan
jakke	jacka	næse	näsa
egern	ekorre	sol	sol
pande	panna	bjerg	berg
sko	sko	lampe	lampa
fod	fot	vind	vind
får	får	kanin	kanin
skorsten	skorsten	mælk	mjölk
hoved	huvud	dør	dörr
kylling	kyckling	mus	mus
eng	äng	øje	öga
ugle	uggla	bi	bi
tallerken	tallrik	hest	häst
spurv	sparv	skov	skog
hjort	hjort	gaffel	gaffel
tag	tak	ged	get
træ	träd	mund	mun
bluse	blus	stol	stol
kød	kött	finger	finger
ko	ko	myre	myra
te	te	skab	skåp
ske	sked	arm	arm
bord	bord	seng	säng
mejse	mes	ben	ben
fugl	fågel	sofa	soffa
myg	mygga	regn	regn
svin	svin	birk	björk
ost	ost	rose	ros
hånd	hand	trøje	tröja
felt	fält	væg	vägg
linned	linne	hagl	hagel
sok	socka	øre	öra
blomst	blomma	ræv	räv
flue	fluga	sne	snö
kælder	källare	gang	gång
eg	ek	fisk	fisk
rum	rum	hund	hund
hamster	hamster	gris	gris
kniv	kniv	bukser	byxor