

*Inger Schoonderbeek Hansen, Tina Thode Hougaard og Peter Widell (udg.):
14. Møde om Udforskningen af Dansk Sprog, Århus 2012*

Polske danskstuderendes tilegnelse af dansk – præsentation af et igangværende projekt*

Thomas Mathiasen og Mikołaj Sobkowiak
Adam Mickiewicz-Universitetet i Poznań

1. Formålet med projektet

Kontakten mellem Danmark og Polen er blevet intensiveret gennem de seneste år, hvilket bl.a. kommer til udtryk ved det store antal polakker, der er indvandret til Danmark. Der er sket over en fordobling af den polske befolkningsgruppe i Danmark siden EU-udvidelsen i 2004, og polakkerne udgør nu den tredjestørste indvandringsgruppe.¹ Det er nok de færreste af de nytilkomne polakker, der har lært dansk inden afrejsen, selv om dansk sprog og kultur også er i vækst i Polen. Man kan således læse dansk på fire polske universiteter², og herudover er der et utal af sprogskoler, der tilbyder danskundervisning. Danskstudierne på universitetsniveau er endvidere meget populære. Vi har således 8-10 kandidater pr. plads på vores institut, når vi optager 25 nye studerende to ud af tre år. Til trods for det store behov for undervisning af polakker i dansk er der ikke gennemført noget større studie af, hvordan polakker tilegner sig sproget, men det forsøger vi at rette op på med dette projekt.

* Vi er vores fagfællebedømmer, Hans Basbøll, meget taknemmelige for hans gode og hjælpsomme kommentarer. Alle fejl og mangler er vores eget ansvar.

¹ Pr. 1. januar 2012 opholdt der sig 31.720 indvandrere og efterkommere fra Polen i Danmark. I 2004 udgjorde tallet blot 13.025, jf. data fra udlændingestyrelsen. (Tal og fakta på udlændingeområdet 2011:73).

² Blot på to af dem kan man dog læse dansk filologi som hovedfag på BA- og MA-studier, nemlig i Poznań og Gdańsk. De to andre steder kan man alene læse dansk som tilvalgsfag (i Wrocław og Warszawa).

Thomas Mathiasen og Mikolaj Sobkowiak

I oktober 2011 påbegyndte vi en længdeundersøgelse af polske danskstuderendes tilegnelse af dansk, hvor vi følger en gruppe på omkring 15 studerende gennem et treårigt studieforløb (BA-studierne). De studerende læser dansk som hovedfag på vores institut og har enten tysk eller engelsk som bifag, men har ingen (eller så godt som ingen) danskfærdigheder, når de starter på studierne. Når de tre år er omme, behersker de dansk på et middelhøjt niveau på grænsen mellem B2 og C1 i henhold til den fælles europæiske referenceramme for sprog. Vi laver optagelser med de studerende ca. en gang om måneden og kommer efter tre år til at råde over et ganske omfattende vidnesbyrd over de studerendes tilegnelse af dansk. Det vil være muligt at foretage analyser af de studerendes sprogtilegnelse på forskellige sprogstadier, både som gruppe og individuelt. Vi håber således at kunne fastlægge den gennemsnitlige polske danskindlærers intersprog, men har også mulighed for at gennemføre en række casestudies. Vores håb er, at dette korpus og analyserne heraf vil bidrage til et større indblik i polakkers tilegnelse af dansk, og at denne viden kan bruges til at målrette sprogundervisningen i dansk både i Danmark og i udlandet. Vi håber ligeledes at kunne bidrage til udforskningen af andetsprogstilegnelse (*Second Language Acquisition (SLA)*), bl.a. om overgangen fra et sprogligt udviklingstrin til et andet eller virkningen af transfer fra L2/L3 på dansk over tid. Der er desuden tale om det første korpus over indlæreres tilegnelse af dansk som fremmedsprog (et indlærerkorpus eller et *leaner corpus*, som det kaldes på engelsk³), og vi håber, at sprogforskere, der er interesseret i dette område, vil benytte vores data i fremtiden. Vi har nemlig planer om at gøre hele korpusset frit tilgængeligt via Talkbank.org inden for en overskuelig fremtid.

2. Korpussets opbygning

Korpusset kan i overensstemmelse med Grangers typologi over indlærerkorpora defineres som et lille, POS-tagget, akademisk, ikke-engelsk longitudinelt talesprogkorpus med både umiddelbare og forsinkede pædagogiske anvendelsesmuligheder (Granger 2008:261-

³ Der findes, så vidt vi er bekendt, ikke en gængs betegnelse for denne korpusstype på dansk, og vi har ladet os inspirere af den norske terminologi ”innlærerkorpus”.

263). Det tæller på nuværende tidspunkt lidt over 5 timers optagelser, der består af monologer og dialoger, og vi regner med, at det kommer til at indeholde mellem 15 og 20 timers optagelser.⁴ Vi bestræber os på at indsamle data cirka en gang om måneden, hvilket svarer til tre gange i hvert semester.

Korpusdeltagerne udgøres som nævnt af studerende, der blev optaget på BA-studier i dansk som hovedfag i oktober 2011. Omkring 15 studerende har indvilget i at deltage i projektet, hvilket de gør i deres fritid. Til gengæld modtager de udførlig feedback på deres sproglige input, hvilket vi vender tilbage til senere. Der er tale om en ret homogen gruppe: langt de fleste var 19-20 år ved studiestart, og de har alle polsk som L1. De har endvidere engelsk og/eller tysk som L2/L3, hvormed dansk typisk er det tredje eller fjerde sprog, som de tilegner sig.

3. Dataindsamling

Optagelserne foretages i et lydtæt studie tilknyttet universitetet⁵, hvor vi har adgang til professionelt studieudstyr. Vi optager ved hjælp af lydbehandlingsprogrammet Audacity og gemmer efterfølgende filerne som mp3.

I forbindelse med dataindsamlingen benytter vi os af det princip, som Ellis og Barkhuizen har kaldt for ”klinisk frembragte almindelige prøver” (*”clinically elicited general samples”*) (Ellis & Barkhuizen 2005:30 ff.). I overensstemmelse med denne metode får deltagerne stillet nogle opgaver, som de skal løse ved hjælp af deres kommunikationsfærdigheder. Det er således vigtigere at få et budskab igennem, end at alle sproglige former er korrekte. Opgaverne er ganske vist konstruerede, men omhandler situationer, som man kan møde til daglig. I overensstemmelse med Ellis og Barkhuizen (ibidem:48) mener vi, at denne dataindsamlingsform minder mest muligt om naturlig forekommende sprogbrug, eftersom deltagerne anvender sproget som kommunikationsredskab. Det er naturligvis lidt svært at tale om naturlig, spontan tale hos vores studerende,

⁴ Til sammenligning tæller det danske talesprogkorpus, DanPASS, knap ti timers optagelser. (Grønnum 2009:596)

⁵ Center for Speech and Language Processing, Adam Mickiewicz-Universitetet i Poznań.

eftersom de sidder i et studie og er pinligt bevidste om, at to af deres lærere optager dem og sidenhen vil analysere deres tale. Ikke desto mindre er dette vist det tætteste, man kan komme på indsamling af spontan tale til et indlærerkorpus.

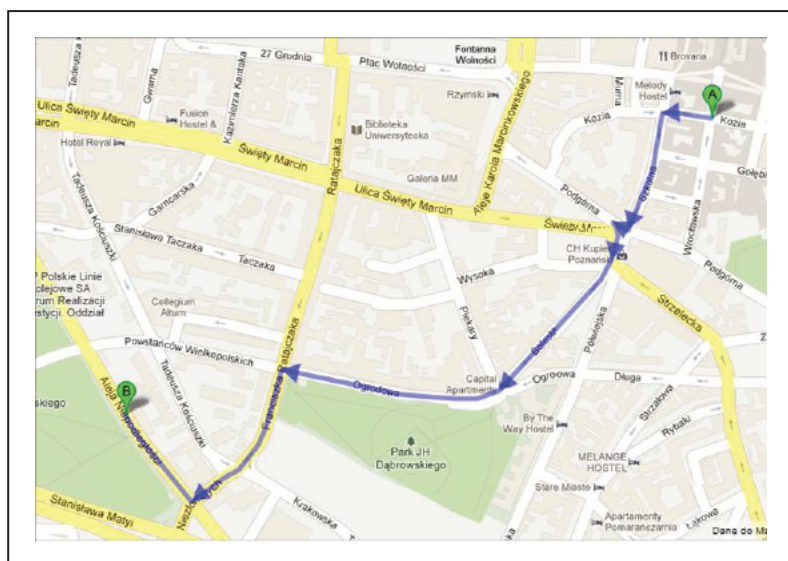
Forud for hver optagelse overvejer projektlederne, hvilken type data, der skal være fokus på. Vi designer herefter nogle opgaver, der sætter nogle situationer i scene, hvor korpusdeltagerne må forventes at producere de ønskede data. Vi forsøger således at skabe en kontekst, der skal stimulere deltagerne til at bruge bestemte konstruktioner eller tale om et bestemt emne uden at angive, at de fx skal anvende datidsformer. Opgaverne har fokus på områder, som vi af erfaring ved, skaber de største problemer for de studerende, hvilket især vedrører sproglige forskelle mellem polsk og dansk.

Vi har fx tilrettelagt flere opgaver, der omhandler udtryk for rum- og retningsopfattelse, eftersom der er stor forskel på, hvordan dette udtrykkes på de to sprog. De polske studerende har især vanskeligheder med dobbelte præpositionsforbindelser såsom *inden i, nede under* eller *oppe over*, og de har endnu sværere ved at tilegne sig konstruktioner med dynamiske eller statiske retningsadverbier efterfulgt af præposition, såsom *han går ned på gaden* vs. *han går nede på gaden* eller *hun løber ud i haven/hun løber ude i haven*. I mange tilfælde kan retningsadverbiet udelades på dansk uden nogen større betydningsforskel, hvilket især gælder de statiske retningsadverbier, der sjældent udtales med tryk. Brugen af retningsadverbier fortæller imidlertid noget om, hvordan taleren opfatter "relationen til den person eller det sted der tales om" (Hovmark 2009:7). Denne opfattelse af rum og retning kan ikke udtrykkes på samme måde på polsk, og det er en af grundene til, at polakker, der lærer dansk, ofte laver fejl på dette område.⁶

Vi har bl.a. undersøgt korpusdeltagernes rum- og retningsopfattelse i forbindelse med nedenstående opgave, som deltagerne har løst parvist, hvilket i øvrigt gælder for langt de fleste optagelser indeholdt i korpusset. Vi bad således deltagerne om at vise vej til forskellige steder, som vi på forhånd havde angivet på et Google-kort over Poznan.

⁶ Se også Szymańska (2011) om polske indlæreres rumopfattelse i forbindelse med tilegnelse af norsk som L2.

Polske danskstuderendes tilegnelse af dansk ...



I en anden opgave fik deltagerne udleveret et billede af en stue, hvor der befandt sig en række genstande, og vi bad deltagerne om at stille hinanden spørgsmål om genstandenes placering. Vi har ligeledes gentaget en af de opgaver, der er blevet anvendt i forbindelse med optagelserne til DanPASS, der vedrører vejvisning gennem en imaginær by ved navn Slotsby (Grønnum 2009:601). Deltagerne i DanPASS-optagelserne har dansk som L1, og vi har gentaget opgaven med vores korpusdeltagere for direkte at kunne sammenligne rum- og retningsopfattelsen hos de to deltagergrupper.

I forbindelse med hver optagegang beder vi korpusdeltagerne om at løse to opgaver. De lægger ud med en opvarmningsøvelse, der typisk består af et rollespil. I forbindelse med den allerførste optagelse, der fandt sted på et tidspunkt, hvor de studerende kun havde studeret dansk i en måned, lavede de således en dialog baseret på nogle retningslinjer (bl.a. på polsk), der i dette tilfælde var ret detaljerede for at hjælpe de studerende på vej.

Thomas Mathiasen og Mikolaj Sobkowiak

PERSON A	PERSON B
Du ringer til Helle. Hun er ikke hjemme. Du spørger, hvad hun laver/hvor hun er henne samt hvornår hun kommer hjem. Den person, du taler med, foreslår at give dig Helles mobilnummer. Du får skrevet numret ned. Du takker pænt og siger farvel.	Du hedder Morten. Du er Helles bror. Telefonen ringer. Du tager den. Det er en, der gerne vil tale med Helle, som ikke er hjemme. Du ved ikke, hvornår hun kommer hjem. Du giver vedkommende Helles mobilnummer.
Dzwonisz do Helle. Nie ma jej w domu. Pytasz, co robi/gdzie jest i kiedy wróci. Rozmówca proponuje Ci podanie numeru tel. kom. Helle. Zapisujesz go. Dziękujesz za pomoc i żegnasz się.	Masz na imię Morten. Jesteś bratem Helle. Dzwoni telefon. Odbierasz. Ktoś chce rozmawiać z Helle, której nie ma w domu. Nie wiesz, kiedy wróci. Proponujesz podać jej numer tel. kom.

Den indledende opgave kan også udgøres af en billedbeskrivelse eller korte monologer/samtaler om forskellige dagligdagsemner, fx hvad de studerende har lavet i weekenden.

Efter opvarmningsrunden rettes blikket mod en fokusopgave, fx de tre nævnte om rum og retning. I forbindelse med en anden optagelse ønskede vi at undersøge korpusdeltagernes tilegnelse af datidsformer. De fik her udleveret et ark med en række tegninger, der beskrev en persons gøremål i løbet af en dag med tilhørende tidsangivelser, hvorefter deltagerne skulle fortælle, hvad personen lavede den pågældende dag, samt hvornår det skete. Vi har desuden undersøgt, hvordan de studerende danner spørgsmål på dansk med samtidig beskrivelse af personers udseende. Her bad vi korpusdeltagerne om at spille en polsk udgave af spillet ”Hvem er hvem”, hvor begge spillere har 24 billeder af en række tegnede personer foran sig. Spillerne udvælger på skift en af personerne uden at røbe, hvem det er, hvorefter den anden spiller via en række spørgsmål skal forsøge at finde frem til den valgte person.

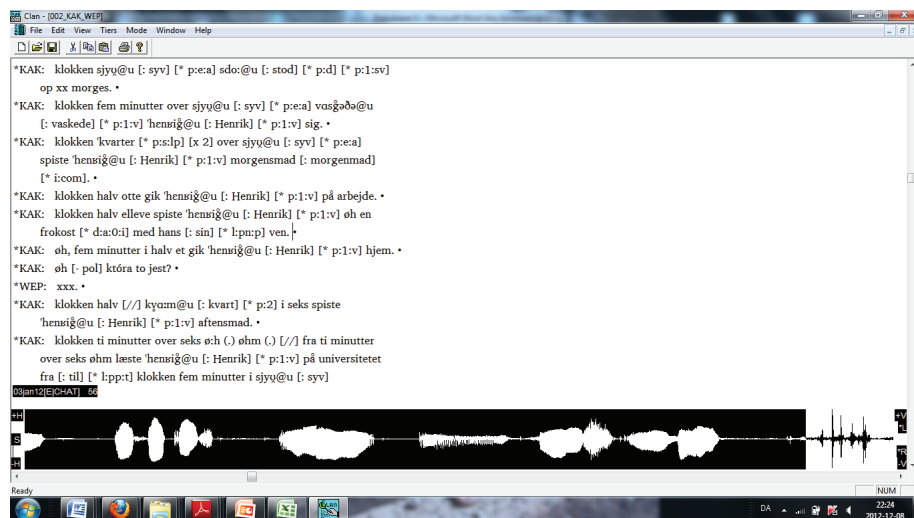
4. Databehandling

Mellem optageperioderne bestræber vi os på at transskribere indholdet af de foregående optagelser. Vi udfører transskriptionerne i CLAN i overensstemmelse med CHAT-formatet (MacWhinney 2012a+b). På lang sigt er vi desuden med til at videreudarbejde en såkaldt ”MOR grammar” (*morphosyntactic coding*), en POS-tagging af dansk, som kan anvendes i CLAN.⁷

⁷ Brian McWhinney har skabt en basisudgave af MOR-grammatikken med de mest anvendte ord på dansk og bestræber sig på at videreudvikle den (personlig kommunikation).

Polske danskstuderendes tilegnelse af dansk ...

Vi markerer fejl i transskriptionerne ved hjælp af et fejl-kodesystem, som vi selv har udarbejdet. Vi har omkring 60 koder på nuværende tidspunkt, men videreudvikler løbende fejlkodesystemet, i takt med at vi foretager transskriptionerne. Fejlkodeerne dækker mange områder og vedrører bl.a. fonologiske (segmentale og prosodiske), grammatiske, leksikalske og syntaktiske fejl samt markering af fejl i bestemte sætningskonstruktioner såsom formulering af spørgsmål eller dannelse af relativsætninger. I det følgende ses et uddrag fra en transskription, hvor en af korpusedtagerne er i færd med at beskrive, hvilke handlinger en person foretager sig ud fra en række tegninger. Transskriptionerne er synkroniseret med lyden, hvormed man kan høre lyden af den pågældende passage, når man klikker på et af udsagnene. Nederst i billedet ses en grafisk fremstilling af lyden.



Der forekommer en lang række fejltypen i dette eksempel, især en del udtalemæssige. Korpusedtageren udtaler fx navnet *Henrik* med [i] frem for [e], hvilket markeres med koden [* p:1:v], hvor p'et markerer, at der er tale om en fonologisk fejl (*phonological error*), 1-tallet angiver, at der forekommer én fejl i udtrykket, og v'et, at fejlen

Thomas Mathiasen og Mikolaj Sobkowiak

forekommer i en vokal. I den tredje ytring udtaler deltageren ordet *kvarter* med tryk på første stavelse, hvilket markeres med [* p:s:lp], hvor s'et viser, at der er tale om en trykrelateret fejl (*stress*), l'et angiver, at det vedrører en leksikalsk trykfejl, og p'et at der er tale om det, vi definerer som en polsk trykfejl, eftersom man på polsk så godt som altid sætter tryk på næstsidste stavelse i ord. Dette trykmønster overfører polakker ofte til dansk. I den femte sætning siger deltageren, at den omtalte Henrik spiser *en frokost* med sin ven. Denne fejl er efterfulgt af koden [* d:a:0:i], hvor d'et angiver, at fejlen vedrører bestemthed (*definiteness*), a'et, at fejlen er forbundet med brugen af artikel, 0'et, at der ikke burde forekomme nogen artikel, og i'et viser, at ubestemt form (*indefinite*) blev anvendt.

Vi anvender først og fremmest de mange fejlmarkeringer for at lette fremtidige analyser af de studerendes intersprog. I transskriptionsprogrammet CLAN kan man nemlig foretage søgninger efter de bestemte fejltyper, og det kan gøres i alle mulige konfigurationer: i en enkelt fil for én eller flere korpusdeltagerere, på tværs af en enkelt korpusdeltagers filer over tid eller på tværs af alle filer, der indgår i korpusset for en enkelt optagelse, et enkelt studieår eller hele den 3-årige periode, som vores projekt varer. Hvis man således ønsker at lave en fejlbaseret analyse af korpusdeltagernes tilegnelse af bestemthed, er det nok, at man søger efter alle fejlmarkeringer forbundet med bestemthed, hvorefter programmet udfærdiger et dokument, hvori alle fejl figurerer inddelt efter fejltype. Ved at klikke på de bestemte fejl kan man desuden se og høre de udsagn, hvor fejlene forekommer. Fejlmarkeringerne udgør således et stærkt redskab til at udføre omfattende søgninger, som man kan bruge til at lave store fejlbaserede analyser.

5. Anvendelse af korpusset til undervisningsformål

Indlærerkorpora kan både anvendes direkte og indirekte til undervisningsformål, hvilket bl.a. Römer (2008:112-131) har beskrevet. Begge anvendelsesmuligheder er gældende for vores korpus (Sobkowiak & Mathiasen 2012). Vi anvender korpusset direkte til at lave undervisningsøvelser, der er baseret på korpusdata, de såkaldte

Polske danskstuderendes tilegnelse af dansk ...

datadrevne øvelser. Man kan fx udvælge særligt velegnede fragmenter fra korpusset, der repræsenterer typisk polsk intersprog for dansk som L2 og dermed rette de sprogstuderendes opmærksomhed mod forhold, som de skal være særligt opmærksomme på. Øvelserne kan fx omhandle brugen af førnutid vs. datid, som er et af de punkter, der generelt volder polske danskstuderende vanskeligheder. Øvelserne kan også omhandle brugen af konkrete ord og udtryk, fx ytringsverbene *tro* og *synes*, som polakker ligeledes har visse besværligheder med at tilegne sig. Opgaverne kan ligeledes vedrøre udtaleforhold, og der er således ufattelig mange muligheder for at designe øvelser, der passer til de studerendes behov.⁸

En anden direkte anvendelse af korpusset udgøres af den feedback, vi som udgangspunkt giver korpusdeltagerne efter hver optagelse. Som nævnt transskriberer vi løbende optagelserne og markerer de fejl, som korpusdeltagerne laver. Vi foretager derpå søgninger på hver enkelt deltagers fejl i den givne optagelse ved hjælp af transskriptionsprogrammet CLAN, der udarbejder et enkelt dokument, hvoraf alle fejlene fremgår opdelt efter type og med sætningskontekst. Projektlederne mødes herefter med korpusdeltagerne én for én og omtaler de fejltyper, der forekommer i den pågældende korpusdeltagers intersprog, og deltagerne har efterfølgende selv mulighed for at lytte til de optagelser, som de har medvirket til. De kan her lytte til optagelserne og samtidig følge med i transskriptionerne, der er synkroniseret med hinanden og kører som som en "film". Korpusdeltagerne har udtrykt stor tilfredshed med at modtage denne feedback, der virker motiverende for dem. Nedenfor har vi anført et eksempel på en korpusdeltagers feedbackdokument i forbindelse med en af de første optagelser. Den anførte side omhandler bestemtheds- og bøjningsfejl.

⁸ Hadley (1997 og 2001) anfører bl.a. en række mulige datadrevne øvelser.

Thomas Mathiasen og Mikolaj Sobkowiak

4 [* d:a:d:0]
File "001_MOS_NAW.cha": line 97.
NAW: oh (.) computene [: computeren] [d:n] [* i:n] ohm (.) står på oh (.) bord [: bordet] [* d:a:d:0] .
File "001_MOS_NAW.cha": line 99.
NAW: hvor er telefon [: telefonen] [d:a:d:0] ?
File "001_MOS_NAW.cha": line 141.
NAW: mand [: manden] [d:a:d:0] [- pol] no .
File "001_MOS_NAW.cha": line 173.
NAW: ohm (.) oh (.) jeg løber to oh (.) 'gana [: gange] [l:in] om u [: ugen] [* d:a:d:0] [* l:in] .
2 [* d:n]
File "001_MOS_NAW.cha": line 22.
NAW: oh (.) det bliver ohm (.) syvogtres@u [: syvogtres] [p:l:c] kronen [: kroner] [* d:n] .
File "001_MOS_NAW.cha": line 97.
NAW: oh (.) computene [: computeren] [d:n] [* i:n] ohm (.) står på oh (.) bord [: bordet] [* d:a:d:0] .
2 [* i:n]
File "001_MOS_NAW.cha": line 97.
NAW: oh (.) computene [: computeren] [d:n] [* i:n] ohm (.) står på oh (.) bord [: bordet] [* d:a:d:0] .
File "001_MOS_NAW.cha": line 147.
NAW: oh (.) <computernen> [: computeren] [i:n] [/] computeren oh (.) ligger xxx <mand> [/] manden [- pol] właściwie .

Data fra indlærerkorpora kan desuden anvendes indirekte til at udarbejde studieplaner og undervisningsmaterialer. Som bl.a. Granger (2002:22 ff.), Meunier (2002:123 ff.) og Römer (2008:117 ff.) har gjort opmærksom på, kan studier af indlærersprog bidrage til viden om, hvordan den pågældende indlærergruppe tilegner sig sproget, altså fortælle noget om det fælles intersprog og de udfordringer, som indlærere med en bestemt sprogbaggrund står over for. Via kontrastive analyser af L1 og L2 kan man udpege forskellige sproglige områder, der må forventes at være særlig vanskelige for en given indlærergruppe, og med tiden udvikler sproglærere en forståelse herfor. Studier af indlærersprog kan imidlertid bruges til at teste denne intuitive viden hos sproglærerne (Granger 2002:22). Vi håber således, at de analyser, som vi (og forhåbentlig andre forskere) vil foretage af korpusdeltagernes sprogbrug, vil bidrage til viden om polakkers tilegnelse af dansk samt til den generelle viden om indlæring af dansk som andet- og fremmedsprog.

I 2011 gennemførte artiklens forfattere et pilotstudie forud for søsætningen af dette longitudinelle projekt, der bekræfter vigtigheden af at analysere indlærersprog. Vi undersøgte her, hvordan polske danskstuderende på B1-niveau placerer leksikalsk tryk og sætningstryk i forbindelse med oplæsning og semispontan tale. Vi

Polske danskstuderendes tilegnelse af dansk ...

indsamlede sammenlagt 3 timers optagelser, som blev foretaget i det samme lydstudie, som vi anvender til længdeundersøgelsen. De studerende blev optaget to gange: Den ene gang oplæste de enkeltvis en ukendt tekst fra en undervisningsbog, der svarede til deres sproglige niveau. Efterfølgende blev de optaget i små grupper, hvor de snakkede om forskellige dagligdagsemner, såsom fritidsinteresser, familieforhold, yndlingsfilm m.m. Vi transskriberede derpå oplæsningen og dialogerne samt markerede trykplaceringen. Baseret på vores mangeårige erfaring som undervisere i dansk udtale sammenholdt med en kontrastiv analyse af trykforholdene på polsk og dansk havde vi udpeget fem konstruktioner, hvor vi forventede, de studerende ville have problemer med trykplaceringen, og det omhandlede især tryktab på verber, fx i forbindelse med nøgent objekt (*han spiller 'fodbold*), verbalpartikel (*hun ser 'godt 'ud*), retningsadverbial (*han løber 'ned til 'bageren*). Det viste sig, at vi havde ret i disse forudsigelser, da forsøgsdeltagerne lavede fejl i omkring halvdelen af tilfældene. Til vores overraskelse fandt vi dog ud af, at der forekom mange flere fejl i en helt anden gruppe: De studerende lavede rigtig mange leksikalske fejl, og denne fejlkategori var den største både i forbindelse med oplæsning (32 %) og semispontan tale (14 %), og en stor del af fejlene lod sig ikke forklare som følge af transfer fra polsk eller tidligere tilegnede fremmedsprog.⁹

Pilotstudiet viste, at indlærersprog kan indeholde uforudsete elementer, og at det er vigtigt at analysere sprogbrugen for at kunne tilpasse undervisningen præcist til målgruppen. Dette lille studie har således medført ændringer i udtaleundervisningen, hvor vi er mere opmærksomme på vanskeligheder ved leksikalsk trykplacering. Samtidig bekræftede studiet dog også vigtigheden af at arbejde med sætningsbaseret tryk samt andre prosodiske forhold såsom reduktion af tryksvage stavelser. Alt dette har vi meddraget i et kursus i

⁹ Studiet viste endvidere, at forsøgsdeltagerne lavede mere end dobbelt så mange trykfejl i forbindelse med oplæsning som i semispontan tale. Vi ved dog ikke, om dette skyldes, at deltagerne blev bedt om at læse en ukendt tekst højt, hvor de måske ikke kendte alle ordene. Vi ønsker således at gentage forsøget ved at bede deltagerne om at læse en egenproduceret tekst højt (Mathiasen & Sobkowiak 2011).

Thomas Mathiasen og Mikolaj Sobkowiak

dansk for udlændinge med arbejdstitlen *Den varme kartoffel*¹⁰, som vi har arbejdet på gennem de sidste tre år, og hvori vi ligeledes løbende inddrager viden, som vi erhverver os i forbindelse med det longitudinale studie.

6. Anvendelse af korpusset til forskningsformål

Korpusset har rige anvendelsesmuligheder: Det kan anvendes til formelle sprogvidenskabelige analyser af udtale og grammatik, men kan ligeledes anvendes til undersøgelser af mere sociolingvistisk karakter såsom kodeskift eller konversationsanalyse (CA). Det er nærmest kun den sprogvidenskabelige fantasi, der sætter grænserne.

Vi følger som nævnt en gruppe polske danskstuderende i perioden 2011-2014, og vi foretager omkring 5-6 optagelser om året. Herved får vi omkring 15 optagelser med hver korpusdeltager jævnt fordelt over tre år, der består af både monologer og dialoger. Optagelserne vil være repræsentative for deltagerens sprogbrug på de forskellige tidspunkter, hvilket sikres ved den omtalte indsamlingsmetode med klinisk frembringelse af tale. Vores korpus tilbyder altså en unik mulighed for at analysere en række danskstuderendes intersprog på forskellige stadier ved hjælp af case studies, og samtidig er det muligt at foretage brede studier af polske danskstuderendes tilegnelse af bestemte forhold. Deltager- og datamængden gør det desuden muligt at fremføre visse generaliserede teser om, hvordan polakker tilegner sig dansk. Hos os vil det desuden være muligt at identificere alle deltagerne i analyserne, selv når de omfatter store datamængder. Dette udgør en stor fordel for vores korpus i forhold til mange indenfor SLA, der består af data, der er indsamlet på én gang på et eller flere udviklingstrin, hvorefter alle dataene blandes sammen i en stor bunke og analyseres samlet uden mulighed for identifikation af de enkelte deltagere. Disse korpusundersøgelser har som regel til formål at fortælle noget generelt om den undersøgte gruppes intersprog, men vil indeholde resultater, der peger i flere retninger, eftersom indlæring er en personlig proces. Ingen lærer imidlertid et

¹⁰ Kurset forventes udgivet i Polen i 2013. Det er skrevet på dansk, og den primære målgruppe er studerende, der læser dansk som fremmedsprog. Kurset har almen karakter og kan altså anvendes af studerende med andre modersmål.

Polske danskstuderendes tilegnelse af dansk ...

sprog på samme måde, hvorved hvert intersprog er noget særligt og unikt (Gass & Selinker 2008:50 ff.). Det kan således være svært at finde forklaringer på bestemte uventede eller ulogiske fejl, da det ikke vil fremgå af oplysningerne, om de kun forekommer hos én bestemt deltager, eller om alle talere af et bestemt L1 er fælles om dem.

Vi har planer om at iværksætte en række undersøgelser af korpusdeltagernes intersprog på forskellige udviklingstrin, og vi har især tænkt os at fokusere på nogle grammatiske kategorier. Som tidligere nævnt ønsker vi at undersøge deltagernes rum- og retningsopfattelse og har derfor designet en række øvelser, der fokuserer på dette. Vi har bl.a. til hensigt at sammenligne de polske danskstuderendes opfattelse af rum og retning med danske modersmålstalendes, hvortil vi vil bruge data fra DanPASS.

Vi ønsker endvidere at undersøge korpusdeltagernes tidsopfattelse, herunder brugen af tempus. I forbindelse med dette vil vi også komme ind på deres brug af aspekt på dansk. Hvor vi på dansk har fortidsformerne førnutid, datid og førdatid, har man på polsk kun én datidsform, nemlig præteritum, der som regel består af ordpar med henholdsvis imperfektivt og perfektivt aspekt. Polakker, der lærer dansk, har ofte problemer med valg af tempusformer på dansk, og det skyldes i vid udstrækning disse forskelle.¹¹ Mange studier for tilegnelse af temporalitet har vist, at verbaltyper tegnes i en bestemt rækkefølge, alt efter om der er tale om verber med perfektivt eller imperfektivt aspekt, den såkaldte aspekthypotese (*the Aspect Hypothesis*), der for SLA først blev foreslået af Andersen (1991). Vi ønsker bl.a. at teste denne hypotese mod korpusdeltagernes intersprog.

Bestemthedskategorien er et af de notorisk sværeste at tilegne sig som polak, eftersom denne kategori ikke findes separat på polsk, der ikke besidder artikler. Bestemthed kan i stedet udtrykkes ved hjælp af de substantiviske leds placering i sætninger (tema vs. rema) (Lyons 1999:278). Der er lavet mange analyser af, hvordan sprogbrugere, der ikke har artikler i deres modersmål, tilegner sig

¹¹ Se Sobkowiak (2009) om en kontrastiv analyse af imperfektivitet og perfektivitet på dansk og polsk.

dem på engelsk som L2, og de viser overordnet, at de mest hyppige fejl omfatter udeladelse på de tidlige udviklingstrin og ombytning på mere avancerede trin, hvilket indebærer en overrepræsentation af bestemte former (Zergollern-Miletić 2010:165). Polske modersmålsbrugeres tilegnelse af danske artikler er ikke tidligere blevet undersøgt, hvilket vi ser frem til at rette op på.

Vi har ligeledes til hensigt at undersøge korpusdeltagernes sætningsopbygning. I modsætning til dansk har polsk en relativt fri ordstilling, og de polske indlærere har især vanskeligheder med placeringen af negationer (specielt i bisætninger) og verballedet. Vi ønsker at undersøge disse tendenser nærmere.

Transfer fra L1 til L2 er et af de områder, der gennem tiderne har fået mest opmærksomhed inden for andet- og fremmedsprogforskningen.¹² I visse tilfælde kan modersmålet støtte indlæringen af nyt sprog, da man med held kan overføre forskellige sproglige strukturer til L2, og man taler her om positiv transfer, men det modsatte kan også være tilfældet, og der er så tale om negativ transfer (Gass & Selinker 2008:94). Der vil oftest forekomme transfer mellem nærtbeslægtede sprog, men meget afhænger dog af, hvordan sprogindlæreren opfatter forbindelsen mellem L1 og L2 (ibid. 141-151). Vi vil gerne undersøge, på hvilke områder deltagerne i vores korpus overfører sprogligt indhold fra polsk til dansk.

Korpusdeltagerne har alle lært minimum ét og ofte flere fremmedsprog, inden de begyndte på danskstudierne, hvormed dansk i de fleste tilfælde er L3 eller L4. Ligesom sprogligt indhold kan overføres fra modersmålet til L2, kan det samme forekomme mellem fremmedsprog. Flere studier har endog vist, at transfer i de tidligere indlæringsstadier i højere grad kan forekomme mellem fremmedsprog end fra L1 til L2. På det leksikalske område er der dog oftest tale om transfer af funktionsord fra et L2, hvorimod mere systematisk transfer stammer fra modersmålet (Gass & Selinker 2008:151-155 og Ringblom 2006:41 ff.). Denne intersproglige transfer udgør yderligere et af de punkter, som vi gerne vil undersøge. Vi har bl.a. planer om at undersøge transfer fra korpusdeltagernes L2/L3 (oftest engelsk og/eller tysk) til dansk på det leksikalske plan over tid. De

¹² Se Gass & Selinker 2008:92-102 og 136-151 om en kort gennemgang heraf.

Polske danskstuderendes tilegnelse af dansk ...

fleste studier vedrører nemlig transfer i de tidlige indlæringsstadier, og transferen er ikke blevet påvist i længdestudier (Gass/Selinker 2008:155).

Vi har som nævnt især til hensigt at undersøge områder, der er forbundet med det grammatiske og leksikalske område. Derudover egner vores korpus sig til undersøgelser af udtaleforhold, og en kort kontrastiv analyse af polsk og dansk (baseret på bl.a. Grønnum 2005, Basbøll 2005 og Gusmann 2007) viser tydeligt, at der er mange interessante forhold, man kan kaste sig over: Polske andet- eller fremmedsprogsstuderende vil ofte have vanskeligheder med det danske vokalsystem, der er meget mere komplekst end det polske, hvor der alene findes 6 orale og 2 nasale vokaler, der er uden distinktiv længde. På dansk er der minimum 15 fuldvokaler, der næsten alle indgår i en længde- og stødmodsætning. Der findes herudover fire ubetonede vokaler [ə ɐ ɪ ʊ], hvoraf de to sidste opstår som assimilationsprodukter. På konsonantområdet er der også en række forskelle: Der findes mange flere stemte konsonanter på polsk end på dansk, og flere af dem, der er stemte på polsk, er ustemte på dansk, hvilket fx gælder for /bdg/. Artikulationsstederne af konsonanterne afviger ligeledes fra hinanden. Hvor /tdns/ er alveolære på dansk, er de tilsvarende lyde dentale på polsk. Aspiration er endvidere ikke en kategori, der findes i det polske konsonantsystem. Derudover findes der på dansk nogle konsonantlyde, der slet ikke har ækvivalenter på polsk, fx det bløde d [ð], det reducerede r [ɐ̯] og det danske drøbel-r [ɾ].

Ud over forskellene på det segmentale plan, der kan medføre problemer for udtalen af de danske lyde eller transfer fra polsk, findes der også store forskelle på det prosodiske plan, og den danske prosodi udgør vitterlig en stor udfordring for polske indlærere. Stødet udgør en helt ukendt kategori for polakkerne, og selv om de med tiden lærer at producere lyden i enkeltord i forbindelse med udtaleundervisning, er det de færreste, der lærer at tale med noget, der minder om naturligt stød. De danske trykforhold forårsager ligeledes store vanskeligheder for de polske modersmålsbrugere, og det gælder såvel leksikalsk tryk som sætningsbaseret tryk. Derudover kommer de mange reduktioner og assimilationsprocesser, der finder

Thomas Mathiasen og Mikolaj Sobkowiak

sted på dansk med almindelig talehastighed. Dette er ligeledes et af de områder, der skaber de største problemer for polske danskindlærere, ligeledes på avancerede sprog niveauer.¹³

6. Planer for den nære fremtid

Projektet har på skrivetidspunktet (december 2012) snart kørt i halvandet år, og vi er således omkring halvvejs med dataindsamlingen. Gennem de næste tre semestre vil vi fortsætte med de månedlige studieoptagelser og løbende udføre transskriptioner, bl.a. med henblik på at yde feedback til korpusdeltagerne.

Selv om transskriptionsarbejdet er en tidskrævende proces, håber vi på at få mere tid til at analysere de indsamlede data og nærstudere korpusdeltagerenes intersprog med en række fokusområder såsom bestemthed, tidsopfattelse og tempusbrug, sætningsopbygning, transfer fra L1 og intersproglig transfer samt opfattelse af rum og retning.

Vores håb er, at vi med disse analyser vil medvirke til at skabe større viden om polske modersmålstalendes tilegnelse af dansk, hvilket kan bidrage til en mere målrettet danskundervisning af polakker i Danmark samt uden for landets grænser. Vi håber ligeledes, at vi mere overordnet kan være med til at videreudvikle kendskabet til, hvordan dansk tilegnes som andet- og fremmedsprog, samt bidrage til den allerede omfattende SLA-viden. Inden for få år har vi desuden til hensigt at gøre korpusset frit tilgængeligt på Talkbank.org, og vi håber, at andre sprogforskere vil ønske at anvende dette talesprogs korpus med polske modersmålstalendes tilegnelse af dansk til undervisnings- og forskningsformål.

Litteratur

Andersen, Roger W (1991) Developmental sequences: the emergence of aspect marking in second language acquisition. *Crosscurrents in Second Language Acquisition and Linguistic Theories*. Thom Huebner & Charles A. Ferguson (ed.). Amsterdam: John Benjamins:305-324.

¹³ Se endvidere Jørgensen & Arnfast (2003) om udviklingen af dansk udtalefærdighed hos polakker

Polske danskstuderendes tilegnelse af dansk ...

- Basbøll, Hans (2005) *The Phonology of Danish*. Oxford/New York: Oxford University Press.
- Ellis, Rod & Barkhuizen, Gary (2005) *Analysing Learner Language*. Oxford: Oxford University Press.
- Gass, Susan M. & Selinker, Larry (2008) *Second language acquisition: an introductory course*. New York/Abingdon: Routledge.
- Granger, Sylvaine (2002) A Bird's-eye view of learner corpus research. *Computer Learner Corpora, Second Language Acquisition and Foreign Language Teaching*. Sylviane Granger, Joseph Hung & Stephanie Petch-Tyson (ed.). Amsterdam/Philadelphia: John Benjamins:3-33.
- Granger, Sylvaine (2008) Learner corpora. *Corpus Linguistics. An International Handbook*. Anke Lüdeling & Merja Kytö (ed.). Berlin/New York: Walter de Gruyter:259-275.
- Grønnum, Nina (2005) *Fonetik og fonologi*. København: Akademisk Forlag.
- Grønnum, Nina (2009) A Danish phonetically annotated spontaneous speech corpus (DanPASS). *Speech Communication* 51:594-603.
- Gussmann, Edmund (2007) *The Phonology of Polish*. Oxford/New York: Oxford University Press.
- Hadley, Gregory (1997) Sensing the Winds of Change: An Introduction to Data-driven Learning. Elektronisk udgave: www.nuis.ac.jp/~hadley/publication/windofchange/windsofchange.htm
- Hadley, Gregory (2001) Concordancing in Japanese TEFL: Unlocking the Power of Data-driven Learning. Elektronisk udgave: <http://www.nuis.ac.jp/~hadley/publication/jlearner/jlearner.htm>
- Hovmark, Henrik (2009) Ude og hjemme - om brugen af retningsadverbier i Jylland og på Øerne. *Ord & sag* 29:5-28.
- Jørgensen, Jens Normann & Arnfast, Juni (2003): ['jæme 'gåd mæn 'u:ð 'bæsd]: Om udvikling af dansk udtalefærdighed hos unge med henholdsvis tyrkisk, amerikansk-engelsk og polsk modersmål. *Veje til dansk: Forskning i sprog og sprogtilegnelse*. Anne Holmen, Esther Glahn og Hanne Ruus (ed.). København: Akademisk Forlag:55-89.
- Lyons, Christopher (1999) *Definiteness*. Cambridge: Cambridge University Press.
- MacWhinney, B. (2012a) *The CHILDES Project: Tools for Analyzing Talk. Part 1: The CHAT Transcription Format*. Elektronisk udgave: <http://childes.psy.cmu.edu/manuals/chat.pdf>.
- MacWhinney, B. (2012b) *The CHILDES Project: Tools for Analyzing Talk*.

Thomas Mathiasen og Mikołaj Sobkowiak

- Part 2: *The CLAN Programs*. Elektronisk udgave: <http://childes.psy.cmu.edu/manuals/clan.pdf>.
- Mathiasen, Thomas og Sobkowiak, Mikołaj, *Trykforhold ved oplæsning og semi-spontan tale. En korpusbaseret analyse af polske danskstuderendes sprogbrug på B1-niveau*. Oplæg præsenteret på 20 Arbeitstagung der Skandinavistik (ATdS).
- Mathiasen, Thomas og Sobkowiak, Mikołaj, *Den varme kartoffel, et kursus i dansk udtale for udlændinge* (manuskript).
- Meunier, Fanny (2002) The pedagogical value of native and learner corpora in EFL grammar teaching. *Computer Learner Corpora, Second Language Acquisition and Foreign Language Teaching*. Sylviane Granger, Joseph Hung & Stephanie Petch-Tyson (ed.). Amsterdam/Philadelphia: John Benjamins Publishing Company:119-141.
- Ringblom, Håkan (2006) The Importance of Different Types of Similarity in Transfer Studies. *Cross-linguistic Influences in the Second Language Lexicon*. Janusz Arabski (ed.). Clevedon: Multilingual Matters:36-54.
- Römer, Ute (2008) Corpora and language teaching. *Corpus Linguistics. An International Handbook*. Anke Lüdeling & Merja Kytö (ed.). Berlin/New York: Walter de Gruyter:112-130.
- Sobkowiak, Mikołaj (2009). *Perfektivitet og imperfektivitet i dansk og polsk - et kontrastivt studium i aspekt og aktionsart*. Ph.d.-afhandling fra Adam Mickiewicz-Universitetet i Poznań. Elektronisk udgave: <http://hdl.handle.net/10593/1082>
- Sobkowiak, Mikołaj & Mathiasen, Thomas (2012). A Learner Corpus of Spoken Danish and its Direct and Indirect Applications in Language Teaching. Accepteret konferenceartikel i forbindelse med 10th Teaching and Language Corpora Conference (TaLC).
- Szymańska, Oliwia (2011) Konseptualisering av rommet hos polske innlærere – utfordringer og diagnoser. En korpusbasert studie med norske påstedspreposisjoner i og på i fokus. Ph.d.-afhandling fra Adam Mickiewicz-Universitetet i Poznań. Elektronisk udgave: <http://hdl.handle.net/10593/1269>
- Udlændingestyrelsen (2012) *Tal og fakta på udlændingområdet 2011*.
- Zergollern-Miletić, Lovorka (2010) English Articles Revisited. *Empirical Studies in English Applied Linguistics*. Magdolna Lehmann, Réka Lugossy & József Horváth (ed.). Pécs: Lingua Franca Csoport:165-172.