

Om polske danskstuderendes skriftlige præstation

Mikołaj Sobkowiak

Adam Mickiewicz Universitetet i Poznań

1. Indledning

Denne artikel og den pilotundersøgelse, der skal beskrives her, er på en måde en opfølger til et tidligere MUDS-bidrag, jeg er medforfatter til (Mathiasen & Sobkowiak 2013). Thomas Mathiasen og jeg indledte vores artikel fra 2013 med en konstatering af, at der ikke rigtig fandtes nogen større længdeundersøgelser af, hvordan polakker tilegner sig det danske sprog.¹ Vi understregede også, at behovet for at undervise polakker i dansk var stort både i og uden for Danmarks grænser. Dette, mente vi, skyldtes først og fremmest, at polakkerne på daværende tidspunkt udgjorde den tredjestørste indvandrergruppe i Danmark, samt at interessen for dansk sprog og kultur var steget markant i Polen (jf. Mathiasen & Sobkowiak 2013).

Siden da er polakker blevet til den største indvandrergruppe i Danmark (IID 2016:13), og interessen for danskundervisning på polske universiteter og sprogskoler er mindst lige så stor, som den var dengang. Det, som er forblevet uændret, er, at polakkernes tilegnelse af dansk i høj grad fortsat er uudforsket.² Udforskningen af det danske sprog i Polen har i de sidste år været domineret af morfologi og især orddannelse (fx Stopyra 2008; Rajnik 2011; Szubert 2012). I Polen er der også blevet skrevet omfattende håndbøger i dansk morfologi

¹ Juni Södeberg Arnfasts og Jens Normann Jørgensen har faktisk beskæftiget sig med polakkernes tilegnelse af den danske udtale (Arnfasts & Jørgensen 2003), men dette var ikke en klassisk længdeundersøgelse.

² Analysen af de indsamlede data, som er beskrevet i Mathiasen & Sobkowiak (2013), er til dags dato ikke blevet fuldendt eller offentliggjort.

(Rajnik 1999, 2014) og udtale (Janikowski 1990; Rasmussen 2005), men de er baseret på forfatterens lingvistiske viden og til en vis grad deres didaktiske erfaringer og ikke systematiske undersøgelser af sprogtilegnelsen. På polske universiteter og sprogskoler går der minimum et par dusin andre danskundervisere med flere års erfaring og en hel del viden om polakkernes tilegnelse af dansk, men den viden er indtil videre stort set ikke blevet udmøntet i tilegnelsesundersøgelser gennemført på grundlag af autentiske lingvistiske data produceret af polske indlærere.³ Dette virker som en niche i udforskningen af tilegnelsen af dansk, som gerne skal udfyldes.

2. Det potentielle korpus

Udover behovet for at undersøge polakkernes tilegnelse af dansk er der også potentiale for at gøre det, for data er der rigeligt af. Fx har Adam Mickiewicz Universitetets Institut for Skandinaviske Studier et omfattende arkiv til rådighed med danskstuderendes eksamensopgaver inden for grammatik, ordforråd og skriftlig fremstilling. Opgaverne går tilbage til ca. 1989, og der foreligger i skrivende stund endnu ikke en fuldkommen opgørelse over, hvor mange opgaver der er i arkivet, men der er potentielt tale om 100-120 opgaver for hvert af de fire studieår.⁴

I øjeblikket er alle opgaverne skrevet i hånden, men der er uden tvivl materiale nok til, at arkivet efter digitalisering kan udvikles til et lille, POS-tagget, akademisk, ikke-engelsk longitudinelt skriftsprogskorpus med potentiale for umiddelbar og forsinket pædagogisk anvendelse (jf. Granger 2008:261 ff.; Römer 2008:112 ff.). Korpuset kunne således anvendes til udforskningen af det danske sprogs sprogtilegnelse, men også, direkte og indirekte, i undervisningsprocessen (jf. også Mathiasen & Sobkowiak 2013).

Der er minimum to grunde til, at dataene i det potentielle korpus kunne være spændende at undersøge nærmere. For det første er der tale

³ Ved Adam Mickiewicz Universitetets Institut for Skandinaviske Studier er der dog en uformel forskningsgruppe, der arbejder med polske studerendes præstation på dansk/svensk/norsk i akademisk kontekst.

⁴ Indtil 2016 skulle instituttets magisterstuderende i dansk op til eksamen i praktisk dansk sidst på de første 3 studieår og efter vintersemestret på 5. studieår. Fra 2016 af skal de også op til tilsvarende eksamen sidst på 4. studieår.

om eksamensopgaver fra fire forskellige studieår, og det virker derfor oplagt at opdele dataene i fire indlæringsprofiler. Da de studerende får over 200 timers danskundervisning hvert studieår, kan man forvente en betydelig niveauvariation mellem de enkelte indlærerprofiler og det fulde spektrum af sproglige kompetencer, herunder basissprogbrugere på første år (ca. B1/B2 ifølge CEFR) og meget kompetente sprogbrugere på femte år (ca. C2/C1). Dataene egner sig i teorien udmærket til tværsnitsundersøgelser af sprogtilegnelsen inden for de fire indlæringsprofiler, men de kan også bruges til længdeundersøgelser. For mange af de studerende er der nemlig i arkivet data fra alle fire danskeksamener, og derfor kan man potentielt også følge de enkelte studerendes sproglige udvikling via *case studies*.

Det nævnte arkiv virker meget interessant som potentielt korpusmateriale, men at få det digitaliseret (dvs. indtastet hver opgave for sig og tagget) lyder til gengæld som en meget tidskrævende opgave. Jeg har derfor besluttet mig for at tage en smagsprøve af materialet og gennemføre en pilotundersøgelse for at se, hvad materialet har at tilbyde, og muligvis få overblik over, hvad det er for nogle undersøgelser, korpusset kan bruges til. Jeg har i første omgang valgt at se på de studerendes stile og korte essays, og valget af netop den del af materialet skyldtes, at jeg håbede på at kunne bruge korpusset til i fremtiden at udarbejde en model (eller snarere fire modeller) af indlærernes syntaks på de forskellige niveauer. Denne pilotundersøgelse beskriver jeg i det følgende.

3. Pilotundersøgelsen

3.1. De undersøgte data

Som data til pilotundersøgelsen har jeg valgt opgaver skrevet af første-, tredje- og femteårsstuderende, for det var mellem netop de 3 indlærerprofiler, jeg ud fra mine didaktiske erfaringer havde forventet de tydeligste niveaunkontraster. Jeg har besluttet mig for at se på 10 opgaver pr. indlærerprofil, som sikkert ikke er nok til, at mine resultater kan betragtes som statistisk relevante eller repræsentative, men dette er heller ikke formålet. Jeg gik nemlig i gang med undersøgelsen med den hensigt at finde ud af, om materialet rent faktisk var så interessant, som jeg havde formodet, og få overblik over, hvad jeg potentielt kunne

bruge arkivet til, hvis jeg besluttede mig for at få det digitaliseret og udviklet til et indlæringskorpus.

Som nævnt får de polske danskstuderende i Poznań næsten 300 undervisningstimer pr. akademisk år i starten, hvorefter antallet falder lidt. De studerende, der kommer op til eksamen på femte år, har haft over et tusind undervisningstimer bare i praktisk dansk og skulle ifølge studieordningen have opnået C2/C1-niveau. De fleste af dem har på dette tidspunkt også tilbragt et eller to semestre på et dansk universitet og/eller en dansk højskole. Oplysninger om de undersøgte indlærerp profiler findes i tabellen nedenfor:

Betegnelse	Årgang	Antal undervisningstimer (cirkaværdier)	Niveau ⁵ (CEFR)
DK1	1	270	A2
DK3	3	840	B1/B2
DK5	5	1030	C2/C1

Inden for hver indlærerp profil valgte jeg at se på opgaver fra minimum to forskellige hold. For hver af profilerne var opgavelængden og -typen forskellig og fremgik af studieordningen og de instruktioner, de studerende havde fået under selve eksamenerne. Førsteårsstuderende skulle som regel skrive en historie ud fra et foto, de fik udleveret, mens DK3-studerende typisk skulle skrive et resumé af en avisartikel og kommentere den samt præsentere deres synspunkter om det pågældende emne. 5.-årsstuderende skulle typisk skrive en mere ambitiøs stil om et mere krævende eller omfattende emne, nogle gange ud fra pensummateriale, som de havde fået udleveret et stykke tid inden eksamen. DK1-opgaverne var oftest på ca. 200 ord, mens længden for DK3- og DK5-grupperne typisk var henholdsvis ca. 450 og 650 ord pr. opgave.

3.2 Forventninger og metode

Da jeg selv har en del erfaring som danskunderviser i Polen, havde jeg nogle forventninger til materialets indhold eller overordnede niveau,

⁵ Antagende/ønskede niveauer ifølge den nuværende studieordning.

som jeg gerne ville få be- eller afkræftet. Jeg forventede nemlig, at der ville være en tydelig forskel mellem DK1- og DK3-opgaverne angående teksternes syntaktiske (eller overordnet set grammatiske) kompleksitet og en betydelig mindre analogisk forskel mellem DK3- og DK5-stilene. Jeg regnede samtidig med, at DK1-opgaverne ville indeholde ret mange grammatiske fejl, der skyldes ”ukendskab til normen” (jf. Togeby 1996:9), for det meste bøjnings- og syntaksfejl. Jeg havde ligeledes på fornemmelsen, at DK5-studerende gerne skulle præstere væsentligt bedre end DK3-gruppen, dvs. lave betydelig færre fejl. Dog havde jeg ikke rigtig nogen forventninger angående fejltyperne blandt DK3- og DK5-studerende.

For at kunne sige noget om teksternes overordnede kompleksitet har jeg besluttet mig for at se på deres lixtal samt på antallet af ord pr. sætning og antallet af lange ord pr. sætning, hvor ord på over 6 bogstaver betragtes som lange ord. Derefter var planen at gennemføre en generel fejlanalyse. På den måde ville jeg både finde ud af, hvad de studerende havde stræbt efter (dvs. hvor komplicerede strukturer de forsøgte at anvende i deres opgaver) og, takket være det fejlanalytiske, hvordan de har præsteret.

3.3. Resultater og diskussion

3.3.1. (De misvisende) tal

Resultaterne fra lixberegneren⁶ har vist, at DK1-teksterne var betydeligt simplere end tekster skrevet af både DK3- og DK5-studerende, som så lå i læsbarhedskategorien ”Middel, dagblade og tidsskrifter”. Forskellen i lixtal mellem DK3- og DK5-opgaverne var som forventet væsentlig mindre end mellem DK1- og DK3-teksterne. I tabellen nedenfor sammenlignes gennemsnitslixtallene for de tre indlærerprofiler med lixtal for udvalgte tekster skrevet af indfødte danskere – opgaver skrevet af danske gymnasieelever (1.g) samt et udvalg af artikler fra *Berlingske* (B) og *Ekstra Bladet* (EB)⁷:

⁶ Jeg har brugt Niels Gamborgs online lixberegner (<http://www.nielsgamborg.dk/indhold/lixberegner.htm>).

⁷ Dataene vedrørende presseartiklerne er baseret på Ida Gacs beregninger (jf. Gac 2012).

Profil	Lixtal	Beskrivelse ⁸
DK1	23,3	Let tekst for alle læsere, børnelitteratur
DK3	34,8	Middel, dagblade og tidsskrifter
DK5	37,2	Middel, dagblade og tidsskrifter
1.g	37,7	Middel, dagblade og tidsskrifter
EB	42,2	Middel, dagblade og tidsskrifter
B	46,5	Svær, saglige bøger, populærvidenskabelige værker, akademiske udgivelser

Interessant nok havde de potentielt mest avancerede polske studerendes tekster (DK5) næsten samme lixtal som opgaver skrevet af danske gymnasieelever, men begge gruppers opgaver havde lavere lixtal end både artiklerne fra *Berlingske* og *Ekstra Bladet* (selvom DK3, DK5, 1.g og EB alle sammen lå inden for samme læsbarhedskategori).

Tallene angående antal ord pr. sætning har generelt vist samme tendens – sætningerne i DK1-teksterne var væsentligt kortere end DK3- og DK5-opgaverne, hvor sætningerne i gennemsnit var næsten lige lange, men dog lidt kortere end sætningerne skrevet af henholdsvis danske gymnasieelever og journalisterne på *Ekstra Bladet* og *Berlingske*, jf. tabellen nedenfor:

Profil	Antal ord pr. sætning
DK1	8,72
DK3	14,88
DK5	14,22
1.g	17,60
EB	16,89
B	19,85

Ganske overraskende var den gennemsnitlige sætningslængde en anelse længere i DK3-studerendes opgaver end hos deres ældre med-studerende (DK5), hvis sproglige niveau i princippet burde være højere. Jeg tror dog, at dette resultat kan være et tilfælde, der skyldes,

⁸ Samtlige beskrivelser stammer fra <http://www.nielsgamborg.dk/indhold/lixberegner.htm>.

at den undersøgte prøve er så lille og dermed ikke repræsentativ, og jeg tvivler stærkt på, at dette skulle være den generelle tendens i resten af det potentielle korpusmateriale.

Angående antallet af lange ord pr. sætning har mine beregninger vist, at DK5-studerende havde brugt en anelse færre lange ord pr. sætning end danske gymnasieelever, men dog klart flere end DK3-studerende, mens DK1-studerende brugte færrest lange ord pr. sætning, jf. nedenfor.

Profil	Antal lange ord pr. sætning
DK1	1,25
DK3	2,94
DK5	3,17
1.g	3,28

Dette kriterium bør man dog se på med visse forbehold, da det ikke tager hensyn til de ”lange ords” egentlige længde (lixberegneren betragter alle ord over 6 bogstaver som lange). Jeg mener, at det i fremtidige undersøgelser ville være hensigtsmæssigt også at inkludere gennemsnitslængden på de såkaldte ”lange ord” blandt de relevante kriterier, da det kan vise sig at være en mere troværdig måde at måle en teksts kompleksitet på (og en potentiel forskel mellem DK3- og DK5-studerendes tekster).

Ovenstående metoder er langt fra optimale, og en grundig og omfattende undersøgelse af indlærerteksternes kompleksitet ville kræve flere kriterier, der er mere detaljerede (jf. fx Lus måder at måle syntaktisk kompleksitet på, Lu 2010). Uanset hvor langt de simple metoder, jeg har brugt, ligger fra det optimale, viser de alligevel, at der er væsentlige strukturelle forskelle i de undersøgte eksamensopgaver på tværs af indlærersprofilerne. Disse forskelle og forsøg på at måle dem er yderst interessante i sprogtilegnelsessammenhængen, fordi de kan danne et billede af, hvordan indlærerne forsøger at udtrykke sig på de forskellige niveauer, hvilke strukturer de bruger, samt hvornår og hvordan de gør det.

3.3.2. Fejlanalyse som supplement

Da der er tale om indlærertekster, mener jeg, at det ikke er nok udelukkende at måle de undersøgte teksters kompleksitet på forskellige niveauer, uanset hvor talrige og præcise kriterier man vælger. På den måde danner man kun en del af de pågældende indlæreres sproglige portræt – man finder godt nok ud af, hvilke strukturer indlærerne forsøger at anvende, men det er lige så vigtigt at undersøge, hvorvidt det (mis)lykkes for dem at bruge dem korrekt. Uden undersøgelsens fejlanalytiske del kan man nemt komme til falske konklusioner, jf. eksemplerne nedenfor.

- (1) Ifølge dem, der synes, at det er mennesket, der skruede op for varmen, vil kloden være varmere og varmere, hvis vi ikke holde op med at producere drivhusgasserne.
- (2) Under besættelsen bragte sange mennesker sammen i parker, på gader og gav mulighed for at tro på bedre morgendagen og understrege deres tilknytning til Danmark.
- (3) Jeg kan ikke lide at løbe eller øve intensivt, men jeg har fundet, at svømmebassinet er min yndlings måde til at spille tiden.

Sætningerne (1), (2) og (3) stammer alle sammen fra opgaver skrevet af DK3-studerende, og de ligger alle sammen over gennemsnittet angående lixtal (hhv. 49, 53 og 36), antal ord pr. sætning (hhv. 28, 25 og 23) samt antal lange ord pr. sætning (hhv. 6, 7 og 3) i forhold til den indlærerprofil, de tilhører. Mens de nævnte eksempler udmærker sig tydeligt i forhold til gennemsnittet inden for DK3-profilen med hensyn til sætningernes kompleksitet, er præstationen næppe så god, fordi sætningerne indeholder flere typer fejl, hvor nogle af dem i øvrigt er ganske grundlæggende. I alle eksemplerne kan man finde fejl forbundet med "[b]rug (valg og kombination) af sprogets elementer (morfemer)" (Togeby 1996:9) – i (1) er der tale om r-drop⁹ (*hvis vi ikke holde op*), (2) indeholder klart nok en bestemthedsfejl i nominalfrasen (*bedre morgendagen*), mens man kan konstatere to præpositionsfejl i (3) i frasen *måde til at spille tiden* (dvs. én overflødig og én manglende præposition). I sidstnævnte frase vækker verbet *spille* også mistanke

⁹ De fleste fejlbetegnelser stammer fra Togeby 1996.

af minimum to grunde. Anvendelsen af netop dette verbum synes ikke at give mening i den givne kontekst, og der kan derfor være tale om forkert ordvalg (Togebys GO-fejl). En anden mulighed er, at afsenderen mente verbet *spilde*, og i så fald er vedkommende skyldig i at have begået en stavefejl, der bygger på ”uoverensstemmelse mellem lyd og skrift” (Togeby 1996:9). Derudover er alle de nævnte eksemplers kvalitet lidt tvivlsom med hensyn til ”afsenderens udtryk af sine tanker” (Togeby 1996:9).

DK5-teksternes kompleksitet kan være misvisende på samme måde. Selvom eksempel (4) bestemt ligger over gennemsnittet inden for DK5-profilen både med hensyn til lixtal, antallet af ord og lange ord pr. sætning, er præstationen bestemt ikke tilfredsstillende, ikke mindst på grund af manglende kongruens og muligvis en bestemthedsfejl i frasen *det ”rigtigt” møde*, ledsaget af en stavefejl (manglende *t*) i ordet *bekendskabet* og et manglende komma, jf. nedenfor.

- (4) Nogle gange er det simpelthen fysisk umuligt at kunne mødes i den virkelige verden, men ellers er det ”rigtigt” møde bevis på at personen, vi mødes med er mere engageret i bekendskabet.
- (5) I vores nuværende verden er det umuligt at leve uden Facebook (selvom findes der nogle som prøver at gøre det).

Til gengæld er der i (5) en ordstillingsfejl (*selvom findes der...*), potentielt overflødig brug af parentes (jf. MB-fejl, Togeby 1996:21) og en forkert bestemmer (*vores*) i frasen *i vores nuværende verden*. Sætningen er efter min mening også et udmærket eksempel på, hvad Ole Togeby kalder for ”manglende korrespondens”, og mere præcist en fejltype, han betegner som ”usandhed” (jf. Togeby 1996:48 f.), men det kan være, at jeg fortolker den pågældende sætning en anelse for bogstaveligt.

Formålet med at analysere ovenstående eksempler er at understrege, hvor relevant det er at supplere kvantitative undersøgelser af indlærertekster med fejlanalyse for dermed at kunne danne et mere detaljeret billede af, hvordan det står til med indlærernes sprog-

kundskaber inden for de forskellige profiler. Jeg har ligeledes forsøgt at vise, at denne fremgangsmåde kan vise sig at være et meget effektivt redskab, der kan anvendes til at identificere niveauforskellene mellem indlærerprofilerne.

3.3.3. Forskelle mellem indlærerprofilerne

I dette afsnit vil jeg komme med nogle flere kommentarer til de undersøgte opgaver med håb om at kunne skitsere nogle generelle (og uden tvivl præliminære) profilkarakteristika for indlærergrupperne DK1, DK3 og DK5.

Som nævnt i 3.3.1. er det typisk for tekster skrevet af DK1-studerende, at sætningerne er meget korte (i gennemsnit under 9 ord pr. sætning), og man kan tydeligt mærke den karakteristiske staccatostil, jf. (6) og (7):

- (6) I Leszno var familien i biografen. De så en dansk film. Børnene kunne ikke lide den. De besøgte også mange musser, men børnen kunne dem ikke lide.
- (7) Familien var lykkelig. En dag gik de en tur. De kunne godt lide at gjorde det, fordi børnene syntes, det var meget fedt og sjov at se butikker og anden ting.

De få komplekse sætninger, jeg har truffet i DK1-opgaverne, var som regel sideordnede sætninger med konjunktionerne *og* eller *men*, mens ledsætninger, der forekommer ret sjældent, som regel indledes vha. *at* eller *fordi*.

Udover de korte sætninger er fejlantallet også et af de mest iøjefaldende træk ved uddragene ovenfor. Det er typisk for DK1-opgaverne, at bøjnings-, syntaks- og stavfejl forekommer meget oftere end leksikalske fejl. Dette skyldes sikkert, at de studerendes grammatiske viden sidst på første studieår generelt er ret ustabil, men sjovt nok sker det også ofte, at ordforrådet i DK1-opgaverne er stærkt inspireret af ordforrådet i de øvrige komponenter af samme eksamen.¹⁰

I flere tilfælde er jeg stødt på fejl, som kun skyldes de pågældende studerendes manglende konsekvens i deres valg af bøjningsformer

¹⁰ Den skriftlige fremstilling kommer som regel sidst på eksamensdagen.

eller ortografisk fremstilling af de pågældende morfemer. Det kan man fx se i (6), hvor den bestemte pluralisform *børnene* også forekommer i en variant uden det finale *-e*, og i (7), hvor vedkommende har husket at tilføje det biordsdannende morfem *-t* til adjektivet *fed*, men dog ikke til adjektivet *sjov*, som er sideordnet med *fed*. Infinitivfrasen *at se butikker og anden ting* er til gengæld et eksempel på inkonsekvent brug af pluralisformen inden for frasens objekt. Nogle gange er det svært at bedømme, om der er tale om en stavefejl eller en anden type, fx *musser* i (6). Det kan godt være, at fejlen skyldes uoverensstemmelsen mellem lyd og skrift, men det er også meget sandsynligt, at dette blot er en bøjningsfejl – vedkommende har husket, at et af bogstaverne skal fordobles i pluralisformen, bare ikke hvilket af dem.

Angående syntaksen har jeg lagt mærke til, at fundamentfeltet i DK1-opgaverne sjældent indeholder andet end subjektet, og hvis det gør det, så er det typisk et simpelt adverbial, der bliver placeret først i sætningen. Til gengæld sker der ofte transfer fra modersmålet hvad angår ordstillingen, og det kan man fx se i (6), hvor *dem* i sætningen *men børnen kunne dem ikke lide* er blevet flyttet til positionen direkte efter det finitte verbal. Man kunne tro, at *dem* blot er blevet placeret ukorrekt i letledspositionen, men for mig ligner det en direkte oversættelse fra polsk.

Der er som sagt en markant forskel mellem DK1- og DK3-opgaver angående teksternes kompleksitet. Man kan tydeligt mærke, at 3.-årsstuderende har haft væsentligt mere undervisning og kontakt med det danske sprog. De forsøger klart at kommunikere på et mere abstrakt niveau, og det gør de ved at bruge meget mere avancerede sproglige midler. Det er ikke sjældent, man støder på komplekse sætninger med et lixtal på 50 som (8):

- (8) For at være sikker på, at de fødevarer, man spiser, er sunde, kan man købe ind hos en lokale købmand, købe økologiske- og sæsonvarer og læse etikketerne på produkterne.

Sætningernes struktur er bestemt mere avanceret, og de studerende prøver klart at eksperimentere med ordstillingen og variere de anvendte konstruktioner. Det ser dog ud til, at forsøg på at formidle mere komplekse budskaber ved at anvende mere avancerede sproglige

midler sker på bekostning af korrektheden.¹¹ Bare i (8) har jeg lagt mærke til stavfejl (*etiketterne, økologiske- og sæsonvarer*) og en kongruensfejl (*en lokale købmand*), som jo er ret grundlæggende fejl forbundet med "[b]rug (valg og kombination) af sprogets elementer (morfemer)" (Togeby 1996:9). (8) er langt fra det eneste eksempel af slagsen i det undersøgte materiale. Det sker faktisk ofte, at fejlene virker mere grundlæggende end man skulle tro, og jeg mener netop, at det skyldes, at DK3-studerende fokuserer mest på indholdet og derfor glemmer, at de også helst må skrive korrekt, jf. fx det følgende:

- (9) Nu om stunder tales der meget om en sund livsstil, som har lige blevet meget populær blandt alle generationer i mange lande. Sund livsstilen er særligt forbundet med kost og motion, men også med sådan livsområder som arbejde og familieliv.

I (9) kan man både se simple stavfejl forbundet med lydforholdene (*sådan*), brug af forkert hjælpeverbum i førnutid (*har lige blevet*) og, ligesom det var tilfældet med DK1-studerende, manglende konsekvens (*en sund livsstil* versus *sund livsstilen*).

Eksempel (10) nedenfor indeholder ligeledes nogle ret basale fejl som hyperkorrektions-r (*man skal fokuserer på...*) og manglende præposition i udtrykket *forsyne noget med noget*, men vedkommende kom fx også til at bruge substantivet *antal* forkert som attribut med utællelige substantiver. Leksikalske fejl forekommer hyppigere blandt DK3- end DK1-studerende, og eksempler derpå kan også ses i (11), hvor udtrykket *at stå i spidsen for noget* bruges, ret atypisk, med et ikke-menneskeligt subjekt.

- (10) Det er ikke uden betydning, hvilke fødevarer, man vælger. Man skal fokuserer på fødevarer, som forsyner mad de fornødne ernæringsstoffer frem for varer med et stort antal fedt og sukker.

¹¹ Mine fejlanalytiske bemærkninger skal anses for præliminære. Jeg er klar over, at sådan en fejlanalyse kræver en dybere udforskning af forholdene mellem form og indhold i indlærernes tekster, samt at det sproglige niveau i en given indlærertekst ofte er afhængigt af opgavetypen (jf. fx Bialystok 2001).

- (11) Det er især regelmæssig indtagelse af mad, som står i spidsen for vores sunde livsstil.

Anvendelsen af avancerede sproglige strukturer og avanceret ordforråd samt hyppigheden af forholdsvis basale grammatiske fejl og ret mange leksikalske fejl er efter min mening tegn på, at de studerende generelt gør sig umage for at markere, at de er ved at bevæge sig op ad sprogstien, men samtidig er deres sprogkundskaber ikke stabile nok til, at det lykkes.

I det undersøgte materiale er forskellen mellem DK3- og DK5-teksterne ikke markant angående lixtal samt antal ord og lange ord pr. sætning. DK5-studerende er ikke bange for at danne komplekse sætninger med 3 eller flere ledsætninger og bruge ret avanceret ordforråd, herunder flere sammensætninger end DK3-studerende. Samtidig begår de relativt få grammatiske fejl.

- (12) Formålet med denne opgave er at gennemgå, hvordan de nævnte netværkssider fungerer generelt og som kommunikationsmidler for derefter at sammenligne denne model at kommunikere sig på med ”traditionelle” mellem menneskelige relationer i den ”virkelige” verden.

Som det fremgår af (12), kommer sætningernes kompleksitet nogle gange på bekostning af småfejl som fx den direkte oversættelse fra polsk, hvor fejlen skyldes, at det polske verbum *komunikować się* er refleksivt, mens dets danske formelle ækvivalent ikke er det.

En anden klar tendens blandt DK5-studerende, jeg har lagt mærke til, er at skrive sætninger, der er så lange og komplicerede, at det påvirker deres forståelighed, jf. (13).

- (13) Da selvom i den reele verden, man aldrig kan slette de ord, der er blevet talt (hvilket man kan gøre med de ord der f. eks. er blevet skrevet på Facebook) og man kan ikke bare ”fjerne” dem, man ikke kan lide, så kun de rigtige mennesker – ikke de ”virtuelle” venner kan være en rigtig støtte for os.

Jeg har også på fornemmelsen, at de studerende nogle gange selv opdager, at deres sætninger er svære at forstå, fordi de er for indviklede, og de prøver derfor at gøre dem mere læservenlige eller nemmere at afkode ved at (mis)bruge parenteser, jf. (14):

- (14) Dér kan man vise grænseoverskridende billeder uden risikoen for at blive kritiseret (man kan altid slette uvenlige kommentarer), man kan præsentere til alle sin eneste ene (og understrege, om man er gift eller forlovet), og man kan ryste posen med hensyn til sit liv ved at sige højt noget, som før var en hemmelighed.

DK5-opgaverne kræver sikkert en dybere udforskning og mere eftertanke, men denne indlærerprofil skiller sig i hvert fald tydeligt ud fra de resterende to ved, at de studerende er i stand til at udtrykke sig relativt frit på dansk, også på et højere abstraktionsniveau. Samtidig begår de leksikalske fejl og fejl, der generelt er forbundet med ”sammenhæng (kohærens) i afsenderens udtryk af sine tanker” (Togeby 1996:9) oftere end rent grammatiske fejl, som til gengæld typisk og hyppigt forekommer blandt DK1- og DK3-studerende.

4. Opsummering og konklusioner

Det overordnede formål med denne pilotundersøgelse var at tjekke, hvorvidt en samling af polske studerendes eksamensopgaver har potentiale til at kunne bruges som korpusmateriale til at udarbejde en model af polske studerendes intersprog. Jeg har taget en smagsprøve på de omtalte data og foretaget en meget simpel analyse af teksternes kompleksitet inden for 3 indlærerprofiler (1.-, 3.- og 5.-årsstuderende), som var opfulgt af en generel fejlanalyse. Jeg havde håbet på at finde tydelige sproglige niveauforskelle mellem opgaverne på tværs af profilerne og dermed på at kunne begrunde, hvorfor det giver mening at digitalisere resurserne og udvikle dem til et indlærerkorpus.

Den undersøgte dataprøve var som udgangspunkt for lille til at kunne udarbejde en troværdig model for de tre indlærerprofiler, men undersøgelsens resultater tyder i hvert fald på, at de sproglige niveauforskelle mellem DK1-, DK3- og DK5-studerendes skriftlige

dansk er tydelige og mangfoldige. Jeg har kunnet konstatere, at der er en væsentlig forskel mellem teksternes kompleksitet (lixtal, antal ord pr. sætning og antal lange ord pr. sætning) mellem DK1- og DK3-studerendes men ikke mellem DK3- og DK5-studerendes opgaver. Mens DK1-sætninger var korte, simple og indeholdt mange grammatiske fejl, var DK3- og DK5-teksterne syntaktisk set mere komplekse med længere sætninger, mere varieret ordstilling samt mere avanceret ordforråd. Fejlanalytisk set var forskellen mellem DK3- og DK5-opgaverne den, at de sidstnævnte for det meste indeholdt flere leksikalske fejl og fejl forbundet med kohærens, mens 3.-årsstuderende fortsat lavede mange grammatiske fejl, til trods for at deres tekster lå tæt på DK5-teksterne angående kompleksitet.

Materialet har efter min mening et stort forskningspotentiale, og det kan vise sig at være en god investering at få alle opgaverne digitaliseret og få udviklet materialet til et ægte akademisk, longitudinelt indlærerkorpus. Analyser gennemført ud fra sådan et korpus kunne udfylde en stor og vigtig niche i udforskningen af udlændinges tilegnelse af dansk. Dataene skulle i så fald som minimum POS-tagges, muligvis ved hjælp af Brill-taggerens fordanskede version¹², og fejlkodes.¹³ Et sådant korpus kunne efterfølgende også bruges som træningskorpus for diverse sprogteknologiske værktøjer (ikke mindst værktøjer til automatisk fejlretning) samt i undervisningsprocessen (jf. Römer 2008; Mathiasen & Sobkowiak 2013).

I pilotundersøgelsen bestod min fremgangsmåde i, at jeg både så på opgaverne med fokus på deres syntaktiske kompleksitet og de begåede fejl. Ser man på indlærertekster ud fra de to synspunkter, får man et billede af, hvordan indlærerne forsøger at udtrykke sig, og hvilke (eller hvor komplicerede) sproglige midler de griber til, men samtidig får man også muligheden for at se, hvad de ikke kan, og bedømme deres succesrate. Man kan dermed prøve at identificere deres subjektive sproglige kompetence ved at måle den syntaktiske kompleksitet af, hvad de har skrevet, og det objektive kompetenceniveau ved at se på, hvilke fejltyper de begår.

¹² Værktøjet er tilgængeligt via Københavns Universitets Center for Sprogteknologi (https://cst.dk/online/pos_tagger/).

¹³ Ole Tøgebys (1996) fejlklassificering kunne bruges som udgangspunkt for det fejlanalytiske tagsæt.

En eventuel udarbejdelse af sådan en model ville dog uden tvivl kræve et mere avanceret og omfattende målesystem vedrørende syntaktisk kompleksitet. De tre metoder, som jeg har brugt i pilotundersøgelsen, viste sig lige at være nok til, at jeg kunne skitsere de mest grundlæggende kompleksitetsforskelle på tværs af indlærerprofilerne, men de er ikke tilstrækkelige til at danne grundlag for en omfattende intersprogsmodel. Der er i denne sammenhæng brug for flere og mere detaljerede måder at måle syntaktisk kompleksitet på. I den forbindelse kunne man f. eks. overveje at fordanske Lus 14 syntaktiske kompleksitetsindekser (jf. Lu 2010) eller muligvis anvende et lignende sæt indekser tilpasset det danske sprog specifikt. Her virker det oplagt at inkludere sætningsskemaet – enten dets oprindelige form (jf. Diderichsen 1962) eller en af de mange modificerede versioner deraf. Beskrivelsen af de forskellige kompetenceniveauer kunne også baseres på Pienemanns processability theory (jf. Pienemann 1998, 2005).

Kombinationen af et velgennemtænkt sæt målekriterier for syntaktisk kompleksitet og et solidt fejlanalytisk tagsæt kunne udgøre et udmærket udgangspunkt for udarbejdelsen af en værdifuld og omfattende model af de studerendes intersprog på de tre nævnte niveauer.

Litteratur

- Bialystok, Ellen (2001) *Bilingualism in development. Language, Literacy, and Cognition*. Cambridge: Cambridge University Press.
- Diderichsen, Paul (1962) *Elementær dansk grammatik*. 3. udg. København: Gyldendal.
- Gac, Ida (2012) *Sammenligning af artiklernes læsbarhed i dansk kvalitets- og populærpresse*. Bachelor-projekt, Adam Mickiewicz Universitetet i Poznań.
- Gamborg, Niels = <http://www.nielsgamborg.dk/indhold/lixberegner.htm> (tilgået oktober 2016).
- Granger, Sylvaine (2008) Learner corpora. *Corpus Linguistics. An International Handbook*. Anke Lüdeling & Merja Kytö (udg.). Berlin/New York: Walter de Gruyter:259-275.
- IID = *Indvandrerne i Danmark 2016*. Udgivet af Danmarks Statistik. <http://www.dst.dk/pubfile/20704/indv2016> (tilgået oktober 2016).

- Janikowski, Krzysztof (1990) *Podręcznik wymowy duńskiej*. Wrocław: Wydawnictwo Uniwersytetu Wrocławskiego.
- Lu, Xiaofei (2010) Automatic analysis of syntactic complexity in second language writing. *International Journal of Corpus Linguistics* 15, 4:474-496.
- Mathisen, Thomas & Mikołaj Sobkowiak (2013) Polske danskstuderendes tilegnelse af dansk – præsentation af et igangværende projekt. *14. Møde om Udforskningen af Dansk Sprog, 11.-12. oktober 2012*:271-288.
- Pienemann, Manfred (1998) *Language Processing and Second Language Development. Processability Theory*. Amsterdam/Philadelphia: John Benjamins.
- Pienemann, Manfred (2005) (udg.) *Cross-Linguistic Aspects of Processability Theory*. Amsterdam/New York: John Benjamins.
- Rajnik, Eugeniusz (1999) *Gramatyka języka duńskiego. Morfologia*. Poznań: Wydawnictwo Naukowe UAM.
- Rajnik, Eugeniusz (2011) *Wortbildung des Substantivs im Dänischen – explizite und implizite Derivation*. Poznań: Wydawnictwo Naukowe UAM.
- Rajnik, Eugeniusz (2014) *Podstawy gramatyki języka duńskiego z ćwiczeniami*. 2. udg. Poznań: Wydawnictwo Naukowe UAM. (1. udg. 2006).
- Rasmussen, Jens (2005) *Sig det med takt og tone. Dansk fonetik i teori og praksis*. Poznań: Wydawnictwo Naukowe UAM.
- Römer, Ute (2008) Corpora and language teaching. *Corpus Linguistics. An International Handbook*. Anke Lüdeling & Merja Kytö (udg.). Berlin/New York: Walter de Gruyter:112-130.
- Södeberg Arnfast, Juni & Jens Normann Jørgensen (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*. A. Holmen et al. (udg.). København: Akademisk Forlag: 55-90.
- Stopyra, Janusz (2008) *Nominale Derivation im Deutschen und im Dänischen*. Wrocław: Wydawnictwo Uniwersytetu Wrocławskiego.
- Szubert, Andrzej (2012) *Zur internen Semantik der substantivischen Komposita im Dänischen*. Poznań: Wydawnictwo Naukowe UAM.
- Togeby, Ole (1996) *Stiltræk*. Aarhus: Aarhus Universitet.