

Strukturalismens grundlag i breve

Caroline H.V. Boolsen
Aarhus Universitet

1. Infrastrukturalisme-projektet

På Det Kongelige Bibliotek i København opbevares en stor samling af de danske strukturalisters efterladte arkiver – herunder Louis Hjelmslevs efterladte papirer, som i arkivet fordeler sig over 167 kapsler. Har man kigget i bare et par stykker af disse kapsler, vil man hurtigt forstå, at Hjelmslev var en af de typer, som gemmer absolut alt, for i kapslerne ligger alt fra afsendte og modtagne breve til hurtigt skribede noter, detaljerede forelæsningsplaner, sirligt udformede skemaer og figurer over indviklede teoretiske hypoteser blandet mellem hotelregninger, posthuskvitteringer, julekort og kopier af konferenceprogrammer i seks eksemplarer. Et livs grammatiske grublerier samlet i foldere med overskrifter, underoverskrifter og under-underoverskrifter efter tema og med henvisninger på tværs af folderne i et metodisk, hjemmelavet arkiveringssystem.

Desværre har det indtil nu kun været muligt at se nærmere på Hjelmslev-arkivet, hvis man befandt sig fysisk på Det Kongelige Bibliotek i København, og selv da havde man kun adgang til én kapsel ad gangen for at undgå sammenblanding kapslerne imellem. Der er derfor tale om et sparsomt udnyttet materiale, som ved sin tilgængeliggørelse vil kunne udgøre et væsentligt bidrag til dokumentationen af den særlige variant af strukturalismen, som blev en dansk intellektuel eksportvare. Samtidig vil en offentlig- og tilgængeliggørelse af Louis Hjelmslevs efterladte papirer gøre det muligt at (om)skrive en væsentlig del af dansk og international sprogvidenskabs historie.

Det har længe været et stort ønske hos mange at få gjort disse væsentlige dokumenter bedre tilgængelige. Una Canger, Frans Gregersen og Lorenzo Cigana har i flere år arbejdet på en dokumentation af glossematikens arbejde, og ved hjælp af en stor bevilling fra Carlsbergfondet lykkedes det i 2019 at etablere et projekt med det formål at skabe en forskningsinfrastruktur for den danske strukturalisme. Navnet måtte selvfølgelig blive 'Infrastrukturalisme-projektet', og det har nu løbet lige siden – trods corona og andre udfordringer

Kernen i projektet er Louis Hjelmslevs værker og brevveksling, men også kredsen *omkring* Hjelmslev bliver inddraget i løbet af projektets levetid: Eli Fischer-Jørgensen, Paul Diderichsen, Harry Wett-Frederiksen, Francis J. Whitfield og (vigtigst af alt) H.J. Uldall. Alle er de ligeledes at finde i Det Kongelige Biblioteks arkiver, men i denne artikel vil jeg se bort fra de andre strukturalister og arkiver for en stund for i stedet at fokusere på de 167 kapsler, som rummer Louis Hjelmslevs efterladte papirer – og heraf særligt kapslerne 20 til 43, som indeholder Hjelmslevs korrespondancer. Men først vil jeg forklare, hvorfor det lige netop er Louis Hjelmslev, som spiller hovedrollen i Infrastrukturalisme-projektet.

2. Dansk strukturalisme

I international sammenhæng udmærker strukturalismen sig ved at være den strømning inden for humanioras videnskabshistorie, hvor dansk forskning har haft størst indflydelse, og i udlandet var specifikt den danske strukturalisme nærmest identisk med Louis Hjelmslevs sprogteori. Særligt var den knyttet til værket 'Omkring Sprogteoriens Grundlæggelse' (på dansk 1943; på engelsk ved Whitfield 1952). Derfor er der også en vis international interesse for at gøre Hjelmslevs arkiv tilgængeligt online. En sådan offentliggørelse vil nemlig ikke bare øge vores viden om Hjelmslev selv og hans egne tanker, men også give adgang til hans korrespondancer med nogle af sin samtids største lingvister, herunder bl.a. Roman Jakobson, Antoine Meillet, Nikolai Trubetzkoy, André Martinet og Otto Jespersen.

Ligesom hans korrespondancer er fyldt med navne, som nok vil vække genklang hos enhver med interesse for lingvistikken, så er også navnelisterne over deltagerne til hans forelæsninger en *who's*

who over danske sprogforskere fra generationen, der fulgte efter ham. Forskere, som han har præget ikke bare gennem sin undervisning på Københavns Universitet, men også i kraft af sin rolle som formand for Lingvistikredsen fra dens oprettelse i 1931 og næsten uafbrudt frem til sin død i 1965.

Hjelmslev var altså ikke alene om at udarbejde sin sprogteori, som han bl.a. udviklede i korrespondancerne med tidens andre store hoveder, og han var heller ikke isoleret som strukturalist i sin samtid. Alligevel indtager Hjelmslevs markeringssystemer, kendt blandt folk som 'rugekasserne' eller 'fuglekasserne', og hans ikoniske tandsæt vist alligevel en førsteplads blandt de ting, som de fleste forbinder med dansk strukturalisme – og alene mindet om førstnævnte vil nok kunne fylde mange lingvistikstuderende med gru.

Hvad der måske også ville kunne vække gru hos mange mennesker, er tanken om, hvor man skulle begynde og ende i så stor en mængde materiale, som Infrastrukturalisme-projektet har kastet sig over, så lad os se lidt nærmere på, hvordan vi har valgt at gribe det an. Først vil jeg gennemgå den praktiske del: sortering og udvælgelse af materialet, og derefter vil jeg gå i detaljer med etableringen af den digitale infrastruktur med særligt fokus på, hvordan vi har arbejdet med at gøre denne database over Louis Hjelmslevs efterladte papirer anvendelig for videre forskning.

3. Materialet

Med et omfang på 167 kapsler er der som sagt nok at vælge og vrage imellem i arkivet over Hjelmslevs efterladte papirer, selvom alt materialet fra arkivet selvfølgelig ikke er af lige stor relevans for eftertiden. Derfor har vi også ført arkivet igennem en udvælgelsesproces, hvor hvert dokument er blevet bedømt som værende enten relevant, interessant eller irrelevant for videre forskning og dermed digitalisering. Således er det samlede materiale i Det Kongelige Biblioteks Hjelmslevarkiv før digitalisering blevet skåret ned fra ca. 71.000 enheder i alt til et håndplukket udvalg på ca. 23.000 enheder. Alle bidrager til det samlede billede af, hvor Hjelmslev har været på hvilket tidspunkt i sit liv, hvad han har arbejdet på hvornår, og hvem han har talt med om hvad.

I sidste ende kommer projektets digitale infrastruktur til at indeholde alle mulige forskellige typer af dokumenter fra kvitteringer og huskesedler til breve, billeder og publikationer. Men ud af disse forskellige dokumenttyper er det ikke alt, som skal udsættes for yderligere bearbejdning end en affotografering hos Det Kongelige Biblioteks digitaliseringsafdeling. Det materiale, som kun bliver en del af infrastrukturen i form af billedfiler – fx fotografier, figurer og tekster med tegn, der ikke kan omsættes til computertekst – bliver datasat så præcist som muligt og derefter registreret i den samlede registrant. Den del af Hjelmslevarkivet, som efter digitalisering skal behandles yderligere, omfatter både publicerede og upublicerede værker fra Det Kgl. Biblioteks arkiver og særligt kapslerne med breve til og fra personer i gruppen af danske strukturalister. Heraf er sidstnævnte opdelt i de centrale korrespondancer:

- Louis Hjelmslev – H.J. Uldall
- Louis Hjelmslev – Jens Holt
- Louis Hjelmslev – Eli Fischer-Jørgensen
- Louis Hjelmslev – Paul Diderichsen
- Louis Hjelmslev – Roman Jakobson
- Louis Hjelmslev – L.L. Hammerich
- M.fl.

Derudover inkluderes korrespondancen mellem Eli Fischer-Jørgensen og Paul Diderichsen, som findes i Det Kongelige Biblioteks arkiver, korrespondancen mellem H.J. Uldall og den engelske fonetiker Daniel Jones, som er i projektets varetægt, og korrespondancen mellem Jens Holt og Louis Hjelmslev fra Jens Holts efterladte papirer på Universitetshistorisk Arkiv, Aarhus Universitet.

Efter således at have håndplukket de vigtigste enheder for projektet fra det store arkiv fulgte det næste skridt: at udvælge de vigtigste fokuspunkter i selve teksterne og finde frem til, hvordan de enkelte punkter på bedst mulig vis kan fremstilles digitalt for at fremme en forskningsmæssig brug af den endelige infrastruktur.

4. En søgbar registrant

I tilgængeliggørelsen af Hjelmslevs efterladte papirer er projektet først og fremmest interesseret i at fremvise selve arkivet, det vil sige materialet som faksimiler. Den medfølgende transskription – udført enten manuelt eller ved hjælp af OCR-genkendelse (Optical Character Recognition) – skal ses som et hjælpemiddel. Et forhold, der særligt gør sig gældende ved håndskrevne dokumenter, som kan variere enormt i kvalitet og læsbarhed. Men derudover bidrager den transskriberede tekst gennem opmærkning og annotering også til, at registranten i sidste ende bliver søgbar. Før jeg gennemgår, hvilke kategorier det vil blive muligt at søge efter i den kommende registrant, vil jeg beskrive den overordnede opsætning af den digitale fil og opdelingen af henholdsvis metadata og tekst. Projektets skabelon for opsætningen af den digitale fil er baseret på Hjelmslevarkivets brevkorrespondancer. Det giver nemlig den fordel, at brevene rummer den fulde mulige opsætning – både brødtekst, brevspecifikke elementer og noter – mens alle andre typer dokumenter, som projektet har udvalgt, kun indeholder brødtekst og eventuelt noter og en dato. Alle andre typer dokumenter end breve kan altså blot behandles på samme måde som brevenes brødtekst.

Hvad end der er tale om et brev eller en anden type dokument, indledes den digitale fil med det beskrivende og oplysende metadata om filens digitale tilblivelse, inklusive en stadfæstelse af filens forhold til dens fysiske arkivariske ophav. Metadataet fordeler sig over følgende kategorier:

- En bibliografisk beskrivelse af den digitale fil.
- En 'Relevant for...' -sektion, hvor dokumentet kædes sammen med andre publikationer eller breve med forskningsmæssig relevans, fx på grund af referencer til bestemte teorier eller udgivelser.
- En detaljeret beskrivelse af de ikke-bibliografiske aspekter ved dokumentet (hvilket sprog er teksten skrevet på, hvilken situation blev det produceret i, og hvem var involveret).

- En log over den digitale fils revisionshistorik (hvem har opmærket, korrekturlæst eller foretaget andre ændringer, hvad har de gjort, og hvornår har de gjort det).
- Links til faksimilen/faksimilerne for det pågældende dokument.

Efter dokumentets metadata følger den faktiske tekst. Modellen for den digitale underinddeling af brødteksten er for den digitale brevfil den samme som inddelingen i det fysiske brev og indeholder alle eller flere af følgende punkter:

- Brevhoved (“Aarhus Universitet, Institut for Lingvistik”).
- Information om tid og sted for produktionen af dokumentet (“Aarhus, D. 12/7 1956”).
- Indgangshilsen (“Kære Hjelmslev!”).
- Brødtekst (“De har ikke hørt fra os siden jul. Jeg kunde nemlig ikke sige noget om...”).
- Afslutningshilsen (“Mange hilsener til jer begge, din hengivne Jens Holt”).
- Postscriptum (“P.S. De skulde drikke et glas vin til middag”).
- Afsender-/modtagernoter (“L. gennemlæste ms. i foråret ’63”).
- Redaktørnoter (“Paragraffen er markeret med en lodret streg i venstre margen”).

Ved at markere indgangs- og afslutningshilsner detaljeret for korrespondancerne bliver det muligt at søge på en bestemt person som afsender eller modtager af et brev, sådan at navnet nævnt i forbindelse med den specifikke rolle ’korrespondancepartner’ kan adskilles fra navnet nævnt i alle andre sammenhænge inde i selve brevtteksten (“Kære Prof. Brøndal” vs. “jeg talte med Brøndal i dag”).

Den digitale version af de resterende dokumenttyper, hovedsageligt Hjelmslevs publicerede og upublicerede værker, indeholder som sagt de samme punkter som ovenfor nævnt med undtagelse af de brevspecifikke elementer indgangs- og afslutningshilsener og postscriptum.

Efter dokumentets digitale og fysiske metadata følger så selve brødteksten, hvor den vigtigste gældende præmis er ønsket om at opmærke simpelt og præcist. For her må fremgangsmåden for opmærk-

ningen af det, vi er interesserede i at fremhæve, også nødvendigvis gå hånd i hånd med, hvad vi kan nå at udføre på den tid, som projektet har til rådighed.

I brødteksten er vi interesserede i at fremhæve følgende:

- Personnavne (Louis Hjeltslev og Jens Holt, men ikke 'min svi-germor' eller 'naboen').
- Tid (datoer, måneder, årstal).
- Sted (stednavne, adresser).
- Publikationer (bøger, artikler, tidsskrifter, udgivne manuskripter).
- Fagtermer (specifikke for Hjeltslev, det vil sige 'plerem' og 'glossematik', men ikke 'maskulinum' eller 'pluralis').
- Konferencer og andre (akademiske) begivenheder ('9. nordiske filologmøde, Helsingfors 1939', men ikke 'Lenes 2-års fødselsdag').
- Sprog (med ISO 639-id-koder).
- Understregninger (vi skelner mellem understregning i hånden eller på maskine og dermed sandsynligheden for, om understregningen er foretaget af afsender eller modtager).
- Ikke-stilgende rettelser af slå-/stavefejl (vi retter kun fejl, som forfatteren ikke selv har rettet. Personlig retskrivning som fx *deraf/dærav* eller brug af *å* i stedet for *aa* før 1948 bliver ikke kommenteret. Rettelser af slå-/stavefejl tager så vidt muligt højde for forfatterens eventuelle personlige retskrivning i det pågældende brev, så et ord bliver rettet til den staveform, som det var forfatterens intention at skrive, eller til den staveform, som forfatteren oftest benytter, så fx *dærva* rettes til *dærav*).

5. En projektspecifik opmærkningsmanual

For at sikre overensstemmelse mellem opmærkning og annotation af den digitale registrant har Infrastrukturalisme-projektet udarbejdet en opmærkningsmanual, som er opbygget og tilpasset specifikt til projektets materiale efter anbefalingerne i '*Fælles praksis for TEI-kodning, Version 5 – 14/12-2017*'. Denne skabelon til projektets

egen opmærkningsmanual kom til verden i en kollaboration mellem medarbejdere fra flere af projektets egne samarbejdspartnere: Det Danske Sprog- og Litteraturselskab, Det Kongelige Bibliotek, DigHumLab (Aarhus Universitet), Center for Sprogteknologi (Københavns Universitet) og Grundtvigs Værker (Aarhus Universitet).

'*Fælles praksis for TEI-kodning, Version 5 – 14/12-2017*' er fuldt ud konform med de officielle TEI P5 guidelines og skal forstås som en anbefalet delmængde af TEI P5. Det vil sige, at der ikke er tale om en ny standard for opmærkning, men om en praksis udformet sådan, at det enkelte projekt, som benytter sig af fællespraksissen, uden problemer kan anvende en mere eller mindre avanceret opmærkning end den, som praksissen anbefaler, så længe kodningen er valid og konform i forhold til de officielle TEI P5 guidelines.

På baggrund af denne allerede eksisterende fælles praksis er opmærkningsmanualen for Infrastrukturalisme-projektet således udformet som en grundig vejledning til, hvordan Hjeltslevarkivet skal opmærkes på både det manuskriptbeskrivende og det filologiske plan. For at kunne transformere '*Fælles praksis for TEI-kodning, Version 5 – 14/12-2017*' om til en opmærkningsmanual, som passer perfekt til Infrastrukturalisme-projektet, er projektets version udarbejdet ved at anvende fællespraksissen på dokumenter fra en forsøgskapsel fra Det Kongelige Biblioteks Hjeltslevarkiv.

Kapsel 24, som blev udvalgt til formålet, indeholder hovedsageligt korrespondance mellem Jens Holt i Aarhus og Louis Hjeltslev i Charlottenlund, men der er også enkelte breve mellem Jens Holt og Louis Hjeltslevs enke, Vibeke Hjeltslev, sendt efter Louis Hjeltslevs død i 1965. Derudover udmærkede den sig også ved at indeholde en hel del af den type forhindringer, som vi ville være nødt til at finde opmærkningsløsninger på, fx skemaer, figurer, stavfejl, noter skrevet i margenen, rettelser, postkort med billeder på den ene side, brevkort med tekst på begge sider o.a. Forhindringer, som erfaringen sagde os, at vi ville støde på i både breve og andre typer dokumenter.

Brevene i kapsel 24 var på forhånd blevet transskriberet af en studentermedhjælper på Infrastrukturalisme-projektet. Disse transskriptioner blev indsat i et XML-dokument, som i første omgang var opbygget udelukkende efter fællespraksissens anbefalinger. Fra dette punkt begyndte vi at justere disse anbefalinger fra at gælde

for et bredt spektrum af materiale (romaner, teaterstykker, poesi) til udelukkende at gælde vores eget specifikke materiale, og gennem gradvise justeringer har vi arbejdet os frem til en indkogt version af den originale fællespraksis. Hvad der manglede i fællespraksissens skabelon, har vi enten fundet direkte i de officielle TEI P5-guidelines, ladet os inspirere til ud fra fremgangsmåden i lignende onlinesamlinger eller opfundet en passende løsning på. Alt i samarbejde med Center for Sprogteknologi for at sikre, at vores nye skabelon ville fungere med de systemer, som de udvikler til projektet.

Formålet med at udarbejde en opmærkningsmanual til Infrastrukturisme-projektet ud fra projektets faktiske materiale var, at vi umuligt på forhånd ville kunne forudse alle de problematikker, som vi kunne tænkes at komme ud for. Derimod gjorde arbejdet med kapsel 24, at størstedelen af de virkelige problematikker blev serveret for os helt gratis.

Retningslinjerne i projektets opmærkningsmanual er selvfølgelig konstrueret sådan, at de ikke bare er universelt anvendelige på alle de typer af dokumenter fra Hjælmslevarkivet, som vi er interesserede i at inddrage i den søgbare registrant. Retningslinjerne vil også være universelt anvendelige for alle følgende korrespondancer inden for projektets rammer. Det sikrer, at vi ikke har brug for et nyt sæt opmærkningsregler ved opmærkningen af de resterende strukturalisters arkiver. Derfor måtte opmærkningsprincipperne også afprøves og justeres ved at anvende dem på et andet materiale end Hjælmslevarkivets. Til det formål brugte vi en lille samling af originale og kopierede breve afsendt mellem Daniel Jones og H.J. Uldall fra 1931 til 1946. Dette lille brevkorpus tilføjede løsninger på flere problematikker til manualen, som vi endnu ikke var stødt på i kapsel 24. Brevene er alle skrevet på engelsk og afsendt til og fra meget forskellige steder i verden: Daniel Jones skriver fast fra England, mens H.J. Uldall i denne periode skriver fra så forskellige steder som New York, København Ø, Udby pr. Ørsted, Aarhus, Vedbæk, Tel Aviv, Jerusalem og Buenos Aires. Det gav ikke bare mange nye tilføjelser til vores lister over stednavne og personnavne, men også udfordringer for en ensformig udformning af adressefeltet. Sværest var dog nok, at Daniel Jones og H.J. Uldall, der begge var fonetikere, har lange diskussioner om fejl og mangler i det internationale fonetiske alfabet,

som det så ud i denne periode, og derfor er deres breve spækket med selvopfundne og ofte kryptiske forslag til fonetiske tegn, som de mener dækker denne og hin lyd og udtale bedre end de allerede eksisterende IPA-tegn.

Efter gentagne afprøvninger og justeringer af projektets opmærkningsprincipper følte vi os klar til at bevæge os videre fra at teste små mængder til at kaste os over hovedparten af vores korpus. I første omgang korrespondancerne.

6. Digitale smutveje

Ethvert projekt må nødvendigvis komme med en tidsbegrænsning, og for Infrastrukturalisme-projektets vedkommende udløber vores opmærkningstid sidst i marts 2022, da vi har været så heldige at kunne forlænge projektet med tre måneder. Tid, vi ellers havde tabt i forbindelse med den første lockdown. Men ligegyldigt, hvilken mængde tid der havde været til rådighed, så ville det naturligvis stadig være en stor opgave, som projektet har påtaget sig, når man tager størrelsen af arkiverne i betragtning. Og hvis alting skulle udføres manuelt, ville det måske også være blevet en umulig opgave.

Derfor måtte vi tænke kreativt i forhold til, hvilke processer vi kunne digitalisere og automatisere uden at sætte kvaliteten på det endelige produkt over styr. Noget af det, som ville have kunnet sluge en masse af vores tid, er den manuelle transskription af vores korpus. Derfor har projektet inddraget OCR-genkendelse som en del af digitaliseringsprocessen, sådan at vi i hvert fald kan minimere behovet for manuel transskription til hovedsageligt at gælde de dokumenter, der er helt eller delvis håndskrevne. Allerede dét er en gevaldig lettelse af arbejdsbyrden, og med OCR-genkendelse af maskinskrrevne breve er fejlprocenten så tilpas lav, at det efter en manuel korrekturlæsning i forbindelse med opmærkningen ikke går ud over den samlede kvalitet af registranten i sidste ende.

En anden ting, der ville kunne spise unødigt meget af vores tid, er opbygningen af hvert enkelt dokument som digital fil, og derfor besluttede vi, at udarbejdelsen af projektets opmærkningsmanual ikke bare skulle udmønte sig i netop manualen, men også i en standardskabelon over opbygningen af den digitale fil ud fra vores

vedtagne regler. Formålet med denne skabelon er, at Center for Sprogteknologi kan tage skabelonens informationer og fusionere dem med transskriptionen af et hvilket som helst dokument og på den måde autogenerere et XML-dokument, hvor teksten samt dokumentets metadata automatisk indsættes i den vedtagne TEI-opbygning.

Ud over præmisserne for dokumentets opbygning fra opmærkningsmanualen, som den autogenererede fil får fra vores generiske TEI-skabelon, bliver den autogenererede fil også fodret med oplysninger fra projektets lister over vigtige informationer, som vi vil fremhæve i den løbende tekst (personnavne, sprog, publikationer osv.). Det betyder, at vi i stedet for at starte hver enkelt digitale fil ud med at opbygge filen fra bunden, kan arbejde ud fra computerens bud på en korrekt opmærket fil ud fra de regler, som vi selv har givet den.

Det kræver selvfølgelig stadig en manuel gennemgang, eftersom en computer fx ikke forstår om 'Hans' skrevet med stort forbogstav efter punktum er et personligt pronomen eller et fornavn, ligesom den heller ikke kan gennemskue, hvornår 'system' bliver brugt som et glossematisk fagterm, og hvornår det bare er et almindeligt substantiv. Der er også brug for et menneske til at oversætte for computeren, når afsenderen af et brev refererer til en dato med et tal efterfulgt af 'ds.' i betydningen 'den nævnte dato i denne måned', ligesom det også kræver et menneske at udlede af konteksten, hvilken person der er tale om, hvis vedkommende kun omtales ved efternavn – det ville jo være et mirakel, om vi kun havde haft én Sørensen og én Jensen på vores navneliste.

Med denne fremgangsmåde minimerer vi altså hængepartierne kraftigt, og eftersom der stadig er et par ting, der kræver en menneskelig hånd til at gå efter i sømmene, så er vi trods alt ikke ved at gøre os selv helt overflødige til fordel for robotter og andre maskiner endnu. I stedet er vi taknemmelige over denne maskinelle håndsækning, som gør det muligt for os at fokusere på den mere forfinede kvalitetskontrol frem for at spille tiden på manuelt gravearbejde.

Et eksempel på, hvorfor der stadig er behov for menneskelig behandling af filerne efter autogenereringen af transskriberet tekst til TEI-fil, kan man se nedenfor i Eksempel 1, der viser, hvordan en autogenereret TEI-fil ser ud *før* den manuelle gennemgang. Efter at

vi har fodret computeren med informationerne fra vores vedtagne skabelon, kommer den med et bud på opbygningen af teksten. Her et brev skrevet af Daniel Jones i hånden på fortrykt brevpapir og sendt til H.J. Uldall fra London den 14. februar 1931. Udsnippet viser transskriptionen af brevets brevhoved, information om tid og sted for brevets produktion samt indgangshilsen til brevets modtager.

```
<text>
  <body>
    <div type="letter">
      <pb n="1" facs="#DJtilHJU-1931-02-14-1.tif"/>
      <fw><!-- brevhoved --></fw>
      <opener xml:id="xx">
        <dateline>
          <placeName ref="#London"><!-- sted --></placeName>
          <date when="1931-02-14"><!-- dato --></date>
        </dateline>
        <salute><persName type="receiver" ref="xx">Hans Jørgen Uldall <!-- modtager --></persName></salute>
      </opener>
      <p xml:id="p1"> </p>
      <p xml:id="p2">STATION : GOLDERS GREEN. </p>
      <p xml:id="p3">TELEPHONE : SPEEDWELL 3497. </p>
      <p xml:id="p4">90 , LYTTETON ROAD <placeName ref="#np1095">LONDON</placeName> , N.W.II. </p>
      <p xml:id="p5"><date when="1931-02-14">14th Feb. 1931</date> </p>
      <p xml:id="p6">Dear Mr <persName ref="#np145">Uldall</persName> , </p>
```

Eksempel 1: Autogenereret TEI-fil før den manuelle gennemgang.

Computeren ved altså, at TEI P5 for brødteksten kræver de overordnede omkransende elementer <text>, <body> og <div> (*division*) med en attribut, der specificerer typen af dokument (*type*=“*letter*”). Elementet <pb> (*page beginning*) markerer starten på en ny side og specificerer sidenummer (*n*=“1”) og faksimile (*facs*=“#DJtilHJU-1931-02-14-1.tif”).

Efterfølgende er vi selv interesserede i et felt til det fortrykte brevhoved: <fw> (*forme work*). Efter <fw> følger det brevspecifikke felt med informationer om brevets tilblivelse i elementet <opener> med underelementet <dateline>, som indeholder informationer om brevets tilblivelse. Men en maskine ved ikke nødvendigvis hvilken del af teksten, der her er tale om, så i stedet er adressens linjer blevet opdelt i hver sit <p>-element (*paragraph*). Autogenereringen har trukket informationer om brevets metadata (sted, dato og modtager) og autoudfyldt informationerne i de korrekte felter, mens den faktiske transskription igen er blevet fordelt i <p>-elementer nedenfor. Den grønne tekst er blot noter og påmindelser til opmærkerne og ikke en blivende del af infrastrukturen.

I den efterfølgende manuelle gennemgang, som ses nedenfor i Eksempel 2, er transskriptionen manuelt blevet rykket ind i de korrekte felter, sådan at <fw> nu indeholder brevpapirets brevhoved, og <dateline> har fået underfeltet <address>. Her er Daniel Jones' adresse indsat i passende elementer – gadenavnet '90, LYTTTELTON ROAD' i <street>, 'LONDON' i <placeName> inklusive sit specifikke reference-id (ref="#np11095") og postnummeret 'N.W.II' i <postCode>. Indgangshilsenen 'Dear Mr Uldall' er placeret under <opener> i elementet <salute> (*salutation*), og H.J. Uldall, som nu er rykket ind i et <persName>-element (*personal name*) har fået tilknyttet en attribut (type="receiver"), der markerer ham som modtager af brevet, samt det personspecifikke reference-id (ref="#np145").

```
<text>
  <body>
    <div type="letter">
      <pb n="1" facs="#DJtilHJU-1931-02-14-1.tif"/>
      <fw>STATION : GOLDERS GREEN. </fw>
      <fw>TELEPHONE : SPEEDWELL 3497. </fw>
      <opener xml:id="xx">
        <dateline>
          <address>
            <street>90 , LYTTTELTON ROAD </street>
            <placeName ref="#np11095">LONDON , </placeName>
            <postCode>N.W.II.</postCode>
          </address>
          <date when="1931-02-14">14th Feb. 1931</date>
        </dateline>
        <salute>Dear Mr <persName type="receiver" ref="#np145">Uldall</persName>,</salute>
      </opener>
    </div>
  </body>
</text>
```

Eksempel 2: Autogenereret TEI-fil *efter* den manuelle gennemgang.

Som sagt, og som eksemplerne viser, er der altså stadig manuelt arbejde at gøre efter den autogenererede oprettelse af en TEI-fil, men i stedet for alternativet – at starte helt fra bunden med hver eneste fil – er der nu blot tale om en smule omrokering. Og heldigvis bliver vi gennem arbejdet klogere dag for dag på både materialet og opmærkningsproceduren, så vi hele tiden kan optimere os selv, vores arbejdsproces og i sidste ende vores digitale infrastruktur.

7. Infrastrukturalisme-projektet fremadrettet

Ved udgangen af 2020 er Hjelmslevarkivet fra Det Kongelige Bibliotek scannet, mens arkiverne for Eli Fischer-Jørgensen, Paul Diderichsen, H.J. Uldall, Harry Wett-Frederiksen og Francis J. Whitfield er

gennemgået og kategoriseret på samme måde som Hjelmslevarkivet ud fra bedømmelserne 'relevant', 'interessant' eller 'irrelevant'.

Digitaliseringen og indlemmelsen af de ovennævnte arkiver i projektet sker med henblik på at udvide perspektivet i projektet, eftersom der her er tale om flere af Hjelmslevs primære korrespondancepartnere. Formålet er så at klarlægge i hvor høj – eller i hvor ringe – grad Hjelmslevs tanker var påvirket af hans nærmeste kolleger. Denne fase er afgørende for, at projektet bevæger sig ud over tilgængeliggørelse og frem til en egentlig forskningsinfrastruktur.

Desværre satte 2020's anden coronanedlukning en midlertidig stopper for digitaliseringen af disse arkiver, men som antydnet har Louis Hjelmslevs hang til at gemme sørget for, at vi har haft masser at tage fat på i mellemtiden.