

Human IT

Tidskrift för studier av IT
ur ett humanvetenskapligt perspektiv

Elektronisk publicering av Wittgenstein's Nachlaß

[Peter Cripps](#) & [Espen S. Ore](#)

Wittgensteinarkivet, Universitetet i Bergen & Humanistisk Datasenter,
Universitetet i Bergen, Norge

Abstract

Wittgenstein's Nachlaß - The Bergen Electronic Edition (WN-BEE) is an electronic version of the 20,000 pages or so of philosophical notes, manuscripts etc. left behind when Ludwig Wittgenstein died in 1951.

The WN-BEE will comprise the transcribed texts in two versions: diplomatic and normalized, and digital facsimiles of the individual pages. The two transcriptions and the facsimiles are linked on the page level.

This paper discusses the production work including text transcription and encoding, presentation software, image processing, and linking of the individual elements.

Innehåll

[1. Sammendrag](#)

[2. Wittgenstein's Nachlaß og utgaver hittil](#)

[3. Den elektroniske utgaven - Bergen Electronic Edition](#)

[3.1 De transkriberte tekstene](#)

[3.2 Fotografier og digitaliserte bilder](#)

[3.3 Presentasjon av transkripsjoner og faksimile](#)

[3.4 Erfaringer fra prosjektet](#)

[Författarna](#)

[Referenser](#)

1. Sammendrag

Wittgenstein's Nachlaß - Bergen Electronic Edition er en elektronisk utgave av transkriberte tekster og digitale faksimiler av ca. 20.000 manuskriptsider. Utgaven skjer i samarbeid mellom Universitetet i Bergen og Oxford University Press og vil komme som fire volumer med siste volum planlagt i 1999. I arbeidet med utgivelsen har man måttet forholde seg til spørsmål angående:

- tekstgjengivelse og tekstkoding
- bildeprosessering og bildeformater
- brukergrensesnitt

([Åter](#) till början av artikeln)

2. Wittgensteins Nachlaß og utgaver hittil

Da Ludwig Wittgenstein døde i 1951 etterlot han seg ca. 20.000 sider med filosofiske arbeider, dels håndskrevne og dels maskinskrevne. Bortsett fra boken *Tractatus Logico-Philosophicus*, enkelte faglige artikler og en ordbok for den østerrikske folkeskole utga Wittgenstein ingenting mens han levde. Mesteparten av dette materialet var derfor upublisert og for en stor del ukjent. I sitt testamente ba Wittgenstein om at hans litterære arvinger og testamentfullbyrdere, G. E. M. Anscombe, Rush Rhees og G. H. von Wright, skulle "publish as many of my unpublished writings as they think fit". (Huitfeldt & Rossvær 89)

I 1953 utga Anscombe og Rhees *Philosophische Untersuchungen* og siden er stadig flere deler av Nachlaß publisert slik at det nå kan hevdes at de viktigste delene av materialet er tilgjengelig. Likevel er det innhold og informasjon i Nachlaß som ikke kan tilgjengeliggjøres med tradisjonelle publiseringsmetoder.

Wittgensteins Nachlaß består av en mengde varianter av til dels samme tekster. En av hans mer spesielle arbeidsmåter var å få maskinskrevet et manuskript, deretter klippet han opp arkene i henhold til det han betraktet som logiske enheter. Disse bitene ("Zettel") kunne så få rettelser og kommentarer før de igjen ble sammenskrevet. For en stor del kan vi derfor si at det ikke eksisterer noen ferdig versjon av en tekst: mange av manuskriptene er varianter der Wittgenstein fortsatt eksperimenterer med nye kombinasjoner og med nye formuleringer. For å kunne forstå Wittgensteins tanker og hvorledes disse utvikler seg er det derfor nødvendig å få tilgang til disse tekstvariantene. Siden tekstene forholder seg til hverandre slik som omtalt ovenfor er det vanskelig, for ikke å si umulig å bruke tradisjonelle utgivelsesfilologiske

formidlingsformer som å etablere en grunntekst og så vise til de andre ved hjelp av et kritisk apparat. Derimot er det fullt mulig å utgi de enkelte manuskripter som egne utgaver, noe som også blir gjort i serien *Wiener Ausgabe* (Nedo93). Men her som med faksimileutgavene som omtales nedenfor, blir det tungvint å manøvrere mellom de forskjellige tekstversjonene og manuskriptene.

I 1968 ble Nachlaß mikrofilmet ved Cornell University. Disse filmene ga til et visst punkt adgang til materialet på en måte som de utvalgte, trykte utgavene ikke ga. Likevel er det flere problemer, både med mikrofilm og andre faksimileutgaver generelt og ved Cornell-filmene spesielt. Faksimiler er i utgangspunktet ikke lettere å finne frem i og å sammenligne enn originalene er. Man savner altså den tilgjengeligheten som ligger i en redigert utgave som gjerne også vil ha hjelpemidler for gjenfinning som innholdsfortegnelser, indekser og kryssreferanser. I tillegg til dette hadde Cornell-filmene sine egne svakheter. Wittgenstein brukte av og til en enkel substitusjonskode og skrev i det han kalte "Geheimschrift". Det ble antatt at disse passasjene bare var personlige bemerkninger, og fordi de kunne gjelde fortsatt levende personer ble det besluttet å "sladde" dem ved å legge et papir over under av fotograferingen. Men det viste seg for det første at slett ikke alle var av en personlig natur og for det andre at det var flere steder man hadde oversett slik at de ikke ble sladdet. Etter avfotograferingen ved Cornell er det også funnet nye manuskripter slik at mikrofilmutgaven heller ikke på denne måten er komplett. At det siden er utgitt mikrofilmer ved Trinity College Library, Cambridge, der mesteparten av Nachlaß finnes, gjør ikke situasjonen mer oversiktlig.

([Åter](#) till början av artikeln)

3. Den elektroniske utgaven - Bergen Electronic Edition

Wittgenstein's Nachlaß - The Bergen Electronic Edition (WN-BEE) er resultatet av et samarbeid mellom Wittgensteinarkivet ved Universitetet i Bergen (WAB) Humanistisk Datasenter ved Universitetet Bergen (HD) og Oxford University Press (OUP). Målet er å gjøre det samlede Nachlaß tilgjengelig på CD-ROM. Utgaven vil omfatte transkriberte, søkbare tekster og faksimiler av manuskriptene lenket til tekstene.

WAB er ansvarlig for de redaksjonelle prinsippene, har stått for transskripsjonene og har utviklet programmer som brukes i arbeidet med dem. HD har arbeidet med å sy sammen delene i et fullstendig system og har stått for redigeringen og tilpasningen av faksimilene. OUP har arrangert fotograferingen av manuskriptene og står for selve utgivelsen. Utgivelsen vil bestå av fire "bind" eller pakker med første del i første halvår 1998. Hele utgivelsen skal etter planen være ferdig innen utgangen av 1999.

([Åter](#) till början av artikeln)

3.1. De transkriberte tekstene

Was bildet mich zu sagen:
 Ich nenne etwas nicht
 "Calculus" wenn ich nicht
 ein Ding ^{gegeben} ~~aus~~ ^{der} ~~beweist~~ ^{geigt} ~~hat~~.
 Ich ~~hoffe~~ ^{hoffe} ~~ich~~ ^{ich} ~~daß~~ ^{daß} ~~man~~
 in ihm ~~ein~~ ^{nicht} ~~gebilde~~ ^{von} ~~etwa~~
~~den~~ ~~den~~ ~~Form~~ ~~erzeugen~~
 kann?

18.10.

Ich will die Formulierung
 entgehen: "ich wird jetzt
 mehr über den Kalkül",
 & statt ihrer die sagen:
 "ich habe jetzt einen anderen
 Kalkül". Der Grund hier von ist,
 die Kluft zwischen einem ma-
 thematischen Wissen & nicht-
 mathematischem Wissen immer
 in ihrer vollen Größe vor Augen
 zu behalten.

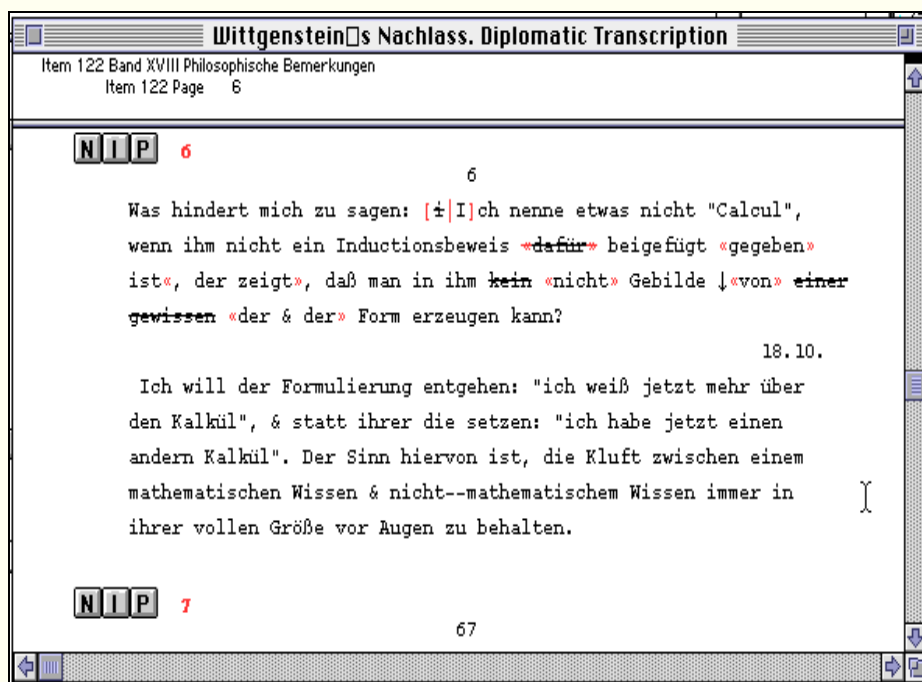
Side 6 i manuskript 122

Manuskriptene, både de håndskrevne og de maskinskrivne, er fulle av redigeringsmarkeringer, overstrykninger, innsetninger, intralineær tekst, marginaltekst, forkortelser, figurer med mer. I utgaven som nå kommer ved OUP skal det være en diplomatarisk og en normalisert tekst. Datatranskripsjonen måtte altså i hvert fall inneholde nok informasjon til at dette skulle la seg produsere. I praksis inneholder originaltransskripsjonene mer informasjon enn det som brukes i denne utgaven. Det skyldes at primærmålet for WAB ikke har vært denne utgaven, men ideelt sett at så mye informasjon kodes fra manuskriptene at det ikke skal være nødvendig å foreta dette arbeidet en gang til.

Primærtransskripsjonene blir kodet i et eget kodesystem, MECS (Multi Element Code System) som er utviklet ved WAB. MECS er et metakodesystem på linje med SGML (Standard Generalized Markup Language) og det konkrete kodesystemet som blir brukt ved WAB kalles MECS-WIT. MECS-WIT inneholder mer enn 400 koder og er under stadig utvikling. Det er flere grunner til at man bygger på et eget kodesystem fremfor å bruke SGML eller det SGML-baserte TEI (Text Encoding Initiative). (TEIP3) Da WAB begynte sitt arbeid i 1990 var det lite tilgjengelig programvare som

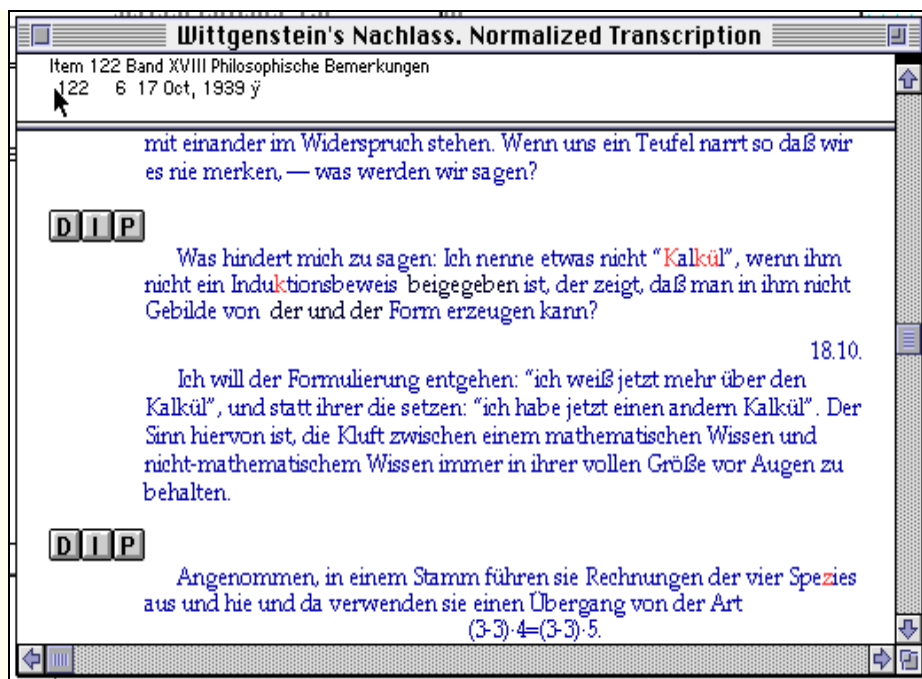
kunne behandle SGML, og TEI var langt fra noen standard. Og i SGML bygger man koderstrukturer som er hierarkiske, det vil si at de kodete elementene er nøstet inni hverandre. I arbeidet med manuskripter slik som det gjøres ved WAB er det også behov for å kode overlappende strukturer, noe som ikke lett lar seg gjøre innenfor SGML, og noe som i praksis ikke støttes av SGML-programmer. En diskusjon om tekst, tekstkoding og hierarkiske og ikke-hierarkiske strukturer finnes i *Monist 80/3* (Biggs & Huitfeldt97).

Ved WAB er det utviklet programmer som kan behandle tekst kodet i MECS. Disse programmene kan brukes under transkripsjon og korrektur til verifisering av kodesyntaks, uttrekk av ord til ordlistesjekk osv. Det er også skrevet programmer som kan filtrere data kodet i MECS til ønskede utdataformater inklusive SGML. Som nevnt ovenfor vil den kommende utgaven inneholde en diplomatarisk og en normalisert tekst. Begge disse er produsert fra den samme primærtranskripsjonen ved hjelp av filterprogrammene.



Siden gjengitt i den diplomatariske versjonen

Den diplomatariske versjonen skal i størst mulig detalj gjengi originalmanuskriptet. Det er grenser, det er for eksempel ingen koding for kaffeflekker eller lignende, slikt får man bruke faksimilen til. Den diplomatariske versjonen gjengir blant annet slettete ord og bokstaver, forkortelser, margkommentarer etc. I utgaven er det valgt en layout og typografiske effekter som i stor grad samsvarer med originalen. Denne versjonen vil være spesielt interessant for forskere som arbeider med tekstens tilblivelse. Det er likevel viktig her å understreke at denne teksten, selv om den etterstreber å gjengi originalen, selvfølgelig vil være et resultat av valg foretatt av transskribenten; den er derfor en redigert tekst på samme måte som den normaliserte versjonen er. (Pichler95)



Siden gjengitt i den normaliserte utgaven

Den normaliserte versjonen gjengir en mer redigert tekst, tilnærmet en tekstutgave. Slettete deler blir ikke vist, og der Wittgensteins anvisninger forklarer entydig hvorledes teksten skal omstruktureres (ord og avsnitt bytte plass etc.), er dette stilltiende gjort. Forkortelser er ekspandert og ortografien er standardisert. Dette er en tekst som er egnet for lesning, og fordi den er standardisert, for ordsøk.

I utgaven er begge disse tekstversjonene og facsimilene lenket på sidenivå slik at det alltid er mulig å hoppe fra den ene til den andre versjonen. Ovenfor ble det også nevnt hvorledes Wittgensteins Nachlaß kan vise til flere forekomster av samme - eller nesten samme - tekststykke i de forskjellige manuskriptene. I denne utgaven er disse parallelle stedene ikke lenket. Dette skyldes ikke bare at dette ville være svært arbeidskrevende. En slik lenking reiser også enkelte metodologiske problemer: med hvilken rett kan vi som utgivere hevde at ett tekststed er det samme som et annet? Er det tilstrekkelig at en setning er lik en annen, eller må et helt avsnitt ord for ord være det samme som et annet? Det er mulig at man under det videre arbeidet kan fastsette en redigeringspolitikk som gjør det mulig å definere et nivå der like tekstdeler blir lenket. Slik tekstene blir publisert i første del av denne utgaven er det i stedet slik at leseren selv kan definere hyperlenker i tekstbasen. Dette gjør det også mulig for en forsker å lenke samme eller like tema selv om de er uttrykt i forskjellig tekst. Lenking på et slikt nivå mener vi ikke kan være en del av arbeidet ved WAB, det er noe den enkelte forsker selv må gjøre.

([Åter](#) till början av artikeln)

3.2. Fotografier og digitaliserte bilder

Eksperimenter foretatt i andre prosjekter tilsier at det ideelle for produksjon av digitale bilder ville være bruk av et digitalt kamera. (Robinson93) Imidlertid har vi ikke bare kunnet la oss styre av ideelle hensyn. Digitale kameraer er dyre, selv om de er blitt billigere i løpet av de siste 4-5 årene er de fortsatt dyre, og da dette prosjektet ble planlagt i 1993 var de langt utenfor våre mulige budsjetter. Digitale kameraer er også

problematiske hvis de ikke allerede er i bruk ved den arkivinstitusjonen som holder materialet som skal digitaliseres. De fleste institusjoner vil være mer åpne for å bruke egne fotoavdelinger enn for at utenforstående skal komme inn og manipulere arkivmaterialet med nytt og ukjent utstyr. Det er rimelig å anta at det ikke er mange årene til de fleste institusjoner har digitale kameraer og kompetanse til å betjene dem, men dette var ikke tilfelle da prosjektet startet. I stedet for direkte digitalisering med digitalt kamera satser vi derfor på bruk av tradisjonell fotografering for deretter å digitalisere filmen.

Av praktiske årsaker valgte vi Kodak PhotoCD som digitalt mellomagringsformat. Dette har sine fordeler i at det eksisterer flere kommersielle firmaer som kan ta hånd om digitaliseringen, men det har sine ulemper ved at fargeformatet i Kodak PhotoCD ikke er spesielt godt. Dette siste er ikke så vesentlig for dette prosjektet. Det viktigste er at brukerne kan se forskjell på fargene i de forskjellige blekktypene som er brukt, ikke at de ser en garantert riktig farge.

Etter at CDene er kommet til Bergen blir de batch-konvertert til JPEG (Joint Photographers' Expert Group) format. Vi tar utgangspunkt i PhotoCD's format Base 4 som tilsvarer 1024 x 1536 pixler og som tar ca. 4,5 MB fullt utpakket. Dette komprimerer vi til mindre enn 170 KB. Tilsynelatende er dette en voldsom komprimering, men etter en del eksperimentering har vi funnet at dette er fullt ut tilfredsstillende.

Etter at bildene er blitt JPEG-komprimert, brenner vi arbeids-CDer der vi grupperer bildene etter de manuskriptene de hører til. Deretter sjekkes de visuelt og opp mot en transkripsjon (så langt denne eksisterer). På dette stadiet avdekker vi feil i fotograferingen som:

- manglende sider
- dobbeltfotograferte sider
- sider ute av sekvens

I noen tilfeller har vi oppdaget at fotografene under arbeidet har glemt igjen papirark oppå sidene under fotograferingen, og noen få ganger har fotografene tatt med tekst som av en eller annen grunn ikke har vært kjent under transkriberingen (som har skjedd ut fra mikrofilm-utgaver med etterfølgende studier av originalen).

For oss har det vært viktig at dette arbeidet skjer mest mulig industrielt. Vi skal behandle ca. 20.000 bilder. For hvert minutt vi bruker i gjennomsnitt per bilde legger vi nesten to personmåneder til prosjektarbeidstiden. For å oppnå raskest mulig gjennomstrømming bruker vi for eksempel et bildevisningsprograms slideshowfunksjon for å vise en jevn strøm av bilder mens vi sjekker med transkripsjon og med paginering på facsimilene i den grad denne finnes.

[\(Åter](#) till början av artikkeln)

3.3. Presentasjon av transskripsjoner og facsimile

Det har alltid vært en forutsetning for dette prosjektet at vi i størst mulig grad ville bruke eksisterende programmer i den publiserte utgaven. Dette skjer av praktiske

årsaker: hvis vi hadde utviklet et system fra bunnen av, ville dette kunne binde oss til vedlikehold og veiledning i flere år fremover. Samtidig har vi også besluttet at det systemet som skulle velges måtte gi oss mest mulig fleksibilitet til å skifte til nye og bedre systemer underveis i prosjektet. Dette siste kriteriet er medvirkende årsak til at faksimilene blir lagret som JPEG bildefiler, ett bilde per manuskriptside. Dette formatet er så generelt at vi uten problemer kunne legge ut den transkriberte teksten i for eksempel HTML (HyperText Markup Language) og fortsatt ha lenkingen mellom tekst og faksimile.

Som presentasjonsverktøy for den transkriberte teksten har vi i første omgang valgt tekstbasesystemet FolioViews. Programmet gir oss et akseptabelt utvalg av typografiske effekter som kan brukes til å gjenspeile originalen i den diplomatariske utgaven, og det har effektive søkeverktøy. Ved WAB har Folio Views vært brukt i mange år, og man har god kompetanse i å tilpasse filtre fra MECS-WIT kodeformat til FolioViews' inputformat (FFF - Folio Flat File). Likevel vil vi understreke at dette valget kan endre seg dersom vi finner noe bedre innen siste bind kommer ut i 1999.

Lenkingen mellom tekst og faksimile blir forenklet ved at alle bildefilene har algoritmisk forutsigbare navn: avbildningen av første ytre omslagsside gis filnavnet 001.jpg, første indre omslagsside 002.jpg, og slik fortsetter det til siste ytre omslagsside. Valget av tre posisjoner for filnavn er gitt fordi vi i utgangspunktet visste at det største manuskriptet - som er maskinskrevet - er på mindre enn 1000 sider.

[\(Åter](#) till början av artikeln)

3.4 Erfaringer fra prosjektet

Tilretteleggingen og gjennomføringen av prosjektet bygger i stor grad på de prinsipper man kom frem til da en "feasability study" ble gjennomført i 1993. Uten en slik forprosjektfase ville vi sannsynligvis hatt store problemer med å gjennomføre den første produksjonsfasen som nå snart er overstått. Men selv om mye var avklart før produksjonen startet, var det også en del nye momenter som kom inn da vi gikk over fra prototyp til produksjon. Vi hadde for eksempel ikke fra starten helt klart for oss hvor mye som kunne skje under fotograferingen og hvor viktig en mest mulig effektiv kontroll av denne ville være. For institusjoner som våre (WAB og HD) som i stor grad har eksistert innenfor den akademiske sfære, har også samarbeidet med erfarne redaktører ved OUP vært lærerikt når det gjelder forskjellen mellom et *funksjonelt* ferdig produkt og et *salgbart* produkt. Vi tenker her både på slike ting som utseende og layout, men også på praktiske forhold som mulige brukerplattformer, hvorledes tilpasse installasjonsprosedyrer osv.

Ved siden av disse praktiske erfaringene ser vi også at hovedprosjektets - Wittgensteinarkivets - målsetning og metode har vært fornuftig. For WAB er denne utgivelsen bare en av mange mulige utgivelser. Forskere som besøker WAB kan bruke verktrøy som arbeider mot primærtransskripsjoner og som gjør det mulig å foreta analyser utover det som er mulig i den utgaven som nå blir publisert. Grunntanken har altså vært å bygge opp et elektronisk arkiv med all den informasjon som det kan ventes at en generell arkivinstitusjon kan bidra med blir lagret, og så kan man derfra produsere utgaver etter behov. Hvis man istedet hadde arbeidet med denne utgaven som mål fra

første stund, er det stor fare for at mye av arbeidet måtte gjøres om igjen om få år.

[\(Åter](#) till början av artikkeln)

Författarna

Peter Cripps har studert filosofi ved University College, London. Siden 1993 har han vært ansatt ved Wittgensteinarkivet ved Universitetet i Bergen der han har arbeidet spesielt med transkripsjon av tekster.

Espen S. Ore har en kombinert bakgrunn i humaniora fra Universitetet i Oslo (bl.a. klassiske fag og religionshistorie) og informatikk/informasjonsvitenskap fra universitetene i Oslo og Bergen. Han har arbeidet ved Humanistisk Datasenter ved Universitetet i Bergen siden 1984. Hans interesser er bl.a.: hypertekst og multimedia, elektroniske tekster og datamaskiner og skriftsystemer.

[\(Åter](#) till början av artikkeln)

Referenser

Biggs, Michael & Claus Huitfeldt (Eds), "Philosophy and Electronic Publishing" in *The Monist*, vol. 80, No 3, La Salle Illinois 1997.

Huitfeldt, Claus & Viggo Rossvær, *The Norwegian Wittgenstein Project Report*, Bergen: The Norwegian Computing Centre for the Humanities, 1988.

Nedo, Michael (Ed), *Ludwig Wittgenstein: Wiener Ausgabe*, Wien, New York: Springer, 1993.

Pichler, Alois, "Advantages of a Machine-Readable version of Wittgenstein's Nachlaß" in *Culture and Value. Pilosophy and Cultural Sciences*, eds. Kjell S. Johannessen & Tore Nordenstam, The Austrian Ludwig Wittgenstein Society, Kirchberg am Wechsel 1995.

Robinson, Peter, *The Digitization of Primary Textual Sources*, Office for Humanities Communication Publications Number 4, Oxford 1993.

TEI P3, eds. Lou Burnard & Michael Sperberg-McQueen, Illinois & Oxford 1994

[\(Åter](#) till början av artikkeln)

© Peter Cripps & Espen S. Ore, 1997

Manuscript facsimile from Wittgenstein's Nachlaß reproduced by permission from the copyright holders, the Trustees of Ludwig Wittgenstein's Nachlaß.

Åter till [Human IT 4/1997](#)

