

EEN CASE STUDY: 18^e eeuws drukwerk automatisch omzetten in digitale tekst

Oktober 2020, door M. Roling (Corts Foundation)

In dit korte en praktische onderzoek zijn ABBYY Finereader™ en Transkribus™ gebruikt om één pagina achttiende eeuwse gedrukte tekst automatisch te transcriberen met een getraind model. Het doel was om te kijken of het mogelijk is een bijna perfecte digitale tekst te verkrijgen uit een gedrukte historische uitgave, met het oog op het doorzoekbaar en bruikbaar te maken voor onderzoekers.

Eén pagina is gekozen uit de facsimile heruitgave van de 'Oud en Nieuw Oost-Indiën' door Francois Valentijn uit 1724. Pagina 120 uit het eerste deel van deze reeks is in twee kolommen, heeft verschillende fontgroottes, normaal en italic, en heeft een aantal marginalia. Het lijkt voldoende representatief voor de gehele achtdelige reeks die ettelijke duizenden pagina's beslaat. Een scan is gemaakt van deze pagina in 300DPI kleur, om zo min mogelijk details te verliezen. Vervolgens is deze pagina in beide genoemde tools omgezet in digitale tekst.

120		B E S C H R Y V I N G D E R	
	Zyn recht op de zelve, in 't jaar 1667 by 't aanweezen van den Heer <i>Speelman</i> aldaar als Commiffaris geoppert, had hy niet grondig konnen bewyzen, hoewel hem egter den 10den Januari in 't jaar 1669 toegeltaan was, een van zyne Babatos, of ryx-raaden met eenige gemagtigden uit Amboina derwaarts te zenden, en van daar naar 't eiland Oebi alle, die zig genegen daar toe betoonden, te mogen overvoeren.	Raad van Justitie, even eens als in 't jaar 1675, ten tyde van den Heer <i>Hurd</i> , van boord gehaalt, en onder drie schootten van de chaloepe na land gevoert, en daar ook zo ontfangen.	
	De naamen van die dorpen, waar op hy cifch maakt, zyn deze: Liffabatta, Hatoewè, Saway, Laulata, Papoeti, Bawoer, Toelocfey, Saleman, en Hatilen.	's Konings brief, die toen mede van boord ging, en in een silvere schotel door hen aan een Corporaal van den Heer <i>van der Stel</i> 's lyf-wagt, <i>Mylraad</i> , overgegeven, en weder in een andere silvere schotel gelegd was, had de lengte van een span, en was drie vingeren breed, zynde (gelyk wy zullen zien) in 't Malcits, en met een Arabische letter, geschreven.	De Prins van Batsjan komt toen in Amboina, om 'er bezit af te nemen:
Wat hy in Amboina heeft.	Haar Edelheden meenden in 't jaar 1672, dat de overgave der negentig huisgezinnen, by den Heer landvoogd in Ternate gedaan, zou konnen dienen, omdezen vorft, in	De brief wierd door dien Corporaal, na dat hy met een geele zyde neusdoek bedekt was, onder een geel Sonnescherm (de Koninklyke verwe) bloots hoofds gedragen.	

Als eerste is de pagina door Transkribus heen gehaald met het daarin beschikbare model genaamd 'Dutch_Romantype_Print' gemaakt door Annemieke Romein (Universiteit Gent). Daarna is in dit onderzoek door de auteur zelf een standaard Nederlands OCR model in ABBYY uitgebreid met een eigen karakter-set die is opgebouwd uit letterpatronen uit een achttal voorbeeld-teksten met ground truth transcriptie. Deze ABBYY karakterset is genaamd 'valentijn.fbt' (en vrij beschikbaar en uitbreidbaar), en kan in ABBYY worden ingeladen alvorens OCR op een pagina-scan wordt uitgevoerd¹. De resultaten van de transcripties met Transkribus en ABBYY zijn hieronder naast elkaar gezet. Verschillen zijn zo goed mogelijk gearceerd en geteld, om tot een Character Error Rate te komen.

Bij ABBYY valt op dat er veel fouten komen doordat de drukletter 'e' vaak niet wordt herkend, en als 'c' in de transcriptie komt. Eveneens is de 'lange s' problematisch en komt veelal als 'f' of 'l' in de transcriptie. Het is de vraag of het verder 'trainen' van het ABBYY OCR model dit nog zou kunnen verbeteren. Als dit voor nu buiten beschouwing wordt gelaten, dan zien we dat het Transkribus model (CER = 0,04) in elk geval al duidelijk veel beter presteert dan de ABBYY (CER = 0,12). Beide tools hebben wat moeite met marginalia en woordfragmenten worden niet altijd herkend in de OCR. Als we hiervoor nog verder zouden corrigeren, dan komt het Transkribus model er alleen nog maar beter uit.

ABBYY	TRANSKRIBUS	ORIGINELE GROUND TRUTH
<p>[120] BESCHRYVING d e r Zyn recht op d e zelve, in 't jaar 1667 by 't aanweezen van den Heer <i>Speelman</i> aldaar als Commiffaris geoppert, had hy niet grondig konnen bewyzen, hoewel hem egter den 10<kn>den Januari in 't jaar 1669 toegel<aa>aan was, een van zyne Babatos, of ryx-raaden met eenige gemagtigden uit Amboina derwaarts te zenden, en van daar naar 't eiland Oebi alle, die zig genegen daar toe betoonden, temo- gen overvoeren.</p>	<p>BESCHRYVING DER 120 Zyn recht op de zelve, in't jaar 1667 by 't aanweezen van den Heer <i>Speelman</i> aldaar als Commissaris geoppert, had hy niet grondig konnen bewyzen, hoewel hem egter den 10den Januari in 't jaar 1669 toegestaan was, een van zyne Babatos, of ryx-raaden met eenige gemagtigden uit Amboina derwaarts te zenden, en van daar naar 't eiland Oebi alle, die zig genegen daar toe betoonden, te mogen overvoeren.</p>	<p>120 BESCHRYVING DER Zyn recht op de zelve, in 't jaar 1667 By 't aanweezen van den Heer <i>Speelman</i> Aldaar als Commissaris geoppert, had hy niet grondig konnen bewyzen, hoewel hem egter den 10den Januari in 't jaar 1669 toegestaan was, een van zyne Babatos, of ryx-raaden met eenige gemagtigden uit Amboina derwaarts te zenden, en van daar naar 't eiland Oebi alle, die zig genegen daar toe betoonden, te mogen overvoeren.</p>

¹ Zie ABBYY menu: Gereedschappen > Opties > OCR > Gebruikerspatronen en talen laden > Gebruikerspatronen gebruiken, incl. ingebouwde patronen

<p>De naamen van die dorpen, waar op hy eisch maakt, zyn deze : Wat hy Lissabatta, Hatoewe, Saway, Laula- in Am- ta, Papoeti, Bawoer, Toeloefey, Sale- [boina] heeft man, en Hatilen.</p> <p>Haar Edelheden meenden in ' tjaar 1672, dat de overgave der negentig huisgezinnen, by den Heer landvoogd in Ternate gedaan , zou konnen dienen, omdezen vorst, in opzicht van zyne andere eischen op de voornoemde dorpen, in toom te houden.</p> <p>Dog des niet te min hebben de Koningen van Batsjan in ' t jaar 1673, 1674, En wat en 1675, onderden Heer Hurdt (nader hy er qualyk) uit schryven harer Edelheden van den cisch. 27ten December 1673 blykende) en 1682, ten tyde van den Heer Commissaris Padbrugge, groote eischen gedaan; en hoewel hun toen duidelyk de onbillykheid, en de ongegrondheit der zelve aangewezen is, zo heeft ook de domme-kracht (dus wierd hy in de wandeling genaamt) Koning Moesom, volgens den aangeboren aart van alle deze vorften, niet konnen rusten, voor en al eer hy haar Edelheden eindelyk bewogen heeft, hem toe te staan, dat hy bezit van deze landen zou mogen nemen, mits dat ook die ingezetenen daar niet tegen moesten hebben, en zig genegen toonen onder hem te willen staan; maar dat zy anders geenzins daar toe zouden mogen gedwongen worden.</p>	<p>De naamen van die dorpen, waar op hy eisch maakt, zyn deze: Wat hy Lissabatta, Hatoewe Saway, Laula- in Am- ta, Papoeti, Bawoer, Toeloese, Sale- [boina] hee, man, en Hatilen. eef.</p> <p>Haar Edelheden meenden in ' t jaar 1672, dat de overgave der negentig huisgezinnen, by den Heer landvoogd in Ternate gedaan, zou konnen dienen, omdezen vorst, in opzicht van zyne andere eischen op de voornoemde dorpen, in toom te houden.</p> <p>Dog des niet te min hebben de Koningen van Batsjan in't jaar 1673, 1674, En wat en 1675, onder den Heer Hurdt (nader hy er qualyk) uit schryven harer Edelheden van den cisch. 27sten December 1673 blykende) en 1682, ten tyde van den Heer Commissaris Padbrugge, groote eischen gedaan; en hoewel hun toen duidelyk de onbillykheid, en de ongegrondheit der zelve aangewezen is, zo heeft ook de domme-kracht (dus wierd hy in de wandeling genaamt) Koning Moesom, volgens den aangeboren aart van alle deze vorsten, niet konnen rusten, voor en al eer hy haar Edelheden eindelyk bewogen heeft, hem toe te staan, dat hy bezit van deze landen zou mogen nemen, mits dat ook die ingezetenen daar niet tegen moesten hebben, en zig genegen toonen onder hem te willen staan; maar dat zy anders geenzins daar toe zouden mogen gedwongen worden.</p>	<p>De naamen van die dorpen, waar op hy eisch maakt, zyn deze: Wat hy Lissabatta, Hatoewe, Saway, Laula- in Am- ta, Papoeti, Bawoer, Toeloefey, Sale- boina heeft man, en Hatilen.</p> <p>Haar Edelheden meenden in ' t jaar 1672, dat de overgave der negentig huisgezinnen, by den Heer Langdvoogd in Ternaten gedaan, zou konnen dienen, omdezen vorst, in opzicht van zyne andere eischen op de voorgenoemde dorpen, in toom te houden.</p> <p>Dog des al niet te min hebben de Koningen van Batsjan in ' t jaar 1673, 1674, En wat en 1675, onder den Heer Hurdt (nader hy er qualyk uit schryven harer Edelheden van den eischt 27sten December 1673 blykende) en 1682, ten tyde van den Heer Commissaris Padbrugge, groote eischen gedaan; en hoewel hun toen duidelyk de onbillykheid, en de ongegrondheit der zelve aangewezen is, zo heeft ook de domme-kracht (dus wierd hy in de wandeling genaamt) Koning Moesom, volgens den aangeboren aart van alle deze vorsten, niet konnen rusten, voor en al eer hy haar Edelheden eindelyk bewogen heeft, hem toe te staan, dat hy bezit van deze landen zou mogen nemen, mits dat ook die ingezetenen daar niet tegen moesten hebben, en zig genegen toonen onder hem te willen staan; maar dat zy anders geenzins daar toe zouden mogen gedwongen worden.</p>
<p>In ' tjaar 1707, schreven haar Edelheden den 22ten February, met het schip Slooten, aan den Heer Van der Stel, en den raad van Amboina, dat zy uit hun schryven aan die van Ternate hunne nader order en begeerte omtrent die negen onwillige Ceramse, en aan den Koning van Batsjan op zekere voorwaarde toegestaane dorpen, nader zouden konnen beoogen.</p> <p>Negen Dit wierd aan dezen Koning in ' t jaar dorpen 1708. toegestaan, die in de maand Au- opCeram g U R us den Kroon-Prins afzond, om van [door haar] Edelhe- die dorpen bezit te nemen. den aan Hy verscheen den vyf en twintigsten hem toe- met de chaloep Larike op de reede van gestaan. Amboina, en werd den zeven en twintigsten door de Gemagtigden, de onder Kooplieden, Piet er Nuyts, Balthasar de Bruynne, en Alexander van Brunei, boek- houder, en geheim-schryver van den</p> <p>Raad van Justitie, even eens als in ' tjaar 1671, ten tyde van den Heer Hardt, van boord ghaalt, en onder drie schooten van de chaloep na land gevoert, en daar ook zo ontfangen.</p>	<p>In ' t jaar 1707, schreven haar Edelheden den 22sten February, met het schip Slooten, aan den Heer Van der Stel, en den raad van Amboina, dat zy uit hun schryven aan die van Ternate hunne nader order en begeerte omtrent die negen onwillige Ceramse, en aan den Koning van Batsjan op zekere voorwaarde toegestaane dorpen, nader zouden konnen beoogen.</p> <p>Negen Dit wierd aan dezen Koning in ' t jaar dorpen 1708. toegestaan, die in de maand Au- op Ceram gustus den Kroon-Prins afzond, om van [door haar] Edelhe- die dorpen bezit te nemen. den aan Hy verscheen den vyf en twintigsten hem toe- met de chaloep Larike op de reede van gestaan. Amboina, en werd den zeven en twintigsten door de Gemagtigden, de onder Kooplieden, Pieter Nuyts, Balthasar de Bruynne, en Alexander van Brussel, Boek- houder, en geheim-schryver van den</p> <p>Raad van Justitie, even eens als in't jaar 1675, ten tyde van den Heer Hurdt, van boord ghaalt, en onder drie schooten van de chaloep na land gevoert, en daar ook zo ontfangen.</p>	<p>In ' t jaar 1707, schreven haar Edelheden den 22sten February, met het schip Slooten, aan den Heer Van der Stel, en den raad van Amboina, dat zy uit hun schryven aan die van Ternate hunne nader order en begeerte omtrent die negen onwillige Ceramse, en aan den Koning van Batsjan op zekere voorwaarde toegestaane dorpen, nader zouden konnen beoogen.</p> <p>Negen Dit wierd aan dezen Koning in ' t jaar dorpen 1708. toegestaan, die in de maand Au- op Ceram gustus den Kroon-Prins afzond, om van door haar Edelhe- die dorpen bezit te nemen. den aan Hy verscheen den vyf en twintigsten hem toe- met de chaloep Larike op de reede van gestaan. Amboina, en werd den zeven en twintigsten door de Gemagtigden, de onder Kooplieden, Pieter Nuyts, Balthasar de Bruynne, en Alexander van Brussel, Boek- houder, en geheim-schryver van den</p> <p>Raad van Justitie, even eens als in ' t jaar 1675, ten tyde van den Heer Hurdt, van boord ghaalt, en onder drie schooten van de chaloep na land gevoert, en daar ook zo ontfangen.</p>

<p>'s Konings brief, die toen mede van p c p rin 5 boord ging, en in een filverc fshotel door van Bats- hen aan een Corporaal van den Heer j n komt van der Stel 's lyf-wagt, Mylraad, over- [toen in] Amboi- gevegen, en weder in een andere filvere na a om 'et ichotel glegt was , had de lengte van bezit af te een lpan , en was drie vingeren breed, nemen: zynde (gelyk wy zullen zien) in ' t Ma- leits, en met een Arabische letter, ge- fchreven.</p> <p>De brief wierd door dien Corporaal, na dat hy met een geele zyde neusdoek bedekt was, onder een geel Sonncfcherm (de Koninklyke verwe) bloots hoofds gedragen.</p> <p>Hier op volgde, als in Processie, de statie eerste gezant, of de Prins van Batsjan omtrent ('s Konings oudste zoon) Kaytsilli Mant- 't inhaa- fber genaamt. Hy was een Heer van een [len van den zel]</p> <p>korte gefaltc, geenzins de vetfte, en ven, en na gissing een of twee en dertig jaa- fan 's ren oud. Hy had een Taffeta geel en rood gebloemt kleed aan , een rooden bnc. gemarmelden Tulband (Diftar of wrongs- gewyze) op ' t hoofd, die met verscheide filverc verdeeling en hier en daar in ' t rond, en wel een vinger breed, gelyk zy omtrent ook zo verre van een Ron- den, onderscheiden was, waar aan men groote filvere quasten met loovers op de linker zyde hangen zag. Ook had hy een grooten rood en geel gestikten neus- doek over zyn fchouder tot achter op zyn rug hangen.</p>	<p>'s Konings brief, die toen mede van De Prins boord ging, en in een silvere schotel door van Bats- hen aan een Corporaal van den Heer jan komt van der Stel 's lyf-wagt, Mylraad, over [toen in] Amboi- gevegen, en weder in een andere silvere na, om 'er schotel gelegd was, had de lengte van bezit afte cen span, en was drie vingeren breed, nemen: zynde (gelyk wy zullen zien) in 't Ma- leits, en met een Arabische letter, ge- schreven.</p> <p>De brief wierd door dien Corporaal, na dat hy met een geele zyde neusdoek bedekt was, onder een geel Sonnescherm (de Koninklyke verwe) bloots hoofds gedragen.</p> <p>Hier op volgde, als in Processie, de Statie eerste gezant, of de Prins van Batsjan omtrent ('s Konings oudste zoon) Kaytsilli Mant- 't inhaa- soer genaamt. Hy was een Heer van een [len van den zel]</p> <p>korte gestalte, geenzins de vetste, en ven, en na gissing een of twee en dertig jaa- van 's ren oud. Hy had een Taffeta geel Konings en rood gebloemt kleed aan, een rooden brief. gemarmelden Tulband (Distar of wrongs- gewyze) op 't hoofd, die met verscheide silvere verdeelingen hier en daar in 't rond, en wel een vinger breed, gelyk zy omtrent ook zo verre van een ston- den, onderscheiden was, waar aan men groote silvere quasten met loovers op de linker zyde hangen zag. Ook had hy een grooten rood en geel gestikten neus- doek over zyn schouder tot achter op zyn rug hangen.</p>	<p>'s Konings brief, die toen mede van De Prins boord ging, en in een silvere schotel door van Bats- hen aan een Corporaal van den Heer jan komt van der Stel 's lyf-wagt, Mylraad, over toen in Amboi- gevegen, en weder in een andere silvere na, om 'er schotel gelegd was, had de lengte van bezit afte een span, en was drie vingeren breed, nemen: zynde (gelyk wy zullen zien) in 't Ma- leits, en met een Arabische letter, ge- schreven.</p> <p>De brief wierd door dien Corporaal, na dat hy met een geele zyde neusdoek bedekt was, onder een geel Sonnescherm (de Koninklyke verwe) bloots hoofds gedragen.</p> <p>Hier op volgde, als in Processie, de Statie eerste gezant, of de Prins van Batsjan omtrent ('s Konings oudste zoon) Kaytsilli Mant- 't inhaa- soer genaamt. Hy was een Heer van een len van den zel-</p> <p>korte gestalte, geenzins de vetste, en ven, en na gissing een of twee en dertig jaa- van 's ren oud. Hy had een Taffeta geel Konings en rood gebloemt kleed aan, een rooden brief. gemarmelden Tulband (Distar of wrongs- gewyze) op 't hoofd, die met verscheide silvere verdeelingen hier en daar in 't rond, en wel een vinger breed, gelyk zy omtrent ook zo verre van een ston- den, onderscheiden was, waar aan men groote silvere quasten met loovers op de linker zyde hangen zag. Ook had hy een grooten rood en geel gestikten neus- doek over zyn schouder tot achter op zyn rug hangen.</p>
<p>Op zyn zyde zag men een fclloon en Cris, diens hegt van uitnemend fchoon rood en wit Acarbahar, of Koraal was, vier duim breed, hoedanig een soort ik noit meer gezien heb.</p> <p>De tweede gezant, Doebo Doebo ge- Zyn ver- naamt, was Hhockom, of 's Lands op- dere per-richter, een lyvig Heer; dog mede Trein ' dc langfte niet. Het was een man van twee of drie en veertig jaaren , die een Citsche rok aan, een enkel-roode Tul- band, of Diftar Maboeka , met filvere loovers op ' t hoofd , en ook een fchoo- nen Cris op zyn zyde had.</p> <p>Deze twee gingen blootsvoets tuflehen onze Gemachtigden in.</p> <p>Daar na volgde de Sjahbandar, of Tollenaar, die de eerste van hun gevolg, en in ' t rood Scharlaken gekleedt was.</p> <p>De tweede van hun gevolg had een gestreepden Armozyne rok aan , die van twee personen, in ' t wit gekleed , met Crissen op zyde, gevolgt wierden; dog een derde, mede in ' t wit gekleed, fchcen wel een Pricfter te zyn.</p> <p>Bchal-</p>	<p>Op zyn zyde zag men een schoonen Cris, diens hegt van uitnemend schoon rood en wit Acarbahar, of Koraal was, vier duim breed, hoedanig een soort ik noit meer gezien heb.</p> <p>De tweede gezant , Doebo Docbo ge- Zyn ver- naamt, was Hhoekom, of 's Lands op- dere per-richter, een lyvig Heer; dog mede Trein. de langste niet. Het was een man van twee of drie en veertig jaaren, die een Citsche rok aan, een enkel-roode Tul- band, of Distar Maboeka, met silvere loovers op 't hoofd, en ook een schoo- nen Cris op zyn zyde had.</p> <p>Deze twee gingen bloots voets tusschen onze Gemachtigden in.</p> <p>Daar na volgde de Sjahbandar, of Tollenaar, die de eerste van hun gevolg, en in 't rood Scharlaken gekleedt was.</p> <p>De tweede van hun gevolg had een gestreepden Armozyne rok aan, die van twee personen, in 't wit gekleed, met Crissen op zyde, gevolgt wierden; dog een derde, mede in 't wit gekleed, scheen wel een Priester te zyn.</p> <p>Behal-</p>	<p>Op zyn zyde zag men een schoonen Cris, diens hegt van uitnemend schoon rood en wit Acarbahar, of Koraal was, vier duim breed, hoedanig een soort ik noit meer gezien heb.</p> <p>De tweede gezant , Doebo Doebo ge- Zyn ver- naamt, was Hhoekom, of 's Lands op- dere per-richter, een lyvig Heer; dog mede Trein. de langste niet. Het was een man van twee of drie en veertig jaaren, die een Citsche rok aan, een enkel-roode Tul- band, of Distar Maboeka, met silvere loovers op 't hoofd, en ook een schoo- nen Cris op zyn zyde had.</p> <p>Deze twee gingen bloots voets tusschen onze Gemachtigden in.</p> <p>Daar na volgde de Sjahbandar, of Tollenaar, die de eerste van hun gevolg, en in 't rood Scharlaken gekleedt was.</p> <p>De tweede van hun gevolg had een gestreepden Armozyne rok aan, die van twee personen, in 't wit gekleed, met Crissen op zyde, gevolgt wierden; dog een derde, mede in 't wit gekleed, scheen wel een Priester te zyn.</p> <p>Behal-</p>
<p>Aantal errors = 219 CER = 0,24</p> <p>Minus c/e en f/s errors [107]: 112 Gecorrigeerde CER = 0,12</p>	<p>Aantal errors = 41 CER = 0,04</p>	<p>Aantal karakters = 901</p>