

Invoerinstructie Crowd Leert Computer Lezen

Leer de computer lezen en help mee aan het woord voor woord doorzoekbaar maken van de Amsterdamse Notariële Archieven.



Leer de computer lezen

Gelijktijdig met het VeleHanden-project Alle Amsterdamse Akten, waarbij de akten van de Amsterdamse notarissen worden geïndexeerd op hoofdkenmerken, gaat Crowd Leert Computer Lezen een stap verder door de miljoenen akten op woordniveau te ontsluiten.

Door de mens gemaakte transcripties dienen als leermateriaal voor de computer. Met behulp van het handschriftherkenningsprogramma Transkribus wordt de computer vervolgens getraind in het lezen van oude handschriften. Door middel van *machine learning* hopen we met behulp van de crowd op termijn alle handschriften uit het notarieel archief – en uiteindelijk natuurlijk *alle* handgeschreven teksten – te kunnen doorzoeken. Het 'googelen' door handgeschreven documenten komt zo een grote stap dichterbij.

De eerste trainingen zijn achter de rug. Wat heeft de computer er na de eerste leeslesjes van gemaakt? Het is aan u om met het rode potlood door de tekst te gaan en alle foutjes te corrigeren. Door de fouten te corrigeren en te hertrainen leert de computer elke keer weer wat bij. Maar voordat de computer zelfstandig kan lezen hebben wij uw hulp hard nodig!

Voor u gaat beginnen

Transcriberen of corrigeren? De computermodellen worden in het begin per handschrift getraind. Nu had een notaris vaak meerdere klerken in dienst, waardoor er verschillende handschriften binnen één boek of zelfs binnen één scan te vinden zijn. Het kan daarom voorkomen dat u een scan voorgeschoteld krijgt met een handschrift waar de computer nog **niet zo ervaren** mee is. Wilt u in dat geval de **computer helpen en intypen** wat er wel moet staan? Hoe meer de computer leert, hoe beter de automatische transcriptie wordt en hoe **minder foutjes** er nog gecorrigeerd hoeven te worden.

Het gaat om het aanleren van handschriften. Blijf daarom zo **dicht mogelijk bij de tekst** zoals die op de scan staat en ga nadrukkelijk zelf **niet interpreteren**. Maak u niet te druk om de weergave van afkortingen, krullen en andere details. De volledige woorden zijn over het algemeen bruikbaar om te trainen en voor de doorzoekbaarheid.

Maak kennis met uw notaris. Vindt u deze notaris maar saai of het handschrift niet fijn? Probeer dan eens een andere. Er zullen in de loop van het project scans van nieuwe notarissen beschikbaar komen.

Houd het forum in de gaten voor de laatste nieuwtjes en tips! Het forum is er óók om elkaar **op weg te helpen** en om de allerleukste **vondsten te delen**.

Voorkeuren

Als u bent ingelogd ziet u een knop 'Voorkeuren', waar u een selectie kunt maken voor een moeilijkheidsgraad of een notaris. U kunt uw voorkeur op elk moment wijzigen en weer opslaan.

Tussen haakjes ziet u het aantal akten dat op dat moment beschikbaar is.

Bent u nieuw, kies dan altijd voor niveau 1 - Beginner. Vanaf niveau 3 hebt u verdergaande kennis van oud schrift nodig.

Let op het cijfer dat staat achter 'Indicatie van het aantal beschikbare scans'. Staat daar 0, dan zijn de scans van die notaris op of hebt u een combinatie aangevinkt waar geen scans aan voldoen. U krijgt dan de eerste willekeurige notaris in de wachtrij.

Berichten	Project info	Statistieken	Voorkeuren	Voortgang
-----------	--------------	--------------	------------	-----------

Kies voorkeuren voor project Crowd leert computer lezen

Kies welke soort scans u te zien wilt krijgen.

Indicatie van het aantal beschikbare scans met deze voorkeuren **200**

moeilijkheidsgraad (verplicht)

<input type="checkbox"/> 1 - Beginner (749)	<input type="checkbox"/> 3 (330)
<input type="checkbox"/> 5 - Expert (660)	

notaris (verplicht)

<input type="checkbox"/> ANTHONY VAN DE VEN 1666-1670 (110)	<input checked="" type="checkbox"/> CORNELIS STAAL 1749-1753 (200)
<input type="checkbox"/> CORNELIS VAN LOON 1700-1743 (110)	<input type="checkbox"/> DANIEL VAN DEN BRINK 1734-1785 (109)
<input type="checkbox"/> DAVID DES POMMARE 1698-1724 (110)	<input type="checkbox"/> DIRK VAN DER GROE 1670-1720 (110)
<input type="checkbox"/> HENDRIK SCHAEF 1636-1665 (110)	<input type="checkbox"/> HERMANUS VAN HEEL 1745-1771 (110)
<input type="checkbox"/> JACOB DE WINTER 1648-1710 (110)	<input type="checkbox"/> JACOB JANSZ WESTFRISIUS 1612-1654 (110)
<input type="checkbox"/> JAN BARELS DE JONGE 1716-1753 (110)	<input type="checkbox"/> JAN FRANSSSEN BRUIJNINGH 1593-1624 (110)
<input type="checkbox"/> JAN VERLEIJ 1741-1775 (110)	<input type="checkbox"/> PALM MATHIJSZ 1610-1641 (110)
<input type="checkbox"/> SIMON VAN SEVENHOVEN 1681-1705 (110)	

Selecteer een notaris of een moeilijkheidsgraad.

Stappen

1. Log in op de website van [Vele Handen](#) en selecteer het project Crowd Leert Computer Lezen.
2. Ga naar het tabblad Voorkeuren en selecteer daar een moeilijkheidsgraad **of** een notaris. Klik op 'Opslaan'. Niet tevreden? Deze selectie kan elk moment worden aangepast. Vergeet ook in dit geval uw voorkeur niet opnieuw op te slaan.
3. Klik op 'Invoeren'. U krijgt een scan op het scherm.
4. Let goed op de segmentatie van de regels. Deze geeft aan welke regel u moet transcriberen.
5. Kunt u niets invoeren omdat u de tekst niet kunt lezen, gebruik dan de knop 'Te moeilijk'. Deze scan wordt dan aan een andere invoerder voorgelegd.
6. Bent u klaar met de scan, klik dan op 'Afronden'. Wilt u op een ander tijdstip verder gaan, klik dan op 'Tussentijds bewaren'. Let op: de scan wordt 48 uur voor u bewaard vanaf het moment dat u de eerste keer op de knop 'Tussentijds bewaren' heeft geklikt.
7. Wanneer de scan op het scherm geen informatie bevat (een koft of een lege pagina is) klikt u op onbruikbaar. LET OP: gebruik deze knop alleen als er niets aan de scan in te voeren valt. Naamindexen worden dus bijvoorbeeld wél ingevoerd.

Scans, geen akten

Met dit project wordt niet gewerkt met volledige akten maar met losse scans. Dit betekent dat u slechts één scan per keer kunt transcriberen en u na afronding een scan uit een andere akte kunt krijgen. Mocht u toch benieuwd zijn naar de volledige akte, dan kunt u klikken op de knoppen 'Bekijk vorige' en 'Bekijk volgende'. Het is ook mogelijk om de akte terug te vinden via de [inventaris](#).

Algemene invoerinstruaties

- De algemene vuistregel is dat de tekst zo letterlijk en volledig mogelijk moet worden overgenomen. Omdat het vroegmodern Nederlands geen standaardspelling kende, wordt er in principe zo min mogelijk gestandaardiseerd en geïnterpreteerd. Ga dus **niet hertalen** in modern Nederlands.
- Diakritische tekens die we tegenwoordig nog kennen (zoals é, ç, ö) en leestekens (ook haakjes) worden overgenomen. Tekens die niet meer in gebruik zijn, zoals een streepje boven de u, worden niet overgenomen.
- Hoofdlettergebruik wordt zoveel mogelijk aangehouden als in de originele bron. Soms is het niet duidelijk of iets een hoofdletter of een kleine letter is. Probeer in zo'n geval de letter te vergelijken met andere letters in de scan en maak een eenduidige keuze. Hoofdletters midden in een woord vormen de uitzondering: die worden niet overgenomen.
- Afkortingen van namen (bijvoorbeeld Wilha, Catha of Jansz) worden letterlijk overgenomen en dus niet 'opgelost'.
- Scan zo beschadigd dat niet alles te lezen is? Dat kan voorkomen, vooral bij de oudere akten. Voer dan in wat mogelijk is. Misschien is ook die informatie waardevol genoeg om gevonden te worden. Ook als er bij een akte staat 'niet gepasseerd' toch graag invoeren.

Regel voor regel

Let goed op de paarse lijnen die de regel aangeven en zorg dat ze (waar mogelijk) altijd gelijk lopen met de getranscribeerde tekst.

Help!

Ik kan het handschrift niet goed lezen

Is het handschrift (veel) te moeilijk? Ga dan niet gokken, maar klik op de blauwe knop 'te moeilijk'. De akte wordt dan uitgedeeld aan een andere invoerder. Gebeurt dit heel vaak, selecteer dan een niveau lager op het tabblad [Voorkeuren](#) of kies voor een andere notaris.

Ik heb een algemene vraag of opmerking over de inhoud van de akte

Stel deze op het forum van Crowd Leert de Computer Lezen, waar we elkaar op weg kunnen helpen. Ook het forum van Alle Amsterdamse Akten kan uitkomst bieden.

Uitleg van de knoppen

Afronden	= Indien u klaar bent met het transcriberen/corrigeren van de scan klikt u op deze knop.
Te moeilijk	= Ik kan het handschrift niet lezen. Pas eventueel uw Voorkeuren aan.
Onbruikbaar	= De scan bevat geen informatie; er valt niks te transcriberen.
Opmerkelijk!	= Attendeer de projectbeheerders op een inhoudelijk heel bijzondere akte.
Tussentijds bewaren	= De akte tussentijds opslaan om er op een ander moment verder aan te werken.
Bekijk vorige	= Bekijk de scan die aan uw getranscribeerde scan vooraf gaat.
Bekijk volgende	= Bekijk de scan die op uw getranscribeerde scan volgt.

Onbruikbaar = Vuilnisbak

Gebruik deze knop alleen als er niets aan de scan in te voeren valt. Gebruik het **niet** voor indexen, of een scan die te moeilijk, niet compleet of verkeerd ingebonden is. Ook als er handtekeningen ontbreken of de scan niet goed laadt, deze knop niet gebruiken.

Aan de slag

Crowd leert computer lezen

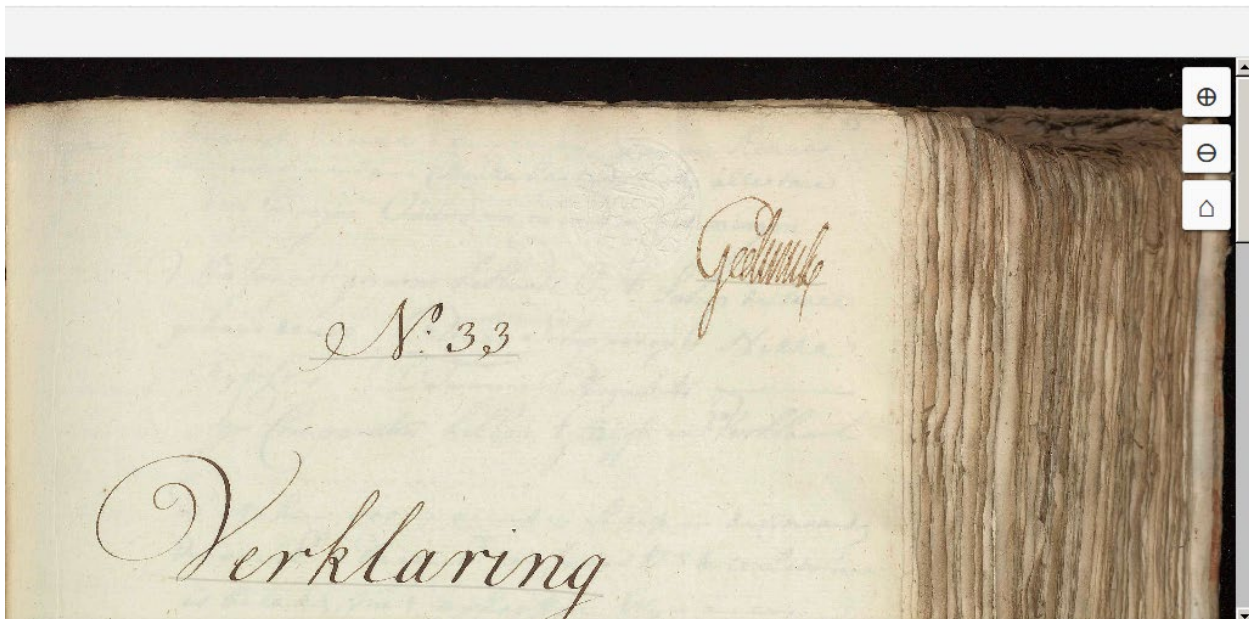
Mijn profiel: Pauline (Stadsarchief Amsterdam) | Uitloggen

Jan Verleij, Inventaris 11925 - 1755 - KLAB05733000056

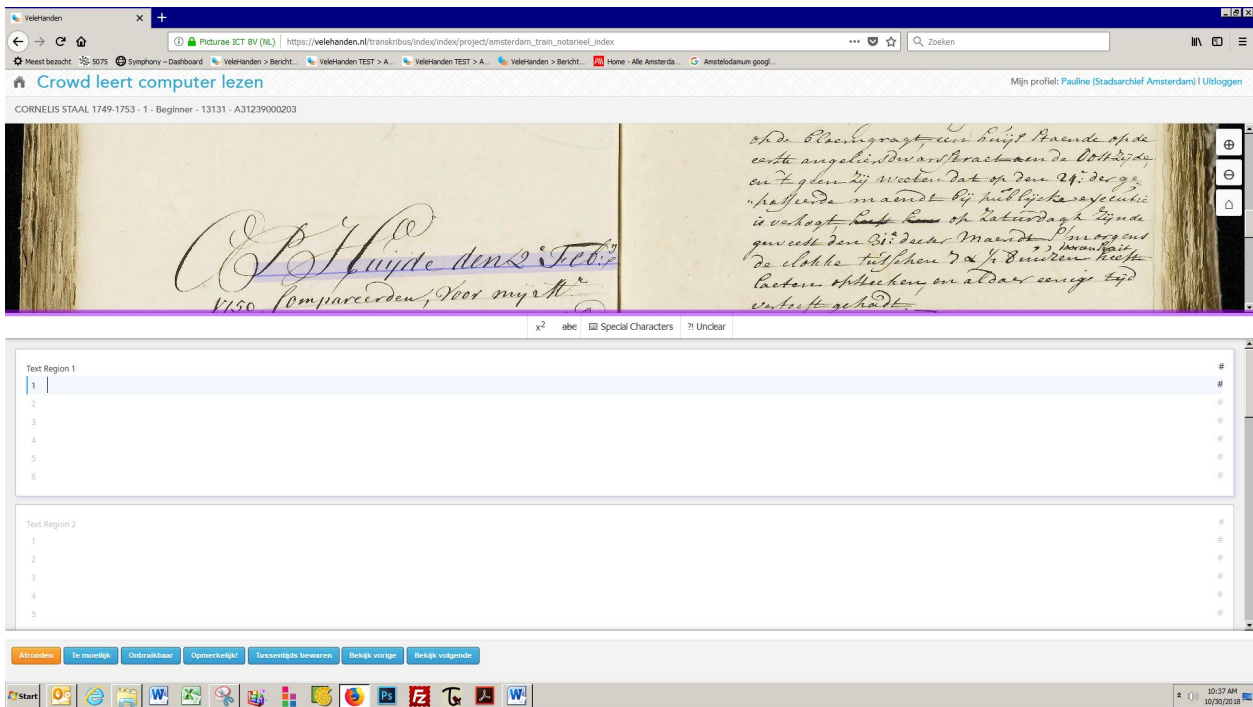


Click to start ...

Klik op het potloodje bij 'Click to start...!' om het invoerscherm in beeld te krijgen. De scan is verdeeld in één of meerdere *text regions* (grotere tekstblokken/alinea's) en regels, te herkennen aan de paarse lijnen. Klik op de eerste regel van het invoerveld en voer in wat u op de regel ziet. Met enter, tab of de pijltjes op het toetsenbord kunt u naar een nieuwe regel.



Helemaal rechts bovenin het beeld kunt u in- of uitzoomen.

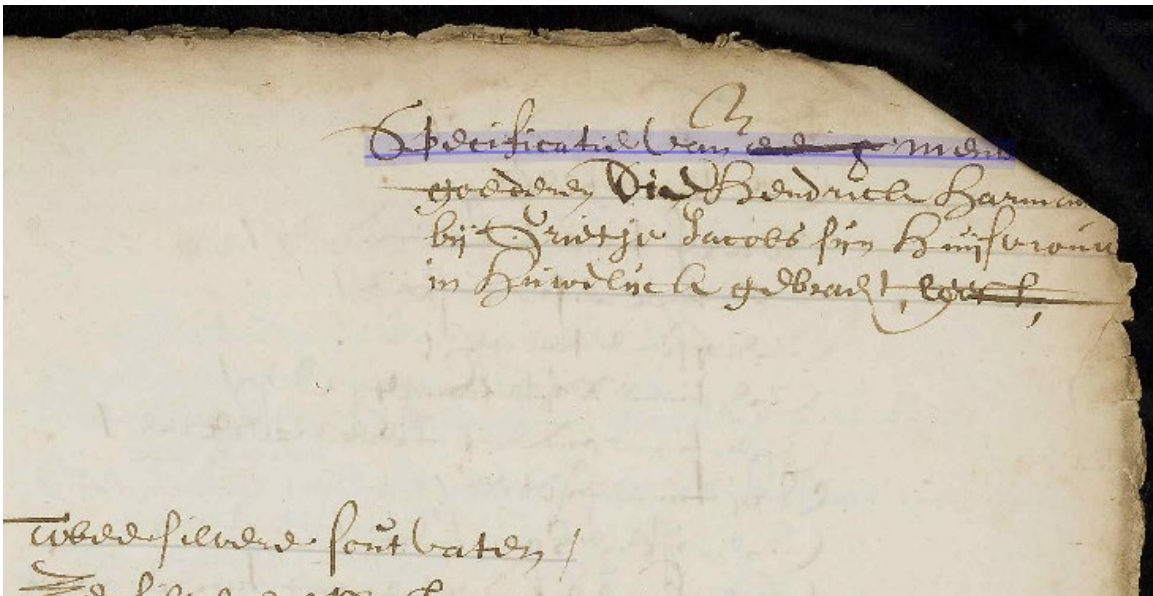


De hoogte van het invoerscherm kunt u naar eigen voorkeur aanpassen door uw cursor op de scheidingsrand te plaatsen. Zodra een paarse lijn op deze rand verschijnt kunt u uw cursor naar boven of beneden slepen.



De gesegmenteerde paarse regel op de scan hoort bij de paarse regel in het transcriptieveld dat zich in de onderste helft van het scherm bevindt. Hier kunt u nu beginnen met het transcriberen van deze regel.

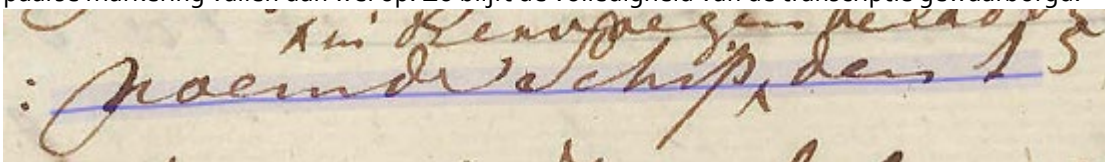
Vuistregel: alle tekstregels die gesegmenteerd zijn worden getranscribeerd. Ze moeten dus ook getranscribeerd worden als ze een ander handschrift hebben, een (gedrukte) bijlage betreffen of een paginanummer, datum, titel, handtekening etc. bevatten.



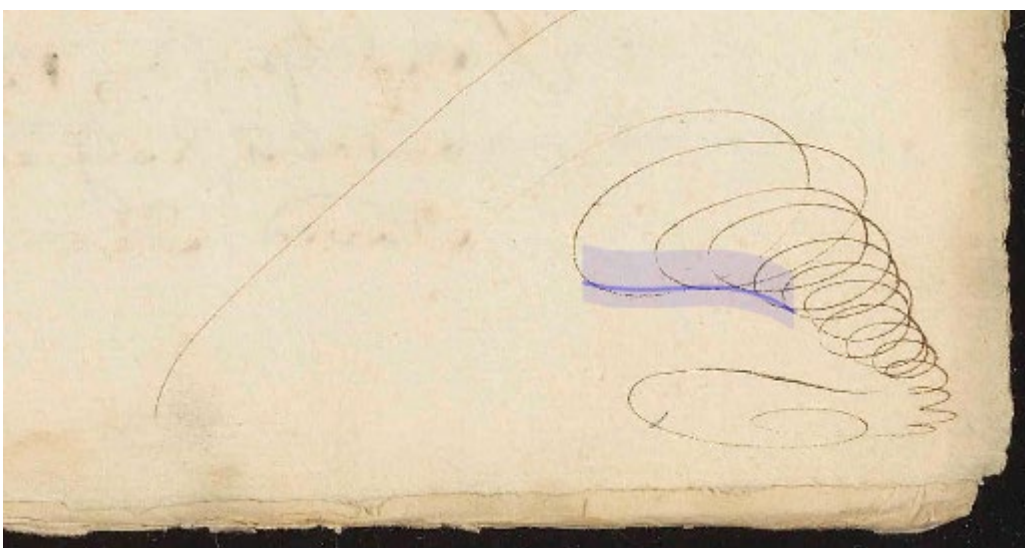
Let op: valt een deel van de tekst op de scan weg, dan wordt die niet aangevuld en transcribeert u alleen het zichtbare deel: 'Specificatie vande eenige meubelen'.

Uitzonderingen

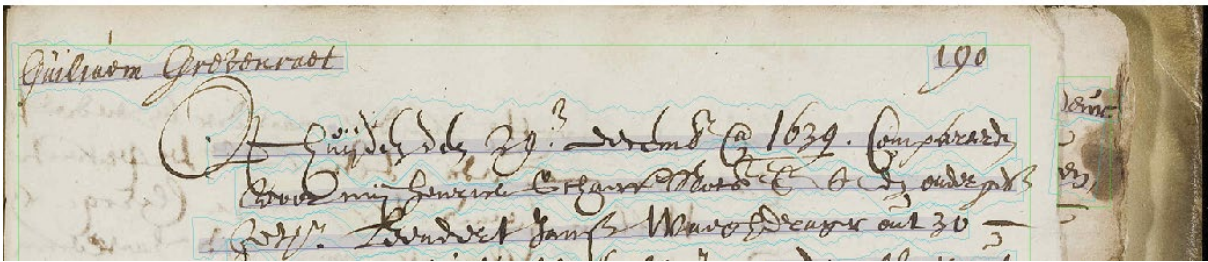
Het kan soms voorkomen dat een regel op de scan niet volledig gesegmenteerd is. Schrijf de letters die buiten de paarse markering vallen dan wel op. Zo blijft de volledigheid van de transcriptie gewaarborgd.



Het nummer 5 valt buiten de lijn, maar wordt wel ingevoerd.



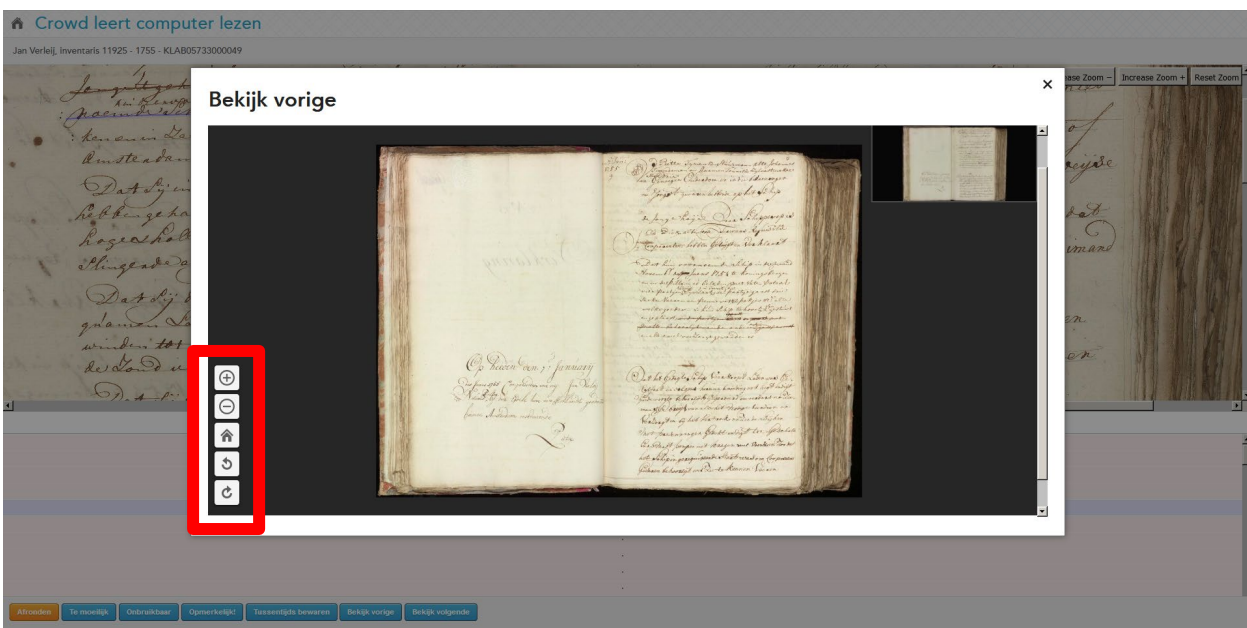
Soms zijn er ten onrechte krabbels, streepjes of doorgedrukte tekst voor een tekstregel aangezien. Die regels kunt u bij de transcriptie gewoon leeg laten.



'Per ongeluk' gesegmenteerde stukjes van andere pagina's, zoals hierboven rechts, worden **niet** getranscribeerd.

Bekijk vorige of volgende scan

Indien u meer van de akte wilt zien, kunt u op de knoppen 'Bekijk vorige' en 'Bekijk volgende' klikken. Hierdoor verschijnt er een nieuw, kleiner scherm voor het invoerscherm. De knoppen links onderin kunt u gebruiken om in en uit te zoomen of om de scan te draaien. U kunt niet meer dan één bladzijde vooruit- of terugbladeren.



Als u helemaal klaar bent met het transcriberen van de scan klikt u op de knop 'Afronden' links onderin het scherm. Het duurt heel even voordat de volgende scan geladen wordt.

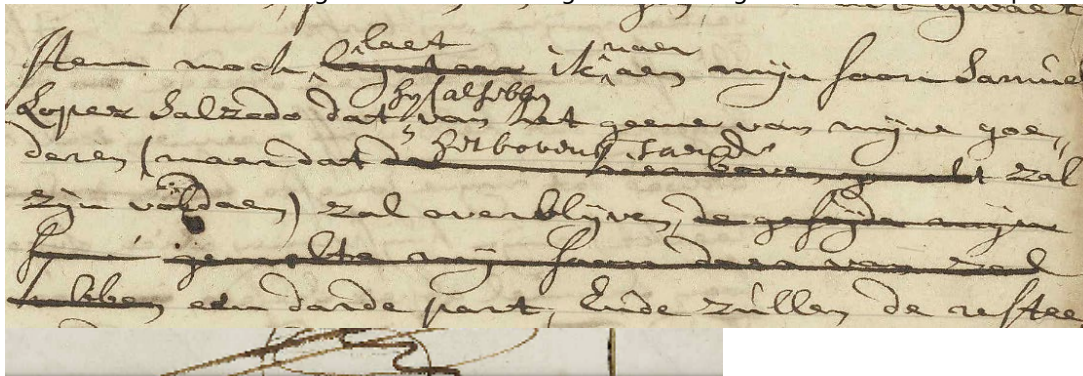
Transcriptieregels

Blijf zo dicht mogelijk bij de tekst

De algemene regel is dat de tekst zo letterlijk en volledig mogelijk moet worden overgenomen. Ga dus niet hertalen naar modern Nederlands, neem hoofdletters zoveel mogelijk over (behalve hoofdletters midden in een woord) en gebruik de juiste letters (staan er puntjes boven een 'ij', gebruik dan ook de 'ij'; staan er geen puntjes, gebruik dan de 'y').

Doorgehaalde tekst?

Indien leesbaar wordt doorgehaalde tekst voluit geschreven en gemarkeerd met de knop **abc**.

x²**abc**

Special Characters

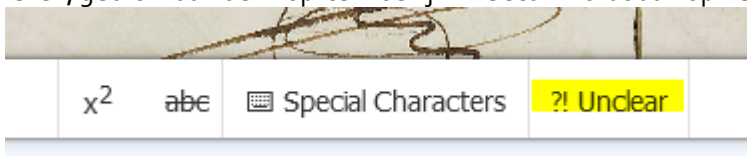
?! Unclear

Onleesbare tekst?

Letters, woorden of regels die door schade, vlekken of doorhaling onleesbaar zijn, worden niet getranscribeerd. Dit geldt ook als een stuk tekst niet leesbaar is omdat deze in de vouw van het notarisprotocol 'verdwijnt'. Probeer in dit geval alsnog wel de letters te transcriberen die u wel kunt lezen.

Onzekere lezing: unclear

Bent u niet helemaal zeker over een stukje uit de transcriptie, dan kunt u deze met de knop 'Unclear' selecteren en markeren. De controleur kijkt er dan naar. Gebruik deze functie niet voor elk detail. Kunt u heel veel niet lezen, gebruik dan de knop 'te moeilijk'. De scan wordt dan opnieuw uitgedeeld.

x²**abc**

Special Characters

?! Unclear

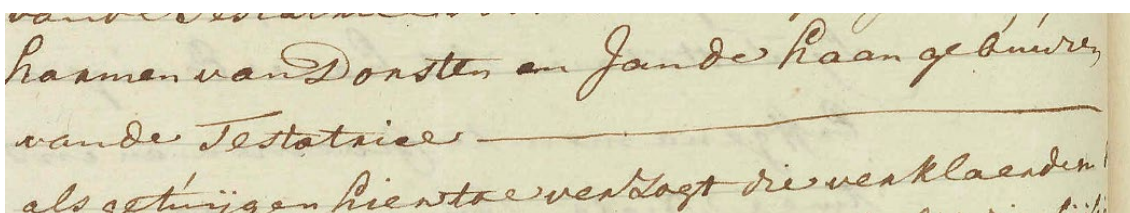
Tekst is onleesbaar of weggevalen: **niets transcriberen**

Tekst is doorgehaald maar dit lijkt er te staan: **abc**.

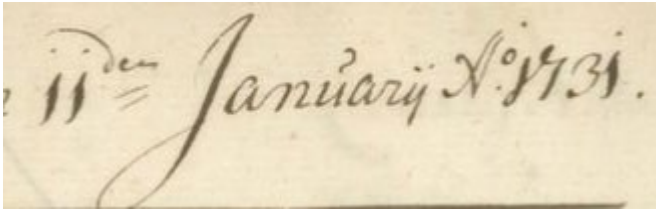
Ik weet niet of ik dit goed lees: **Unclear**.

Opmaak

De opmaak van de scan wordt niet overgenomen. Tabs, meer dan één spatie of algemene opmerkingen horen dus niet in het transcriptieveld. Daarnaast komen op enkele scans veel puntjes of streepjes voor (bijvoorbeeld in inhoudsopgaven en boedels). In principe wordt interpunctie overgenomen, maar als deze overduidelijk geen functie heeft en slechts dient om een bepaalde lay-out aan te geven, wordt deze niet overgenomen.



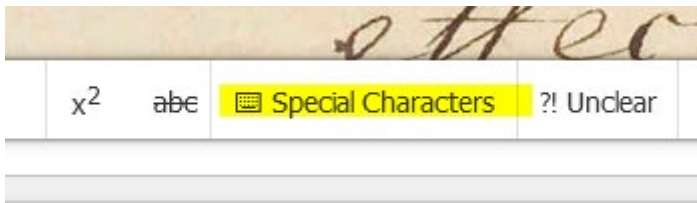
Interpunctie die geen doel dient, zoals hier een streep, wordt niet overgenomen.



Uitzondering: superscript mag wel worden ingevoerd d.m.v. de knop **x²**: 11^{den} Januarij A^o. 1731.

Sommigen notarissen hebben de neiging bijna alles aan elkaar te schrijven. In zo'n geval mag het spatiegebruik omwille van de leesbaarheid enigszins worden genormaliseerd.

Special Characters



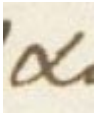
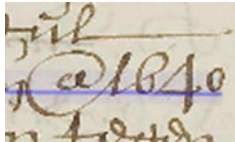
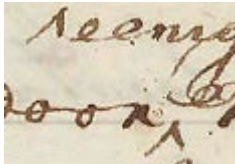

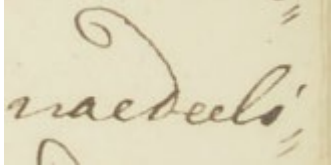
Afbrekingstekens aan eind van regel

Voor de afbrekingstekens aan het eind van de regel (vaak een afbrekingsstreepje of dubbele lage aanhalingstekens) gebruiken we altijd één hetzelfde teken: ¬ (deze unicode is te vinden onder de knop 'Special Characters' dat zich in de werkbalk tussen het scan- en transcriptieveld bevindt).

De onderstaande regels gelden vooral voor de zestiende- en zeventiende eeuwse handschriften.

Leestekens met unicode

Leestekens waarvoor een eigen unicode is, mogen als zodanig worden getranscribeerd met behulp van uw toetsenbord of de knop 'Special characters'. Enkele voorbeelden:

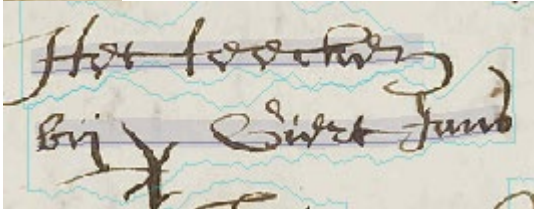
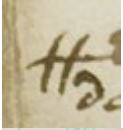
&	@	^	f (guldenteken)	¬ (regeleinde)
				
Alt 38	Alt 64	Alt 94	Alt 159	Alt 170
Shift+7	Shift+2	Shift+6		Alt+CTRL+\

Bijzondere tekens en afkortingen komen vooral voor in 17^e-eeuwse teksten. In het algemeen geldt: bestaat er een veelgebruikt teken voor (&, @) dan mag dat in de transcriptie gebruikt worden. Sommige duidelijke afkortingen mogen opgelost worden, zie hieronder wat voorbeelden.

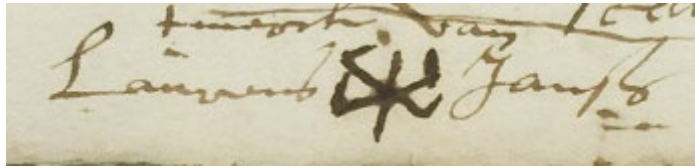
Maak u vooral niet te druk om de afkortingen en concentreer u op alle andere woorden die de computer óók moet leren.

Leestekens zonder unicode / vreemde tekens

Komt u een teken tegen waarvoor geen unicode bestaat en dat niet staat voor een specifieke afkorting? Transkribeer dan niks.



'het teeken
bij Giert Jans'





'Laurens Jansz'

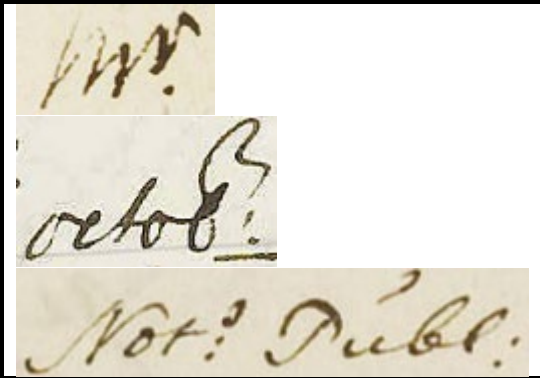
Afkorting

Afkorting die eenduidig te spellen zijn maar waarvoor geen unicode-karakter beschikbaar is, mogen worden uitgeschreven. Hieronder worden enkele voorbeelden getoond. Voor meer voorbeelden kunt u het hoofdstuk [Voorbeelden](#) raadplegen.

    	<p><u>Versoecke</u></p> <p><u>voorsz.</u> (dit is een uitzondering op de regel waarbij het 'ver'-teken wordt gebruikt voor 'voor', het kan ook gebeuren bij afkortingen voor 'voorn'.)</p> <p><u>Persoon</u></p> <p><u>Ende</u></p> <p><u>Vande</u></p>
--	---

Er zijn afkortingen waarvoor zowel geen unicode als geen eenduidige spelling is. Deze afkortingen zijn vaak te herkennen aan een streep boven het woord of het eindigen van het woord aan de bovenzijde van de regel. Volgens de transcriptieregel worden deze afkortingen **niet** uitgeschreven en beëindigd met een punt (of ander afbrekingsteken dat achter de afkorting zelf staat). Enkele voorbeelden:

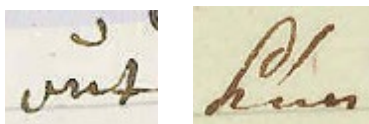
 	<p>voorn.</p> <p>voorss.</p>
--	------------------------------

	<p>Mr.</p> <p>octob:</p> <p>Not^s: Publ:</p>
---	--

Uitzonderingen, krullen en instinkers

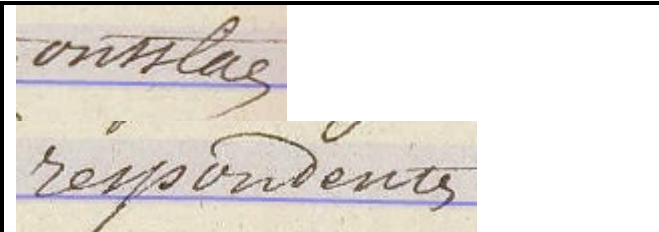
u/n

In vroegmodern Nederlands wordt vaak een streepje boven de letter 'u' genoteerd om deze te onderscheiden van de letter 'n', aangezien deze in vroegmodern handschrift gemakkelijk met elkaar verward konden worden. Omdat we met deze letters met moderne karakters wel duidelijk van elkaar kunnen onderscheiden, mag de letter 'u' met een streepje erboven genormaliseerd worden tot een 'u'.



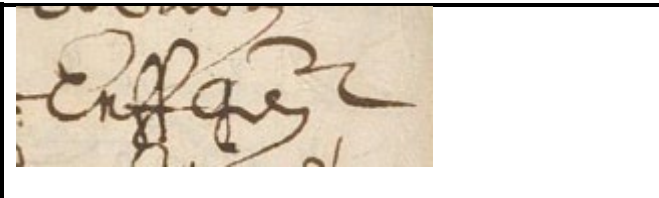
Afgeraffelde woorden

Het komt vaak voor dat in de akten woorden afgeraffeld worden. Dit zijn geen afkortingen, maar complete woorden die vaak aan het einde van het woord niet netjes afgemaakt zijn. Deze woorden moeten volledig uitgeschreven worden. Meestal gaat het hier om een samentrekking van de letters **-en**.

	<p>ontsla<u>en</u> (≠ ontlae / ontslas)</p> <p>Respondent<u>en</u> (≠ respondente / respondents)</p>
---	--

Halen en krullen aan einde afkorting/woord

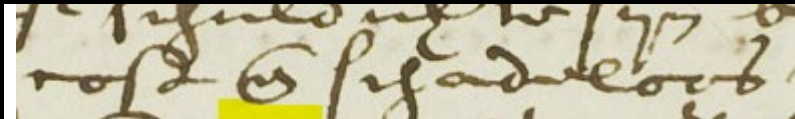
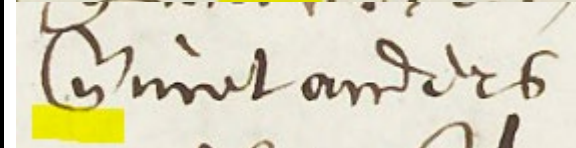
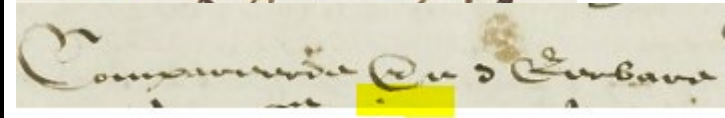
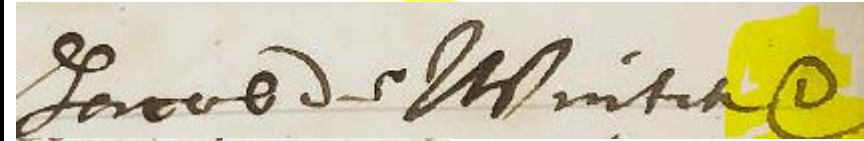

Onduidelijke krullen en overige haaltjes aan het eind van een woord of afkorting kunnen in veel gevallen genegeerd worden.

	<p>Erffgen (≠ Erffgenamen / Erffgenⁿ)</p>
---	--

Afkortingen 'etc.' en 'ende'

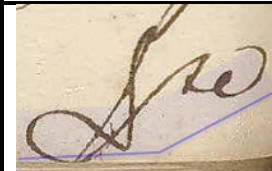
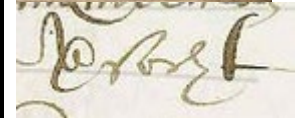
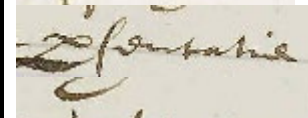
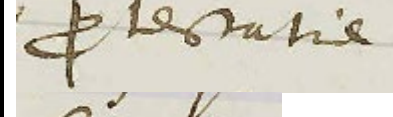
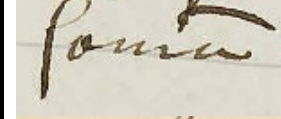
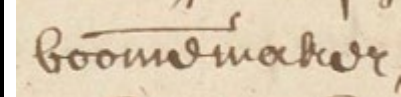
Voorals in zeventiende-eeuwse akten staan vaak afkortingstekens voor 'etc.' ('et cetera') en 'ende'. Vaak lijken

deze tekens op elkaar maar doorgaans valt uit de context op te maken om welke afkorting het gaat. Let op: vaak lijken deze afkortingen op het @-teken, maar ze zijn anders. Transcribeer ze dus **niet** met '@'.

	cost ende schadeloos
	ende niet anders
	Compareerde etc. d Eerbare
	Jacob de Winter etc.
	Nots etc. ende den

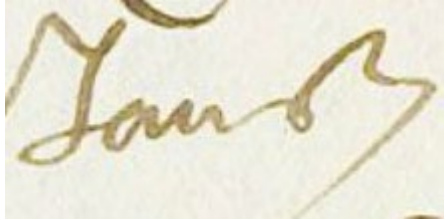
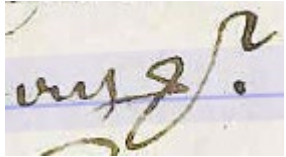
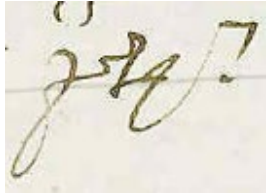
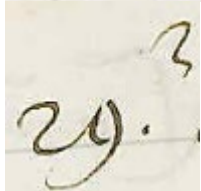

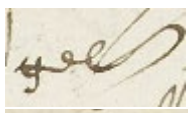
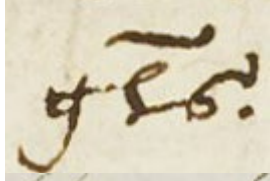
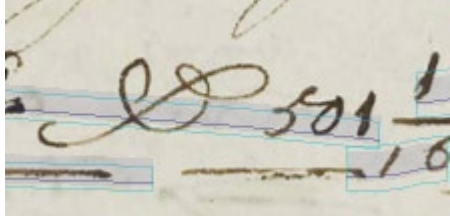
Voorbeelden

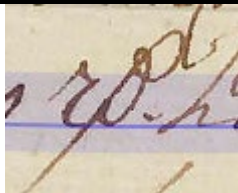
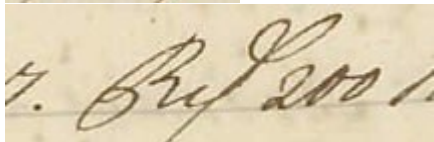
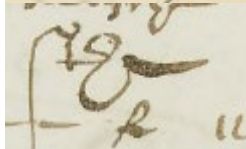
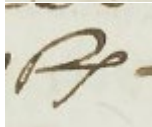
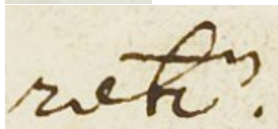
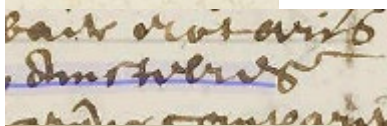
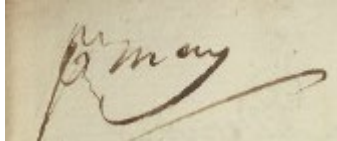
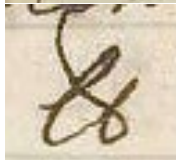
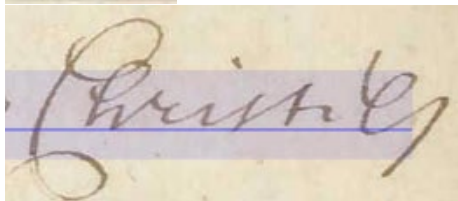
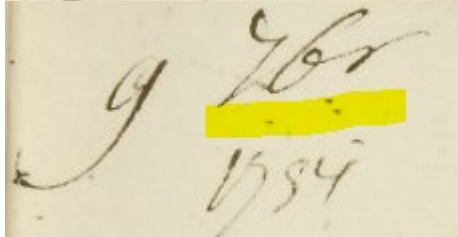
Afkortingen die uitgeschreven worden

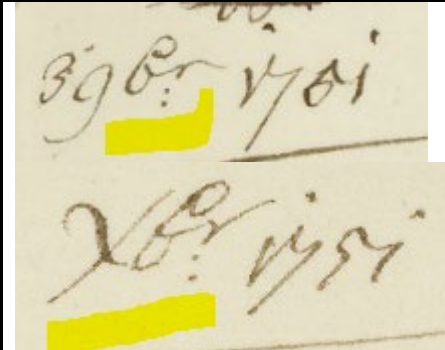
	<u>verte</u>
	<u>versocht</u>
	<u>presentatie</u>
	<u>protestatie</u>
	som <u>ma</u>
	boom <u>ma</u> maker

	leggende
	hierover

Afkortingen die niet uitgeschreven worden

	Jansz	<div style="border: 2px solid red; padding: 10px;"> <p>De Ringel-S aan het eind van patroniemen wordt getranscribeerd als 'sz' of 'ss' en niet uitgeschreven als 'ssen' of 'soon'.</p> </div>
	ont.	
	get.	
	29. ^e	
	a ^o	
	gel.	
	gl.(guldens)	
	D (daalder)	

	rd (rijksdaalder)
	Rd (rijksdaalder)
	stver (stuiver)
	stver (stuiver)
	rek ⁿ .
	Amstelre.
	P. May
	lb (=pondteken)
	Christel.
	9 7br : 1754 (9 <u>september</u> 1754)
	14 8b^r : 1751 (14 <u>oktober</u> 1751)

	<p>3 9b^r: 1751 (3 <u>november</u> 1751)</p> <p>Xb^r: 1751 (<u>december</u> 1751)</p>
---	---