

Protocol voor Orthografische Transcriptie Colleges Universiteit van Nederland

HoMed versie 1.0, augustus 2021¹

A1. Algemene inleiding

In opdracht van de Nederlandse Taalunie zullen we zo'n 50 colleges van de Universiteit van Nederland transcriberen. De aanpak die we hiervoor kiezen is een eerste transcriptie door een automatische spraakherkenner die vervolgens met de hand wordt gecorrigeerd. De tool die we voor de transcriptie gebruiken is ASRcorrector. Daarvoor is een aparte handleiding beschikbaar.

Transcriberen wil zeggen dat je letterlijk opschrijft (in gewone spelling) wat je de spreker hoort zeggen. Daarnaast moeten er in de tekst sprekerannotaties plaatsvinden die aangeven welke spreker op welk moment aan het woord is. De regels voor die laatste annotatie staan in de handleiding bij ASRcorrector beschreven. Dit protocol bevat de regels waaraan de orthografische transcriptie moet voldoen. Hierbij worden zoveel mogelijk de regels van de transcriptie van het CGN² aangehouden.

Aanvullingen en aanpassingen ten opzichte van het originele protocol in het kader van onderzoeksproject HoMed worden met groene tekstmarkering aangeduid.

A2. Definitie van de orthografische transcriptie

Een orthografische transcriptie is een *woordelijke* neerslag van hetgeen er gezegd is. De transcriptie sluit nauw aan op het *geschreven* Nederlands. Op bepaalde punten wordt er echter afgeweken van de *Nederlandse spellingsregels* of wordt er extra informatie toegevoegd. Op deze manier wordt de transcriptie maximaal consistent en bruikbaar voor allerlei verschillende wetenschappelijke doeleinden.

De orthografische transcriptie bevat tevens een *oplijning op segment-niveau*³. Ook de aanduiding van achtergrondgeluiden, en een veld voor opmerkingen, zijn voorzien in de transcriptie.

¹ Deze HoMed-versie is uitbreiding van Versie 2.2 (oktober 2020) van het oorspronkelijke protocol.

² Corpus Gesproken Nederlands: een groot nationaal project waarin honderden uren spraak werden opgenomen en getranscribeerd.

³ *een segment*: Plaats segmentgrenzen rond sprekerbeurten. Binnen segmenten komen een of meer uitingen voor die je met interpunctie markeert.

B. Hoofdregels

- B1. Gebruik de nieuwe spelling (volgens Groene Boekje⁴). Pas dus ook de nieuwe regels toe voor het plaatsen van koppeltekens (bijv.: zee_engte), de tussen-n (bijv.: pannennkoek), etc. (zie ook F2).
- B2. Werk zo nauwkeurig mogelijk, maar **verknoei geen tijd** op moeilijk verstaanbare stukken of stukken waar mensen door elkaar praten; duid ze aan met **xxx** of **Xxx** (zie C9 en C10).
- B3. Schrijf **een hoofdletter** aan het begin van een zin en bij eigennamen (bijv.: Bert Jansens, de Tweede Kamer), bij gespelde letters en bij titels van boeken, platen, etc. (bijv.: De Naam Van De Roos) (zie ook C1-3 en F3-5).
- B5. Gebruik *alleen punten en vraagtekens* als leestekens.
- B6. Schrijf getallen uit (dus niet: 2018 maar: **tweeduizend achttien**). Zie voor de conventies: <http://woordenlijst.org/leidraad/6/9>. Echter, schrijf elk getal als één aaneengesloten woord, zonder spaties, **behalve bij duizendtallen en miljoentallen**. Ook rangtelwoorden schrijf je aaneen en zonder cijfers. NB: de tekst moet de uitspraak volgen. Dus als iemand de postcode als 30 81 KP uitsprekt, moet je niet drieduizendeenentachtig schrijven maar dertig eenentachtig.

Bij twijfel: In de gevallen waar het protocol ontoereikend is, kies je voor de schrijfwijze die het meest aansluit bij de conventionele schrijfwijze.

⁴ Zie ook <http://woordenlijst.org/> en <http://woordenlijst.org/leidraad>

C. Nadere toelichting bij de gebruikte conventies

C1. **Gebruik hoofdletters voor eigennamen.** Onder eigennamen wordt verstaan: plaatsnamen, persoonsnamen, merknamen, bedrijfsnamen, verenigingsnamen, namen van commissies, ...

- ik woon in de Dorpsstraat in Den Haag.
- een Vlaamse schrijver; Nederlandstalig
- Porsche; Mercedes; Ford; Citroën
- Transport Vereecken; het transportbedrijf Vereecken
- Fokker Services; de serviceafdeling van Fokker
- Club Brugge; de voetbalclub Ajax
- de Tweede Kamer
- Zovirax⁵

Als een eigenaam uit meerdere woorden bestaat, schrijf dan ieder woord met een hoofdletter, ook al zou je dat volgens de Nederlandse spellingsregels niet doen⁶.

- Romain De Coninck; professor Van Veen
- Karel De Kale; Karel De Vijfde; Jan Zonder Vrees
- Roel In 't Veld (evt.: Roel-in-'t-Veld);

Waar 's-, d', l', ... (een kleine letter samen met een weglatingsteken) voorkomen in eigennamen, schrijf je deze met een kleine letter.

- 's-Gravenhage; mevrouw d'Ancona; de regel van l'Hopital

Wees consequent in je schrijfwijze: als je eenmaal een bepaalde spelling gebruikt hebt, gebruik dan dezelfde wanneer de eigenaam nogmaals voorkomt in hetzelfde fragment.

C2. **Gebruik hoofdletters voor titels van boeken, platen, rapporten, richtlijnen etc.** Als een titel uit meerdere woorden bestaat, worden alle woorden met een hoofdletter geschreven.

- ik heb Hugo Claus' boek Het Verdriet Van België gekocht.
- Heb jij De Naam Van De Roos al gelezen?
- De nieuwe richtlijn Cardiovasculair Risicomanagement is nu beschikbaar.
- NHG-Standaard⁷

C3. **Gebruik hoofdletters voor gespelde letters.** Namen en kentekens worden regelmatig gespeld. Bij gebruik van een (quasi) pilotenalfabet moeten de betreffende woorden worden gebruikt; deze dienen wel te beginnen met een hoofdletter (zie ook F3-5).

- Ik sta op de sneekkade S N E E K K A D E
- Ik sta op de sneekkade Sierra November Echo Echo Karel Karel Alpha Delta Echo

C8. **Gebruik xxx, -xxx, xxx- of -xxx- voor onverstaanbare uitingen.** Gebruik deze code wanneer je een woord, een deel van een woord of een reeks woorden niet goed heeft verstaan en je er ook niks van kunt maken.

- ik heb al xxx bladzijden gelezen.

⁵ Let op: gebruik alleen een hoofdletter voor merknamen van geneesmiddelen, niet voor stoffen en generieke aanduidingen. Bijvoorbeeld: Het middel aciclovir wordt verkocht onder de merknaam Zovirax.

⁶ Met deze spelling wijken we dus nadrukkelijk af van de conventionele spelling; we doen dit om een zo consistent mogelijke transcriptie te verkrijgen en om eigennamen gemakkelijk als een geheel te herkennen.

Standaarden worden wel met hoofdletter geschreven, maar modules niet. Dus: NHG-Standaard maar FTO-module. Dit is omdat dit de meest gangbare manier van schrijven is, al wordt dit in de praktijk niet consequent gedaan.

- ik heb al **xxx**-enzeventig bladzijden gelezen.
- ik heb al achten-**xxx**-tig bladzijden gelezen.

Transcribeer ook uitingen die niet bij het eigenlijke gesprek horen, die moeilijk te transcriberen zijn of veel achtergrondlawaai bevatten met **xxx**. Zet segmentgrenzen om **xxx**.

- C9. **Gebruik Xxx bij onverstaanbare eigennamen of titels.** Gebruik deze code wanneer je een woord niet goed hebt verstaan en er ook niets van kunt maken, maar toch zeker weet dat er een eigennaam of een titel gezegd is.

- dit truitje heb ik bij **Xxx** gekocht.
- ze zei dat **Xxx** nog zou komen.

D. Hoe gesproken taal neerschrijven?

- D1. **Schrijf woordelijk neer wat er gezegd is.** Hiermee bedoelen we dat je niets verandert aan wat er gezegd is; ook niet wanneer de spreker **foutieve zinsconstructies** gebruikt of **grammaticale fouten** maakt:

<i>je hoort en schrijft</i>	<i>je verbetert dit dus NIET tot</i>
de koe dat graast. de open venster. hun zijn er al.	de koe die graast. het open venster. zij zijn er al.

- D2a. D1a. In sommige gevallen is het **niet duidelijk** of **foneemdeleties** een grammaticale fout of een uitspraakfout zijn. Bijvoorbeeld ‘een leuk uitstap’ waarbij de doffe ‘e’ op het einde van ‘leuke’ verdwenen is. Schrijf op wat je hoort.

<i>je hoort en schrijft</i>	<i>je schrijft dus NIET</i>
Een leuk uitstap. De gemeente zorgt voor straten en fietspad .	Een leuke uitstap. De gemeente zorgt voor straten en fietspaden .

- D2. **Gebruik enkel woorden uit de Nederlandse spelling.** Dit lexicon zou in principe alle Standaardnederlandse woorden uit de *geschreven* taal moeten bevatten. We vermelden hier expliciet een aantal vormen die wel gebruikt moeten worden hoewel eventueel betwist kan worden of ze wel tot het Standaardnederlands behoren. Het gebruik van deze vormen is verplicht als ze door de spreker worden gebruikt.

- D2b. De volgende **gereduceerde vormen** (‘k, ‘t, ‘s, ‘m, ‘r, ‘ns, d’r, m’n, z’n, da’s, zo’n, ie) zijn goedgekeurd, schrijf ze dus ook wanneer ze gebruikt worden.

<i>je hoort</i>	<i>dus je schrijft</i>	<i>je schrijft dus NIET</i>
/kmoet hier weg/ /tis waar/ /savonds/	‘k moet hier weg ‘t is waar ‘s avonds	ik moet hier weg. het is waar. des avonds.
/ik hepðm gezien/ ⁸ /ik hepðr gezien/ /kijk ðs hier/ of /kijk ðns hier/ /dðr is geen plaats/ /dðrtussen, dðruit/ /mðn huis, zðn auto/ /das nie waar/	ik heb ‘m gezien ik heb ‘r gezien kijk ‘ns hier d’r is geen plaats d’rtussen, d’ruit m’n huis, z’n auto da’s niet waar	ik heb hem gezien. ik heb haar gezien. kijk eens hier. er/daar is geen plaats. ertussen, eruit . mijn huis, zijn auto. dat is niet waar.

⁸ Het symbool ð staat voor de schwa (de stomme of doffe “e”).

/zoon tas heb ik ook/ /daar is ie /	zo'n tas heb ik ook daar is ie	zo een tas heb ik ook. daar is hij
---	---	---

Gebruik deze vormen slechts waar ze van toepassing zijn, dus **niet** in volgende gevallen:

<i>je hoort</i>	<i>dus je schrijft</i>	<i>je schrijft dus NIET</i>
zo een tas heb ik ook/	zo een tas heb ik ook	zo'n tas heb ik ook.

D2c. De veelgebruikte Zuid-Nederlandse⁹ vormen “ge” en “gij” (en alle bijhorende werkwoordsvormen) worden aanvaard. Verander ze dus **niet** in “je” of “jij”.

<i>je hoort en schrijft</i>	<i>je schrijft dus NIET</i>
gij loopt snel zeg. gij moogt weggaan. zijt ge weg?	jij loopt snel zeg. jij mag weggaan. ben je weg?

D2d. De **verkleinwoorden op -ke, -ske, -eke** zijn toegelaten.

<i>je hoort en schrijft</i>	<i>je schrijft dus NIET</i>
meis ke , tikker ke . dings ke , boeks ke . huize ke , kapelle ke .	meisje, tikkert je . dinget je , boek je . huis je , kapellet je .

D2e. De volgende **uitspraakvarianten** waarbij de “d” gereduceerd wordt of verdwenen is, zijn **voldoende ingeburgerd in de schrijftaal**. Schrijf deze woorden en soortgelijke gevallen dan ook waar ze van toepassing zijn.

<i>je hoort en schrijft</i>	<i>je schrijft dus NIET</i>
ik hou van jou snij nu een ui fijn ik rij met de wagen	ik houd van jou snijd nu een ui fijn ik rijd met de wagen
een ouwe man dat gaat je niet in de kouwe kleren zitten een goeie grap een dooie man een kwaai e kerel rooie haren	een oude man dat gaat je niet in koude kleren zitten een goede grap een dode man een kwade kerel rode haren

Let op: Sommige varianten zijn **onvoldoende ingeburgerd in de schrijftaal**. Als je dergelijke varianten hoort, schrijf dan de Standaardnederlandse vorm.

<i>je hoort</i>	<i>maar toch schrijf je</i>
/t is lang geleje/ /ik ben niet echt tevreden/	't is lang geleden. ik ben niet echt tevreden

Als uitzondering hierop beschouwen we **effe**, dat je wel zo moet transcriberen. Andere typische spreektaaluitingen moet je transcriberen zoals je ze hoort: **oh, hè, hé, ah, ja ja, okee**.

D3. **Probeer niet weer te geven hoe de woorden zijn uitgesproken**. Dit zou immers al snel leiden tot een soort fonetische transcriptie. Dit is **niet** de bedoeling van een **orthografische** transcriptie.

⁹ Alle typisch Zuid-Nederlandse fenomenen worden aangeduid met een grijze achtergrond.

D3a. Duid **verbindingsklanken** tussen twee woorden **niet** aan. (In de conventionele spelling gebeurt dit ook niet).

<i>je hoort</i>	<i>maar toch schrijf je</i>
/het leuk <u>e</u> nis/ /speeld <u>e</u> nij goed?/ /liep <u>t</u> ie snel?/ /wadist <u>e</u> r hier aan de hand?/	het leuke is. speelde hij goed? liep ie snel? wat is er hier aan de hand?

D3b. Duid **kleine klankveranderingen niet** aan in de transcriptie.

<i>je hoort</i>	<i>maar toch schrijf je steeds</i>
/o <u>o</u> too/ (NI) of /o <u>t</u> too/ (VI) /terap <u>u</u> it/ (NI) of /terap <u>e</u> ut/ (VI) /tr <u>e</u> m/ (NI) of /tr <u>a</u> m/ (VI) /m <u>a</u> rt/, /s <u>r</u> ift/, /a <u>s</u> jebli <u>e</u> ft/	Auto Therapeut Tram markt, s chrift, a lsjebli <u>e</u> ft

F. Aanvullende regels en tips

- F1. Schrijf **getallen** steeds voluit **zoals ze worden uitgesproken**. De getallen worden aan elkaar geschreven. Scheiden met een spatie mag je alleen bij miljoentallen en **duizendtallen**. Verder wordt alles gescheiden door spaties.

<i>je hoort en schrijft</i>
drieënvijftig, ee nentwintig. tweehonderddrie. tweehonderd en drie. achtthonderdzevenendertig. duizend achthonderdzevenendertig. twaalfduizend tweehonderddrieënvijftig. vijfhonderdduizend. driemiljoen vijfhonderdtweeëndertigduizend. vijftien komma zesentwintig twaalf.

Als je het woord /één/ hoort, schrijf je ook “één”, als je /ɔn/ of /n/ hoort, schrijf dan “een” (zonder accenttekentjes).

<i>je hoort</i>	<i>je schrijft</i>
/ɔn auto/ /één auto/ /er is één en ander gaande/	een auto één auto er is één en ander gaande

Schrijf rangtelwoorden en getallen die deel uitmaken van een woord steeds voluit.

<i>je hoort en schrijft</i>	<i>je schrijft</i> NIET
de tweede keer Een veertienjarige	de 2de keer; de 2^{de} keer. een 14 -jarige.

Tenzij ze deel uitmaken van een naam, **medische term of geneesmiddel** (zie ook F4).

<i>je hoort</i>	<i>je schrijft</i>
/de politieke partij dee zesenzestig/ /de snelweg aa drieënzeventig/ /diabetes mellitus type één/ /gee el pee één antagonist/	de politieke partij D66. de snelweg A73. diabetes mellitus type 1 GLP1-antagonist

- F2. Gebruik het **koppelteken** zoals in geschreven taal. Bijvoorbeeld in de volgende gevallen:

<ul style="list-style-type: none"> * een in- en uitgang; meelopen of -rennen; * een zee-engte; de West-Europese economie; * een glas-in-loodraam; vraag-en-aanbodssituatie; * 's-Gravenhage.

Heel wat samenstellingen worden aan elkaar geschreven, dus zonder koppeltekens (zie ook F5 en C4).

Gebruik **geen** gedachtestreepje tussen twee woorden of bij .

<i>je schrijft</i>	<i>je schrijft</i> NIET
de stand was één twee de match Brugge Anderlecht	de stand was 1 – 2. de match Brugge – Anderlecht.

Gebruik geen weglatingsteken aan het begin of einde van woorden.

je schrijft	je schrijft NIET
hart en vaatziekten	hart- en vaatziekten

F3. Schrijf **gespelde letters (of klanken)** met hoofdletters, zoals in onderstaande tabel. Meerdere gespelde letters die na elkaar komen, worden gescheiden door een spatie.

je hoort	je schrijft	je hoort	je schrijft	je hoort	je schrijft	je hoort	je schrijft	je hoort	je schrijft
/aa/	A	/gee/	G	/em/	M	/es/	S	/zet/	Z
/bee/	B	/haa/	H	/en/	N	/tee/	T	/au/	AU of OU
/see/	C	/ie/	I	/oo/	O	/uu/	U	/ij/	IJ of EI
/dee/	D	/jee/	J	/pee/	P	/vee/	V	/eu/	EU
/ee/	E	/kaa/	K	/kuu/	Q	/wee/	W	/oe/	OE
/ef/	F	/el/	L	/er/	R	/iksl/, /ieks/	X	/ui/	UI

Schrijf nooit 'Y', maar schrijf wat je hoort: **i-grec, IJ, ypsilon, Griekse IJ**, etc.

<ul style="list-style-type: none"> * yoghurt wordt gespeld als i-grec O G H U R T. * X IJ Z zijn de drie laatste letters van het alfabet. * hij is een lijder met lange IJ. * baan met twee A's. * mevrouw d'Ancona met een klein D'tje. * meneer A M De Groot.

F4. Schrijf **afkortingen** waar die gebruikt worden door de spreker. Schrijf de letters van een afkorting aan elkaar vast, dus zonder spatie ertussen. Schrijf alle afkortingen altijd met hoofdletters, ook wanneer dat in de conventionele spelling niet (altijd) gebeurt.

je hoort	je schrijft
/bee tee wee/, /el pee gee/, /gee es em/ /ex wee weeër/, /ee haa bee oojer/ of /ee haa bee ooër/ /bee tee wee heffing/, /gee es em gesprek/ de /wee see/, een /tee vee programma/ /aa uu beel/, /tee zet tee/	BTW, LPG, GSM. ex-WW'er, EHBO'er. BTW-heffing, GSM-gesprek. de WC, een TV-programma. AUB, TZT (dus niet "a.u.b.", "t.z.t." zoals in de schrijftaal).
/uval/ OF /uu vee aa/ /er uu gee/ OF /rug/ /vee bee oo mavo klas/ doaks en es aa ie dees es es er ies	UVA RUG Vmbo-mavo-klas DOAC's NSAID's SSRI's
de groep /innekses/, de partij /dee zesenzestig/	de groep INXS , de partij D66

Schrijf **nooit** een afkorting wanneer er **geen** gezegd werd:

je hoort en schrijft	je schrijft NIET
Dat wil zeggen eventueel met andere woorden	d.w.z. evt. m.a.w.

F5. Schrijf **letterwoorden** zoals in geschreven taal, maar steeds volledig met hoofdletters.

<ul style="list-style-type: none"> * een experiment van de NASA. * een uitzending van de TROS of van de VARA.
--

F6. Probeer te letten op de regels met betrekking tot het al dan niet **aaneenschrijven** van woorden. Schrijf bij twijfel de woorden aaneen, de spellingschecker zal dan automatisch corrigeren.

<i>je twijfelt tussen</i>		<i>je schrijft</i>	<i>na spellingscontrole wordt dit</i>
woensdag morgen. Hij gaat er van uit.	woensdagmorgen. hij gaat ervanuit.	woensdagmorgen. hij gaat ervanuit.	woensdagmorgen. hij gaat ervan uit.

In sommige gevallen is er een betekenisverschil tussen het aaneengeschreven woord en het niet aaneengeschreven woord. Let dus goed op:

<i>aaneen</i>	<i>niet aaneen</i>
dat zei ik toch nietwaar ? Hij is tenslotte net vader geworden (= per slot). Ze is nu tenminste tevreden (= in ieder geval).	dat is niet waar . dan heb ik ten slotte nog een opmerking (= tot slot). ik wil ten minste 90 jaar worden (= op z'n minst).

Heel wat woorden worden aaneengeschreven:

* <u>erv</u> andaan, <u>er</u> bovenop, <u>erna</u> artoe, <u>er</u> middenin.
* d' <u>r</u> vandaan, d' <u>r</u> bovenop, d' <u>r</u> maartoe, d' <u>r</u> middenin.
* hij gaat <u>er</u> van uit dat; die ziet <u>er</u> uit als; het zit <u>er</u> op.
* vijf <u>pr</u> ocentseis, lagel <u>o</u> nenland, hogesnelheids <u>l</u> ijn.

Probeer te letten op de regels met betrekking tot het al dan niet aaneenschrijven van woorden. Schrijf bij twijfel de woorden aaneen.

F7. Schrijf websites uit op de volgende manier:

- Health Base punt NL
- RIVM punt NL