Protocol voor Orthografische Transcriptie

(Laatste wijziging: 27 september 2000)

A1. Algemene inleiding

Het project *Corpus Gesproken Nederlands* (CGN) is een Vlaams-Nederlandse samenwerking met als doel ongeveer 10 miljoen woorden gesproken Nederlands te verzamelen. Het project wordt gefinancierd door de Vlaamse regering, de Nederlandse regering en de Nederlandse Organisatie voor Wetenschappelijk Onderzoek (NWO).

Op technologisch vlak is een dergelijk corpus essentieel voor de verdere ontwikkeling van Nederlandstalige taal- en spraaktechnologie, en helpt het op deze manier de toekomst van het Nederlands als cultuurtaal in het veeltalig Europa veilig te stellen. Het Corpus Gesproken Nederlands is ook voor andere onderzoeksgebieden van bijzonder belang. Voorbeelden zijn lexicografie, taalonderwijs, spraak- en taalontwikkeling bij kinderen, sociolinguïstiek, psycholinguïstiek, fonetiek, fonologie en conversatieanalyse.

Om het corpus bruikbaar te maken voor de verschillende onderzoeksgebieden is het noodzakelijk dat elk stukje opgenomen spraak vergezeld is van een orthografische transcriptie. Dit protocol bevat de regels waaraan deze orthografische transcriptie moet voldoen.

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*[1] op *chunk-niveau*[2] en een aanduiding van de verschillende sprekers. Ook de aanduiding van achtergrondgeluiden en een veld voor opmerkingen, zijn voorzien in de transcriptie.

---

[1] *oplijnen:* het koppelen van een stukje transcriptie aan een stukje geluidssignaal door middel van het plaatsen van grenzen. Deze grenzen bepalen de begin- en eindtijd van de stukjes transcriptie. Zie verder de handleiding bij het programma Praat voor meer details over het plaatsen van grenzen.

[2] *een chunk:* een stukje spraak van ongeveer 2 à 3 seconden dat aan beide zijden begrensd wordt door een (korte) zichtbare en hoorbare pauze. Chunks hoeven niet overeen te komen met zinnen of zinsdelen, ze worden enkel bepaald door de pauzes in het spraaksignaal. Het wordt aangemoedigd om (indien het geluidssignaal het toelaat) de chunks zo kort mogelijk te houden. Verdere richtlijnen hieromtrent worden gegeven in de handleiding bij het programma Praat.

B. Hoofdregels

B1. Gebruik de nieuwe spelling. Pas dus ook de nieuwe regels toe voor het plaatsen van koppeltekens (bijv.: zee**-**engte), de tussen-n (bijv.: panne**n**koek), 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 **geen hoofdletter** aan het begin van een zin. Schrijf **wél hoofdletters** bij eigennamen (bijv.: Bert Jansens) en bij titels van boeken, platen, etc. (bijv.: De Naam Van De Roos) (zie ook C1 en C2).

B4. Gebruik waar nodig of zinvol, de volgende codes:

- gebruik ***v** voor woorden uit een **vreemde taal** (bijv.: tomorrow*v) (zie ook C3);
- gebruik ***d** voor **dialectwoorden** (bijv.: kortewagen*d; keuje*d) (zie ook D4);
- gebruik ***z** voor **woorden** uit de standaardtaal die **zwaar dialectisch zijn uitgesproken** (zie D5);
- gebruik ***n** voor **nieuwe woorden** (bijv.: zerotolerantie*n) (zie ook C4);
- gebruik ***t** voor **nieuwe tussenwerpsels** (bijv.: amaai*t; hoppa*t) (zie ook C5);
- gebruik ***a** voor **afgebroken woorden** (bijv.: uitges*a; verpr*a) (zie ook C6);
- gebruik ***u** voor het weergeven van de **uitspraak** en **versprekingen** (bijv.: boink*u; gespreken*u) (zie ook C7);
- gebruik ***x** bij woorden waarvan u **niet** zeker bent of u ze **goed** heeft **verstaan** (zie C8).

Gebruik nooit meer dan één dergelijke code per woord. Kies eventueel voor de meest toepasselijke code.

B5. Controleer of de transcriptie enkel woorden bevat die in het CGN-lexicon[3] staan. Gebruik hiervoor de speciale CGN-spellingschecker.

B6. Gebruik *enkel* de leestekens **punt**, **vraagteken** en **beletselteken** (**.** / **?** / **...**). Gebruik geen andere leestekens, zoals komma, uitroepteken, aanhalingstekens, e.d. (zie ook sectie E).

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

B7. Gebruik **ggg** voor **duidelijk hoorbare *spreker*geluiden** (zie ook C11).

B8. Gebruik de ***speaker_unknown***-tier[4] voor stukken waarvan u niet (zeker) weet door welke van de sprekers ze gezegd zijn, of voor spraak van een spreker die slechts enkele woorden zegt en waarvoor geen aparte tier gecreëerd moet worden.

B9. Gebruik de ***background***-tier voor localiseerbare **achtergrondgeluiden** die zich duidelijk manifesteren en/of het verloop van het gesprek beïnvloeden (zie C12).

B10. Gebruik het ***comment***-veld voor alle soorten commentaar over de opname en voor opmerkingen over achtergrondgeluiden die over de gehele opname te horen zijn (zie C13).

---

[3] *CGN-lexicon:* een woordenlijst met alle woorden die we als "Standaardnederlands" beschouwen. De lijst is echter onvolledig en zal in de loop van het project verder uitgebreid worden (zie "*n").

[4] *een tier:* een tier lijkt op een notenbalk of een lijn waarop alle spraak van één bepaalde spreker getranscribeerd wordt (inclusief de oplijning op chunk-niveau (zie A2)). Voor iedere spreker wordt een dergelijke tier voorzien. Al deze tiers of (noten)balkjes visualiseren samen het verloop van het gesprek en de interacties. Dit is vergelijkbaar met een muziek- of toneelpartituur (zie ook handleiding bij het programma Praat).

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 **d**e **D**orpsstraat in **D**en **H**aag.
- een **V**laamse schrijver; **N**ederlandstalig
- **P**orsche; **M**ercedes; **F**ord; **C**itroën
- **T**ransport **V**ereecken; het **t**ransportbedrijf **V**ereecken
- **F**okker **S**ervices; de **s**erviceafdeling **v**an **F**okker
- **C**lub **B**rugge; de **v**oetbalclub **A**jax

Als een eigennaam uit meerdere woorden bestaat, schrijf dan ieder woord met een hoofdletter, ook al zou u dat volgens de Nederlandse spellingsregels niet doen[5].

- Romain **D**e Coninck; professor **V**an Veen
- Karel **D**e Kale; Karel **D**e Vijfde; Jan **Z**onder Vrees
- **R**oel **I**n **'t** **V**eld (evt.: **R**oel-in-'t-Veld);

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

- **'s-G**ravenhage; mevrouw **d'A**ncona; de regel van **l'H**opital

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

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

- ik heb Hugo Claus' boek **H**et **V**erdriet **V**an **B**elgië gekocht.
- Heb jij **D**e **N**aam **V**an **D**e **R**oos al gelezen?

C3. **Gebruik *v voor woorden uit een vreemde taal.** Hieronder wordt verstaan: woorden die niet (in die betekenis of met die uitspraak) voorkomen in de woordenschat van het Standaardnederlands maar wel in een vreemde taal.

- **he's*v a*v rebel*v** is Engels voor hij is een rebel.
- vertrek nu maar snel. **see*v you*v tomorrow*v**.
- **cave*v canem*v** is Latijn voor pas op voor de hond.

Ook vreemde woorden in titels van boeken, films, etc. krijgen een ***v**.

- de band **New*v Kids*v On*v The*v Block*v**.
- heb je **La*v Vie*v En*v Rose*v** al eens gehoord?

Eigennamen uit een vreemde taal krijgen **geen** ***v**.

- het liedje **Josephine** van **Chris Rea**.
- heb je **Grace Jones'** laatste hit al gehoord?

Ook vreemde woorden die in het Nederlands zijn ingeburgerd, krijgen **geen** ***v**.

- bij het opstarten van de **computer** krijg je een **jingle** te horen. **save** jij die **file** eens?
- zal ze **überhaupt** wel te weten komen waar **Freetown** ligt?

[5] 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.

C4. **Gebruik *n voor nieuwe woorden**. Hieronder wordt verstaan: woorden die volgens u correct Nederlands zijn maar die, bij spellingscontrole, (nog) niet in het lexicon blijken te staan. Hieronder vallen onder andere neologismen die langzaam ingeburgerd raken in ons taalgebruik.

- **zerotolerantie*n, topprioriteit*n, millenniumprobleem*n, vijfprocentseis*n.**

C5. **Gebruik *t voor tussenwerpsels.** Enkel de tussenwerpsels die (nog) niet in het CGN-lexicon staan, moeten een ***t** krijgen.

- **tsjakka*t** die zit.
- **amaaikes*t** wat gaat er nu gebeuren.

Heel wat tussenwerpsels staan reeds in het CGN-lexicon; deze moeten **geen *t** krijgen. Probeer een tussenwerpsel eerst zonder ***t** te schrijven. Voeg pas ***t** toe wanneer de spellingschecker protesteert (of aanvaard de suggestie van de spellingschecker).

- **nou** dat denk ik niet **hoor**.
- **sjonge** wat hebben we nou?
- **allee** dat meen je toch niet **zeker**.
- **awel** 't is goed **zulle**.
- **hé** kom jij 'ns even hier alsjeblieft.
- ik **uh** weet het niet goed **zuh**.
- **mmm** dat weet ik nog zo niet.

Om het transcriberen van tussenwerpsels een beetje **uniform** te maken, geven we hier een lijst van veelvoorkomende, korte tussenwerpsels. Schrijf ze **enkel** zoals in de lijst.

ah	bwa	hu	oesje	tut (ook "tut tut")
aha	eih, eikes	hum (ook "hum hum")	oh, o	uh, uhm
ai	goh	jee (ook "o jee")	oho	uhu
au	ha, haha	mm-hu	poeh	wauw
bah	hé, hè, hei	mmm	pst	woh
boe	ho	oeh, oei	sjt, sst	zuh, zulle

Als een woord onderbroken wordt door een tussenwerpsel (doorgaans door "uh"), gebruik dan koppeltekens om de samenstellende delen en het tussenwerpsel aan elkaar te koppelen en voeg ***u** toe (zie ook C7c):

- heb jij het geld **over-uh-gemaakt*u?**
- ze heeft haar ring **ver-uh-uh-patst*u**.

C6. **Gebruik *a voor afgebroken woorden.** Gebruik deze code wanneer een woord, om welke reden dan ook, onvolledig is uitgesproken. Dit geldt ook voor woorden die zijn afgebroken maar daarna nog eens volledig zijn uitgesproken, zoals in het derde voorbeeld hieronder. Gebruik *a ook bij versprekingen die zijn afgebroken en niet worden hervat (vergelijk C7c). Als maar één klank is uitgesproken, zoals bijv. in <m*a ik weet het niet>, beschouw die klank dan als een afgebroken woord en gebruik *a, al weet u niet *welk* woord er is afgebroken. Gebruik als een woord in z'n geheel is herhaald, dus bijv. <ik ik ik weet het niet> **geen** code en **geen** koppeltekens.

- ik ga **mo*a** overmorgen naar de tandarts.
- neem de **link*a** uh de rechter deur.
- ik moet **mo*a** morgen naar de tandarts.
- ik moet **overm*a**... ik moet overmorgen naar Tilburg.
- **m*a**… ik weet het niet.
- ze heeft **bel*a** boulimia nervosa.
- het goede weer zal een **voelpr*a voorproefje** zijn.

C7. **Gebruik *u voor het weergeven van de uitspraak.** Gebruik deze code wanneer iemand, om welke reden dan ook, een woord zegt dat niet (in die betekenis) in het CGN-lexicon staat, en ook niet tot één van de andere types (*v, *d, *n, *t, *z, *x, *a) behoort. Typische voorbeelden zijn:

C7a. klanknabootsingen en (opzettelijke) vervormingen:

- **boink*u** hoorde ik. hij was tegen de muur gevlogen.
- ik **spreuk*u** een beetje **Zweuds*u**.
- hij zegt voor de grap **toekenbas*u** in plaats van boekentas.

C7b. versprekingen:

- ik heb me **verspraakt*u**. dit was een **versprekeling*u**.
- mmm ik lust wel een **broekje*u** kaas.
- KPN annuleert fusie met Spaans bedrijf **alduns*u aldus** een kop in de krant.

C7c. alle gevallen van versprekingen en hernemingen **binnen een woord**. Gebruik in deze gevallen *u in combinatie met koppeltekens om de verschillende woorddelen aan elkaar te koppelen. Ook woorden die onderbroken worden door een tussenwerpsel of een sprekergeluid (bijv. door gelach) vallen in deze categorie.

- hij is nog niet naar de kapper **gewee-weest*u**
- de verzetsstrijders van het **Kosovo-bevrijdingslever-leger*u**.
- de **jeugdlied-liefdes*u** van de schrijver.
- dat is de **mogelijkhee-heid*u**.
- heb jij dat spul nog over het gras **uitges-ge-uh-gesproeid*u**?
- dat is een **ongeluks-uh-baby*u**.
- ze heeft haar ring **ver-uh-uh-patst*u**.
- ze heeft haar ring **ver-ggg-patst*u**.

Let op: gebruik dus ***u** wanneer de spreker zich verspreekt, maar uiteindelijk wel alle delen van het desbetreffende woord uitspreekt. Wanneer de spreker een woord afbreekt en het niet hervat, of het volledig hervat, dus bij onvolledig uitgesproken woorden, gebruikt u ***a** (zie C6).

C8. **Gebruik *x bij woorden waarvan u niet zeker bent of u ze goed heeft verstaan.** Gebruik deze code wanneer u, ondanks de moeilijke verstaanbaarheid van een uiting, bijna zeker weet wat er gezegd is.

- hij kan **dertig*x** bladzijden per uur uit het hoofd leren.
- ze zei dat **Jan-Peter*x** nog zou komen.

C9. **Gebruik xxx, -xxx, xxx- of -xxx- voor onverstaanbare uitingen.** Gebruik deze code wanneer u een woord, een deel van een woord of een reeks woorden niet goed heeft verstaan en u er ook niks van kunt maken.

- ik heb al **xxx** bladzijden gelezen.
- 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**.

C10. **Gebruik Xxx bij onverstaanbare eigennamen of titels**. Gebruik deze code wanneer u een woord niet goed heeft 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.

C11. **Gebruik ggg voor duidelijk hoorbare *spreker*geluiden** zoals gelach, gehuil, gegil, gekuch, etc. Enkele voorbeelden:

- even m'n keel schrapen hoor. **ggg**. zo dat is beter.
- ken je die mop van die man die naar Parijs ging? nou hij ging niet. **ggg**.
  (transcribeer dus niet "haha" of iets dergelijks wanneer er gelachen wordt).
- spreker A: **ggg**. (transcribeer dus niet "hatsjie" of iets dergelijks bij niezen).
  spreker B: gezondheid.

Let op: Het is nadrukkelijk **niet** de bedoeling om allerlei **zwakke sprekergeluiden** aan te duiden, zoals:

- in- en uitademen, zacht kuchje, gesmak met de lippen, klikje met de tong.
- zachtjes schrapen van de keel, licht krakend keelgeluid bij het begin van een uiting, etc.

C12. **Gebruik de *background*-tier voor duidelijk hoorbare achtergrondgeluiden** waarvan u meent dat het nuttig kan zijn om te weten dat ze zijn opgetreden en/of geluiden die duidelijk het verloop van het gesprek beïnvloeden.

Het is beslist **niet** de bedoeling dat u zich **inspant** om het begin- en eindtijdstip van deze geluiden exact aan te duiden. Zorg ervoor dat de grenzen **ongeveer** goed staan.

Een reeks aparte geluiden die (inhoudelijk) samenhoren duidt u aan als één geheel, met één beginpunt en één eindpunt.

- U hoort een applaus tijdens of na een lezing.
  → Maak dan een chunk met "**applaus**", de exacte positie van het begin- en eindpunt doen weinig ter zake.
- U hoort duidelijk voetstappen en daarna een sleutel die in het slot gestoken wordt. Eén van de sprekers zegt : "ha, mijn vrouw komt thuis.".
  → Transcribeer "**voetstappen + sleutel in slot**" als één geheel, de exacte tijdstippen doen weinig ter zake
- Tijdens een gesprek van een half uur, is er vijf minuten een drilboor te horen. Het geluid wordt nu en dan enkele seconden onderbroken.
  → Maak slechts één chunk van ongeveer vijf minuten met de vermelding "**drilboor**".
- U hoort een luid blaffende hond, dit geblaf manifesteert zich duidelijk in het gesprek.
  → Transcribeer dit blaffen als één geheel met "**blaffende hond**", ook al zwijgt de hond nu en dan.

Let op: Het is nadrukkelijk **niet** de bedoeling dat u zich inspant om **zwakke en weinig informatieve geluiden**, net zoals **kleine artefacten in de opname**, aan te duiden, zoals:

- het klikken van een balpen tijdens een vergadering
- het vallen van bijvoorbeeld een balpen
- het kraken van een stoel
- het tikken van een klok
- een piepje of gekraak in de opname

C13. Gebruik het ***comment***-veld voor geluiden waardoor de gehele opname wordt gekenmerkt of voor zaken die u opvielen bij het beluisteren van de opname. In het comment-veld wordt geen tijdsinformatie opgenomen (er worden geen begin- en eindpunten aangeduid). Enkele voorbeelden van commentaar:

- er is achtergrondmuziek te horen tijdens de hele opname
- nu en dan is er geroezemoes te horen van andere sprekers, verspreid over de hele opname
- het hele interview is opgenomen in een rijdende wagen
- in het tweede deel van de opname is er regelmatig een parkiet te horen

D. Hoe gesproken taal neerschrijven?

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

u hoort en schrijft	*u verbetert dit dus **NIET** tot*
de koe **dat** graast. **de** open venster. **hun** zijn er al.	de koe **die** graast. **het** open venster. **zij** zijn er al.

Wanneer u echter zeker weet dat de spreker zich *verspreekt*, gebruikt u ***u** (zie C7b).

D2. **Gebruik enkel woorden uit het CGN-lexicon.** Dit lexicon zou in principe alle Standaardnederlandse woorden uit de *geschreven* taal moeten bevatten. We vermelden hier expliciet een aantal vormen die wel degelijk in het CGN-lexicon zitten (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.

D2a. De volgende **gereduceerde vormen** (’k, ’t, ’s, ’m, ’r, ’ns, d’r, m’n, z’n, da’s, zo’n, ie) behoren tot het lexicon, schrijf ze dus ook wanneer ze gebruikt worden.

u hoort	*dus u schrijft*	*u schrijft dus **NIET***
/**km**oet hier weg/ /**tis** waar/ /**sa**vonds/	**’k** moet hier weg. **’t** is waar. **’s** avonds.	**ik** moet hier weg. **het** is waar. **des** avonds.
/ik hep**∂m** gezien/[6] /ik hep**∂r** gezien/ /kijk **∂s** hier/ of /kijk **∂ns** hier/ /**d∂r** is geen plaats/ /**d∂r**tussen, **d∂r**uit/ /**m∂n** huis, **z∂n** auto/	ik heb **’m** gezien. ik heb **’r** gezien. kijk **’ns** hier. **d’r** is geen plaats. **d’r**tussen, **d’r**uit. **m’n** huis, **z’n** auto.	ik heb **hem** gezien. ik heb **haar** gezien. kijk **eens** hier. **er**/**daar** is geen plaats. **er**tussen, **er**uit. **mijn** huis, **zijn** auto.
/**das** nie waar/ /**zoon** tas heb ik ook/	**da’s** niet waar. **zo’n** tas heb ik ook.	**dat is** niet waar. **zo een** tas heb ik ook.
/daar is **ie**/	daar is **ie.**	daar is **hij.**

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

u hoort	*dus u schrijft*	*u schrijft dus **NIET***
/**zo een** tas heb ik ook/	**zo een** tas heb ik ook.	**zo’n** tas heb ik ook.

D2b. De veelgebruikte Zuid-Nederlandse[7] vormen “**ge**” en “**gij**” (en alle bijhorende werkwoordsvormen) worden aanvaard binnen het CGN-lexicon. Verander ze dus **niet** in “je” of “jij”.

u hoort en schrijft	*u schrijft dus **NIET***
gij loopt snel zeg. **gij moogt** weggaan. **zijt ge** weg?	**jij** loopt snel zeg. **jij mag** weggaan. **ben je** weg?

[6] Het symbool ∂ staat voor de schwa (de stomme of doffe “e”).
[7] Alle typisch Zuid-Nederlandse fenomenen worden aangeduid met een grijze achtergrond.

D2c. De **verkleinwoorden op -ke, -ske, -eke** zijn toegelaten. Het is mogelijk dat verschillende van deze vormen bij spellingscontrole nog niet in het CGN-lexicon blijken te zitten. Het is in dat geval de bedoeling dat u ze van een "***n**" voorziet.

u hoort en schrijft	*u schrijft dus **NIET***
meis**ke**, tikker**ke.** ding**ske**, boek**ske.** huiz**eke**, kapell**eke.**	meis**je**, tikker**tje.** ding**etje**, boek**je.** huis**je**, kapell**etje.**

D2d. De volgende **uitspraakvarianten** waarbij de "d" gereduceerd wordt of verdwenen is, zijn opgenomen in het CGN-lexicon omdat ze **voldoende *ingeburgerd* zijn in de schrijftaal**. Schrijf deze woorden en soortgelijke gevallen dan ook waar ze van toepassing zijn.

u hoort en schrijft	*u schrijft dus **NIET***
ik **hou** van jou. **Snij** nu een ui fijn. ik **rij** met de wagen.	ik hou**d** van jou. snij**d** nu een ui fijn. ik rij**d** met de wagen.
Een **ouwe** man. dat gaat je niet in de **kouwe** kleren zitten. Een **goeie** grap. Een **dooie** man. Een **kwaaie** kerel. **rooie** haren.	een ou**d**e man. dat gaat je niet in kou**d**e kleren zitten. een goe**d**e grap. een do**d**e man. een kwa**d**e kerel. ro**d**e haren.

Let op: Het is **niet** de bedoeling om **álle** uitspraakvarianten op te nemen in het CGN-lexicon. Sommige varianten zijn **onvoldoende *ingeburgerd* in de schrijftaal**. Als u dergelijke varianten hoort, schrijf dan de Standaardnederlandse vorm.

u hoort	*maar **toch** schrijft u*
/'t is lang gele**j**e/ /ik ben niet echt tevre**j**en/	't is lang gele**den.** ik ben niet echt tevre**den.**

D3. **Probeer niet weer te geven hoe de woorden zijn uitgesproken.** Dit zou immers al snel leiden tot een soort fonetische transcriptie[8]. Dit is **niet** de bedoeling van een **orthografische** transcriptie; deze mag alleen woorden uit het CGN-lexicon bevatten (tenzij bij *v, *d, *n, *t, *z, *x, *a, *u, of als er een hoofdletter wordt geschreven).

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

u hoort	*maar **toch** schrijft u*
/het leuke<u>n</u>is/ /speelde<u>n</u>ij goed?/ /liep<u>t</u>ie snel?/ /wadis<u>t</u>er hier aan de hand?/	het leuke is. speelde hij goed? liep ie snel? wat is er hier aan de hand?

[8] Voor een deel van de geluidsfragmenten zal er naast de orthografische transcriptie ook een échte fonetische transcriptie (met fonetische symbolen) worden gemaakt. Beide zaken blijven echter strikt gescheiden.

D3b. Duid **niet** aan in de transcriptie wanneer klanken **niet** of **onvolledig** zijn **uitgesproken** (tenzij het duidelijk om dialectische invloeden gaat, bijv. diplom***d**, zie onder D4d).

u hoort	*maar **toch** schrijft u*
/ik ben no nie klaar/ /da vin ik nie/ /’k moe ier weg/ /wa is er nie goe?/ /**jeb**t niets gezegd/ /ik eb **dan**twoorden/ /’t is **to**pen dat…/	ik ben no**g** nie**t** klaar. da**t** vin**d** ik nie**t.** ’k moe**t** **h**ier weg. wa**t** is er nie**t** goe**d**? **je heb**t niets gezegd. ik **h**eb **de an**twoorden. ’t is **te ho**pen dat…

D3c. Duid **kleine klankveranderingen niet** aan in de transcriptie. (Vergelijk dit met regel D5, waar het gaat om zwaardere, dialectische klankveranderingen, bijv. /bruudruuster/ geschreven als broodrooster***z**).

u hoort	*maar **toch** schrijft u steeds*
/**oo**too/ (Nl) of /**o**ttoo/ (Vl) /terap**ui**t/ (Nl) of /terap**eu**t/ (Vl) /tr**e**m/ (Nl) of /tr**a**m/ (Vl) /ma**rt**/, /**sr**ift/, /**as**jeblie**f**/	**au**to. therap**eu**t. tr**a**m. ma**rkt**, **schr**ift, **als**jeblie**ft.**

D4. **Duid alle dialectwoorden en -constructies die niet bestaan in het Standaardnederlands maar wel in dialect aan met *d.** We sommen hier een aantal mogelijke afwijkingen op:

D4a. De typisch Zuid-Nederlandse verbuiging van lidwoorden, voornaamwoorden, adjectieven en substantieven:

- **ne*d** cola, **nen*d** auto, **den*d** boom.
- zijn**e*d** cola, mijn**en*d** auto, Jan, mijn**e*d** man.
- **nen*d** blauwen***d** auto, **nen*d** dikke**n*d** boom, mijn**en*d** goeie**n*d** vriend.
- die**nen*d** dikke**n*d**, die**jen*d** zwarte**n*d**.

D4b. De typische Zuid-Nederlandse combinatie /**∂kik**/. Duid deze aan met **’k*d ik**. De vorm **’k*d** komt enkel voor in de dialectische constructie **’k*d ik** en kan uitgesproken worden als /**∂k**/ en als /**k**/. Deze vorm is niet te verwarren met de vorm **’k** die voorkomt in de gewone schrijftaal en uitgesproken wordt als /**k**/. (zie onder D.2.a.)

u hoort	*dus u schrijft*
/alz∂**kik** weg moet/ /da weet∂**kik** niet/ /dat is de informatie **diekik** heb/ /ik denk **dakkik** weg ben/	als **’k*d ik** weg moet. dat weet **’k*d ik** niet. dat is de informatie **die ’k*d ik** heb. ik denk **dat ’k*d ik** weg ben.

D4c. De vormen /**de**/ en /**me**/. Beschouw deze als dialectische vormen van **ge** en **we**. Noteer ze als **de*d** en **me*d**, en gebruik de werkwoordsvorm die bij **ge** en **we** hoort.

u hoort	*dus u schrijft*
/**waardegij** dat?/ /**hebde** geen tijd?/ /nu **moede** gaan slapen. da **weettegij** wel/	waart **de*d** gij dat? hebt **de*d** geen tijd? nu moet **de*d** gaan slapen. dat weet **de*d** gij wel.
/ik denk **damme** toffe jongens zijn/ /**gamewij** nu al weg?/	ik denk dat **me*d** toffe jongens zijn. gaan **me*d** wij nu al weg?

D4d. Dialectische getinte woorden of constructies uit de spreektaal die teveel afwijken van de geschreven taal om ze te vervangen door de Standaardnederlandse vorm.

- **wulle*d, wulder*d**, **wijle*d** (= wij); **gulle*d**, **gulder*d**, **gijle*d** (= jullie); **zulder*d**, **zulle*d**, **zijle*d** (= zij).

- in **dees*d** geval, **dees*d** doos (= deze).
- een mand**e*d**, een doz**e*d**, een schijv**e*d**.
- deze morge**nd*d**, een dipl**om*d**, ’tzelf**st**e***d**.

- ze zaten daar allemaal **tope*d** (= tesamen, bij elkaar).
- /tis tope datniejen regent/ → ’t is te hopen dat ’t niet **en*d** regent.
- /die-en dag emmewij geen tijd/ → **dienen*d** (evt. **dieën*d**/**diejen*d**) dag **he*d me*d** (evt. **emme*d**) wij geen tijd.
- /kénnukkik genen tiet jong/ → **’k hen*d ’k*d** ik (evt. **kennukik*d**) **genen*d tijd*z** jong.

D4e. Echte dialectwoorden (die al dan niet dialectisch zijn uitgesproken):

- Nederland: keuje***d** (= varken); trulen***d** (= rollen); verrampeneren***d** (= ruïneren).
- Vlaanderen: kortewagen***d** (= kruiwagen); savatten***d** (= sloffen, pantoffels); persienne***d** (= rolluik).

D5. Duid woorden die **wél tot het Standaardnederlands behoren maar zwaar dialectisch zijn uitgesproken**, aan met ***z**.

u hoort	*u schrijft*
/kap n**âh**/ /**laa**tsepl**aa**n/	kap nou***z.** Leidseplein***z.**
/br**uu**dr**uu**ster/ /ks**aaj**n terug th**oa**s/ /in m**ie**n t**ie**d waz dad **alsan** zo/	broodrooster***z.** ’k zijn***z** terug thuis***z.** in mijn***z** tijd***z** was dat alsan***d** zo. (ter info: alsan*d = altijd)

E. Het gebruik van leestekens

Gebruik bij het transcriberen alleen punten, vraagtekens en beletseltekens (**. / ? / …**). Andere leestekens worden niet gebruikt! Plaats deze leestekens volgens de regels en gebruiken van de geschreven taal: waar zou u een punt of vraagteken zetten wanneer u de uitingen zelf had *geschreven*?

E1. Gebruik een **punt** aan het einde van een uiting. Maak geen onnodig lange zinnen: als u volgens uw intuïtie een punt *kunt* zetten, plaats er dan ook een. Het gebeurt vaak dat sprekers niet pauzeren voordat ze een nieuwe uiting beginnen, of op een gramaticaal onlogische plaats pauzeren. Laat u er niet van weerhouden in dat soort gevallen toch een punt te zetten op plaatsen waar dit in geschreven taal ook zou gebeuren, en waar uw intuïtie zegt dat een punt gezet *kan* worden.

- ik ga 'ns naar de Kamerleden kijken of er op dit moment een brandende vraag is. anders vraag ik de organisaties weer op elkaar te reageren.

E2. Gebruik een **vraagteken** om een vraag aan te duiden. Vragen kunnen worden ingeleid door een vraagwoord (wie, welke, waar, etc.) of een voorop geplaatst werkwoord.

- wie moet er nog koffie?
- en wie moet alles weer opruimen?
- komt Jan vandaag ook?

Soms is alleen de intonatie van de spreker een indicatie dat het om een vraagzin gaat:

- Jan komt ook?
- nog koffie?

E3. Gebruik het **beletselteken** wanneer een zin, om welke reden dan ook, niet wordt afgemaakt.

- mmm niet als het… ik zal zien hoe mijn stages meevallen.
- *spreker A:* ik ben woensdag naar…
  spreker B: ja ja ik weet het al. je bent naar Amsterdam geweest.

Gebruik dit teken alleen om een (niet afgemaakte) zin of uiting mee af te sluiten. Gebruik het **nooit** aan het begin of in het midden van een zin **om een pauze aan te duiden**.

- ik denk dat het waar is. (gebruik ook bij een lange pauze tussen bijv. *"denk"* en *"dat"* geen beletselteken)

Wanneer een zin wordt afgebroken en daarna wordt hernomen, plaats dan **enkel** een beletselteken indien de daaropvolgende zin **volledig** wordt hernomen en dus op zichzelf kan bestaan.

- ik ben niet goed... ik ben niet goed wakker. (*"ik ben niet goed wakker."* kan op zichzelf bestaan)
- ik ben niet goed uit*a... ik ben niet goed wakker. (*"ik ben niet goed wakker."* kan op zichzelf bestaan)
- ik ben niet goed uitgesl*a niet goed wakker. (*"niet goed wakker."* kan **niet** op zichzelf bestaan)
- ik ben niet goed niet goed wakker. (*"niet goed wakker."* kan **niet** op zichzelf bestaan)

E4. Zorg ervoor dat **iedere** uiting steeds eindigt met één van de drie leestekens.

- akkoord.
- Jan. telefoon.
- een biertje graag.
- nog koffie?
- leuke grap zeg. ggg.
- niet als het… ik zal zien hoe mijn stages meevallen.

F. Aanvullende regels en tips

F1. Schrijf **getallen** steeds voluit **zoals ze worden uitgesproken**. De getallen van 0 tot 99, de 100-tallen, de 1000-tallen en de 100.000-tallen worden aan elkaar geschreven. Verder wordt alles gescheiden door spaties. (Om het tikwerk te verlichten is het ook toegelaten om de getallen van 0 tot 99, de 100-tallen, de 1000-tallen en de 100.000-tallen als cijfers te schrijven. Ze zullen automatisch omgezet worden naar de voluit geschreven vorm.)

u hoort en schrijft	*OFWEL u schrijft*
drieënvijftig, **een**entwintig. tweehonderd drie. tweehonderd **en** drie. achttienhonderd zevenendertig. duizend achthonderd zevenendertig. twaalfduizend tweehonderd drieënvijftig. vijfhonderdduizend. Drie miljoen vijfhonderd tweeëndertigduizend. vijftien komma zesentwintig twaalf.	53, 21. 200 3. 200 **en** 3. 1800 37. 1000 800 37. 12000 200 53. 500000. 3 miljoen 500 32000. 15 komma 26 12.

Als u het woord /één/ hoort, schrijft u ook "één" (of "1"), als u /∂n/ of /n/ hoort, schrijf dan "een" (zonder accenttekentjes).

u hoort	*u schrijft*
/∂n auto/ /één auto/ /er is één en ander gaande/	een auto. één auto. (ofwel: 1 auto.) er is één en ander gaande.

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

u hoort en schrijft	*u schrijft* ***NIET***
de tweede keer. Een veertienjarige.	de **2de** keer; de **2**$^{\text{de}}$ keer. een **14-**jarige.

Tenzij ze deel uitmaken van een naam (zie ook F4).

u hoort	*u schrijft*
/de politieke partij dee zesenzestig/ /de snelweg aa drieënzeventig/	de politieke partij D66. de snelweg A73.

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

- een **in-** en uitgang; meelopen of **-rennen**;
- een **zee-engte**; de **West-Europese** economie;
- een **glas-in-lood**raam**; vraag-en-aanbod**situatie;
- **'s-Gravenhage**.

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

Gebruik **geen** gedachtestreepje tussen twee woorden.

u schrijft	*u schrijft* ***NIET***
de stand was 1 2 **OF** de stand was één twee. de match Brugge Anderlecht.	de stand was 1 – 2. de match Brugge – Anderlecht.

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

u hoort	*u schrijft*	*u hoort*	*u schrijft*	*u hoort*	*u schrijft*	*u hoort*	*u schrijft*	*u hoort*	*u 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	/iks/, /ieks/	X	/ui/	UI

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

- 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.

Wanneer letters op een afwijkende manier gespeld of uitgesproken worden, gebruikt u ***u**.

- op de lagere school spelt men laf als **le*u a*u fe*u**.
- ik heb hopen hout. met de **hasj*u** van hallo. (Vl.)

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.

u hoort	*u 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 bee/, /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).
/uva/ **OF** /uu vee aa/ /er uu gee/ **OF** /rug/ /vee bee oo mavo klas/	UVA RUG VBO-MAVO-klas
de groep /innekses/, de partij /dee zesenzestig/	de groep **INXS**, de partij **D66**

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

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

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

- een experiment van de **NASA**.
- een uitzending van de **TROS** of van de **VARA**.

Als een afkorting of letterwoord nog niet in het CGN-lexicon blijkt te staan, voegt u ***n** toe aan de afkorting of het letterwoord.

F5. 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.

u twijfelt tussen		*u schrijft*	*na spellingscontrole wordt dit*
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:

aaneen	*niet aaneen*
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:

- ervandaan, erbovenop, ernaartoe, ermiddenin.
- d'rvandaan, d'rbovenop, d'rnaartoe, d'rmiddenin.
- hij gaat ervan uit dat; die ziet eruit als; het zit erop.
- vijfprocentseis, lagelonenland, hogesnelheidslijn.

G. Index

In de tweede kolom staan de hoofdstukaanduidingen; de getallen in de laatste kolom verwijzen naar de bladzijden.

H. Aanvullingen

(februari 2004)

H1. Woorden uit een vreemde taal (bij C3).

Het eerste voorbeeld in het tweede opsommingsblok onder C3 <de band **New*v Kids*v On*v The*v Block*v**.> is een onjuist voorbeeld, aangezien het hier om een eigennaam van een band gaat, die dus **geen** codering zou moeten krijgen.

Ter verduidelijking van de verschillende aanpak van titels en eigennamen voor wat betreft vreemdtalige woorden, zou de voorbeeldzin <de hit **Step*v By*v Step***v van de band **New Kids On The Block**> opgenomen kunnen worden ná het derde opsommingsblok onder C3.

Ook vaste Latijnse uitdrukkingen of Italiaanse muziektermen worden opgevat als woorden die in het Nederlands zijn ingeburgerd. Ze krijgen dus geen *v. Als deze "ingeburgerde woorden" nog niet in het lexicon staan, krijgen ze uiteraard wel een *n.

- ik heb **in extremis** de trein nog gehaald.
- hij speelde een **adagio** op de piano.
- de gemachtigde ambtenaren zouden een **status questionis*n** opmaken.

H2. Tussenwerpsels (bij C5).

Wanneer iemand zegt <de ramen> en hierbij de schwa van het woordje <de> lang aanhoudt, zodat het eigenlijk op het tussenwerpsel <uh> begint te lijken, transcriberen we gewoon <de>en dus **niet** <d-uh*u>, <de-uh*u> of <de uh>.

H3. Afgebroken woorden en versprekingen (bij C6 en C7).

De code *a mag ook gebruikt worden bij afgebroken eigennamen. Ook bij eigennamen die worden afgebroken maar daarna nog eens volledig worden uitgesproken. Als voorbeeld is hieronder <Antwer*a Antwerpen> opgenomen. Dus in plaats van <Antwer-Antwerpen*u> gebruik je < Antwer*a Antwerpen >.

- de Universiteit **Antwer*a** Antwerpen is de vragende partij.
- Het is de **Belgi*a** de Vlaamse overheid die daar moet op toezien.

Ook gebruiken we *a voor afgebroken versprekingen, al dan niet gevolgd door een herneming ; dus : *a heeft voorrang op *u

- ze keken door het **vren*a** raam.
- ze keken door het **vren*a** venster.
- het is daar **zer*a** erg rustig.

In het laatste voorbeeld kan er sprake zijn van een verspreking van het woord <erg>, <zeer> of misschien wel <zeker>. Omdat nog niet duidelijk is om welk woord het gaat, kiezen we in zo'n geval voor *a. In tegenstelling tot dit voorbeeld kiezen we in de zin <het is daar zerg*u erg rustig> wel voor *u, omdat het hier om een verspreking van een volledige woord ('erg' of 'zeer') gaat.

Zoals in het protocol aangegeven staat onder C7c. gebruiken we *u, in alle gevallen van versprekingen en onderbrekingen **binnen een woord**, waarbij het “goede” woord daarna **niet** alsnog in z’n geheel wordt uitgesproken. Het laatste voorbeeld, **<**ze heeft haar ring **ver-ggg-verpatst*u**>, moet precies geïnterpreteerd worden zoals het er staat: als een woord in z’n geheel lachend wordt uitgesproken transcriberen we dat **niet**, alleen als het woord **onderbroken** wordt door gelach gebruiken we deze conventie.

- hij is nog niet naar de kapper **gewee-weest*u**
- de verzetsstrijders van het **Kosovo-bevrijdingslever-leger*u**.
- de **jeugdlied-liefdes*u** van de schrijver.
- dat is de **mogelijkhee-heid*u**.
- heb jij dat spul nog over het gras **uitges-ge-uh-gesproeid*u**?
- dat is een **ongeluks-uh-baby*u**.
- ze heeft haar ring **ver-uh-uh-patst*u**.
- ze heeft haar ring **ver-ggg-patst*u**.

H4. Hoe gesproken taal neerschrijven (bij D)

Regel D1. uit het protocol luidt als volgt:

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

u hoort en schrijft	*u verbetert dit dus **NIET** tot*
de koe **dat** graast. **de** open venster. **hun** zijn er al.	de koe **die** graast. **het** open venster. **zij** zijn er al.

Wanneer u echter zeker weet dat de spreker zich *verspreekt*, gebruikt u ***u** (zie C7b).

In de praktijk komen we zinnen tegen zoals <ik **heb** ben daar niet zo voor te vinden>. In dit geval is <heb> duidelijk een verspreking, die door de spreker verbeterd wordt tot <ben>. Vanwege regel D1 wordt <heb> echter niet als verspreking gecodeerd. De laatste opmerking <Wanneer u echter..> heeft in de praktijk echter zeker voor verwarring gezorgd.

H5. Het Koppelteken (bij F2–F5);

documentatie door Raffaëla Vlot d.d. 27-03-2001.

ALGEMEEN:

SCHRIJF SAMENSTELLINGEN IN HET NEDERLANDS ALS ÉÉN WOORD

computertafel, rekenmachine, koffiezetapparaat, loonkostenbeheersing, stoeltjeslift etc.

Wanneer schrijf je dan een koppelteken?

- bij samenstellingen met klinkerbotsing tussen de delen (De klinkers zouden samen één klank vormen):

aa	ee	ie	oe	ui
ae	ei	ii	oi	uu
ai	eu	iij	oo	
au		ij	ou	
		iji		
		ijij		

dus: *dia-avond, zee-eend, sproei-installatie, tosti-ijzer, glij-ijzer,radio-uitzending, menu-idee*

maar dus niet bij: *massaontslag, milieueffect, naijlen, parapluantenne*

- bij samenstellingen met nevengeschikte, gelijkwaardige delen (een *journalist-cabaratier* is zowel een journalist als een cabaratier):

dus: *chef-kok, journalist-cabaretier, minister-president, rood-wit-blauw, sociaal-democratisch, koningin-moeder, Naarden-Bussum*

UITZONDERING: niet bij werkwoorden, dus: *zigzaggen, hiphoppen, pingpongen*
niet bij twee dezelfde delen, dus: *beriberi, blabla, tamtam*

- bij samenstellingen waarbij het linkerlid de kern is:

dus: *cola-tic, rekening-courant, Staten-Generaal, Antwerpen-Oost*

[normaal gesproken is het rechterlid de kern van de samenstelling: een *tomatensoep* is een soort *soep*. Een *cola-tic* daarentegen is niet een soort *tic* maar een soort *cola*]

- bij samenstellingen met een naam als tweede deel en betekenis *genoemd naar*:

dus: *voorstel-Den Uyl, commissie-Geerts*

[pas op: komt er een spatie voor in de naam dan blijft die gehandhaafd. Niet nog een extra koppelteken. Zie *voorstel-Den Uyl*]

- ook bij gelegenheidssamenstellingen <u>waarin een eigennaam voorkomt</u>:

  dus: *nep-Rembrandt, Nuis-optreden*

  UITZONDERING: níét als de eigennaam een <u>geografische naam</u> is (die gecombineerd wordt met een soortnaam (= “niet-eigennaam”):

  dus: *Greenwichtijd, Schipholverkeer*

  Samenstellingen met een eigennaam erin die in woordenboeken zijn opgenomen (en dus geen gelegenheidssamenstelling zijn), verliezen meestal hun koppelteken.
  Het is dan ook *Mariabeeld* en *Nobelprijs*.

- ook bij samenstellingen met namen van <u>talen</u> en samengestelde geografische <u>bijv.nw</u> van het volgende type:

  dus: *Oud-Nederlands, Nieuw-Grieks*

  en ook bij geografische namen met (een van de) <u>windstreken</u> en hun afleidingen:

  dus: *Zuid-Frankrijk, Oost-Vlaanderen, Noord-Nederland* en ook *Oost-Vlaming, Noord-Nederlander, Noord-Nederlandse*

  ook bij: *Centraal-Azië, Midden-Amerika*

  UITZONDERING: als de geografische naam van oorsprong zonder koppelteken geschreven wordt, dan blijft dat zo, bijv. *Zuidhorn*

  Opm: gebonden morfemen als *oer-* en *on* krijgen géén hoofdletter:
  dus: *oer-Amerikaans, on-Engels*

- bij samenstellingen met een <u>afkorting</u>, <u>letteraanduiding</u> of <u>cijfers</u>:

  dus: *80-jarige, kleuren-TV*, meervouds-S, PVDA-voorstel, top-100, TL-buis, X-benen, S-vormig, contra-VN-beleid, @-teken*

  * Binnen het CGN-project geldt de afspraak dat alle afkortingen met hoofdletters worden geschreven, dus *kleuren-TV*, terwijl je normaal eigenlijk zou schrijven *kleuren-tv*

  Opm: - indien het getal uitgeschreven wordt, dan geen koppelteken dus: *tachtigjarige*. <u>Schrijf binnen het CGN-project *tachtigjarige* voluit</u>

- bij een afleiding met een afkorting komt er een <u>apostrof</u> i.p.v. een koppelteken, dus: *D66-leider*, maar: *D66'er*

- als het <u>woord bestaat uit een woordgroep</u> óf als er <u>binnen het woord een woordgroep</u> is, komen er koppeltekens tussen de woorden van die woordgroep. Niet tussen de woordgroep en een (eventueel) volgend rechterlid (dus niet tussen …*-smijt* en *werk,* zie onder):

  dus: *laag-bij-de-gronds, doe-het-zelver, kruidje-roer-mij-niet, jantje-van-leiden, manusje-van-alles,* een *staakt-het-vuren,* een *blik van wat-moet-je-van-me*

  doe-het-zelfzaak, gooi-en-smijtwerk, , hink-stapsprong, mond-tot-mondreclame, nek-aan-nekrace, peper-en-zoutstel

---

- bij samenstellingen met een <u>persoonlijk voornaamwoord</u>:

  dus: *ik-figuur, hij-vorm, het-woord*

- na de volgende <u>voorvoegsels</u> (in díé betekenis) (of <u>woorden</u>):

adjunct	*adjunct-commies* (in de betekenis 'hulp-, plaatsvervangend')
archi	*archi-dom* ('uitermate')
aspirant	*aspirant-lid* ('toekomstig')
assistent	*assistent-arts* ('hulp-')
bijna	*bijna-botsing*
co	alléén in *co-counsel(er/ing), co-ouder, co-schap* (betere leesbaarheid) maar: *coassistent*
concept	*concept-tekst* (rechterlid van samenstelling geeft 'tekst' aan: *~begroting, ~tekst,*

	~verdrag)
demi	*demi-finale*
ex	*ex-premier* ('voormalig')
interim	*interim-manager*
kandidaat	*kandidaat-notaris*
leerling	*leerling-verpleger*
loco	*loco-burgemeester* ('plaatsvervangend')
niet	*niet-roker* ('ontkennend')
non	*non-probleem* ('ontkennend')
oud	*oud-bestuurslid* ('voormalig')
pro	*pro-westers*
pseudo	*pseudo-kroep* ('lijkend op, schijnbaar')
quasi	*quasi-technisch* ('schijn…')
semi	*semi-bungalow*
sint Sint	*sint-juttemis, sint-janskruid* *Sint-Nicolaas*
substituut	*substituut-griffier* ('plaatstvervangend')
vice	*vice-voorzitter*

Opm: geen koppelteken als het geheel geen samenstelling is of het voorvoegsel een andere dan de gegeven betekenis heeft, dus: *locomotief, pseudoniem, oudtante*

- het trema wordt gebruikt bij klinkerbotsing binnen niet-samengestelde woorden (*patiënt, ruïne*) en bij afleidingen (*beëindiging, ongeëvenaard*). Omdat veel uitheemse voorvoegsels zoals *bio, macro, micro, mini, multi, neo, socio, stereo, tele, ultra* e.d. min of meer als zelfstandige voorvoegsels worden gezien, krijgen zij bij klinkerbotsing géén trema, maar een koppelteken:

  dus: *bio-industrie, macro-economie, micro-organisme, mini-emmer, multi-etnisch, neo-expressionisme, socio-economisch, stereo-opname, tele-informatie, ultra-actief*

  maar: *biobak, macrobiotisch, microkosmos, minibusje, multicultureel, neoclassicisme, sociolinguïstiek, stereotoren, telewerk, ultrageluid*

  UITZONDERING: de voorvoegsels *de-, pre-, re-* vallen niet onder deze regel en krijgen wel een trema
  dus: *deëscalatie, preïndustrieel, reïncarnatie*

  Opm: het voorvoegsel *a-* krijgt bij klinkerbotsing een *n* erbij, dus: a+organisch wordt *anorganisch, analfabeet*

- bij samenstellingen van drie of meer delen waarvan twee of meer anderstalige delen komen koppeltekens tussen de anderstalige delen

  dus: *a-capellakoor, ad-hoccommissie, human-resourcesmanagement, in-vitrofertilisatie, on-lineverbinding* (maar: "*hij is on line*")

  maar:
  samenstellingen die uit twee delen bestaan waarvan één anderstalig, worden aaneen geschreven:

  dus: *jazzmuziek, jockeypet, jobstudent, successtory*

  Zijn beide delen anderstalig en is het woord ingeburgerd, schrijf aaneen:

  dus: *airbus, bottleneck, coffeeshop, compactdisc, freelance, parttime*

  maar:
  bij woordcombinaties van een bijv. nw. en een zelfst. nw.: schrijf los indien beide delen klemtoon kunnen krijgen:

  dus: *black box, happy end, heavy metal, top secret*

- Engelstalige samenstellingen die eindigen met een voorzetsel dat met een klinker begint, krijgen een koppelteken:

  dus: *bottom-up, drive-in, lay-out, spin-off* maar: *topdown, playback*

Er zijn gevallen waarin een samenstelling lastig leesbaar wordt als geen koppelteken gezet zou worden, bijv:

bas-aria, jazz-zanger i.p.v. basaria, jazzzanger

Een koppelteken kan ook duidelijk maken welk woord bedoeld wordt in een geval als (het koppelteken is hier overigens niet verplicht):

kwarts-lagen en *kwart-slagen*

Verder mág een schrijver in samenstellingen, m.n. bij erg lange, voor de duidelijkheid een koppelteken zetten volgens de spellingregels. Is het woord zonder koppelteken ook goed te lezen, wordt er zonder koppelteken geen onduidelijkheid over uitspraak of bedoelde woord geschapen, ZET HET KOPPELTEKEN DAN NIET, ook al lijkt het je misschien iets duidelijker!!! Anders krijgen we allerlei mogelijke vormen van woorden, met op verschillende plaatsen soms wel of niet een koppelteken.
Het is dus gewoon:

achtertuinpolitiek, derdewereldland, heteluchtkachel, kortebaanwedstrijd, teraardebestelling, tienrittenkaart

(Uit de files)

brilletjes-toestand	is dus gewoon	*brilletjestoestand*
bulk-product	is dus gewoon	*bulkproduct*
conservatorium-diploma	is dus gewoon	*conservatoriumdiploma*

H6. Aanvullingen en opmerkingen voor wat betreft vreemtalige woorden, eigennamen en titels, tussenwerpsels en dialectische woorden, Verslag vergadering d.d. 17-10-2001 door Richard Piepenbrock.

1 Probleemstelling

Een aantal punten is aan de orde:
1. de markering van vreemdtalige woorden (bij C3).
2. het onderscheid tussen eigennamen en titels (bij C3).
3. de behandeling van tussenwerpsels (bij C5).
4. de markering van dialectische woorden (bij D4).

2 Het onderscheid tussen vreemdtalige en Standaardnederlandse woorden

2.1 De behandeling van vreemdtalige eenledige niet-eigennamen

Voorbeelden: bonnetterie, bonsai, burn-out, callgirl, calvados, computer, cool, disk, entremets, date, hardware, nerd.

Deze worden als Standaardnederlands beschouwd als ze voorkomen in het driedelig Van Dale Groot Woordenboek der Nederlandse Taal (13e druk, 1999). Dit lijkt arbitrair, maar het opnamecriterium was bij Van Dale dat woorden zeker drie jaar voor moesten komen in meer dan één algemene, dus niet-vakspecifieke bron. Betere criteria zijn binnen de opzet van CGN niet haalbaar.

Anders krijgen ze '*v' en als tag 'SPEC(vreemd)'.

Uitzondering:

Als de als Standaardnederlands geclassificeerde vreemdtalige woorden voorkomen in een duidelijk vreemdtalige context, dan krijgen ze toch '*v'.

Voorbeelden:

"Ik heb al weer een nieuwe computer moeten kopen."
"Man ze hebben het hier over sensational*v computer*v bargains*v."

"Toen was de disk volgelopen."
"Het is zo'n rewritable*v optical*v disk*v."

"My*v Way*v is oorspronkelijk een Frans chanson van Claude François."
"Hij kreeg vaak het verwijt van het schrijven van chansons*v maudites*v."

Let goed op Van Dale, want de zin 'De lowbrow internet user is vooral gebaat bij up-to-date plug-and-play hardware' bevat geen enkel vreemdtalig woord!

Noten:

1. Als een als vreemdtalig uitgesproken niet-eigennaam toevallig homoniem is met een Nederlands woord wordt het gemarkeerd als '*v', ook al zal de spellingchecker het niet als bijzonder signaleren:

"Ik had een raar virus*v op mijn pc." (/vaIr@s/)

"Dat optreden van die lui was echt super*v." (/su:p@r/ met de klank van 'you')

2. Forse versprekingen in vreemdtalige woorden (die resulteren in afwijkende transcriptie) krijgen '*u' en afgebroken vreemdtalige woorden krijgen '*a', zoals bij Nederlandse woorden. Beide krijgen echter wel de POS-tag 'SPEC(vreemd)'. Het is dus niet nodig '*u' te plaatsen bij gangbare verkeerde uitspraken als 'sweater' /swit@r/ of 'corned /kOrnEt/ beef'.

2.2 De behandeling van vreemdtalige meerledige niet-eigennamen

Voorbeelden: anorexia nervosa, chili con carne, crème fraîche, eau de cologne, prima donna, sine qua non, vice versa.

Deze worden eveneens als Standaardnederlands beschouwd als ze als complete uitdrukking voorkomen in de 13e druk van de Grote Van Dale en minstens één van de afzonderlijke delen niet in Van Dale voorkomt (anders zou het gewoon gaan om een toevallige reeks afzonderlijk geassimileerde leenwoorden).
Ze krijgen dus nergens '*v', maar wel 'SPEC(vreemd)' omdat de delen niet volgens de Nederlandse zinsbouw te benoemen zijn.

Noten:

1. Als één of meer van de woorden ook los voorkomen in een Nederlandstalige context worden ze daar opgevat als inheems en dus getagd met een relevante tag volgens hun woordsoort in de zin.

2. Als de uitdrukking als geheel voorkomt in een duidelijk vreemdtalige context, dan krijgen ze toch '*v'.

Vergelijk:

"Voor de smaak heb ik nog wat chili N(soort,ev,basis,zijd,stan) toegevoegd."
"We aten die avond chili SPEC(vreemd) con SPEC(vreemd) carne SPEC(vreemd)."
"Op de menukaart stond chili*v SPEC(vreemd) con*v SPEC(vreemd) carne*v SPEC(vreemd) y*v SPEC(vreemd) patatas*v SPEC(vreemd) fritas*v SPEC(vreemd)."

"Tegen schrale huid kan ik crème N(soort,ev,basis,zijd,stan) met goudsbloem aanbevelen."
"Neem crème SPEC(vreemd) fraîche SPEC(vreemd) want die kun je meekoken."
"Dan zeggen de Fransen natuurlijk weer dat je eigenlijk een crème*v SPEC(vreemd) fraîche*v SPEC(vreemd) épaisse*v SPEC(vreemd) moet gebruiken."

2.3 De behandeling van vreemdtalige eigennamen

Vreemdtalige eigennamen worden nooit gemarkeerd als '*v'.

Titels zijn GEEN eigennamen. Dus:

1. Vreemde woorden die deel uitmaken van een titel behouden hun '*v' (maar krijgen wel een hoofdletter omwille van regel C2)
2. Eigennamen die deel uitmaken van een titel krijgen nog steeds geen '*v' (het blijven immers eigennamen)

Onder eigennaam verstaan we de 'naam' van:

1. een persoon (George Bush, Nick Leeson, François Mitterand, Calamity Jane, Billy The Kid))
2. een groep/(muziek)band/toneelgezelschap/... (New Kids On The Block, The Mama's And The Papa's, The Beatles, Procol Harum)
3. een plaats (Rio De Janeiro, Orlando, Berlin/Berlijn, Nice, Praha/Praag)
4. een straat/plein/park (Fifth Avenue, Route Sixty Six, Parc De La Villette, Madison Square, Baker Street, Hyde Park)
5. bergen, vulkanen, geisers (Vesuvius, Matterhorn, The Old Faithful (geiser in USA))
6. een tunnel/brug/gebouw (Pfandertunnel, Mont Blanctunnel, Sweet Track Bridge, An-Lan Bridge, Hoover Dam, Empire State Building, La Tour D'Eiffel)
7. een bedrijf (General Motors, Toyota, Associated Press/AP, NATO)
8. een (culturele) organisatie of instelling (New York Philharmonic (Orchestra), United Nations/UN, de Neue Nationalgalerie, Ny Carlsberg Glyptotek)
9. een festival of evenement (North Sea Jazz, Crossing Border, Axion Beach Rock)
10. een krant/weekblad (Bild Am Sonntag, Le Monde, New York Times)
11. dieren etc. (fauna en ora): Flipper, Skippy, Donald Duck etc
12. een internetadres (WWW Spot Dot Com, WWW RUG Punt NL, WWW RUG AC BE, WWW BBC CO UK Slash Radio One)

(zie ook de opsomming onder C1 (Nederlandstalige eigennamen)

Onder titels verstaan we de 'naam' van:

1. een film (Girlfight*v, The*v Hunter*v, Leaving*v Las Vegas)
2. een boek (Chocolat*v, The*v Bonesetter's*v Daughter*v, Yoga*v For*v Dummies*v, The*v Adventures*v Of*v Huckleberry Finn, Max Et*v La*v Poule*v En*v Chocolat*v)
3. een plaat/liedje (I*v Can't*v Stand*v Myself*v, Ein*v Deutsches*v Requiem*v, Amsterdam, New York New York)
4. een musical (A*v Chorus*v Line*v, Cats*v, Les*v Misérables*v, The*v Phantom*v Of*v The*v Opera*v, Annie, Evita, Saturday Night*v Fever*v)
5. een radio- of tv-programma (Tatort*v, La*v Dessous*v Des*v Cartes*v, Ein*v Fall*v Für*v

1. Zwei*v, Buffy The*v Vampire*v Slayer*v, Siska, Il*vCommissario*v Rex) etc.

(zie ook de opsomming en voorbeelden onder C2 (Nederlandstalige titels)

Wanneer je vooral in een meer exotische taal twijfelt of er een *v moet worden toegekend, ken er dan een toe. B.v. in

dipues*v vienes*v delhaldea*v (Portugees liedje)
u*v lamentu*v de*v filicone*v (Corsicaans liedje)
sheva*v b'rachot*v (Jiddisch liedje)

wanneer het niet duidelijk is of de titels een eigennaam bevatten.

Wanneer echter uit de context duidelijk is af leiden dat een bepaalde brug/tunnel/gebouw/krant o.i.d. is bedoeld, ken dan geen *v toe:

De Gamla Tjörnbron is in 1960 ingestort (Zweeds)
De Akashi Kaikyo in Japan is de brug met de hoogste pilonen ter wereld (Japans)
Vanaf het dakterras van de Kuningan Persada heeft men een prachtig uitzicht (Indonesisch)

Hier volgen nog een paar extra voorbeeldjes om de praktijk te toetsen:

Chris Rea, Grace Jones' laatste hit, ...
de band New Kids On The Block
Loyola University, Mechanics Institute, Massachusetts Institute Of Technology, ...
Metropolitan Opera, het New York Philharmonic, ...
heb je La*v Vie*v En*v Rose*v al eens gehoord?
het liedje Josephine.
het boek The*v Name*v Of*v The*v Rose*v.
Shakespeare In*v Love*v.
The*v Faculty*v ook die kun je vandaag zien.

Dus ook geen '*v' krijgen:

1. Vreemdtalig uitgesproken letters in namen
('George W (/dVbl ju:/) Bush', 'J (/dZeI/) Edgar Hoover').
2. Vreemdtalig uitgesproken afkortingen
('FBI' (/Ef bi: aI/), 'BBC' (/bi: bi: si:/, 'USA' /ju: Es eI/).

Noten:

1. Gespelde vreemdtalige letters buiten de context van eigennamen krijgen wel '*v'.
Voorbeeld:

"De letters P*v (/pi:/) SPEC(vreemd) D*v (/di:/) SPEC(vreemd) en A*v (/eI/) SPEC(vreemd) staan hier voor personal*v SPEC(vreemd) digital*v SPEC(vreemd) assistant*v SPEC(vreemd). Je weet wel zo'n soort elektronische agenda."

2. Forse versprekingen in vreemdtalige eigennamen (die resulteren in afwijkende transcriptie) krijgen '*u'. Dat geldt ook voor al dan niet opzettelijk als vreemdtalig uitgesproken inheemse eigennamen, b.v. 'Jambers' /dZEmb@rs/, alsof het een Amerikaanse show is, of in een zin als: "Door al die Hollywoodfilms moeten we onderhand zeker spreken over het dagboek van Anne*u Frank*u." /En frENk/

Afgebroken vreemdtalige eigennamen krijgen '*a', zoals bij Nederlandse woorden. Beide krijgen echter wel de POS-tag 'SPEC(vreemd)'. Het is dus niet nodig '*u' te plaatsen bij b.v. 'Arkansas', abusievelijk uitgesproken als /ArkEns@s/.

3. Namen van weekdagen en maanden worden binnen CGN gezien als eigennamen, alhoewel ze zonder hoofdletter worden gespeld. Dat zou betekenen, dat ook vreemdtalige weekdagen en maanden in titels uitgezonderd moeten worden van de '*v'-markering, dus "Hotter*v Than*v July van Stevie Wonder" of "Vivement*v Dimanche van Truffaut". Omdat dat niet werkbaar is, krijgen vreemdtalige weekdagen en maanden ook '*v'.

3 Het onderscheid tussen eigennamen en titels

Het bovenstaande overzicht van eigennamen en titels, hoewel opgesteld voor vreemdtalige woorden, geldt ook voor Nederlandstalige woorden. Hierop is nog een aantal aanvullingen te vermelden:

Onder eigennamen verstaan we ook nog:

1. een getal dat een vast onderdeel uitmaakt van een eigennaam (Amstel Achttien Zeventig, Plein Vierenveertig, Cinema Tweeduizend, Registratiewet Negentien Zeventig). Wanneer het gewoon een huisnummer betreft wordt het echter klein geschreven: Reigerstaat vierenveertig.
2. een naam van een wet, koninklijk/ministerieel besluit of algemene maatregel van bestuur (Algemene Arbeidsongeschiktheidswet, wet op de Erkende Onderwijsinstellingen, de wet Waardering Onroerende Zaken, het besluit Provinciale Opcenten Motorrijtuigenbelasting)

Onder titels verstaan we ook nog:

1. een naam van een nota of rapportage, die niet de kracht heeft van een wet (de vierde nota Waterhuishouding, de nota Ruimte Maken Ruimte Delen, de voortgangsrapportage Weer Samen Naar School).

Alle soortnamen en andere woorden die voorafgaan aan een eigennaam en die eigennaam nader specificeren worden niet gezien als eigennamen of titels en derhalve klein geschreven: het ministerie van Binnenlandse Zaken, het ministerie van Middenstand En Landbouw, de vakgroep Taal En Spraak, de afdeling Personeel En Organisatie, het landelijk bestuur Scouting Nederland, de faculteit der Letteren. Uitzonderingen zijn soortnamen die overduidelijk deel uitmaken van de naam van een organisatie: Orkest Van Het Oosten, Club Brugge, de Raad Van State, de Kamer Van Koophandel.

4 De behandeling van tussenwerpsels

De markering '*t' is indertijd geïntroduceerd met het oog op problemen met de vaststelling van de juiste schrijfwijze van tussenwerpsels en het feit dat ze bij de POS-tagging binnen de zin geen wezenlijke syntactische functie vervullen. In feite echter bevat het label '*t' nauwelijks extra informatie ten opzichte van het label '*n' voor nog niet eerder gesignaleerde woorden. We besluiten dan ook de markering '*t' automatisch te vervangen door '*n'.

5 De behandeling van dialectwoorden

We constateren dat veel als dialectisch (*d) aangemerkte woorden in feite informele spreektaalvarianten zijn van gangbare Nederlandse woorden, waarbij een vaak regelgestuurde klankverschuiving in een consonant of vocaal, ofwel bepaalde assimilatie- en deletieverschijnselen zijn opgetreden.

Voorbeelden:

"as (= als) Kristel hier slaapt"
"hij het (= heeft) niks gezegd"
"hej (= heb je) nog niet naar gekeken"
"wat heb je d'r dan an (= aan)"

We geven in dit soort gevallen de code '*z', waarbij we dan wel de orthografie moeten normaliseren naar de standaardvorm. Als niet duidelijk is welk woord bedoeld wordt, kan de code '*z' niet worden toegekend en markeren we het woord als '*d' met een uitspraaktransliteratie.

Een werkwoordsvorm als 'wij wouden' valt ook volgens de ANS onder het informeel taalgebruik van punt D2d van het orthografisch protocol en hoeft dus niet gemarkeerd te worden. Noord-Nederlandse diminutiefvormen op -ie ('stukkie', 'meissie', 'wijffie') krijgen in navolging van Zuid-Nederlandse vormen op -ke ('stukske', 'meiske', 'wijveke') geen markering als '*d' of '*z'.

Bij twijfelgevallen, zoals de -s-vormen 'derdes', '(ja)zekers', 'veels' en 'ieders', markeren we de woorden als '*d'.

Als het gaat om woorden die geheel niet passen in het standaardvocabulaire, zoals 'naber' voor 'buurman', 'kappes' voor 'kool' of 'boks' voor 'broek', dan wordt het label '*d' toegekend.

Ter referentie: deze woorden staan óf niet in de Grote Van Dale, óf krijgen daar het kenmerk 'gew.' voor 'gewestelijk'.

Specifiek Vlaamse woorden die in Van Dale opgenomen zijn als 'Belg.N.' en/of in het CGN-lexicon als status 'B' hebben gekregen (nl. 'Belgicisme' volgens de bron RBN) worden gezien als equivalent aan de Noord-Nederlandse vormen en krijgen daarom geen '*d'.