

Documentatie
Eindhoven Corpus
(oorspronkelijke versie & VU-versie)

TST-Centrale, 2009

Het Eindhoven Corpus (VU-versie) is sinds 2006 beschikbaar via de TST-Centrale (Centrale voor Taal- en Spraaktechnologie). De TST-Centrale is een initiatief van en wordt gefinancierd door de Nederlandse Taalunie (NTU). Voor meer informatie over de TST-Centrale zie <http://www.tst-centrale.org>. Voor meer informatie over het Eindhoven Corpus of andere producten die bij de TST-Centrale beschikbaar zijn, zie <http://www.tst-centrale.org/producten> of stuur een e-mail naar servicedesk@tst-centrale.org. Het Eindhoven Corpus is na registratie gratis te downloaden van de website.

Inhoudsopgave

Inhoudsopgave	3
1. Inleiding	4
2. Totstandkoming van het Eindhoven Corpus	4
3. Gebruik van het Eindhoven Corpus	4
4. Verschillende versies van het Eindhoven Corpus	5
5. Beschrijving van de verschillende corpuscomponenten	6
6. Voorbeeldfragmenten	7
Referenties	11

1. Inleiding

Het *Eindhoven Corpus* – ook wel *Corpus Uit den Boogaart* (1975) genoemd – is de eerste verzameling van Nederlandstalige geschreven en (getranscribeerde) gesproken teksten. Het corpus dateert uit de periode van 1960 tot 1973. Het geschreven deel bevat fragmenten van in totaal 600.000 woorden uit de periode 1964-1971. Het gesproken deel is aanzienlijk kleiner en bevat circa 120.000 woorden. Zie paragraaf 5 voor een compleet overzicht van de inhoud van het corpus.

Oorspronkelijk is het materiaal in 1975 in boekvorm verschenen als *Woordfrequenties in geschreven en gesproken Nederlands*. De tekst bestaat uit allerlei frequentielijsten die van het corpus zijn afgeleid. De inleiding van het boek bevat aanvullende informatie over de samenstelling van het corpus.

2. Totstandkoming van het Eindhoven Corpus

Het Eindhoven Corpus is tot stand gekomen in het kader van een onderzoek naar statistische verschijnselen in hedendaags geschreven en gesproken Nederlands. Aanvankelijk is het corpus dus opgebouwd om een frequentielijst van het Nederlands – niet gecombineerd met het Vlaams – te kunnen samenstellen. Het onderzoek is gedaan onder supervisie van de Werkgroep Frequentie-Onderzoek van het Nederlands en gefinancierd door de toenmalige Nederlandse Organisatie voor Zuiver-Wetenschappelijk Onderzoek (Z.W.O.), het huidige NWO. Het onderzoek vond plaats aan de toenmalige Technische Hogeschool te Eindhoven, de huidige Technische Universiteit Eindhoven, en aan het toenmalige Instituut voor Dialectologie, Volks- en Naamkunde van de KNAW (nu: Meertens Instituut).

De leden van de Werkgroep Frequentie-Onderzoek van het Nederlands waren: prof. dr. B. Th. Tervoort, dr. J. Daan, ing. D. Zwama, prof. dr. J.J.M. Bakker, prof. dr. B. van den Berg, dr. H. Brandt Cortius, dr. J.P.M. Eggermont, dr. A. Feitsma, dr. J. de Rooij, prof. dr. ing. A. van Wijngaarden en drs. A.M. Zwaneveld

De uitvoerders van het onderzoek dat ten grondslag lag aan het Eindhoven Corpus waren drs. P.C. Uit den Boogaart, E.D. de Jong, A.M. van Mierlo.

3. Gebruik van het Eindhoven Corpus

Door de jaren heen is het Eindhoven Corpus ingezet bij allerlei vormen van onderzoek binnen de taalkunde en de taaltechnologie. De handmatige tagging maakt het corpus geschikt als training- en testcorpus bij de ontwikkeling van part-of-speechtaggers. Het corpus is daarnaast ook van belang voor diachroon taalkundig onderzoek.

4. Verschillende versies van het Eindhoven Corpus

4.1 Originele versie

In juli 2005 heeft Eric Akkerman (Faculteit der Letteren, Vrije Universiteit (VU) Amsterdam) een tape aan de TST-Centrale gegeven in de veronderstelling dat hierop de originele versie van het Eindhoven Corpus zou staan. Dit bleek achteraf niet het geval te zijn. Een beperkte zoektocht (met dank aan Boudewijn van den Berg, werkzaam bij het Meertens Instituut) heeft uitgewezen dat de oorspronkelijke geluidsbanden en de bijbehorende transcripties zijn gearriveerd bij het Meertens Instituut. Of de originele tekstfiles nog bestaan, is niet bekend.



4.2 VU-versie

Het Eindhoven Corpus dat bij de TST-Centrale gedownload kan worden – de VU-versie – is een opgeschoonde versie, gemaakt en gedocumenteerd in 1989 door Dave van Grootheest in het kader van een afstudeerscriptie (VU Amsterdam, Faculteit der Letteren). Voordat deze versie van het Eindhoven Corpus bij de TST-Centrale terecht kwam is deze jarenlang door de VU beheerd en gedistribueerd. In de bijbehorende gebruikershandleiding staat uitgebreid beschreven in welk opzicht deze versie verschilt van de oorspronkelijke: het gaat voornamelijk om verbeteringen van de morfosyntactische annotatie en om verandering van weergave van de codering. Daarnaast is het Renkema Corpus als component toegevoegd.

Het Renkema Corpus is samengesteld t.b.v. een promotieonderzoek naar ambtenarentaal, De Taal van Den Haag gepubliceerd in 1981. Het corpus bevat tekstfragmenten van correspondentie tussen de regering en de Staten-Generaal uit het parlementaire jaar 1975 – 1976. De correspondentie bestaat uit o.a. door Kamerleden gestelde schriftelijke vragen met daarop door de regering gegeven antwoorden, wetsontwerpen, vergaderstukken etc. De tekstfragmenten die opgenomen zijn in het Renkema Corpus bestaan uit telkens vier zinnen. In totaal bevat dit corpus ruim 48.000 woorden. Het Renkema Corpus kon eenvoudig toegevoegd worden aan het Eindhoven Corpus omdat beide hetzelfde codeersysteem hanteren (C3C): een lexicale codering met als doel woorden te onderscheiden naar grammaticale functie en eventueel morfologische eigenschappen. Dit codeersysteem is door de Werkgroep Frequentie-Onderzoek ontworpen in het kader van samenstelling van het Eindhoven Corpus.

4.3 Overige versies

Andere versies van het Eindhoven Corpus zijn te vinden bij verschillende onderzoeksinstituten en bedrijven in Nederland en België. In sommige gevallen gaat het om versies die nog verder verrijkt zijn dan de oorspronkelijke versie. Zo heeft men bij de Universiteit Groningen de Alpino Treebank die o.a. de syntactische annotatie van de dagbladencomponent (CDBL) uit het Eindhoven Corpus bevat.

5. Beschrijving van de verschillende corpuscomponenten

Het Eindhoven Corpus (VU-versie) bestaat uit verschillende componenten waarvan de inhoud in onderstaande tabel te vinden is. De teksten bestaan uit fragmenten die in lengte variëren van 75 tot 250 woorden.

regel-nummers	naam	omschrijving	Inhoud	aantal woorden	periode
1 – 6632	CDBL	dagbladen	Nieuwsblad van het Noorden, NRC, Het Vrije Volk, De Telegraaf, De Tijd en Trouw	120.000	1969-1970
6633 – 13445	COBL	opiniebladen	Elseviers Magazine, HP Magazine, Vrij Nederland, Accent, De Groene, Hervormd Nederland en De Nieuwe Linie	120.000	1970-1971
13446 – 21216	CGBL	gezinsbladen	Avenue, Televizier, Elegance, Eva, Libelle, Margriet, Nieuwe Revue, Panorama, Prinses, Televizier en Sextant	120.000	1970-1971
21217 – 31479	CRNO	romans en novellen	52 werken door Nederlandse auteurs die in 1970 als eerste druk verschenen zijn	120.000	1970
31480 – 37234	CPWE	populair wetenschappelijke boeken	beeldende kunsten, geneeskunde en psychiatrie, godsdienst en filosofie, onderwijs en opvoeding, natuurwetenschappen, plant- en dierkunde, land- en volkenkunde, verkeer en techniek, geschiedenis, psychologie, recht en economie, sociale wetenschappen (5 uit elke categorie)	120.000	1964-1973
37235 – 51077	CGTL	gesproken taal 1 (CGT1)	academici of daarmee gelijk te stellen personen die geacht worden ABN te spreken: groepsdiscussies, interviews, gesproken brieven	60.000	1960-1973
		gesproken taal 2 (CGT2)	mensen met weinig schoolopleiding uit een aantal plaatsen in Noord- en Zuid-Holland: vrije conversaties in ongedwongen sfeer	60.000	1960-1973
51078 – 53035	CAMB	Renkema Corpus	ambtenarentaal: correspondentie tussen regering en Staten-Generaal	48.000	1975-1976

6. Voorbeeldfragmenten

corpuscomponent :	cpwe (populair wetenschappelijke publicaties)
fragment :	31480
<p><jeraaktnoo-1-cpwe><<></p> <p>31480 < een 450 verrassende 203 situatie 000 : op 600 25 470 februari 010 vlogen 246 , hoewel 710 het 440 dus 510 officieel 150 nog 500 winter 000 was 275 , in 600 mijn 330 soester 100 tuin 000 en 700 boven 600 de 370 belendende 203 erven 001 kriskras 500 een 450 groot 100 deel 000 van 600 de 370 dag 000 , toen 710 het 370 blauw 120 van 600 de 370 hemel 000 slechts 500 tijdelijk 150 wat 450 lichte 103 sliertlijnen 001 van 600 wolken 001 vertoonde 255 , tal 000 van 600 citroentjes 001 , die 420 als 720 imagines 001 , dus 510 als 720 volkomen 100 vlinders 001 , hadden 276 overwinterd 206 , maar 700 die 420 door 600 de 370 temperatuurstijging 000 uit 600 hun 330 ' slaap 000 ' waren 276 gewekt 216 . ></p> <p><i>..een verrassende situatie : op 25 februari vlogen, hoewel het dus officieel nog winter was, in mijn Soester tuin en boven de belendende erven kriskras een groot deel van de dag, toen het blauw van de hemel slechts tijdelijk wat lichte sliertlijnen van wolken vertoonde, tal van citroentjes, die als imagines, dus als volkomen vlinders, hadden overwinterd, maar die door de temperatuurstijging uit hun 'slaap' waren gewekt ..</i></p>	

corpuscomponent :	cgbl (gezinsbladen)
fragment :	elegance 12-12-3
<p><elegan 12-12-3-cgbl><<></p> <p>14466 < een 450 huwelijk 000 heeft 253 geen 450 eigen 450 stem 000 . ></p> <p>14467 < de 370 man 000 en 700 de 370 vrouw 000 die 420 voelen 254 dat 710 hun 330 huwelijk 000 vermoeid 216 raakt 273 – meestal 500 herkennen 254 ze 300 de 370 symptomen 001 niet 500 terzelfdertijd 500 of 700 niet 500 beiden 441 even 500 duidelijk 150 – durven 274 zich 340 er 5400 niet 500 over 6105 uit 6200 te 650 spreken 2005 . ></p> <p>14468 < ze 300 hopen 254 maar 500 dat 710 zwijgen 200 ook 500 in 600 dit 370 geval 000 goud 000 is 273 en 700 dat 710 de 370 moeheid 000 op 600 den 373 duur 000 vanzelf 500 wel 500 zal 273 verdwijnen 200 . ></p> <p>14469 < ze 300 wachten 264 er 5400 zich 340 wel 500 voor 6105 een 450 onderzoek 000 in 6200 te 650 stellen 2105 naar 600 de 370 oorzaken 001 . ></p> <p>14470 < ze 300 houden 264 zich 340 blind 100 voor 600 het 370 feit 000 , dat 710 hen 303 toch 500 uit 600 lichamelijke 103 ervaring 000 bekend 100 moet 273 zijn 230 : dat 710 die 370 vermoeidheidverschijnselen 001 een 450 aanduiding 000 kunnen 274 zijn 230 van 600 een 450 heimelijke 103 ziekte 000 , die 420 zich 340 nog 500 niet 500 heeft 273 geopenbaard 206 , in 600 ieder 450 geval 000 een 450 aanwijzing 000 zijn 274 dat 710 er 510 wat 440 ontbreekt 243 ; dat 710 vermoeidheid 000 een 450 mens 000 minder 454 weerbaar 100 maakt 253 tegen 600 kwalen 001 en 700 ongelukken 001 ; >></p> <p><i>Een huwelijk heeft geen eigen stem. De man en de vrouw die voelen dat hun huwelijk vermoeid raakt – meestal herkennen ze de symptomen niet terzelfdertijd of niet beiden even duidelijk – durven zich er niet over uit te spreken. Ze hopen maar dat zwijgen ook in dit geval goud is en dat de moeheid op den duur vanzelf wel zal verdwijnen. Ze wachten er zich wel voor een onderzoek in te stellen naar de oorzaken. Ze houden zich blind voor het feit, dat hen toch uit lichamelijke ervaring bekend moet zijn: dat die vermoeidheidverschijnselen een aanduiding kunnen zijn van een heimelijke ziekte, die zich nog niet heeft geopenbaard, in ieder geval een aanwijzing zijn dat er wat ontbreekt; dat vermoeidheid een mens minder weerbaar maakt tegen kwalen en ongelukken;</i></p>	

corpuscomponent :	cgtl (gesproken taal – hoogopgeleide sprekers)
fragment :	abn band 571
<p><abn-1-band 571><<></p> <p>37235 < l 998 ></p> <p>37236 < en 700 ... 991 de 370 ellende 000 je 440 hebt 252 van 600 te 500 voren 500 dan 510 heb 272 je 440 ... 991 dus 510 allerhande 450 begripen 001 gehad 216 in 600 in 600 de 370 filosofie 000 . ></p> <p>37237 < bijvoorbeeld 500 als 710 je 440 daar 510 het 370 woord 000 mens 090 noemt 253 of 700 of 700 sociaal 090 of 700 zoiets 360 . ></p> <p>37238 < m 998 ></p> <p>37239 < ja 800 . ></p> <p>37240 < ja 800 , ja 800 . ></p> <p>37241 < l 998 ></p> <p>37242 < ... 991 in 600 de 370 sociologie 000 ... 991 hebben 254 die 360 'n 450 heel 500 andere 453 betekenis 000 dan 720 ... 991 dan 720 in 600 filosofie 000 en 700 dat 360 vind 251 ik 300 in 600 't 370 begin 000 hardstikke 500 lastig 100 , want 700 dan 510 ... 991 ja 800 je 440 moet 272 'n 450 hele 503 andere 453 denkmentaliteit 000 omschakelen 210 om 650 deze 370 studierichting 000 weer 500 mee 6200 te 650 kunnen 230 komen 2105 . ></p> <p>37243 < x 998 ></p> <p>37244 < maar 700 ik 300 dacht 255 dat 710 je 440 toch 500 ... 991 sociologisch 100 werd 275 ingesteld 216 of 700 ? ></p> <p>37245 < l 998 ></p> <p>37246 < ja 800 maar 700 't 440 zal 273 waarschijnlijk 150 wel 500 uit 600 de 370 filosofie 000 ook 500 vooral 500 komen 200 ... 991 dat 710 je 440 daar 510 dus 510 die 370 begripen 001 , die 370 begripen 001 die 360 zijn 274 veel 450 breder 104 dan 720 ... 991 in 600 de 370 sociologie 000 dus 510 ... 991 . ></p> <p>37247 < zo 510 ik 300 tenminste 510 de 370 eerste 493 maanden 001 zo 510 ervaren 206 heb 271 . >>></p>	
<p><i>< en ... de ellende je hebt van te voren dan heb je ... dus allerhande begripen gehad in in de filosofie . ></i></p> <p><i>< bijvoorbeeld als je daar het woord mens noemt of of sociaal of zoiets . ></i></p> <p><i>< m ></i></p> <p><i>< ja . ></i></p> <p><i>< ja , ja . ></i></p> <p><i>< l ></i></p> <p><i>< ... in de sociologie ... hebben die 'n heel andere betekenis dan ... dan in filosofie en dat vind ik in 't begin hardstikke lastig , want dan ... ja je moet 'n hele andere denkmentaliteit omschakelen om deze studierichting weer mee te kunnen komen . ></i></p> <p><i>< x ></i></p> <p><i>< maar ik dacht dat je toch ... sociologisch werd ingesteld of ? ></i></p> <p><i>< l ></i></p> <p><i>< ja maar 't zal waarschijnlijk wel uit de filosofie ook vooral komen ... dat je daar dus die begripen , die begripen die zijn veel breder dan ... in de sociologie dus ></i></p> <p><i>< zo ik tenminste de eerste maanden zo ervaren heb . >>></i></p>	

corpuscomponent :	cgtl (gesproken taal – laag opgeleide sprekers)
fragment :	warmenhuizen-II-1
<p><warmenhuizen-II-1-><<></p> <p>46196 < mm 998 ></p> <p>46197 < me 330 oudste 103 kleindochter 000 was 275 jarig 100 , enig 100 vanzelf 500 om 650 heen 6200 te 650 gaan 2005 . ></p> <p>46198 < maar 700 ja 800 , zakenmensen 001 he 800 , die 360 hebben 254 zoveel 450 drukte 000 aan 600 d'r 330 hoofd 000 , ik 300 moest 245 eerst 157 's 372 ochtends 002 na 600 alkmaar 010 , ze 300 woont 243 in 600 schagen 010 dus 510 ik 300 dacht 255 dan 510 's 372 ochtends 002 met 600 de 370 bus 000 na 600 alkmaar 010 en 700 dan 510 als 710 ik 300 daar 510 klaar 100 ben 271 met 600 't 370 spoor 000 na 600 schagen 010 . ></p> <p>46199 < 't 440 begon 245 al 500 slecht 150 want 700 op 600 't 370 oudhof 010 kocht 255 ik 300 'n 450 retourtje 000 nou 580 dat 360 vond 255 ik 300 zonde 000 en 700 vroeg 255 in 600 alkmaar 010 voordat 710 ik 300 uit 600 de 370 bus 000 ging 245 , of 710 't 300 nog 500 inruilen 210 kon 275 , dan 510 moest 275 ik 300 effentjes 500 na 600 't 370 kantoor 000 gaan 200 , nou 580 ik 300 dacht 255 als 710 ik 300 dan 510 straks 500 met 600 't 370 spoor 000 moet 241 ben 271 ik 300 er 510 vlakbij 500 en 700 dan 510 moet 243 dat 360 gebeuren 200 , dan 510 krijg 251 ik 300 bepaald 219 veertig 470 centen 001 terug 500 . ></p> <p>46200 < in 600 alkmaar 010 me 330 zaken 001 gedaan 216 , 't 370 kleindochtertje 000 had 275 d'r 330 kadootje 000 eigenlijk 150 al 500 gehad 216 , maar 700 met 600 lege 103 handen 001 kom 2420 je 440 niet 500 graag 150 aan 6205 dus 510 ik 300 dacht 255 dan 510 koop 251 ik 300 voor 600 alle 453 vier 460 een 450 klein 100 ... 991 he 800 , en 700 zeg 251 dat 710 ik 300 dat 360 maar 500 gedaan 216 had 275 omdat 710 ik 300 niet 500 goed 150 meer 500 wist 255 wie 400 er 510 eigenlijk 150 jarig 100 was 275 van 600 de 370 vier 460 . ></p> <p>46201 < ik 300 had 255 er 5400 bar 150 veel 450 verbeelding 000 van 6105 , ik 300 nog 500 effentjes 500 naar 600 't 370 n.a.c.o.-kantoor 000 , om 600 me 330 veertig 470 centen 001 . > 46202 < maar 700 ik 300 moest 275 me 330 adres 000 en 700 me 330 kaartje 000 geven 210 en 700 dan 510 zou 275 't 440 me 303 met 600 de 370 post 000 gestuurd 216 worden 230 . ></p> <p>46203 < nou 580 ik 300 dacht 255 al 500 , dat 360 zal 273 'n 450 klap 000 geven 210 . ></p> <p>46204 < en 700 ik 300 kreeg 255 dan 510 ook 500 later 154 drie 470 postzegels 001 van 600 vijf 470 centen 001 . > 46205 < afijn 800 toen 510 na 600 't 370 spoor 000 , dat 360 ging 245 best 107 , ik 300 zat 245 lekker 100 en 700 ik 300 was 275 in 600 'n 450 reuze 103 stemming 000 , schagen 010 werd 275 er 510 geroepen 216 , vlakbij 500 de 370 deur 000 zat 245 'n 450 jonge 103 vrouw 000 die 360 wipte 245 d'ruit 540 . ></p> <p>46206 < ik 300 zat 245 'n 450 heel 100 eind 000 verder 154 en 700 voordat 710 ik 300 bij 600 de 370 deur 000 was 275 stroomden 246 de 370 mensen 001 d'r 5400 al 500 weer 500 in 6105 . > 46207 < toen 510 ik 300 na 600 de 370 deur 000 maar 700 ik 300 kon 275 'm 303 onmogelijk 150 open 100 krijgen 210 ik 300 douwde 245 , ik 300 trok 245 , ik 300 wou 275 schuiven 200 , niks 440 ging 245 er 510 , eerst 500 was 275 ik 300 te 500 groots 100 om 650 'n 450 ander 440 te 650 vragen 200 , maar 700 't 440 moest 275 wel 500 , die 360 had 255 't 440 direkt 150 voor 600 mekaar 340 , met 600 twee 470 handen 001 in 600 twee 470 kanten 001 inschuiven 200 . ></p> <p>46208 < 't 440 is 273 maar 500 'n 450 weet 000 . ></p> <p>46209 < 't 370 spoor 000 begon 275 wel 500 effentjes 500 heel 500 zachtjes 500 al 500 te 650 rijen 200 , ik 300 had 275 er 5400 nog 500 best 157 uit 6105 gekend 206 maar 70 d'r 510 liep 245 'n 450 man 000 met 600 'n 450 pet 000 die 360 zeide 255 , binnenblijven 200 , we 300 bennen 274 zo 510 terug 500 . ></p> <p>46210 < ik 300 dacht 255 , zeker 150 effentjes 500 rangeren 200 . ></p> <p>46211 < maar 700 hij 300 reed 245 en 700 hij 300 reed 245 en 700 hield 2450 maar 500 niet 500 op 6205 , ik 300 zei 255 tegen 600 de 370 conducteur 000 , waar 5500 gaan 2441 we 300 na 6105 toe 6206 ? ></p> <p>46212 < na 600 anna-paulownapolder 010 zei 255 hij 300 . ></p> <p>46213 < nou 580 joh 800 , ik 300 kon 275 wel 500 janken 200 , daar 510 zat 245 ik 300 met 600 me 330 kadootjes 001 en 700 me 330 verlangen 000 . ></p> <p>46214 < in 600 anna-paulowna 010 moest 275 ik 300 aan 600 de 370 chef 000 m'n 330 kaartje 000 maar 500 zien 210 laten 230 , zei 255 de 370 conducteur 000 , dan 510 kon 245 ik 300 over 600 'n 450 dik 150 half 100 uur 000 wel 500 weer 500 voor 600 niks 440 terug 500 , nou 580 ik 300 direkt 150 op 600 de 370 chef 000 af 630 ... 991 die 420 daar 510 buiten 500 stond 245 , maar 700 't 440 stormde 245 en 700 dat 710 ik 300 'm 303 't 370 kaartje 000 geven 210 wou 275 , vloog 2550 't 300 zo 510 hoog 100 de 370 lucht 000 in 6205 dat</p>	

710 we 300 't 303 alle 453 twee 460 nooit 510 meer 500 gezien 216 hebben 274 , hij 300 zou 275 schagen 010 wel 500 opbellen 210 dan 510 kon 275 ik 300 bij 600 de 370 controle 000 zo 510 doorlopen 200 , zei 255 die 360 want 700 hij 300 had 275 't 303 gezien 216 . >

46215 < 'n 450 goeie 103 man 000 hoor 287 , want 700 daar 5400 geloofde 255 ik 300 niks 440 van 6105 . >

46216 < dat 370 half 100 uur 000 heb 273 ... 991 heb 273 lang 150 geduurd 206 hoor 287 in 600 't 370 wachthuis 000 van 600 anna-paulownapolder 010 , waar 530 'n 450 paar 000 ruiten 001 door 600 de 370 storm 000 uitgewaaid 206 waren 276 . >>>

< mm >

< me oudste kleindochter was jarig, enig vanzelf om heen te gaan. >

< maar ja, zakenmensen he, die hebben zoveel drukte aan d'r hoofd, ik moest eerst 's ochtends na alkmaar, ze woont in schagen dus ik dacht dan 's ochtends met de bus na alkmaar en dan als ik daar klaar ben met 't spoor na schagen. >

< 't begon al slecht want op 't oudhof kocht ik 'n retourtje nou dat vond ik zonde en vroeg in alkmaar voordat ik uit de bus ging, of 't nog inruilen kon, dan moest ik effentjes na 't kantoor gaan, nou ik dacht als ik dan straks met 't spoor moet ben ik er vlakbij en dan moet dat gebeuren, dan krijg ik bepaald veertig centen terug. >

< in alkmaar me zaken gedaan, 't kleindochtertje had d'r kadootje eigenlijk al gehad, maar met lege handen kom je niet graag aan dus ik dacht dan koop ik voor alle vier een klein... he, en zeg dat ik dat maar gedaan had omdat ik niet goed meer wist wie er eigenlijk jarig was van de vier. >

< ik had er bar veel verbeelding van, ik nog effentjes naar 't n.a.c.o.-kantoor, om me veertig centen.>

< maar ik moest me adres en me kaartje geven en dan zou 't me met de post gestuurd worden. >

< nou ik dacht al, dat zal 'n klap geven. >

< en ik kreeg dan ook later drie postzegels van vijf centen. >

< afijn toen na 't spoor, dat ging best, ik zat lekker en ik was in 'n reuze stemming, schagen werd er geroepen, vlakbij de deur zat 'n jonge vrouw die wipte d'ruit. >

< ik zat 'n heel eind verder en voordat ik bij de deur was stroomden de mensen d'r al weer in. >

< toen ik na de deur maar ik kon 'm onmogelijk open krijgen ik douwde, ik trok, ik wou schuiven, niks ging er, eerst was ik te groots om 'n ander te vragen, maar 't moest wel, die had 't direkt voor mekaar, met twee handen in twee kanten inschuiven. >

< 't is maar 'n weet. >

< 't spoor begon wel effentjes heel zachtjes al te rijden, ik had er nog best uit gekend maar d'r liep 'n man met 'n pet die zeide, binnenblijven, we bennen zo terug. >

< ik dacht, zeker effentjes rangeren. >

< maar hij reed en hij reed en hield maar niet op, ik zei tegen de conducteur, waar gaan we na toe ? >

< na anna-paulownapolder zei hij. >

< nou joh, ik kon wel janken, daar zat ik met me kadootjes en me verlangen. >

< in anna-paulowna moest ik aan de chef m'n kaartje maar zien laten, zei de conducteur, dan kon ik over 'n dik half uur wel weer voor niks terug, nou ik direkt op de chef af... die daar buiten stond, maar 't stormde en dat ik 'm 't kaartje geven wou, vloog 't zo hoog de lucht in dat we 't alle twee nooit meer gezien hebben, hij zou schagen wel opbellen dan kon ik bij de controle zo doorlopen, zei die want hij had 't gezien. >

< 'n goeie man hoor, want daar geloofde ik niks van. >

< dat half uur heb... heb lang geduurd hoor in 't wachthuis van anna-paulownapolder, waar 'n paar ruiten door de storm uitgewaaid waren. >>>

Referenties

- Dave van Grootheest (1992) Handleiding bij het Eindhoven Corpus (VU-versie), Vrije Universiteit Amsterdam.
- P.C. Uit den Boogaard (red.), 1975, Woordfrequenties in geschreven en gesproken Nederlands, Utrecht: Oosthoek, Scheltema en Holkema
- J. Renkema, 1981, De taal van "Den Haag": een kwantitatief-stilistisch onderzoek naar aanleiding van oordelen over taalgebruik, 's Gravenhage: Staatsuitgeverij