

# EUR Research Information Portal

## Een model voor juridische informatica

### Publication status and date:

Published: 01/01/1984

### Document Version

Publisher's PDF, also known as Version of record

### Citation for the published version (APA):

De Mulder, RV. (1984). *Een model voor juridische informatica*. [Doctoral Thesis, Erasmus University Rotterdam]. Erasmus Universiteit Rotterdam (EUR).

[Link to publication on the EUR Research Information Portal](#)

### Terms and Conditions of Use

Except as permitted by the applicable copyright law, you may not reproduce or make this material available to any third party without the prior written permission from the copyright holder(s). Copyright law allows the following uses of this material without prior permission:

- you may download, save and print a copy of this material for your personal use only;
- you may share the EUR portal link to this material.

In case the material is published with an open access license (e.g. a Creative Commons (CC) license), other uses may be allowed. Please check the terms and conditions of the specific license.

### Take-down policy

If you believe that this material infringes your copyright and/or any other intellectual property rights, you may request its removal by contacting us at the following email address: [openaccess.library@eur.nl](mailto:openaccess.library@eur.nl). Please provide us with all the relevant information, including the reasons why you believe any of your rights have been infringed. In case of a legitimate complaint, we will make the material inaccessible and/or remove it from the website.

## INLEIDING TOT DE HOOFDSTUKKEN 9 T.M. 12

In de nu volgende vier hoofdstukken zullen vier projecten worden besproken, alle uitgevoerd aan de Erasmus Universiteit, die een toepassing vormen van de in vorige hoofdstukken gegeven theoretische overwegingen.

Het in de projecten uitgevoerde onderzoek heeft betrekking op pragmatische, semantische en syntactische aspecten van juridisch taalgebruik en de relaties daartussen. Juridische wetenschap heeft tot taak hiernaar empirisch onderzoek te doen en de eerste drie van de te beschrijven projecten zijn een poging in deze richting. Computers bleken bij deze projecten een handig, zo niet onmisbaar hulpmiddel te zijn, alleen al omdat de betrokken tekst- c.q. databestanden steeds van vrij grote omvang waren.

Het eerste project (hoofdstuk 9) concentreert zich op de syntaxis, de onderzochte teksten zijn Nederlandse wetten. Het doel is na te gaan of het woordgebruik van het Wetboek van Strafrecht en het Wetboek van Strafvordering verschilt van ander geschreven Nederlands.

Het tweede project (hoofdstuk 10) concentreert zich op de semantiek, de onderzochte teksten zijn artikelen in het "Tijdschrift voor Strafrecht" en "Delikt en Delinkwent". Doel is met behulp van de techniek van inhoudsanalyse te meten aan welke onderwerpen in beide tijdschriften gedurende een bepaalde periode aandacht werd besteed, en hoeveel.

Het derde project (hoofdstuk 11) richt zich vooral op de pragmatiek, de onderzochte teksten zijn Nederlandse rechterlijke uitspraken (afkomstig van de Hoge Raad der Nederlanden en de Centrale Raad van Beroep). Het doel is door analyse van de uitspraken te onderzoeken welke factoren een rol spelen in de onderzochte rechterlijke beslissingen en te trachten deze beslissingen te voorspellen, waarbij tevens wordt getracht uitspraken te doen over de bijdrage die individuele rechters aan de onderzochte beslissingen hebben gehad. (Dit project is, in tegenstelling tot de drie andere, nog niet geheel afgerond op het moment van publikatie van deze studie. De beschrijving ervan is daarom voorlopig en in bepaalde opzichten niet volledig.)

Het vierde project (hoofdstuk 12) heeft tot doel na te gaan of het mogelijk is bepaalde juridische beslissingen (in dit geval straftoematingsbeslissingen) met behulp van een computer te nemen. Het eindprodukt van dit

projekt bestaat uit een systeem voor het nemen van beslissingen met behulp van een computer. De activiteiten die hiertoe geleid hebben kan men rekenen tot de "artefactkunde", een term waarop ik in hoofdstuk 13 nader zal ingaan.

## HOOFDSTUK 9. PROJEKT: ANALYSE VAN JURIDISCH WOORDGEBRUIK

### 9.1 DOEL VAN HET PROJEKT

Het eindverslag van het project is gepubliceerd onder de titel "Het woordgebruik van het Wetboek van Strafrecht en het Wetboek van Strafvordering, een vergelijking van het woordgebruik van de Nederlandse wetboeken met dat van gewone Nederlandse schrijftaal" (De Mulder en Oskamp, 1979). De beschrijving hier is uiteraard minder gedetailleerd dan die in het onderzoeksrapport.

#### 9.1.1 Vergelijking Van Het Woordgebruik

Doel is het woordgebruik van het Wetboek van Strafrecht en het Wetboek van Strafvordering te vergelijken met dat van gewone Nederlandse schrijftaal. In dit project is getracht uitsluitend op het syntactisch niveau te opereren, en het semantisch niveau zoveel mogelijk te mijden en zich te beperken tot vergelijkingen op woordniveau.

Aangezien noch in Nederland, noch in het buitenland, voorzover bekend, het probleem aldus is gesteld, betreft het hier een onderzoek van exploratieve aard. Dit betekent onder meer, dat aan het onderzoek geen direkt te toetsen hypothesen ten grondslag liggen. Het is slechts de bedoeling te beschrijven en - eventueel - interessante hypothesen te formuleren.

#### 9.1.2 Instrumentontwikkeling

Een tweede doel is een instrument te ontwikkelen om een dergelijke vergelijking van het woordgebruik uit te voeren. Daartoe zal onder meer een techniek gevonden moeten worden om in het woordgebruik zekere patronen, een zekere structuur te ontdekken. Gepoogd is een enigszins exacte wijze van beschrijven van woordgebruik te vinden.

#### 9.1.3 Illustratie Van Technische Mogelijkheden

De technologische ontwikkelingen op het terrein van elektronische data-verwerkende apparatuur bieden naar onze mening aan de rechtswetenschap allerlei nieuwe mogelijkheden om tot verdere ontplooiing te komen. Een van de doelen van dit onderzoek is, een illustratie te geven van enkele van die mogelijkheden.

## 9.2 MATERIAALVERZAMELING EN EERSTE BEWERKING

Het materiaal waarop de analyses worden uitgevoerd bestaat voornamelijk uit frequentielijsten: voor ieder woord (type), dat in een tekstbestand (verder ook wel "corpus" genoemd) voorkomt is een getal bekend, dat aangeeft hoe vaak dat woord in het betreffende corpus voorkomt.

### 9.2.1 Het Materiaal Van De Juridische Teksten

De frequentielijsten voor het Wetboek van Strafrecht en het Wetboek van Strafvordering werden speciaal voor dit onderzoek vervaardigd door de firma Kluwer. Gewerkt werd met frequentielijsten, geordend naar aflopende frequentie en zo'n lijst ziet er voor het Wetboek van Strafrecht voor de eerste vijftig types b.v. als volgt uit:

Tab. 9a

|      |                 |     |             |
|------|-----------------|-----|-------------|
| 2628 | VAN             | 317 | OP          |
| 1941 | DE              | 310 | DAT         |
| 1900 | OF              | 283 | DOOR        |
| 1203 | HET             | 280 | ZIJN        |
| 1063 | EEN             | 268 | VOOR        |
| 991  | TEN             | 262 | BIJ         |
| 934  | IN              | 259 | NIET        |
| 876  | MET             | 253 | TOT         |
| 731  | DIE             | 235 | ENE         |
| 730  | HOOGSTE         | 235 | GELDBOETE   |
| 713  | WORDT           | 227 | HEEFT       |
| 646  | ART             | 215 | ARTIKEL     |
| 569  | HIJ             | 213 | GULDEN      |
| 498  | IS              | 212 | KAN         |
| 488  | TE              | 211 | DER         |
| 486  | GESTRAFT        | 197 | ALS         |
| 470  | GEVANGENISSTRAF | 184 | OPZETTELIJK |
| 467  | EN              | 169 | 3           |
| 399  | 1               | 161 | MAANDEN     |
| 396  | AAN             | 156 | FEIT        |
| 369  | 2               | 151 | MISDRIJF    |
| 359  | JAREN           | 135 | ZES         |
| 343  | DEN             | 134 | HEM         |
| 340  | INDIEN          | 129 | HECHTENIS   |
| 326  | WORDEN          | 126 | DAN         |

Deze lijsten van het Wetboek van Strafrecht en het Wetboek van Strafvordering werden in een computer ingevoerd. De types met zeer kleine frequentie (voor het W.v.Sr.  $< 2$  en voor het W.v.Sv.  $< 3$ ) werden om praktische redenen weggelaten. Uit deze twee lijsten werd een nieuw bestand samengesteld: een voor het woordgebruik van het W.v.Sr. en het W.v.Sv. gezamenlijk. Daarbij werden om redenen van betrouwbaarheid typen van een frequentie  $< 6$  weggelaten.

uit praktische overwegingen - alleen al deze, dat er geen bestaand vergelijkbaar alternatief bestond voor de lijsten van de werkgroep - lag het gebruik van de resultaten uit 'Woordfrequenties' voor de hand. Aangezien het hier ging om de vergelijking met teksten van geschreven strafrechtelijke taal, werd de zesde afdeling van het corpus, gesproken taal, niet in het onderzoek betrokken.

De frequentielijsten afkomstig uit "Woordfrequenties" zien er b.v. (voor de ca. eerste vijftig van het corpus "Nederlandse schrijftaal") als volgt uit:

Tab. 9b

|          |       |           |      |
|----------|-------|-----------|------|
| de 370   | 32843 | door 600  | 2298 |
| van 600  | 17284 | naar 600  | 2179 |
| een 450  | 14641 | maar 700  | 2149 |
| en 700   | 13560 | zich 340  | 2065 |
| het 370  | 12308 | als 720   | 2057 |
| in 600   | 11429 | bij 600   | 1963 |
| is 273   | 6944  | er 510    | 1911 |
| te 650   | 6370  | zijn 274  | 1853 |
| niet 500 | 5582  | tot 600   | 1811 |
| met 600  | 5092  | om 650    | 1648 |
| op 600   | 5082  | Ik 300    | 1645 |
| dat 710  | 5049  | dan 510   | 1609 |
| die 420  | 4349  | of 700    | 1586 |
| voor 600 | 4090  | heeft 273 | 1574 |
| De 370   | 3825  | over 600  | 1533 |
| hij 300  | 3574  | dat 360   | 1527 |
| ik 300   | 3392  | uit 600   | 1490 |
| zijn 330 | 3183  | wel 500   | 1460 |
| was 275  | 3042  | zo 510    | 1454 |
| het 440  | 2998  | geen 450  | 1440 |
| aan 600  | 2951  | die 370   | 1412 |
| ook 500  | 2593  | deze 370  | 1394 |
| ze 300   | 2308  | Hij 300   | 1393 |
| nog 500  | 2306  | wordt 273 | 1388 |

### 9.2.3 Eerste Bewerking En Selectie

De lijsten, zoals verschaft door "Woordfrequenties" verschilden nogal van de Kluwerlijsten. In de eerste wordt onderscheid gemaakt tussen hoofd- en kleine letters ("De" is niet hetzelfde als "de"), en worden homoniemen (aan de hand van grammaticale codes) zoveel mogelijk uit elkaar gehouden. Types die bij elkaar horen (zoals Den Haag) worden als een woord beschouwd. Leestekens (zoals in "m'n") worden niet genegeerd.

De gekozen lijsten uit "Woordfrequenties" werden na invoering in een computer zodanig bewerkt dat ze vergelijkbaar waren met de lijsten van het Wetboek van Strafrecht en het Wetboek van Strafvordering. De hier-

bovengenoemde verschillen werden opgegeven, de lijsten opnieuw geordend naar dalende frequentie. Als laagste frequentie werd om praktische redenen en om redenen van betrouwbaarheid dezelfde laagste frequentie genomen als die welke de werkgroep frequentieonderzoek koos voor de lijsten, geordend naar dalende frequentie.

Om het juridische woordgebruik niet alleen te kunnen vergelijken met gewone Nederlandse schrijftaal, maar ook met wellicht meer vergelijkbare bestanden, werd op de bestanden van

1. de Nederlandse dagbladen,
2. de opiniebladen en
3. de populair-wetenschappelijke boeken

dezelfde bewerking en selectie toegepast.

#### 9.2.4 Het Basismateriaal Na Eerste Bewerking

Het eindproduct van deze bewerkingen en selectie bestond dus uit zeven verschillende, voor de computer leesbare frequentielijsten, geordend naar dalende frequentie (en binnen gelijke frequenties alfabetisch):

Tab. 9c

| Bestand  | afkorting | steekproefgrootte | aantal tekens in lijst | aantal typen in lijst | laatste nog opgenomen frequentie |
|--|-----------|-------------------|------------------------|-----------------------|----------------------------------|
| Wetboek van Strafrecht                               | SR        | 47406             | 45799                  | 2339                  | 2                                |
| Wetboek van Strafvordering                           | SV        | 59488             | 57209                  | 1705                  | 3                                |
| Wetboek van Strafrecht en Strafvordering gezamenlijk | GW        | 106849            | 98401                  | 1543                  | 6                                |
| Nederlandse schrijftaal                              | SC        | 605733            | 407923                 | 1068                  | 40                               |
| Nederlandse dagbladen                                | DB        | 120313            | 78927                  | 1012                  | 10                               |
| Opiniebladen   | OP        | 123873            | 83082                  | 978                   | 10                               |
| Populair-wetenschappelijke boeken                    | PW        | 119979            | 82085                  | 1073                  | 10                               |

## 9.3 ENKELE ANALYSES EN RESULTATEN

### 9.3.1 De Tabellen

Uit het materiaal zijn een aantal tabellen vervaardigd, voor de inhoud waarvan korthedshalve verwezen wordt naar De Mulder en Oskamp, 1979, blz. 91 e.v..

Voor elk van de zeven bestanden, geordend naar dalende frequentie zijn voor ieder type uitgerekend:

1. rangnummer,
2. frequentie,
3. relatieve frequentie (frequentie gedeeld door steekproefgrootte),
4. cumulatieve frequentie,
5. relatieve cumulatieve frequentie,
6. het produkt van rangnummer en frequentie
7. het relatieve produkt van rangnummer en frequentie.

Onder "rangnummer" van een type wordt verstaan het getal dat aangeeft op welke plaats dat type voorkomt in de de frequentielijst, geordend naar dalende frequentie. Het meest voorkomende type heeft dus rangnummer 1, het een na vaakst voorkomend heeft rangnummer 2, etc.. Typen die even vaak voorkomen in een bepaald bestand hebben een gelijk rangnummer.

De relatieve grootheden zijn uitgerekend om vergelijking tussen de bestanden mogelijk te maken. Als eenheid is gebruikt het aantal malen dat een type voorkomt per miljoen tekens.

Bovendien is een omvangrijke tabel, de "overzichtstabel relatieve woordfrequenties" vervaardigd. Deze geeft voor ieder type (in alfabetische volgorde) aan hoe groot in ieder bestand de relatieve frequentie is.

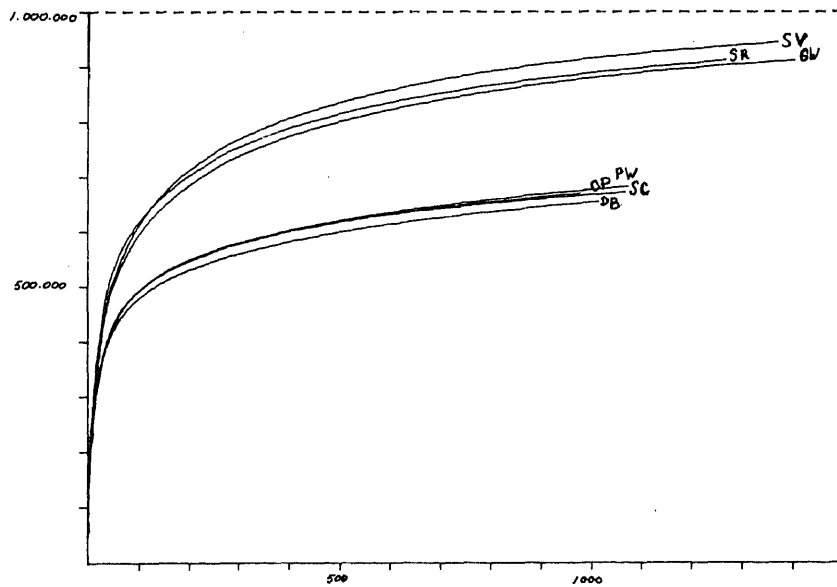
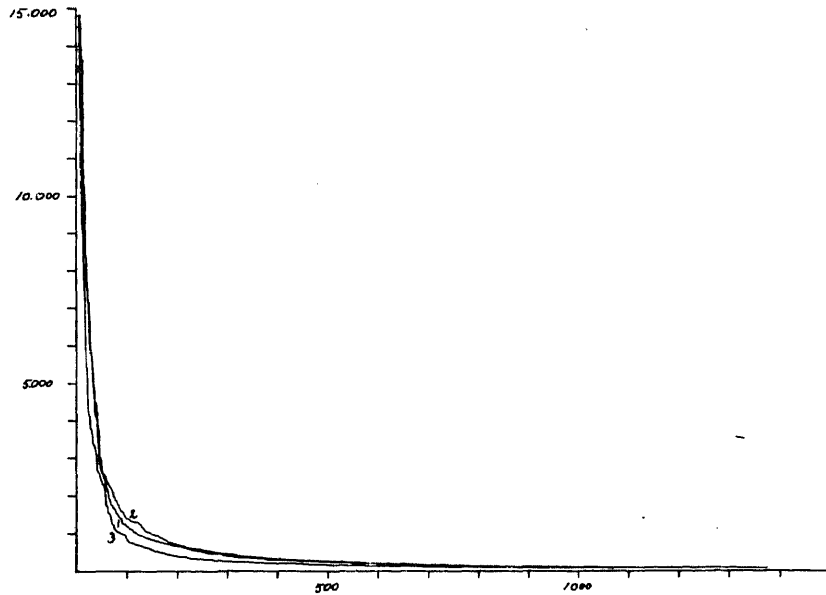
### 9.3.2 De Grafieken

Om een inzicht te krijgen in de structuur van het woordgebruik van het strafwettelijk woordgebruik in vergelijking met gewone Nederlandse schrijftaal zijn uit het cijfermateriaal in de tabellen verschillende grafieken - eveneens per computer - vervaardigd.



Op de relatieve frequentiegrafieken is nog weinig verschil te zien tussen de bestanden (fig. 9a). Voor de meeste rangordes geldt, dat de strafrechtelijke taal (SR (1) en SV (2)) een hogere relatieve woordfrequentie heeft dan gewone schrijftaal (SC (3)).

Fig. 9a en 9b



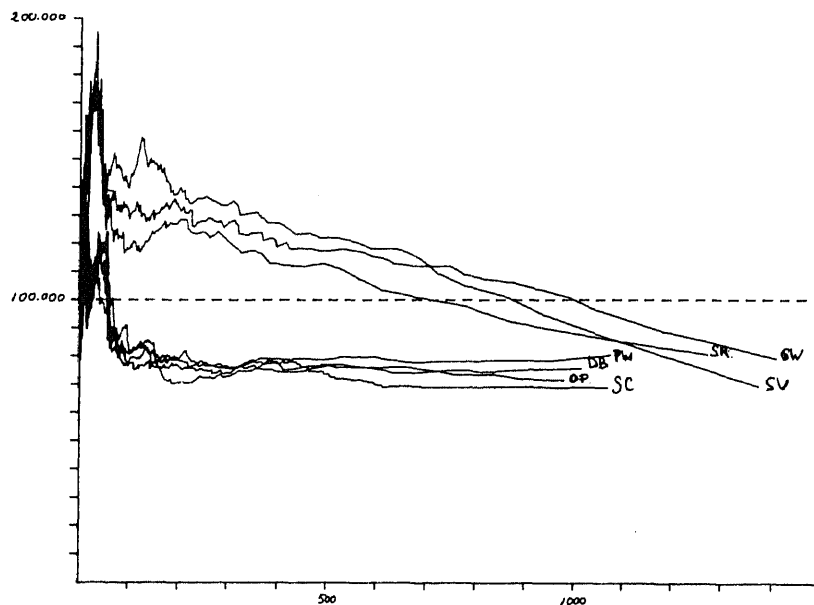
Op de cumulatieve (relatieve) frequentiegrafieken (fig. 9b) blijkt uit de relatief snelle stijging van de lijnen van SR, SV en GW veel duidelijker hetzelfde, maar voor rangordes kleiner dan 100 valt nog weinig te zien, omdat de lijnen praktisch samenvallen. De typen van kleine rangnummers zijn echter juist interessant, omdat dit de meest voorkomende zijn. (Voor het Wetboek van Strafrecht geldt b.v. dat de 200 meest voorkomende typen tezamen 70 % van het aantal tekens voor hun rekening nemen en ook bij Nederlandse schrijftaal is dit een hoog percentage n.l. ca. 50 %.)

Voor de derde soort grafiek, die over dit materiaal is vervaardigd, is gebruik gemaakt van een verschijnsel, dat in de statistische taalkunde bekend staat als de "Wet van Zipf". Deze luidt als volgt (Zie Woordfrequenties, 1975, blz. 25):

Bij natuurlijke talen is het produkt van rangnummer en frequentie voor ieder type steeds ongeveer constant.

In de grafieken is nu steeds op de verticale as uitgezet het produkt van rangnummer en (relatieve) frequentie en als de Wet van Zipf zou kloppen, dan moet het resultaat dus op een horizontale rechte lijn lijken. Zie fig. 9c.

Fig. 9c



N.B. Met deze toepassing wordt er niet van uitgegaan, dat de Wet van Zipf juist is. Evenmin wordt de Wet van Zipf aldus getoetst. (Hetgeen ook in strijd zou zijn met het exploratief karakter van dit onderzoek.) De ver-

wachting is slechts dat de verschillen tussen de bestanden op deze manier op een duidelijker manier in beeld gebracht kunnen worden.

Aan de aldus vervaardigde figuur valt onder meer het volgende op:

1. De "Wet van Zipf" gaat beter op voor voor gewoon Nederlands dan voor de strafrechtelijke taal: de krommen SC, DB, OP en PW lijken meer op een horizontale rechte dan SR, SV en GW.
2. De eerder getrokken konklusie dat althans voor de rangorden waarvoor de grafiek is vastgesteld, de typen van gelijke rangorde bij de strafrechtelijke bestanden vaker voorkomen bij gewoon Nederlands, wordt bevestigd. (Alleen voor zeer kleine rangnummers (< ca. 10) is dat nog niet te zien.)
3. De krommen hebben, vergeleken met de frequentie- en cumulatieve frequentiekrommen een veel grilliger verloop. Onderlinge verschillen tussen de strafrechtelijke bestanden zijn b.v. duidelijk te zien.
4. De krommen hebben een duidelijke structuur. Eerst een deel waarbij de berekende ("Zipf-")grootheden een snelle stijging en na enkele tientallen rangnummers weer een sterke daling vertonen, gevolgd door een deel, dat beter beantwoordt aan de verwachtingen volgens de Wet van Zipf. (Het eerste deel zal ik verder de "kop" van de kromme noemen.)
5. Behalve "hoger" (zie punt 1) is de kop van SR, SV en GW ook "korter" dan van de andere bestanden (rangorde tot ca. 50 vs. rangorde tot ca. 70). Dit betekent dat in de strafrechtelijke bestanden de "zeer vaak voorkomende types" niet alleen vaker voorkomen dan in de gewone schrijftaal maar ook kleiner in getal zijn.

Het grillige verloop van de karakteristieken heeft vooral betrekking op de typen met rangordes tot ca. 300. Het lijkt de moeite waard de kop nog wat nader te bekijken door verder "uitvergrooten". Vgl. fig. 9d voor SV en SC.

Fig. 9d

| Bestand | rangnummer | rel.freq. |
|---------|------------|-----------|
| SR      | 46         | 3185      |
| SV      | 51         | 2908      |
| GW      | 51         | 2788      |
| SC      | 70         | 1342      |
| DB      | 65         | 1346      |
| OP      | 61         | 1655      |
| PW      | 57         | 1650      |

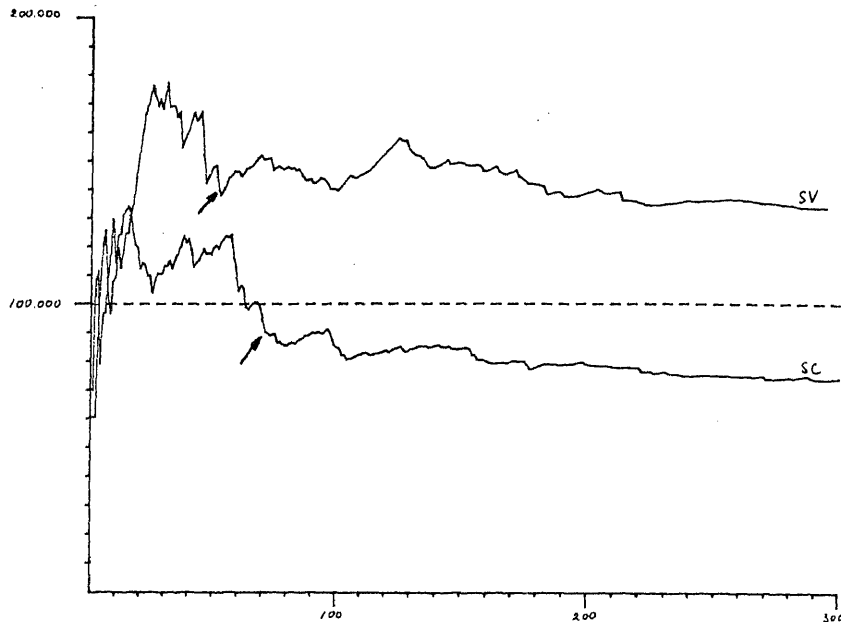
Grafieken als fig. 9d en de equivalenten voor de andere bestanden zijn geschikt om het punt te vinden waar de "kop" eindigt, m.a.w. om uit te maken welk type nog juist als "zeer vaak voorkomend" kan worden beschouwd. (Gezien de duidelijke verschillen in structuur lijkt zo'n keuze minder arbitrair dan b.v. het kiezen van een bepaalde minimale relatieve frequentie.) De gekozen punten zijn in de figuur met pijltjes aangegeven.

Uit de grafiek blijkt tenslotte ook, dat met name voor de rangordes ca. 10 niet altijd geldt dat types van dezelfde rangorde in de strafrechtelijke bestanden een grotere relatieve frequentie hebben dan die in de gewone schrijftaal. Die laatste konklusie wordt uiteraard bevestigd door de tabellen.

### 9.3.3 Iets Over Het Woordgebruik Inhoudelijk

Nu d.m.v. de grafieken een inzicht is verkregen in bepaalde aspecten van de structuur van het woordgebruik verschaffen de tabellen verschillende onderzoeksresultaten. Op basis van de volgende lijst van typen die nog juist tot de "zeer vaak voorkomende" gerekend mogen worden:

Tab. 9d



kunnen uit de tabellen de lijsten van "zeer vaak voorkomende" types worden afgelezen (zie De Mulder en Oskamp, 1979, blz. 91 e.v.). Aan die lijsten valt o.m. het volgende op:

De kop van gewoon Nederlands, van Nederlandse dagbladen, van Opiniebladen en van Populair-wetenschappelijke bladen bevat geen zelfstandige naamwoorden (uitzonderingen: de woorden "jaar" en "heer") en geen bijvoeglijke naamwoorden (uitzonderingen: "andere" en "veel").

De kop van de juridische bestanden bevat eveneens geen bijvoeglijke naamwoorden (uitzonderingen: "eerste" en "tweede"), doch wel talrijke zelfstandige naamwoorden. (De vraag of een bepaald type als bijvoeglijk naamwoord moet worden beschouwd schept problemen van semantische aard. Immers de vraag kan lang niet altijd alleen op syntactisch niveau worden beantwoord (vgl. "hoogste", "opzettelijk", zelfs: "rechter").)

Door onderlinge vergelijking van de gevonden lijsten van zeer vaak voorkomende types kan de vraag beantwoord worden welke zeer vaak voorkomende woorden nu als specifiek voor het Wetboek van Strafrecht, het Wetboek van Strafvordering resp. voor de beide wetboeken gezamenlijk moeten worden beschouwd.

Allereerst blijkt het zo te zijn, dat niet alle woorden die de kop van het Wetboek van Strafrecht, van het Wetboek van Strafvordering, of van beide wetboeken gezamenlijk uitmaken, ook in die van gewoon Nederlands voorkomen. Deze woorden zou men specifiek voor het betreffende bestand kunnen noemen.

1. Woorden die zeer vaak voorkomen in het Wetboek van Strafrecht, maar niet in gewoon Nederlands, zijn (in volgorde van frequentie):

TEN  
HOOGSTE  
ART  
GESTRAFT  
GEVANGENISSTRAF  
1  
2  
JAREN  
DEN  
INDIEN  
EENE  
GELDBOETE  
ARTIKEL  
GULDEN  
DER  
OPZETTELIJK  
3  
MAANDEN  
FEIT  
MISDRIJF

2. Woorden die zeer vaak voorkomen in het Wetboek van Strafvordering, maar niet in gewoon Nederlands, zijn (in volgorde van frequentie):

DEN  
ART  
VERDACHTE  
2  
1  
ARTIKEL  
INDIEN  
JUSTITIE  
LID  
RECHTER  
OFFICIER  
ONDERZOEK  
DER  
3  
RECHTBANK  
TOEPASSING  
TERECHTZITTING  
COMMISSARIS  
TER  
EERSTE  
TWEEDE  
GEVAL

3. Woorden die zeer vaak voorkomen in de Wetboeken van Strafrecht en Strafvordering gezamenlijk, maar niet in gewoon Nederlands zijn (in volgorde van frequentie):

ART  
DEN  
TEN  
1  
2  
HOOGSTE  
INDIEN  
ARTIKEL  
DER  
GESTRAFT  
GEVANGENISSTRAF  
LID  
3  
EENE  
RECHTER  
JAREN  
JUSTITIE

FEIT  
TOEPASSING  
OFFICIER  
TER  
ONDERZOEK

Deze verschillen bestaan dus uit:

- zelfstandige naamwoorden (inklusief "ART")
- archaische woorden of woorden van ouderwetse spelling ("DEN", "DER", "EENE", "TEN", "TER")
- telwoorden ("1", "2", "3", "EERSTE", "TWEEDE")
- overige: "GESTRAFT", "HOOGSTE", "INDIEN".

Men kan dergelijke vergelijkingen - ter vergelijking - ook maken voor de bestanden DB, OP en PW:

4. Woorden die zeer vaak voorkomen in de Nederlandse dagbladen, maar niet in gewoon Nederlands zijn (in volgorde van frequentie):

NA  
TEGEN  
TWEE  
HEER

5. Woorden die zeer vaak voorkomen in de Opiniebladen, maar niet in gewoon Nederlands zijn er niet.
6. Woorden die zeer vaak voorkomen in de Populair-wetenschappelijke bladen, maar niet in gewoon Nederlands zijn (in volgorde van frequentie):

DER  
NADERE

Aan de onder 4, 5 en 6 gegeven lijsten valt op, dat de onderzochte subcorpora van gewoon Nederlands, voor wat betreft de zeer vaak voorkomende woorden, en het inhoudelijke woordgebruik duidelijk minder van gewoon Nederlands verschillen dan de Wetboeken van Strafrecht en Strafvordering.

In de tweede plaats kunnen wij de lijst van specifieke vaak voorkomende woorden in het Wetboek van Strafrecht vergelijken met die van het Wetboek van Strafvordering.

7. Woorden die zeer vaak voorkomen in het Wetboek van Strafrecht, maar niet in gewoon Nederlands, en ook niet in het Wetboek van Strafvordering, zijn (in volgorde van frequentie):

TEN  
HOOGSTE  
GESTRAFT  
GEVANGENISSTRAF  
JAREN  
GELDBOETE  
GULDEN  
OPZETTELIJK  
MAANDEN  
FEIT  
MISDRIJF

8. Woorden die zeer vaak voorkomen in het Wetboek van Strafvordering, maar niet in gewoon Nederlands, en ook niet in het wetboek van Strafrecht, zijn (in volgorde van frequentie):

VERDACHTE  
JUSTITIE  
LID  
RECHTER  
OFFICIER  
ONDERZOEK  
RECHTBANK  
TOEPASSING  
TERECHTZITTING  
COMMISSARIS  
TER  
EERSTE  
TWEEDE  
GEVAL

9. De woorden, die overblijven wanneer men de onder 7. en 8. genoemde specifieke strafrechts- en strafvorderingswoorden wegstreept uit de onder 3. genoemde woorden van de Wetboeken van Strafrecht en Strafvordering tesamen zijn ook interessant, omdat zij misschien beschouwd mogen worden als de woorden die in het algemeen kenmerkend zijn voor strafrechtelijke wetstaal. Het zijn (in volgorde van frequentie):

ART  
DEN  
1  
2  
INDIEN



ARTIKEL  
DER  
DRIE  
EENE

Met het konstrueren van de lijsten als hierboven onder 1 t/m 9 gedaan, zijn de vergelijkingsmogelijkheden van het inhoudelijk woordgebruik nog geenszins uitgeput. De vergelijkingen zijn beperkt tot de "kop" van de steekproeven; de zeer vaak voorkomende woorden. Ook de vergelijking van de middenstukken van de Zipf-karakteristieken ("de tamelijk vaak voorkomende woorden") en van de "staart" (de weinig frequent voorkomende woorden) kan interessant zijn. Andere mogelijkheden zijn denkbaar. Ook wanneer men zich tot de zeer vaak voorkomende types beperkt zijn nog andere vergelijkingen denkbaar. Zo is b.v. in het bovenstaande niet nagegaan, welke types wel zeer vaak voorkomen in gewoon Nederlands, maar niet in de strafrechtelijke bestanden.

#### 9.3.4 Enkele Nadere Analyses

In het bovenstaande is een aantal konklusies geformuleerd over de structuur van het woordgebruik. Daarbij kwamen opvallende verschillen aan het licht tussen enerzijds het woordgebruik van de Wetboeken van Strafrecht en Strafvordering en anderzijds gewoon Nederlands. De bestanden, die daarbij vergeleken werden, vertoonden echter onderlinge verschillen, die de gevonden verschillen in structuur van het woordgebruik en daarmee misschien ook die in het inhoudelijke woordgebruik zouden kunnen verklaren. Teneinde sommige vragen t.a.v. de betrouwbaarheid en de validiteit van de getrokken konklusies te beantwoorden, zijn in het projekt enkele nadere analyses uitgevoerd. Daarbij werd ingegaan op het volgende:

1. In hoeverre zijn de gevonden verschillen in structuur van het woordgebruik terug te voeren op de factor "steekproefgrootte"? Het bestand van de Nederlandse schrijftaal (605733 woorden) was immers vele malen groter dan dat van de beide strafrechtelijke wetboeken (47406 resp. 59488 tekens).
2. In hoeverre zijn de gevonden verschillen in structuur van het woordgebruik terug te voeren op de factor "oude spelling"? De teksten van de beide wetboeken dateren immers ten dele uit tijden van voor verschillende spelling-herzieningen, terwijl in het gewoon Nederlands bestand uitsluitend de nieuwe spelling A.B.N. werd gehanteerd. Dit kan consequenties hebben gehad voor de gevonden verschillen in de structuur van het woordgebruik.

De eerste vraag werd beantwoord door de bestanden OP en GW nogmaals met elkaar te vergelijken en door het kunstmatig homogeen inkorten van het SC-bestand. De tweede vraag werd beantwoord door alle types van oude spelling door die van de nieuwe te vervangen.

De eerder gevonden onderzoeksresultaten bleken niet terug te voeren te zijn op de factoren "steekproefgrootte" of "oude spelling".

#### 9.4 KONKLUSIES

In het rapport (De Mulder en Oskamp, 1979, blz. 75 e.v.) wordt een aantal konklusies geformuleerd, waarvan ik er hier een aantal zal weergeven.

##### 9.4.1 Konklusies Met Betrekking Tot Het Woordgebruik

1. Er is een aantal duidelijke verschillen in woordgebruik tussen enerzijds het Wetboek van Strafrecht, het Wetboek van Strafvordering en de beide wetboeken gezamenlijk en anderzijds Nederlandse schrijftaal, Nederlandse dagbladen, Nederlandse opiniebladen en populair-wetenschappelijke boeken.
2. De structuur van het woordgebruik van het Wetboek van Strafrecht, het Wetboek van Strafvordering en beide wetboeken gezamenlijk enerzijds, en anderzijds Nederlandse schrijftaal, Nederlandse dagbladen, opiniebladen en populair-wetenschappelijke bladen, verschilt in tenminste twee opzichten.
  1. In de strafrechtelijke bestanden komen de vaak voorkomende woorden (typen) - met een enkele uitzondering - aanmerkelijk vaker voor in de tekst dan vergelijkbaar vaak voorkomende woorden in gewoon Nederlands. (Onder "vergelijkbaar vaak" wordt hier verstaan: "van gelijke rangorde".)
  2. De strafrechtelijke bestanden kennen een wat kleiner aantal "zeer vaak voorkomende" typen dan gewoon Nederlands.
3. De zogenaamde "Wet van Zipf" waarin gesteld wordt dat voor natuurlijke talen het produkt van rangorde en frequentie van de woorden (typen) nagenoeg constant is, gaat voor de strafrechtelijke bestanden dan ook minder goed op dan voor gewoon Nederlands.
4. Er is een relatief grote overeenkomst in de structuur van het woordgebruik van Nederlandse schrijftaal, Nederlandse dagbladen, opiniebladen en populair-wetenschappelijke boeken. Tussen de strafrechtelijke bestanden is er in dit opzicht een groter onderling verschil.
5. De gevonden verschillen in de structuur van het woordgebruik zijn niet terug te voeren op de factor "steekproefgrootte" of op de factor "oude spelling".

6. Er zijn twintig typen die zeer vaak voorkomen in het Wetboek van Strafrecht, maar niet zeer vaak in Nederlandse schrijftaal. Men zou deze woorden "specifiek voor het Wetboek van Strafrecht" kunnen noemen.
7. Op overeenkomstige wijze gedefiniëerd zijn er tweentwintig specifieke woorden voor het Wetboek van Strafvordering en drientwintig specifieke typen voor de Wetboeken van Strafrecht en Strafvordering gezamenlijk.
8. Op dezelfde manier gedefiniëerd zijn er vier specifieke woorden voor Nederlandse dagbladen, geen voor de opiniebladen en twee specifieke woorden voor populair-wetenschappelijke boeken.

#### 9.4.2 Konklusies Met Betrekking Tot Het Instrument

1. Het trekken van konklusies omtrent het inhoudelijke woordgebruik wordt vergemakkelijkt door inzicht in de structuur van het woordgebruik.
2. De gehanteerde "Zipf-karakteristieken" vormen een middel om een zekere structuur in het woordgebruik te ontdekken. Zo kan men met behulp van deze karakteristieken min of meer objectief beoordelen welke woorden (typen) als "zeer vaak voorkomend" moeten worden beschouwd en welke niet.
3. De uitgevoerde analyses (inclusief het vervaardigen van de tabellen en het tekenen van de grafieken) waren zonder het gebruik van een computer praktisch onuitvoerbaar geweest.
4. De gebruikte computerprogramma's waren van eenvoudige aard. De omvang van de gehanteerde bestanden vereiste geen buitensporig groot achtergrondgeheugen.

### 9.5 DISKUSSIE

#### 9.5.1 Betrouwbaarheid En Validiteit

Onder dit hoofd (vgl. hoofdstuk 3) kunnen o.m. de volgende vragen aan de orde gesteld worden:

1. Hoe zeker kan men er van zijn dat bij het overbrengen van de woorden en bij de bewerkingen van de bestanden geen fouten zijn gemaakt?
2. In hoeverre komen de frequenties van de typen in de lijsten overeen met hun werkelijke frequentie in een bepaald soort Nederlands?

3. In hoeverre zijn de getrokken konklusies geldig o.m. in verband met het feit dat de gehanteerde lijsten van de bestanden voor "strafrechtelijke wetstaal" en "gewoon Nederlands" op verschillende wijzen tot stand zijn gekomen?

Voor de beantwoording van de eerste vraag wordt verwezen naar het rapport (De Mulder en Oskamp, 1979); handmatige en computercontroles op de bestanden spelen een belangrijke rol.

De tweede vraag veronderstelt, dat de term "werkelijke frequentie" verwijst naar een begrip, geschikt voor gebruik in een op kennisvermeerdering gerichte discussie. Met betrekking tot vele typen (de werkgroep spreekt van "woorden waarvan het gebruik erg afhankelijk is van de context") is dat de vraag. Met name voor de minder frequente woorden zou men naar het oordeel van de werkgroep niet zozeer moeten kijken naar frequenties van een woord apart, maar naar die van een hele klasse woorden, b.v. grammaticale categorieën of betekenis categorieën. Anderzijds is voor z.g. "functiewoorden" (hulpwerkwoorden, voorzetsels, voegwoorden, lidwoorden e.d.) de kans groot, dat in een nieuwe, evengrote steekproef ongeveer dezelfde frequenties zullen worden gevonden.

Voor de strafrechtelijke bestanden lijkt de beantwoording van deze vraag eenvoudiger te zijn. De bestanden zijn immers niet gebaseerd op een steekproef maar op de gehele wetstekst, die zeker representatief voor zichzelf is.

Een eerste opmerking hierover betreft het feit, dat de wetsteksten, zoals ze door Kluwer gepresenteerd worden niet exact gelijk zijn aan die, welke in het Staatsblad verschijnen. In het Staatsblad stonden bijvoorbeeld bij de afkondiging van het Wetboek van Strafvordering een aantal zinnen, beginnend met:

"15 januari 1921. Wet houdende vaststelling van een Wetboek van....."

waarin verder voorkwam:

"Wij Wilhelmina, bij de gratie Gods, Koningin der Nederlanden ....."

en:

"Gelijk wij goedvinden en verstaan bij deze: Artikel 1. Worden vastgesteld de navolgende bepalingen welke zullen uitmaken het Wetboek van Strafvordering. Eerste Boek. Algemene Bepalingen. Titel I. Strafvordering ....."

en aan het eind:

"Wilhelmina. De Minister van Justitie, Heemskerk. Uitgegeven den 28 januari 1921. De Minister van Justitie. Heemskerk."

Ook de talrijke wijzigingswetten bevatten dergelijke op- en naschriften. Kluwer laat deze weg. Wat dan overblijft is niet dezelfde verzameling tekens als die welke middels het Staatsblad in de loop der tijd is uitgegaan van de Nederlandse staatsorganisatie. Behalve dat men de op- en naschriften mist, ontbreken ook de door Kluwer in de tekst door nieuwe bepalingen vervangen artikelen.

Dit zijn opmerkingen, die de betrouwbaarheid, als ook de geldigheid van de konklusies raken. Men kan tegenwerpen, dat althans de tekst is onderzocht, waar de meeste Nederlandse juristen thans gebruik van maken.

Een tweede punt is, dat bij de bewerking die Kluwer uitvoerde voordat de telling werd verricht nog enige andere simplificaties werden toegepast: Verwijzingen en voetnoten werden weggelaten. Hoofdletters, leestekens en cursivering werden genegeerd.

De geldigheid van de getrokken konklusies in meer algemene zin is hierboven reeds aan de orde geweest toen de invloed van mogelijke "artefacten" zoals steekproefgrootte en oude spelling besproken werden. In exploratief onderzoek zal in het algemeen de vraag van de validiteit niet voldoende beantwoord kunnen worden. Daarvoor is nu juist nader - hypothesetoesend - onderzoek nodig. Hieronder zullen daarvoor enkele suggesties gedaan worden.

## 9.5.2 Suggesties Voor Verder Onderzoek

### 9.5.2.1 Toetsing van hypothesen

De konklusies die naar aanleiding van dit onderzoek over het juridische en gewone woordgebruik getrokken werden zijn bedoeld als hypothesen, die in verder onderzoek getoetst en bijgesteld kunnen worden. Men kan daarbij denken aan het analyseren van een andere steekproef voor Nederlandse schrijftaal, zodat meer zekerheid wordt verkregen bij de beantwoording van de representativiteitsvraag. Ook is het interessant, een oudere versie van de onderzochte wetten, b.v. het Wetboek van Strafvordering in zijn oorspronkelijke versie van 1921, aan dezelfde tellingen en bewerkingen te onderwerpen.

### 9.5.2.2 Verklaring van de resultaten

Met de konstatering dat het woordgebruik van de strafrechtelijke wetboeken verschilt van gewoon geschreven Nederlands, en in welke opzichten is nog maar een eerste begin gemaakt van theorievorming, van het maken van een model in de zin van Lave en March (vgl. hoofdstuk 2). Een volgende stap in het modelbouwproces zou het vinden van een proces zijn, dat ten grondslag zou kunnen liggen aan de gevonden feiten. Uitspraken m.b.t. zo'n proces kunnen dan de konklusies van het exploratief onderzoek verklaren. Het lag echter buiten de doelstelling van dit project om



mende typen (het middenstuk) groter zou moeten zijn, dan bij een verzameling steekproeven over verschillende onderwerpen. De kop van de karakteristieken zou dan naast de veel voorkomende neutrale woorden immers bovendien de specifieke, zelfstandige naamwoorden moeten bevatten. De gevonden resultaten wijzen op het omgekeerde. Wat in deze verklaring dan in feite als zeer opvallend overblijft, is de grote overeenkomst in structuur tussen de strafrechtelijke bestanden en gewoon Nederlands, zoals de duidelijke kop tot rangordes van 50-70.

#### 9.5.2.3 Verder exploratief onderzoek

Het onderhavige materiaal leent zich voor verdere exploratieve analyse, zoals eerder aangegeven. Soms kan men in juridische literatuur vragen vinden die met behulp van dit materiaal beantwoord kunnen worden. (B.v. Delikt en Delinkwent, 1972, blz. 399 e.v. waarin melding wordt gemaakt van een veronderstelling van Bianchi, op grond van de z.g. "Sapir-Whorf hypothese", dat het "stigmatiserende effect" van woorden samenhangt met de korthed ervan. Een berekening over de uit dit project beschikbare bestanden laat echter een grotere gemiddelde woordlengte zien voor de strafrechtelijke bestanden dan voor gewoon Nederlands. Doch dit terzijde, omdat de bestanden niet volledig zijn en de veronderstelling geen gebruik maakt van een zo simpel begrip als gemiddelde woordlengte.)

De tabellen bieden allerlei materiaal dat nader beschreven kan worden. Een voorbeeld ligt in de overzichtstabel relatieve woordfrequenties. Het woordje "de" komt in het W.v.Sr. aanzienlijk minder vaak voor dan in gewoon Nederlands, in het Wetboek van Strafvordering daarentegen aanzienlijk vaker. (Is dit te verklaren uit een neiging als beschreven in hoofdstuk 7, nl. dat men over "de" (b.v. Officier van Justitie) spreekt terwijl men "alle" of "een" bedoelt?)

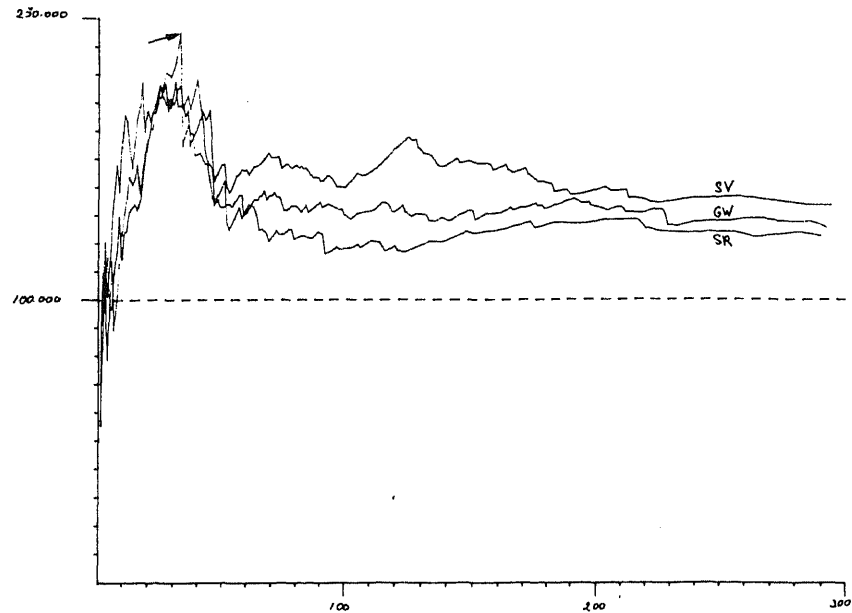
Een interessant verschijnsel is nog de "piek" in de kop van het gezamenlijke bestand van de beide strafrechtelijke wetboeken. (Vgl. fig. 9e.) Het woord dat precies op de top ligt is "verdachte", aldus de toepasselijke tabel.

Tenslotte nog enkele suggesties, waarvoor ander materiaal verzameld moet worden. Het kan interessant zijn, na te gaan of bepaalde generalisaties van gevonden resultaten opgaan dan wel of bepaalde kondities omkeerbaar zijn. B.v. "Wanneer in een tekst de woorden "1", "2", "3", "art", "artikel", "den", "der", "indien" en "eene" zeer vaak voorkomen, dan betreft het een juridische wetstekst".

Op syntactisch niveau blijvend, maar zich niet beperkend tot het woordgebruik, kan men verder onderzoek doen naar b.v. de vraag in hoeverre in juridisch taalgebruik de zinslengte verschilt van gewoon Nederlands.

Of men kan nagaan of bepaalde vaste combinaties van woorden (b.v. "wordt gestraft met", "in de eerste plaats") in juridische teksten vaker voorkomen dan in andere.

Fig. 9e



### 9.5.3 Beperking Tot De Syntaxis

Volgens Brandt-Corstius (1974) luidt de "Tweede wet van de computertaalkunde":

Bij elke toepassing van computertaalkunde, op hoe laag niveau ook, komt de semantiek roet in het eten gooien.

Is de werking van deze "wet" dit project bespaard gebleven? Op verschillende momenten hebben vragen van semantiek, ondanks de doelstelling van het onderzoek toch een rol gespeeld:

- In hoeverre is het geoorloofd leestekens in de woordenlijsten weg te laten? (B.v. "m'n" en "m.n." worden beide "mn".)
- In hoeverre is het geoorloofd woordvormen die "bij elkaar horen", zoals "Den" en "Haag" te splitsen?
- In hoeverre was het in dit onderzoek wenselijk geweest homoniemen zoveel mogelijk van elkaar te scheiden (zoals de werkgroep deed m.b.v. de grammaticale codes)?



- In dit projekt is ondanks de beperking in de opzet tot de syntaxis, gebruik gemaakt van een onderscheid als tussen "zelfstandige naamwoorden" en andere woorden. Dit onderscheid kan men dikwijls niet maken aan de vorm van woorden alleen.
- Onderzocht is de factor "oude spelling". Alleen op grond van de vorm kan men dikwijls niet nagaan of een type van de oude spelling is, of niet (b.v. "den", vgl. "een slanke den"). Voorts: mag men van een woord als "eenigen" eenvoudig "enige" maken of moet het in de nieuwe spelling "enigen" worden?

#### 9.5.4 Slotopmerkingen

Het beantwoorden van de vraag of en in hoeverre sommig juridisch taalgebruik verschilt van gewoon Nederlands blijkt, ook als men zich zoveel mogelijk beperkt tot het niveau van de syntaxis, en tot het woordgebruik, aanleiding te geven tot een groot aantal vragen. Veel van die vragen zijn beantwoordbaar door empirisch onderzoek van in beginsel eenvoudige aard. Het benodigde materiaal (juridische teksten) is gemakkelijk toegankelijk en hetzelfde geldt in steeds sterkere mate voor de benodigde technische hulpmiddelen zoals computers.

In het projekt dat in het volgende hoofdstuk beschreven wordt, staat de semantiek centraal. Een computer is er niet bij gebruikt, al zou dat wel handig zijn geweest.