



Universiteit
Leiden
The Netherlands

Recente ontwikkelingen op het gebied van de geautomatiseerde inhoudsanalyse

Saris-Gallhofer, I.N.; Saris, W.E.

Citation

Saris-Gallhofer, I. N., & Saris, W. E. (1975). Recente ontwikkelingen op het gebied van de geautomatiseerde inhoudsanalyse. *Acta Politica*, 10: 1975(4), 485-505. Retrieved from <https://hdl.handle.net/1887/3451509>

Version: Publisher's Version

License: [Leiden University Non-exclusive license](#)

Downloaded from: <https://hdl.handle.net/1887/3451509>

Note: To cite this publication please use the final published version (if applicable).

- welke de Britse Karel II (1660-1685) rond zich had samengelezen, zie: Waleffe, B., *Some Constitutional Aspects of Recent Cabinet Development in Great Britain and in Belgium*, Brussel, Bruylant, 1968, 210 pp., p. 11.
- 105 Dahrendorf, R., *Gesellschaft und Demokratie in Deutschland*, München, Piper, 1965, 516 pp., p. 39.
- 106 Ibid., p. 37.
- 107 Zie ook: Dahl, R. A., *The Science of Public Administration*, in Hawley, C. E., Weintraub, R. G., eds., *Administrative Questions and Political Answers*, New York, Van Nostrand, 1966, 604 pp., p. 33; Waldo, D., *The Administrative State*, New York, Ronald Press, 1948, 227 pp., p. 189; deze meer sociologische opening komt ook tot uiting in de zgn. ecologische bestuurskunde van Gaus, J. M., *Reflections on Public Administration*, Alabama University Press, 1947, 153 pp., tot Riggs, F. W., *Agraria and Industria, Toward a Typology of Comparative Administration*, Siffin, W. J., ed., *Toward a Comparative Study of Public Administration*, Bloomington, Indiana U.S., 1959, 330 pp., p. 23-116; Riggs, F. W., *The Ecology of Public Administration*, Bombay, Asia Publishing House, 1961, 152 pp.; Riggs, F. W., *Administration in Developing Countries, The Theory of Prismatic Society*, Boston, Houghton Mifflin, 1964, 477 pp.

Recente ontwikkelingen op het gebied van de geautomatiseerde inhoudsanalyse

door I. N. Saris-Gallhofer en W. E. Saris

Sinds de tijd dat Berelson (1952) inhoudsanalyse introduceerde als een specifieke waarnemingsmethode hebben zich vele ontwikkelingen voorgedaan. Een van deze ontwikkelingen was het programma 'The General Inquirer' (Stone e.a., 1962), hetgeen een computer programma is voor het analyseren van geschreven teksten. In de loop van de tijd is dit programma evenveel geprezen als bestreden.¹ Ook nu is het nog niet duidelijk of geautomatiseerde inhoudsanalyse voordelen heeft boven de analyse m.b.v. een jury.

Toch is in de laatste jaren gebleken dat de geautomatiseerde inhoudsanalyse steeds meer terrein wint. Er zijn nu meerdere programma's ontwikkeld voor verschillende talen, er kunnen verschillende soorten analyses worden uitgevoerd, en het oude General Inquirer programma heeft belangrijke verbeteringen ondergaan.

In dit artikel willen we een overzicht geven van deze ontwikkelingen. De informatie hierover is voornamelijk ontleend aan de bijeenkomst van inhoudsanalysten die in het najaar van 1974 in Pisa heeft plaatsgevonden.

We zullen deze uiteenzetting illustreren* aan de hand van voorbeelden van analyses die we hebben uitgevoerd met het programma RIQS dat het enige in Nederland beschikbare inhoudsanalyseprogramma is, voor zover ons bekend. Daarnaast willen we aangeven welke mogelijkheden voor geautomatiseerde inhoudsanalyse bestaan voor Nederlandse teksten in het bijzonder met behulp van het programma RIQS.

Inleiding

Onder geautomatiseerde inhoudsanalyse kunnen we verstaan het uitvoeren van een analyse met de computer van een voor de computer leesbare tekst met behulp van een door de onderzoeker geformuleerd programma.

Belangrijke punten in deze omschrijving zijn: (1) Het leesbaar zijn van de tekst voor de computer, (2) het programma dat door de onderzoeker wordt geformuleerd. We hebben expres de term 'analyse' niet verder gespecificeerd omdat er verschillende vormen van analyse mogelijk zijn. Hierop zullen we in de volgende paragraaf ingaan. Wat betreft de leesbaarheid van de tekst voor de computer raken we meteen aan een van de grootste problemen van de geautomatiseerde analyse van dit moment. Geschreven teksten zijn momenteel nog niet direct leesbaar voor de computer. Dit betekent dat de teksten gepost moeten worden op kaarten of op tape om voor de computer leesbaar te zijn. Dit is veelal een zeer tijdrovende bezigheid. Een voordeel is echter dat wanneer de teksten eenmaal op kaart of op tape staan, deze teksten ook op verschillende manieren kunnen worden geanalyseerd, hetgeen bij jury-proce-

* Dank zijn wij verschuldigd aan de heer B. Kooter die ons geassisteerd heeft bij het zoeken naar mogelijkheden om deze analyses met RIQS uit te voeren.

dures meestal onmogelijk is. In de toekomst zal echter dit probleem steeds minder belangrijk worden aangezien men uitgebreid bezig is met de ontwikkeling van optische lezers zodat op den duur het ponsen overbodig wordt. Het tweede punt is het meest interessante deel van de geautomatiseerde inhoudsanalyse. In het programma dat de onderzoeker ontwikkelt legt hij namelijk zijn theorie vast. Dit is tevens het moeilijkste deel want hierin moet hij procedures die anders de juryleden impliciet gebruiken bij hun beoordeling van de tekst, expliciteren. Door vele inhoudsanalysten wordt betwijfeld dat een onderzoeker een dergelijk programma kan formuleren. Sommigen hebben daarom eenvoudiger programma's ontwikkeld zonder deze interpretatieprocedures, anderen zien om deze reden af van inhoudsanalyse per computer.²

Wanneer men zich echter tot taak stelt om een meetinstrument te ontwikkelen dat zodanig is geëxpliciteerd dat 'de instructie kan worden uitgevoerd door een klerk, die geheel onwetend is op het gebied in kwestie' (De Groot 1961, p. 181) dan ligt in het ontwikkelen van een programma voor geautomatiseerde inhoudsanalyse een grote uitdaging. Stone e.a. (1966, p. 68) sluiten bij deze opmerking van De Groot aan wanneer zij zeggen: 'It is useful to liken the Inquirer to an energetic, compulsive, but stupid clerk who has been trained in content analysis mechanics. This clerk has no ideas of his own but waits for the specification of categories and scoring procedures supplied by the investigator. Once these instructions are received and not found to be self-contradictory, the clerk is able to apply them systematically to endless amounts of data'.

Het is echter zeker zo dat in de eerste fase van het ontwikkelen van computerprogramma's voor inhoudsanalyse deze poging tot objectiviteit te zeer op de voorgrond heeft gestaan en dat aan de validiteit van de procedures te weinig aandacht is besteed.

Het is dan ook precies op dit punt dat er zich in de 'wereld' van de inhoudsanalysten een splitsing heeft voorgedaan, en dat er zich twee verschillende richtingen in de aanpak van inhoudsanalyse per computer hebben ontwikkeld. Hierop zullen we in de volgende paragraaf ingaan.

Alvorens we hierop echter overgaan willen we nog één voordeel van de geautomatiseerde inhoudsanalyse noemen, nl. het feit dat men als de tekst eenmaal leesbaar is en als men een programma heeft ontwikkeld, de telwerkzaamheden in uiterst korte tijd en zeer accuraat kan laten uitvoeren. Dit soort werk geeft bij inhoudsanalyse met behulp van een jury vaak aanleiding tot vele fouten en grote frustraties.

Twee stromingen in de geautomatiseerde inhoudsanalyse

Zoals in veel sociaal wetenschappelijk onderzoek kan men bij inhoudsanalyse

een onderscheid maken tussen waarnemingseenheden en analyse-eenheden. In veel gevallen neemt men kenmerken waar bij individuen, maar men doet uitspraken over groepen, waarbij dus op een of andere manier vanuit het niveau van het individu is geaggregeerd naar het niveau van de groep.

Hetzelfde ziet men bij inhoudsanalyse: men kiest als waarnemingseenheid meestal een kleine teksteenheid, bijv. een woord of een zin, en doet vervolgens uitspraken over hele teksten of over teksten over een bepaalde periode. In beide vormen van inhoudsanalyse is er sprake van een keuze voor een bepaalde waarnemings- en analyse-eenheid.

Waar deze twee stromingen echter in verschillen is de wijze waarop ze gebruik maken van deze waarnemingseenheid om tot een score te komen voor de analyse-eenheid.

De analisten die met 'dictionaires' werken kennen aan de waarnemingseenheden een betekenis toe die niet direct besloten ligt in het stuk tekst zelf. Zij veronderstellen namelijk dat bepaalde teksteenheden bepaalde betekenissen dragen bijv.: het Engelse woord 'pure' heeft een positieve affectieve betekenis en 'scandal' een negatieve affectieve betekenis, volgens de Stanford Political dictionary.³ Zo kan men aan hetzelfde woord zelfs meerdere verschillende betekenissen toekennen.

De analisten die zonder dictionaires werken kennen dergelijke betekenissen niet toe aan de waarnemingseenheden, maar gaan slechts na hoe frequent binnen hun analyse-eenheid een bepaalde waarnemingseenheid voorkomt. Wanneer men een frequentietelling uitvoert voor alle waarnemingseenheden kan men vervolgens nagaan in welke mate bepaalde waarnemingseenheden gezamenlijk voorkomen of afwezig zijn, en op basis daarvan de analyse-eenheid karakteriseren wat betreft haar betekenis. Nadat we deze twee soorten analyses hier globaal van elkaar hebben onderscheiden zullen we nu beiden afzonderlijk meer in detail bestuderen en ze illustreren aan de hand van voorbeelden.

De dictionaire procedure

Zoals hierboven is aangegeven is het meest essentiële deel van de dictionaireprocedures dat de waarnemingseenheden zijn ingedeeld volgens één of ander klassificatie-schema in verschillende betekenis categorieën. Hoewel dit een immense taak lijkt, en het uiteraard ook niet zo is dat alle woorden ingedeeld zijn maar alleen de meest voorkomende, waren er toch in 1965 al 17 verschillende dictionaires ingebouwd in het General Inquirer programma (Stone, e.a. 1966, p. 140). De procedure werkt dan zo dat deze lijst van woorden uit de dictionaire wordt gelegd naast de woorden die voorkomen in de te analyseren tekst. De computer hecht dan aan elk woord uit de tekst de labels (tags) die de verschillende categorieën karakteriseren in de dictionaire en kan dan

vervolgens nagaan, hoe frequent woorden van de verschillende categorieën in de tekst voorkomen. Op deze wijze kan men dan teksten karakteriseren, men kan echter ook conclusies proberen te trekken betreffende de auteur van de desbetreffende tekst.

Aan een procedure zoals hiervoor beschreven kleven nog vele grote problemen: 1 – een woord kan in verschillende contexten verschillende betekenissen hebben; 2 – woorden met dezelfde stam maar met verschillende uitgangen kunnen een gelijke betekenis hebben; 3 – soms is men niet zo zeer geïnteresseerd in het enkele woord maar in combinaties van verschillende woorden (thema's); 4 – Men zal zich net als in ander onderzoek de vraag moeten stellen naar de validiteit van de verschillende dictionaires want daarmee staan of vallen deze analyses. Aan alle vier problemen is in de loop der tijd meer of minder aandacht besteed. Betreffende het eerste punt heeft men procedures ontwikkeld die kunnen nagaan aan de hand van combinaties van woorden rond het woord waarvan men de betekenis wil bepalen, welke van de mogelijke betekenissen in dit specifieke geval de juiste is. Deze procedures worden 'disambiguation' genoemd (Stone 1969). Wat betreft het tweede punt was reeds in de eerste versie van de General Inquirer de mogelijkheid voorzien om de uitgangen van woorden af te hakken en niet zo zeer op basis van de hele woorden maar slechts op basis van de stammen de betekenis te bepalen (Stone 1966).

Ook betreffende het toekennen van betekenissen op basis van thema's in plaats van enkele woorden als eenheden bestond reeds een mogelijkheid in de eerste versie van General Inquirer.

De toetsing van de validiteit van vele van de vermeide dictionaires is nog niet uitgebreid geschied. Op dit punt ligt dan ook naar onze mening één van de grote zwaktes van deze procedures. Van de andere kant moet men zeggen dat confrontatie van scores verkregen via inhoudsanalyse met data verkregen op volledig andere wijze in sommige gevallen opmerkelijk goede resultaten heeft opgeleverd (Namenwirth 1969). Men zou dit als een indicatie van de validiteit mogen beschouwen. Desondanks zijn we van mening dat op dit gebied nog veel onderzoek gedaan zou moeten worden. Dat hiermee een begin is gemaakt, zal blijken wanneer we spreken over de laatste ontwikkelingen.

We zullen nu de procedure illustreren aan de hand van een voorbeeld. Het programma dat we hierbij gebruiken is RIQS. Het is oorspronkelijk ontwikkeld voor 'retrieval' doeleinden, maar kan ook voor simpele inhoudsanalyse opdrachten gebruikt worden.⁴

We analyseren teksten van redevoeringen van wijlen President Kennedy. Twee uit de beginperiode van zijn presidentschap, twee uit de middenperiode en één uit de periode na de Cuba-crisis.⁵ Ter illustratie zullen we nagaan of Kennedy's visie op het gedrag van de Sovjet-Unie is veranderd in bovenver-

melde periodes. Men kan gebruik maken van de Stanford Political dictionary (Holsti, 1964) die uitspraken indeelt naar de dimensies die door Osgood e.d. (1957) ontwikkeld zijn, nl.: een evaluatieve dimensie, een machts-dimensie en een activiteitsdimensie. Alle drie de dimensies worden gekarakteriseerd door een zeven-puntsschaal. We zullen hier alleen van de evaluatieve dimensie gebruik maken.

Voor een dergelijke analyse wordt over het algemeen de tekst voorgecodeerd zoals te zien is in het stuk tekst dat is afgedrukt in tabel 1.

Tabel 1:

Een stuk voorgecodeerde tekst, leesbaar voor de computer en geschikt voor de analyse met behulp van de Stanford Political Dictionary

(1) 1 (2) cd (3) soviet/* and east german leaders have followed the recent soviet aide memoire with speeches/3 which were apparently/4 designed/4 to heighten/4 tension/4* (berlin/7*) +

(1) 1 (2) cd (3) (east-german/3* leaders/3 speeches/3 apparently/4 designed/4 heighten/4 tension/4* berlin/7*).

(1) 1 (2) cd (3) It is of the greatest importance that the american people understand the basic issues involved and the threats/4* to the peace/7 and security/7 of europe /7* and of ourselves posed/4 by the soviet/3* announcement/3 +

(1) 1 (2) cv (3) american/3* people must/4 understand/4* soviet/7* threat/7) +

(1) 1 (2) cd (3) (soviet/3* announcement/3 that they intend/4 to change/4* unilaterally/4 the existing/7 arrangements/7 for berlin/7*.

De nummers (1), (2) en (3) duiden op verschillende 'variabelen'. De eerste variabele (1) is het nummer van de tekst, de tweede variabele (2) duidt aan de 'mode of expression' zoals dit door Holsti (1969) wordt genoemd waarbij 'cd' betekent 'current descriptive', de derde variabele (3) is het thema van de desbetreffende zin. In dit soort onderzoek wordt een thema gevormd door een actor (voorzien van een /3*) een action (voorzien van een /4*) en een target (voorzien van een /7*), terwijl alle andere woorden door de codering met een /3, /4 of /7 worden toegewezen aan de verschillende delen van de zin. Over het algemeen maakt men echter van dergelijke woorden geen gebruik bij een analyse zoals we die hier willen illustreren. Wanneer we geïnteresseerd zijn in de mening van Kennedy over het gedrag van de Sovjets dan moeten we die zinnen bekijken waarin de Sovjets actor zijn. Met een simpel programma zoals afgedrukt in tabel 2 kan iets dergelijks worden gerealiseerd.

In deze simpele opdracht staat dat als de variabele (3) het woord 'Soviet' of 'Soviet-Union' etc. als actor bevat (die voorzien is van een /3*) dan moet dit thema in een aparte set geplaatst worden. (Dat zijn namelijk de zinnen waarin we verder geïnteresseerd zijn om te zien wat Kennedy van het gedrag van de Sovjets denkt).

Tabel 2:

De opdracht in de terminology van het programma RIQS om zinnen met Soviets als actor uit te selecteren en in een aparte file te zetten

if (3) eq ('soviet/3*' or 'soviet-union/3*' or 'khrushchev/3*' or 'russia/3*' or 'russia/*') then place record in set 1.

Vervolgens moet dan voor de zinnen in set 1 worden nagegaan in welke klasse ze vallen van zeer positief naar zeer negatief. Nu worden deze zinnen vergeleken met de classificatie die vastgelegd is in de dictionaire. Een pagina van deze dictionaire hebben we weergegeven in appendix 1.

Tabel 3:

De opdracht in de terminology van het programma RIQS om een nieuwe variabele te creëren die de visie van Kennedy weergeeft op het gedrag van de Soviets volgens de Stanford Political Dictionary

if (3) eq ('tension/4*' or 'threat/4*' or 'surrender/4*' or 'agression/4*' or 'killer/4*' or 'fright/4*' or 'doom/4*' or 'executed/4*') then let v3 = -3.,
 if (3) eq ('blockaded/4*' or 'refuse/4*' or 'mistake/4*' or 'undermine/4*' or 'refusal/4*' or 'disappointing/4*' or 'lower/4*' or 'differently/4*' or 'explosion/4*' or 'regret/4*' or 'violation/4*' or 'prevennt/4*') then let v3 = -2.,
 if (3) eq ('withdraw/4*' or 'stopping/4*' or 'concealed/4*') then let v3 = -1
 if (3) eq ('understand/4*' or 'commitment/4*' or 'realize/4*' or 'facilitated/4*' or 'interest/4*' or 'returned/4*' or 'discuss/4*') then let v3 = 1.,
 if (3) eq ('assurance/4*' or 'progress/4*' or 'sustain/4*' or 'improve/4*' or 'agree/4*' or 'progress/4*') then let v3 = 2.,
 if (3) eq ('freeing/4*') then let v3 = 3.,
 spssfile 1 of (1), v3

Voor ons simpele onderzoekje hebben we slechts gebruik gemaakt van een beperkt onderdeel van deze dictionaire. Het programma voor deze scoring staat afgedrukt in tabel 3. We hebben een variabele v3 gedefinieerd die de waarden +3 tot -3 kan aannemen. In het programma staat welke woorden welke waarden hebben op deze variabele. Nu wordt met dit opdrachtje aangegeven dat elk thema (3) moet worden nagelopen en wanneer in een thema een woord staat vermeld dat verbonden is met een bepaalde waarde op de variabele v3 dan krijgt dit thema deze waarde op de variabele v3. Wanneer in een thema geen woord voorkomt dat ook in de lijst vermeld staat dan wordt de waarde 0 toegekend op de variabele v3. Dit gebeurt automatisch. Vervolgens wordt de opdracht gegeven om van deze variabele v3 en variabele (1) een SPSS-file te maken.⁶

Wanneer deze opdracht is uitgevoerd dan hebben we voor elk thema een score op de variabele (1) het nummer van de tekst en op de variabele v3 het

gedrag van de Soviets volgens Kennedy.

We kunnen vervolgens vergelijkingen maken tussen de verschillende teksten. We hebben dit gedaan m.b.v. SPSS en hebben het resultaat weergegeven in tabel 4.

Tabel 4:

Het verband tussen het volgordenummervariabele (1), en de frequenties van meer of minder positieve uitspraken van Kennedy over het gedrag van de Soviets

variabele (1)	variabele v3							totaal
	-3	-2	-1	0	+1	+2	+3	
1	5	3	1	5	0	8	0	22
2	1	9	0	0	0	0	0	10
3	0	0	0	0	0	1	1	2
4	0	0	0	0	0	1	0	1
5	3	0	0	0	0	1	7	11
totaal	9	12	1	5	1	17	1	46

chi = 608, df = 24, prob < .001.

Uit deze tabel is te zien dat Kennedy het gedrag van de Soviets op den duur steeds positiever gaat beoordelen. Dit beeld zou echter nog duidelijker zijn uitgevallen wanneer het programma dat wij gebruikt hebben ook in staat zou zijn geweest om 'disambiguation' uit te voeren. In de laatste tekst komt namelijk de term 'execute' voor die als -3 wordt gescored. Kennedy gebruikt deze term echter in een zin die spreekt van 'executed agreements' of te wel in dit geval zou 'executed' niet de betekenis van 'neerschieten' moeten hebben maar 'uitvoeren' en zou dus positief moeten zijn. Dit is een fraai voorbeeld van het belang van disambiguation. Hiermee hebben we deze procedure geïllustreerd. Voordat we de paragraaf afsluiten lijkt het nog zinvol op te merken dat coderen zoals bij dit onderzoek is gebeurd, in de meeste gevallen niet meer noodzakelijk is. De laatste versies van de General Inquirer zijn zelf in staat deze codering aan te brengen.

Procedures werkend zonder dictionaires

In 1965 werd door Iker en Harway een computer procedure voor inhoudsanalyse geïntroduceerd die niet uitging van vooraf opgestelde dictionaires. Schematisch zou men deze procedure als volgt kunnen beschrijven.

Men gaat uiteraard weer uit van een voor de computer leesbare tekst. Men splitst deze tekst op in analyse-eenheden. Deze kunnen hele teksten bevatten of paragrafen van teksten of zinnen. Vervolgens neemt men als waarnemings-

Tabel 5:

Een stuk tekst van een redevoering van Nixon, leesbaar gemaakt voor de computer en geschreven in een vorm die voor het programma RIQS analysemogelijkheden biedt

+ text of president Nixon's statement on watergate +
 (1) 01 (2.1) allegations regarding the watergate affair have escalated that I feel a further statement from the president is required at this time.
 (2.2) the climate of sensationalism has developed in which even second or third-hand hearsay charges are headlines as fact and repeated as fact.
 (2.3) important national security operations which themselves had no connection with watergate have become entangled in the case.
 (2.4) as a result, some national security information has already been made public through court orders, trough the subpoenaing of documents and through testimony witnesses have given in judicial and congressional proceedings.
 (2.5) other sensitive documents are now threatened with disclosure.
 (1) 02 (2.1) continued silence about those operations would compromise rather than protect them, and would also serve to perpetuate a grossly distorted view which recent partial disclosures have given, of the nature and purpose of those operations.
 (2.2) the purpose of this statement is threefold:
 (2.3) first, to set forth the facts about my own relationship to the watergate matter.
 (2.4) second, to place in some perspective some of the more sensational - and inaccurate - charges that have filled the headlines in recent days, and also some of the matters that are currently being discussed in senate testimony and elsewhere.
 (2.5) third, to draw the distinction between national security operations and the watergate case.

eenheid het enkele woord en gaat na hoe frequent de verschillende woorden voorkomen in analyse-eenheden. Op deze wijze ontstaat een datamatrix met als eenheden de analyse-eenheden en als variabelen de verschillende woorden. In de datamatrix staat dan aangegeven hoe frequent de woorden in de verschillende stukken tekst voorkomen. Op deze datamatrix kunnen dan weer allerlei analyses worden uitgevoerd om de data te reduceren. Bekende procedures om na te gaan welke woorden frequent samen voorkomen zijn cluster- en factoranalyse. Op basis van deze analyse kan men dan achteraf proberen om de teksten te interpreteren. We zullen deze procedure ook illustreren aan de hand van een voorbeeld. De analyse wordt weer uitgevoerd met het programma RIQS.

Stel dat we voor een redevoering willen nagaan hoe de spreker zijn betoog opbouwt en we willen dit doen met een geautomatiseerde methode, dan zouden we van een procedure als hierboven beschreven gebruik kunnen maken. We zullen dit illustreren aan de hand van een redevoering van Nixon in het midden van het onderzoek naar zijn rol in de 'Watergate affaire'. Een stuk

van de tekst is weer uitgeponst op ponskaarten, waarbij een indeling is gemaakt in analyse-eenheden van 5 zinnen. In deze analyse zijn de volgende variabelen gebruikt: (1) het nummer van de analyse eenheid; (2.n) de ne zin van deze analyse-eenheid. In tabel 5 zijn twee analyse-eenheden van 5 zinnen afgedrukt.

De volgende stap is het tellen van de frequenties waarmee de verschillende woorden voorkomen in de verschillende analyse-eenheden. Een probleem hierbij is echter dat het niet zo duidelijk is wat men verstaat onder hetzelfde woord. Bij dergelijke analyses bestaat niet zo zeer het probleem van meerdere betekenissen voor hetzelfde woord omdat niet met van te voren vastgestelde betekenissen gewerkt wordt. Belangrijk is daarentegen dat er rekening wordt gehouden met de talrijke vormen die één en dezelfde woordstam kan aannemen door vervoeging of verbuiging. Bij deze analyse is een procedure om de stammen van de woorden te bepalen dan ook essentieel. Het programma RIQS bevat echter een dergelijke procedure niet en daarom hebben we vooraf nagegaan of verschillende woordstammen in verschillende vormen voorkwamen. Daarnaast bestaat nog het probleem van synoniemen. Ook die hebben we vooraf gecontroleerd. Deze hebben we dan als één woord opgevat en hiermee is rekening gehouden bij het bepalen van de frequenties van de verschillende woorden. De opdracht daartoe hebben we weergegeven in tabel 6.

Tabel 6:

Een procedure geschreven in de terminology van het programma RIQS om de frequenties van woorden per paragraaf te bepalen

```
let no = 0., for n = 1 to 5.,
if (2.n) eq ('not' or 'no' or 'none') then let no = no + 1., loop., let clemenc = 0, for
n = 1 to 5.,
if (2.n) eq ('clemency') then let clemenc = clemenc + 1., loop., let fact = 0., for
n = 1 to 5',
if (2.n) eq ('fact' or 'facts') then let fact = fact + 1., loop., let unit = 0., for n = 1 to 5.,
if (2.n) eq ('group' or 'unit') then let unit = unit + 1., loop., let menixon = 0., for
n = 1 to 5.,
if (2.n) eq ('i' or 'me' or 'my') then let menixon = menixon + 1., loop.,
let leak = 0., for n = 1 to 5.,
if (2.n) eq ('leak' or 'leaks' or 'leaked') then let leak = leak + 1., loop.,
spssfile of (1), no, clemenc, fact, unit, menixon, leak.,
```

Deze opdracht houdt in dat de computer een nieuwe variabele creëert bijv. de variabele 'no' die zijn waarde krijgt op basis van de frequentie waarmee het woord 'not' of 'no' of 'none' in de paragraaf voorkomt. Deze opdracht wordt gegeven in een 'loop procedure' die inhoudt dat de computer voor elke

zin per paragraaf overnieuw dezelfde opdracht uitvoert d.w.z. dat hij nagaat of deze woorden in de tekst voorkomen. Het aantal zinnen per paragraaf is steeds vijf dus staat er 'for n = 1 to 5.' Voordat deze procedure wordt uitgevoerd wordt eerst de waarde van de variabele op 0 gesteld anders zou doorgeteld worden over alle paragrafen terwijl we een score willen hebben per paragraaf. Wanneer dus één van de woorden voorkomt in de eerste zin dan wordt de waarde van de variabele verhoogd met 1, anders blijft de waarde gelijk aan 0. In het totaal kan de waarde van de variabele dus oplopen tot de waarde 5. Voor de andere woorden ziet de opdracht er identiek uit. Op deze wijze kan men een groot aantal variabelen creëren die als waarden hun frequentie per paragraaf krijgen. Daarna kan er weer een SPSS-file van gemaakt worden, wat betekent dat men op een dergelijke datafile daarna weer alle bewerkingen kan uitvoeren die men normaliter met SPSS op datafiles kan uitvoeren.

Uit deze opdrachten blijkt ook duidelijk het nadeel wanneer men geen automatisch programma heeft voor het opsporen van de stammen.

Bij gebrek aan een dergelijke procedure hebben we namelijk gebruik moeten maken van alle mogelijke alternatieve vormen waarin hetzelfde woord voorkomt. Heeft men een automatisch programma voor het opsporen van de stammen dan hoeft men alleen de stammen te specificeren.

Op de op deze wijze verkregen datafile hebben we vervolgens factoranalyse uitgevoerd om na te gaan welke woorden per paragraaf frequent voorkomen. We hebben met SPSS een principal component factoranalyse uitgevoerd met varimax rotatie en zijn op deze wijze tot het resultaat gekomen dat vermeld staat in tabel 7.

In deze tabel kunnen we zien dat sommige factoren gekarakteriseerd worden door woorden die samen zinnen of delen van zinnen van Nixon vormen. Om enkele voorbeelden te geven: Op factor 1 laden het hoogst de woorden: 'effort' en 'clemency'. Deze factor verwijst naar het feit dat hij spreekt over 'the efforts' (to get) executive clemency'. Voor degenen die beschuldigd zijn van deelname aan de Watergate inbraak.⁷

Uit factor 2 zou men de volgende Nixoniaanse zin kunnen halen: 'I (had) no knowledge (of) illegal (actions) (during) the campaign'. Het woord 'I' komt van de variabele 'menixon'. Het woord 'intelligence' is weggelaten omdat dit woord een negatieve lading heeft. Dit betekent dat dit woord juist weinig voorkomt in zinnen waarin de andere woorden frequent voorkomen. Het woord 'should' is weggelaten omdat het niet in deze zin past. Dat het desondanks positief kan laden op een dergelijke factor zullen we uitleggen aan de hand van andere factoren waar zich hetzelfde geval in versterkte mate heeft voorgedaan. Uit factor 3 kunnen we de tekst halen 'leaks (of) secrets'. Hetgeen één van de hoofdthema's van Nixon is in deze redevoering om aan

Tabel 7:
De resultaten van de factoranalyse en varimax rotatie op de data verkregen uit de redevoering van Nixon

woorden	factor 1	factor 2	factor 3	factor 4	factor 5	factor 6	factor 7	factor 8	factor 9	factor 10
activity	-.19	-.01	-.24	-.14	-.15	.03	.44	.22	.43	.58
approval	-.10	.11	-.03	-.04	-.15	-.03	-.12	-.05	.67	.04
watergate	-.02	-.19	-.05	.05	-.01	.43	.61	.28	-.15	-.12
campaign	.34	.53	-.09	.04	-.27	.10	-.05	.43	.13	.10
intelligence	-.30	-.39	-.07	-.22	-.04	-.31	.30	.08	.46	.01
committee	.20	-.05	-.25	-.15	.14	-.32	-.09	-.08	.37	-.21
cover-up	.19	-.12	-.12	-.20	.03	.85	-.08	.06	-.04	-.08
covert	-.06	-.05	-.02	-.09	-.06	-.05	.76	.15	.04	.10
document	-.11	-.12	-.15	-.06	.06	-.12	-.18	-.09	-.27	.14
effort	.89	-.05	-.08	-.07	.04	.14	-.05	.12	-.04	-.08
foreign	-.03	-.08	.09	.09	.19	-.10	-.10	-.14	-.06	.54
government	-.06	.13	.12	-.16	.73	-.12	-.09	-.03	-.10	.13
illegal	-.14	.85	-.16	-.07	-.05	-.12	-.06	.35	.13	.01
information	-.09	-.23	.12	.68	-.02	-.09	-.12	-.05	-.05	.02
should	.26	.41	-.08	.14	-.01	.57	.21	-.03	.03	-.07
no	.10	.58	-.05	-.12	.49	.40	.25	.04	-.02	-.22
clemency	.93	.07	-.04	-.06	-.07	.11	-.04	-.02	-.05	-.04
fact	-.13	.08	-.34	.43	.13	-.04	.06	-.01	-.17	.31
unit	-.10	.14	.30	-.08	-.07	-.03	.56	-.27	-.11	-.10
secret	-.06	-.09	.75	-.05	-.06	-.05	.26	.00	-.08	.06
menixon	.26	.34	-.19	.49	.04	.14	.61	.27	.11	-.26
involvement	.06	.11	.13	-.04	.00	.05	.18	.73	-.04	-.11
knowledge	.01	.82	-.05	-.14	.16	-.03	.02	-.11	.01	-.07
leak	-.10	-.13	.85	.09	.09	-.11	-.25	.11	.02	-.01
questions	-.01	-.04	-.15	.28	.58	.12	-.07	-.01	-.09	.08

te geven wat in factor 9 staat, nl. de noodzaak van de 'approval of intelligence activities', waarmee hij legale inbraken en af luisteractiviteiten bedoelt. Factor 5 levert dan nog de zin op '(the government) should not cover up Watergate' iets waarvan hij voortdurend werd beschuldigd en factor 4 spreekt van de wens van 'Nixon (to look for) information and the facts'. Wat betreft de andere factoren 5, 7, 8, 10 is de interpretatie niet zo duidelijk en waarschijnlijk hebben we te maken met het samengaan van woorden uit verschillende zinnen. Het kan zijn dat bepaalde woorden in het begin van een paragraaf steeds worden gevolgd door bepaalde woorden in het eind van de paragraaf. Daardoor kan wel een hoge correlatie ontstaan, de waarde van dergelijke factoren voor dit soort analyses is echter erg gering. Dit is ook een van de kritiek-punten op deze methode en de reden waarom bepaalde onderzoekers met kleinere analyse-eenheden werken waardoor dergelijke problemen geminimaliseerd worden.⁸

Men kan een dergelijke analyse voortzetten door na te gaan in welke volgorde deze thema's voorkomen in de tekst om op deze wijze inzicht te krijgen in de opbouw van de redevoering nadat inzicht is verkregen in de onderwerpen die aan de orde zijn gesteld.

In deze tekst blijkt de volgorde zodanig te zijn dat Nixon begint te spreken over de 'leaks of secrets' en vervolgens aangeeft dat hij in verband met 'national security' genooddaakt was tot 'approval of intelligence activities'. Hij vervolgt dan met de herhaalde vaststelling dat hij 'no knowledge of illegal actions during the campaign' had. Toen hij ervan hoorde is hij gaan zoeken naar 'information and facts' en hij eindigt met de constatering dat hij alles in het werk zal stellen dat de regering niet meewerkt aan de 'coverup of the Watergate', maar de waarheid aan het licht zal brengen.

Een resultaat als hier vermeld is kan op verschillende manieren verkregen worden. Wij hebben dit resultaat gekregen door de factorscores te berekenen en de tijdreeksen voor de verschillende factoren te vergelijken. Men zou ook met RIQS de combinaties van de verschillende woorden kunnen laten zoeken net als eerder naar enkele woorden is gezocht door het programma en vervolgens hiervan tijdreeksen maken. Het zou echter te ver voeren om hier nog op in te gaan.

Hiermede hebben we de twee hoofdstromingen in de geautomatiseerde inhoudsanalyse geïllustreerd en kunnen we overgaan op een overzicht van de recente ontwikkelingen zoals die de laatste jaren hebben plaats gehad.

Ontwikkelingen van de laatste jaren

De belangrijkste verbetering is wel dat men een procedure heeft kunnen ontwikkelen voor 'disambiguation'. We hebben hiervoor reeds vermeld dat een van de grote problemen van dictionaireprocedures daarin bestaat dat dezelf-

de woorden in verschillende contexten andere betekenissen kunnen hebben. Wanneer men dus een goede dictionaireprocedure wil ontwerpen dan is een facet daarvan dat men de verschillende betekenissen voor woorden in context moet kunnen achterhalen. In het kader van het onderzoek naar verbeteringen van de General Inquirer heeft men voor Engelse teksten een dergelijke procedure kunnen ontwikkelen. Deze procedure is ook getest op een databestand en bleek slechts in 2 % van de gevallen fouten op te leveren (Stone 1969, p. 217). Voordat men over deze procedure beschikte was het foutenpercentage 25 %. Deze 'disambiguation' procedure lijkt dus een zeer grote verbetering op te leveren. Hierbij moet worden opgemerkt zoals de schrijver ook vermeldt dat het nog getest moet worden of met andere teksten gelijke resultaten zouden worden bereikt.

Het principe van deze procedure is dat men vóór en achter het woord waarvan men de betekenis wil gaan bepalen, zoekt naar andere woorden die de betekenis van het betreffende woord kunnen verduidelijken. We zullen hierop niet verder ingaan omdat deze procedure reeds uitvoerig is beschreven door Stone. Een tweede belangrijke ontwikkeling bestaat daarin dat men zich meer is gaan richten op de validiteit van de verschillende dictionaires. Op de bestaande Harvard-dictionary is door een onderzoeksgroep uit Australië (Dunphy e.a., 1975) een jurytest uitgevoerd. Hieruit is een aantal correcties voortgekomen van de Harvard IV dictionary. Ook de Stanford Political dictionary is aan een validiteits-test onderworpen met een enigszins negatief resultaat (Morton, Saris, Saris 1974). Het bleek dat slechts één dimensie van de dictionary, nl. de evaluatieve dimensie voldoende valide is, om er verdere studies op te baseren. Op dit gebied zou nog veel onderzoek gedaan kunnen en moeten worden om de procedures voor geautomatiseerde inhoudsanalyse valider te maken. Wij zijn het eens met Holsti (1969) dat 'careful attention to these questions must be an integral part of every research design if it is to meet meaningful standards of systematic investigation'. Tot nu toe is aan dit punt te weinig aandacht besteed.

Een derde belangrijke ontwikkeling van de laatste jaren is de nieuwe opzet van het General Inquirer programma welke uitgewerkt is in Edinburgh (Coxon, 1974). Het belang van deze ontwikkeling is dat in dit programma de dictionaires en de disambiguation losgemaakt zijn van het gehele programma. Hierdoor wordt het voor de individuele onderzoeker mogelijk om zijn eigen dictionaires in te voeren in het programma. Dit was bij de 'Harvard' versie niet mogelijk omdat toen het programma, de dictionaires en de disambiguation geheel met elkaar verweven waren. Voor onderzoekers met specifieke onderzoeksobjecten kan dit grote voordelen hebben.

Een vierde belangrijke ontwikkeling is dat men ook voor andere talen dan het Engels dictionaires heeft ontwikkeld. In Duitsland is men hiermee in de

laatste jaren het verst gevorderd (Deichsel, 1974). Tot voor kort hebben alle ontwikkelingen op het gebied van de geautomatiseerde inhoudsanalyse zich afgespeeld in de Verenigde Staten en zijn dus de dictionaires ook alleen voor de Engelse taal ontwikkeld. Ook de disambiguation procedures zijn uiteraard taalgebonden en dus momenteel alleen beschikbaar voor Engelse teksten. Men is echter momenteel in Keulen bezig een versie van de General Inquirer te ontwikkelen die Duitse teksten kan analyseren.⁹

De vijfde en laatste ontwikkeling die we hier willen vermelden is het ontstaan van vele nieuwe programma's voor inhoudsanalyse. In verschillende landen is men hiermee bezig geweest in de afgelopen jaren. Om enkele te noemen: Spence heeft een programma ontwikkeld dat ongeveer dezelfde bedoeling heeft als het programma van Iker e.a. hoewel het minder ver ontwikkeld is (1974). Door Meunier (1974) is een programma (Sato) ontwikkeld dat in het bijzonder geschreven is om Franse teksten te analyseren en na te gaan in welk verband verschillende woorden door één of een aantal teksten heen gebruikt worden. Een interessant aspect van dit programma is dat het tevens beschikt over een automatische procedure voor syntactische analyse van de zinnen. Een ander programma voor de analyse van Franse teksten is het programma Forcod (Couturier, 1974). Hoewel dit programma meer bedoeld is als een soort 'retrieval program' zijn er misschien ook mogelijkheden voor gewone inhoudsanalyse. Hetzelfde geldt voor het programma RIQS. Ook dat is eigenlijk een 'retrieval program', het is echter ook voor simpele inhoudsanalyse te gebruiken hierop komen we nog terug in de volgende paragraaf.

In Duitsland zijn de laatste jaren twee programma's ontwikkeld, nl. EVA (Holzscheck, 1974) en Textpack (Höhe e.a., 1973) beide zijn echter nog duidelijk minder geavanceerd dan de General Inquirer (er ontbreken bijvoorbeeld disambiguation procedures).

Dan is er in Noorwegen door Fonnes ook een programma voor Noorse teksten ontwikkeld, ook dit programma is echter nog niet al te ver ontwikkeld.¹⁰

Het laatste programma dat te vermelden valt is het programma Quester dat ontwikkeld is door Cleveland e.a. (1974) en tamelijk sterk afwijkt van alle andere programma's. Dit programma beschikt ook over een disambiguation procedure die slechts gebaseerd is op de frequenties van de woorden rond het woord waarvan men de betekenis wil achterhalen. Een dergelijke procedure lijkt tamelijk taal-onafhankelijk, hoewel het uiteraard een uitgebreid onderzoek vooraf vraagt naar de frequenties van woorden. Over dit programma is echter tot nu toe niet veel bekend. De validiteit zou dus nog bestudeerd moeten worden. Uit deze lijst blijkt dat er tamelijk veel interesse is voor het ontwikkelen van programma's voor geautomatiseerde inhoudsanalyse. Voorlopig lijkt echter de Edinburgh versie van de General Inquirer de meest geavanceerde van de dictionaireprocedures. Dit programma is echter alleen ge-

schikt voor Engelse teksten. In het kader van de programma's die zonder dictionaires werken is het programma 'Words' het verst ontwikkeld, dit programma lijkt ook tamelijk taal-onafhankelijk wanneer men per taal-gebied een eigen programma maakt voor het opsporen van de stammen van de woorden. Ook het programma RIQS zou waarschijnlijk dergelijke mogelijkheden hebben wanneer hieraan enkele verbeteringen zouden worden aangebracht. We zullen hierop nog ingaan in de volgende paragraaf.

Geautomatiseerde analyse van Nederlandse teksten

In deze laatste paragraaf willen we ons bezighouden met de vraag welke mogelijkheden er bestaan voor geautomatiseerde inhoudsanalyse van teksten die in het Nederlands geschreven zijn. Aangezien het programma RIQS het enige programma hiervoor is dat in Nederland aanwezig is, zullen we dit programma in dit verband ook evalueren. Zoals we reeds eerder gezegd hebben is RIQS eigenlijk ontwikkeld als een 'retrieval program' en daarom mist het faciliteiten zoals het automatisch opsporen van de stammen van woorden en disambiguation procedures. Van de andere kant heeft het de mogelijkheid om inhoudsanalytische resultaten direct in te brengen in SPSS-files waardoor de verdere analyse van de data zeer eenvoudig kan verlopen.

Bovendien beschikt RIQS nog over verschillende handige procedures die de bovenvermelde problemen kunnen verzachten, nl. de mogelijkheid om een alfabetische lijst te krijgen van de woorden van de tekst d.w.z. dat alle woorden in volgorde van het alfabet worden afgedrukt met de frequenties waarmee ze voorkomen en de plaats. Daarnaast is er een procedure om bepaalde woorden te laten afdrucken in hun context. Deze twee procedures kunnen de controle op de betekenis van woorden met een gelijke stam vergemakkelijken. Wij hebben in de hiervoor vermelde analyses ook van deze procedures gebruik gemaakt.

Na deze korte inleiding zullen we nagaan welke eisen men zou moeten stellen aan programma's voor de twee soorten geautomatiseerde inhoudsanalyse die we hierbij vermeld hebben en vervolgens vermelden in welke mate RIQS hiervoor geschikt is.

Inhoudsanalyse met behulp van dictionaires

Wanneer men een inhoudsanalyse met behulp van een uitgebreid dictionaire zou willen uitvoeren waarin de meest frequent voorkomende woorden zijn ingedeeld in verschillende voor het onderzoek relevante betekenis categorieën, dan zou men de volgende eisen aan een dergelijk programma moeten stellen: 1 – het programma zou in staat moeten zijn op efficiënte wijze de dictionaires die 4 tot 5 duizend woorden bevatten, te vergelijken met de woorden in de tekst;

2 – het programma moet in staat zijn om in geval van meerdere betekenissen van een woord afhankelijk van de context waarin het voorkomt de juiste betekenis te achterhalen (disambiguation);

3 – het zou mogelijk moeten zijn om met verschillende dictionaires te kunnen werken al naar gelang het doel van het onderzoek d.w.z. dat men zelf dictionaires moet kunnen invoeren en niet verplicht moet zijn de 'universele dictionaires' te gebruiken zoals dit het geval is bij de Harvard General Inquirer¹¹;

4 – het programma moet ook in staat zijn om ingewikkelde woord-combinaties op te sporen op een redelijk efficiënte wijze;

5 – het programma moet in staat zijn om de stammen van de woorden automatisch op te sporen;

6 – het programma moet een goede aansluiting geven op andere data-analyse programma's zoals SPSS zodat op eenvoudige wijze verdere analyses kunnen worden uitgevoerd.

Aan de eerste, tweede en vijfde eis voor een programma voor inhoudsanalyse m.b.v. een dictionaire kan RIQS zeker niet voldoen. Het grootste probleem ligt in het eerste punt. Het is namelijk onmogelijk om op een enigszins efficiënte wijze met RIQS een vergelijking van duizenden woorden met de tekst te verrichten. Zelfs een simpele opdracht die we het programma wilden laten uitvoeren op de teksten van Kennedy, kon niet in één run worden geanalyseerd hoewel we slechts 54 woorden in een keer wilden laten opzoeken. Het programma heeft namelijk een limiet op het aantal letters tekst dat in één run kan worden opgezocht, nl. 300.¹²

Daarnaast hebben we al gezegd dat aan de eisen 2 en 5 ook niet kan worden voldaan. Hieruit blijkt dat voor dergelijk onderzoek nieuwe programma's zouden moeten worden geschreven. Dit betekent dat deze mogelijkheid nog wel enige jaren niet beschikbaar zal zijn.

Men kan nog wel van mening verschillen of een dergelijke situatie betreurenswaardig is of niet, onze mening is echter dat het interessant zou zijn om reeds gedane analyses waarbij men met juryleden werkte te herhalen met geautomatiseerde procedures om na te gaan of men de gebruikte analyse expliciet kan maken, bovendien zou dan het grote voordeel kunnen ontstaan dat replicatie-onderzoek of secundaire analyse mogelijk zou worden.

Inhoudsanalyse zonder dictionaires

Inhoudsanalyse zonder dictionaires stelt uiteraard minder zware eisen aan de te gebruiken programma's. We zouden hier de volgende punten willen noemen.

1 – het programma zou in staat moeten zijn om de stammen van de woorden op te sporen;

2 – het programma moet in staat zijn op efficiënte wijze de frequentie van 200 tot 500 woorden (stammen) op te sporen in teksten;

3 – het programma moet een goede aansluiting hebben met andere data-analyse programma's zoals SPSS voor verdere verwerking van de data.

Ook aan deze eisen voldoet RIQS niet volledig. Het grote voordeel van RIQS ligt echter in de derde eis, nl. de aansluiting bij het SPSS programma, hierdoor lijkt het programma toch goede mogelijkheden te hebben voor de hier behandelde analyses. Een afzonderlijk programma dat de tekst zodanig verandert dat in de tekst slechts de stammen van de woorden komen te staan lijkt namelijk niet al te ingewikkeld te zijn om te schrijven. Daarmee kan men ook het aantal woorden dat men moet laten opsporen aanmerkelijk reduceren en het lijkt ons niet onmogelijk dat ook de limieten van het programma enigszins kunnen worden aangepast zodat wat grotere opdrachten gegeven kunnen worden. Ook op het moment kan men dergelijke analyses al tamelijk eenvoudig met dit programma uitvoeren.

In het voorbeeld van de analyse van de tekst van Nixon hebben we een dergelijke analyse uitgevoerd. Het probleem dat het programma niet aan de eerste eis voldoet hebben we omzeild door voor de meest frequente woorden die verschillende vormen hadden een analyse uit te voeren m.b.v. de subroutines Invert en KWIC (keyword in context) waarmee we konden nagaan of woorden met dezelfde stam ook in dezelfde betekenis in de tekst voorkwamen.

Wat betreft de beperking in de zoekmogelijkheden van het programma betekent het op dit moment alleen dat men meerdere malen een dergelijke zoekprocedure moet herhalen en daarna de verschillende files aan elkaar moet koppelen. Dit zijn echter geen onoverkomelijke problemen en het grote voordeel is dan dat men meteen data in een vorm heeft die door SPSS kunnen worden geanalyseerd. Wat deze analyseprocedure betreft heeft RIQS ook nu reeds goede mogelijkheden voor Nederlandse teksten.

Noten

1 Hiervoor zou men de discussies kunnen nalopen die op de congressen van de inhoudsanalysten werden gehouden en die zijn samengevat in de volgende readers.

Pool I de Sola, 1959, Trends in content analysis. Urbana: University of Illinois Press. Gerbner G., O. R. Holsti, K. Krippendorff, W. J. Paisley and P. J. Stone Eds The analysis of communication content; developments in scientific theories and computer techniques, Wiley New York 1964.

Draghi, S., e.a., Proceedings of the Pisa contextanalysis conference.

2 Zie bijvoorbeeld de recente kritiek van Marhoff, Shapiro en S. R. Weilman (1975). Op p. 2 doen ze de volgende uitspraak: 'Yet it is apparent to us that context analysis methodology today more often impoverishes documentary work than it enriches it'.

- 3 Deze voorbeelden zijn ontleend aan de Stanford Political dictionare (zie Stone e.a., 1966, p. 189).
- 4 De suggestie om RIQS voor inhoudsanalyse te gebruiken is gedaan door B. Kooter en Van Schuur (1974).
- 5 Deze teksten waren op ponskaart aanwezig daar ze door Dr. E. Morton in haar proefschrift waren gebruikt (1973).
- 6 SPSS is een pakket computerprogramma's dat de meest gebruikte analyse technieken bevat (Nie e.a.). SPSS vereist echter een specifiek soort datafiles. In RIQS is een procedure opgenomen om dergelijke datafiles te maken zodat de overgang van het ene programma naar het andere makkelijk verloopt.
- 7 De woorden die tussen haakjes staan zijn door ons toegevoegd om enigszins volledige zinnen te krijgen. Deze woorden zijn in de analyse weggelaten omdat ze te onbelangrijk waren: bijwoorden als 'in', 'a', etc. of omdat ze zijn weggevalen daar vaak synoniemen zijn gebruikt en de individuele woorden een lage frequentie hadden.
- 8 K. Krippendorf heeft hierop gewezen tijdens de Conferentie in Pisa. Om die reden werkt hij met zinnen als analyse eenheid.
- 9 Dit gebeurt in het Zentral Archiv met medewerking van P. Stone.
- 10 Fonnes heeft hierover gesproken tijdens de conferentie. Het is ons onbekend of hierover gedrukte informatie bestaat. Als het bestaat zou deze verkrijgbaar moeten zijn bij J. Fonnes, Institute of History, University of Oslo. Oslo.
- 11 Hoewel er gestreefd is naar een zo algemeen mogelijk meetinstrument in de Harvard IV dictionary lijkt een dergelijke dictionaire toch niet zo geschikt voor alle soorten onderzoek en zal men vaak zijn toevlucht moeten nemen tot het ontwerpen van eigen dictionaires.
- 12 Zie het RIQS manual, p. c-2.

Literatuur

- Berelson, B., *Content Analysis in Communication Research*. Glencoe, Illinois, The Free Press, 1952.
- Borman, L., Chalice, R., Dilleman, D., Dominich, W., Kron, P., *RIQS: Remote Information Query System*. Niet gepubliceerd, verkrijgbaar bij het Vogelback Computer Center North Western University, Evanston, Illinois, 1974.
- Cleveland, C., Tavish, D. M. C., Pirro, E., *Contextual Content Analysis Methodology*. Niet gepubliceerd, verkrijgbaar bij Department of Sociology, Drake University, Des Moines, Iowa, 1974.
- Couturier, M., *Le Système Forcod*. Niet gepubliceerd, een lezing gehouden tijdens de Pisa-conferentie, meer informatie bij M. Couturier, Maison des Sciences de l'Homme, 54 Boulevard Raspail, Paris VI., 1974.
- Coxon, D. R., Chalmers, A. D., *Content Analysis as a Social Science Resource*. Annual Progress Report. Niet gepubliceerd, verkrijgbaar bij het Sociology Department of the University of Edinburgh, 1974.
- Coxon, D. R., *The General Inquirer: Introduction to a Computer Based System of Content Analysis*. Niet gepubliceerd, tekst is verkrijgbaar bij het General Inquirer Project, Department of Sociology, University of Edinburgh, 1974.
- Deichsel, A., *The Social Structure of Symbolic Words: Content Analysis and the Sociology of Knowledge*. Niet gepubliceerd, verkrijgbaar bij het Seminar für Sozialwissenschaften, Universität Hamburg, 1974.
- Dunphy, D. D., Dullard, D. G., Crossing, E., *Validation of the General Inquirer*

- Harvard IV Dictionary*. Publicatie in voorbereiding, 1974.
- Gerbner, G., Holsti, D. R., Krippendorf, K., Paisly, W. J., Stone, P. J., *The Analysis of Communication Content; Developments in Scientific Theories and Computer Techniques*. J. Wiley, New York, 1969.
- Hoehle, J., Klingeman, H. D., Rademacher, K., Fuchs, D., *Textpack*. Niet gepubliceerd, verkrijgbaar bij het Zentral Archiv für Empirische Sozial Forschung, Universität Köln, 1973.
- Holsti, O. R., *Content Analysis for the Social Sciences and Humanities*. Addison and Wesley. Reading, Mass., 1969. An Adaptation of the General Inquirer for Systematic Analysis of Political Documents, Behavioral Sciences, Vol. 9, pp. 382-388.
- Holzschek, K., *Eva: A System for Electronic Verbal Analysis*. Niet gepubliceerd, verkrijgbaar bij het Seminar für Sozialwissenschaften, Universität Hamburg, 1974.
- Iker, H. P., Harway, N. J., A Computer Approach Towards the Analysis of Content. *Behavioral Science*, Vol. 10, pp. 173-183, 1965.
- Kooter, B., Van Schuur, W., *Some Possibilities of RIQS as a Content Analysis Program*. Niet gepubliceerd, verkrijgbaar bij het Technisch Centrum van de Faculteit der Sociale Wetenschappen van de Universiteit van Amsterdam, 1974.
- Marhoff, J., Shapiro, G., Weilman, S. R., Towards the Integration of Content Analysis and General Methodology. Heise, D., *Sociological Methodology*, pp. 1-59, 1975.
- Meunier, J. G., *Système d'analyse des textes par ordinateur*. Niet gepubliceerd, verkrijgbaar bij het Departement de Philosophie, of the University du Quebec, Montreal, 1974.
- Morton, E., *The Dynamic of Diadic Conflict: The United States and the Soviet Union, 1961-1963*. Ongepubliceerd Phd. Dissertatie, Stanford University, 1973.
- Morton, E., I. Saris-Gallhofer, and W. E. Saris, *A Validation of the Stanford Political Dictionary*. Niet gepubliceerd, verkrijgbaar bij Afd. Methoden en Technieken van de Subfaculteit der Sociaal-Culturele Wetenschappen van de Vrije Universiteit. Amsterdam, 1975.
- Namenwirth, J. Z., *Some Long and Short Term Trends in one American Political Value: a Computer Analysis of Concern with Wealth in 62 Party Platforms*. Gerbner, G., e.a., *The Analysis of Communication Content*, pp. 233-243, 1969.
- Nie, N., Bent, D. H., Hull, C. H., *SPSS: Statistical Package for the Social Sciences*. McGraw Hill Book Company, New York, 1970.
- Osgood, C. E., Suci, G. J., Tannenbaum, P. H., *The Measurement of Meaning*. University of Illinois Press, Urbana, Illinois, 1957.
- Pool de Sola, J., *Trends in Content Analysis*. University of Illinois Press, Urbana, Illinois, 1959.
- Stone, P. J., Bales, R. F., Namenwirth, J. Z., Ogilvie, D. M., *The General Inquirer, A Computer System for Content Analysis and Retrieval Based on the Science and Unit of Information*. Behavioral Science, Vol. 7, no. 4, pp. 484-498, 1962.
- Stone, P., Smith, M. S., Ogilvie, D. M., *The General Inquirer: A Computer Approach to Content Analysis*, M.I.T. Press, Cambridge, Mass., 1966.
- Stone, P. J., *Improved Quality of Content Analysis Categories: Computerized Disambiguation Rules for High Frequency English Words*. Gerbner, G. e.a., *The Analysis of Communication Content*, pp. 199-233, 1969.

Appendix I:

Een pagina van de Stanford Political dictionary

civic = -2,2,2 civin = 3,3 civisicn = -2,-2 civorc = -2,-1 oc = ,1,2 dcdg = ,,3
 dominant = ,3 dominat = -2,3,- dominaticn = -3,3 don = ,1 doom = -3, -3,-3
 doubl = ,1 doubt = -2,-1,13 doubtful = -2,-2 doubtles = 2,- down = -2 draft = ,1,1
 drag = ,,2 drain = -1,-2 drank = ,,1 draw = ,,1 drawn = -1 drealm = -1,-2,-3
 dreadfil = 3,-2 dream = ,,3 dreamt = ,,3 dreary = -2,-1 drift = ,,-2,-3
 drink = ,,1 driv = ,,2 driven = ,,2 drop = -2,-1,-3 drov = ,,2
 drown = -2,-3,3 drunk = -3,-2 drunken = -3,-3 dug = ,,3 dull = -1,-1,-3
 dunb = -2,-3,3 dup = -2,-3,-3 duplicat = ,,2 duplication = ,,2 durabl = 2,3
 dures = -3,2 duti = 1,1 duty = 1,1 dwell = ,,3 dynamic = ,3,3
 earnqestly = 2
 eager = 2,2 earnest = 2,1 eas = 2 easy = 2 eat = ,,1
 ecstatic = 3 educat = ,,1 effect = ,1,1 effectiv = 2,3,2 efficiency = 2,3
 effort = ,,2 elaborat = ,1,-1 elaps = ,,3 elder = ,1 elderly = ,2
 elect = ,,2 eleganc = 3,1 elevat = ,2,3 eliminat = -2,3,3 effectively = 1,2
 elimination = -2,,2 eloquenc = 2,1 eloquent = 2,1 embarras = -2,-2,1
 embarrassment = 2,-2 embodi = ,1,-2 embody = ,1,-2 embrac = 3,,2 emerg = ,1,2
 emergency = -2 eminent = 3,2 emphasiz = ,1,-2 emphatic = ,3 empir = ,2
 employ = ,,2 empti = -2,-2 empty = -2,-2 emulat = 3,,1 enabl = 2,2
 enact = ,2,3 enamor = 3,,-1 enchant = 3,,-1 enclos = ,2,2 embark = ,,2
 empty-hand = -2,-2
 erroneou = -2
 encounter = ,1,3 encourag = 3,1,-1 encouragement = 2 encroach = -2,2,2
 encroachment = -2,2 end = ,,1 endanger = -3 endear = 3,,-2 endearment = 3
 endeavor = ,,3 endles = -1,1,-2 encors = 2,1 endure = -1,3,-3 endoranc = ,3
 enemi = -3 enemy = -3 energetic = 1,3,3 energetically = 1,3,3 energi = ,2
 energy = ,2 enforc = -1,3,3 engag = 1,1,3 enhanc = 1,3,1 enjoy = 3,,-2
 enlarg = ,2 enligten = 2 enlist = ,,2 ennity = -3 ennobl = 3,2
 enormou = ,3 enough = 3,2 enrag = -3,-1,-1 enraptor = 3 enrol = ,,2
 enrol = ,,2 enroll = ,,2 enshrin = 3 enslav = -3,3,2 enslavement = -3
 ensu = 1,2,1 ensur = 1 entent = 3,3 enter = ,,2 enterpris = ,,3
 entertain = 2,,2 enthusiasm = 3 enthusiastic = 3 envelop = ,3,2 envi = -3,-1,-3
 enviou = -3,1,-3 envy = -3,-1,-3 enqual = ,1 equanimity = 2,2 equip = ,2,1
 equipment = ,2 equit = 3 equity = 3 eracicat = -3,2,2 eradication = -3,2,2
 erect = ,2,3 erection = ,2,2 error = -2,-1 escap = ,2,3 escort = ,,1
 essenc = ,1 establish = ,2,2 esteem = 3,2,-2 estimat = ,,2 estrang = -2,-2
 estrangement = -2,-2 eternal = 3,2 ellogy = 3 evacuat = ,,-3,2 evacuation = ,,-3,2
 evac = -2,-2,3 evaluat = ,,2 evaluation = ,,2 evil = -3 evolv = ,,1
 exacerbat = -2 exacerbation = -2 exaggerat = ,,2 exalt = 3,,-1 examin = ,,1
 examination = ,,1 exce = 1,2,3 excel = 3 excellent = 3 exceptional = 3
 excess = -1 exchang = ,,2 excit = 2,,2 excitement = 2,,3 exclud = -2,2,1
 exclusion = -2,2,1 exclusiv = ,2 excursion = ,,1 execut = -3,2,3 execution = -3,2,3
 exempt = ,,1 exercis = ,2,3 exert = -2,2,3 exhaust = ,,3 exhibit = ,,1
 exil = ,1 expansion = ,3,3 expansionist = ,3,3 expect = ,,2 expectation = -2
 expedient = -1 expedition = ,,2 expel = -2,3,2 expenc = ,,2 expens = -1
 experienc = 1,1 exdept = 3,1 expip = -3,-3 explain = ,,2 exploit = -3,2,3
 exploitation = -3,2,3 explosion = -2,,3 export = ,,2 expos = ,,2 expres = ,,2

expression = ,,2
 expressiv = ,,2 expulsion = 3,3,2 exquisit = 3 extenc = ,,2 exterminat = -3,3,3
 extermination = -3,3,3 extinction = -3,2,2 extinguish = -2,2,2 extirrat = -2,2,3
 extirpation = -2,2,3 extol = 3 extract = ,,2 extraordinary = 2 extravagant = -1
 extreem = 1 facilitat = 1,1 facility = ,2 faction = -1 faculti = ,2
 faculty = ,2 fac = -1,12,-3 fail = -3,-3,-3 failur = -3,-3,-3 faini = -3,-2,-3
 fair = 3,1 faith = 3,2 faithful = 3,2 faithfully = 3,2 fal = -2,-3,-3
 fall = -2,-3,-3 fallen = -2,-3,-3 fals = -2,-1 falsifi = -3,-,-1 falsify = -3,1,-1
 falter = -2,-3 falsehood = -2
 falsely = -2
 fam = 2,1 familiar = 2 famou = 2,1 fan = ,,1 fanatic = -3
 fantastic = 3,1 farc = -2,-1 fascinat = 2 fashionabl = 2 fast = ,2,2
 fat = -2,-2,-3 fatal = -3,-3,-3 fateful = -1 fatigo = -2,-2,-3 fault = -2,-1
 favor = 2,1 favorabl = 2,1 favorit = 2 fealty = 2,,-1 fear = -3,-2,-3
 fearful = -3,-3 feast = 3,1,2 federation = 1,2 feebl = -2,-3,-3 feel = ,,3
 fel = -3,- felicitation = 3,,-1 fell = -2,-3,-3 fellow = 2 fellowship = 2,1
 felt = ,,3 ferocico = -3,3,2 ferociousnes = -3,3,2 fertil = 1,3 fervent = 3,2,3
 fervor = 3,2,3 festival = ,,2 fetch = ,,2 fetter = -2,-3,2