

Naar een modulair model voor natuurwetenschappelijke informatie in elektronische artikelen*

Frédérique Harmsze

Vakgroep Experimentele Fysica, Faculteit der Wiskunde, Informatica, Natuur- en Sterrenkunde, UvA, Valckenierstraat 65, 1018 XE Amsterdam, email: harmsze@phys.uva.nl

Maarten van der Tol

Vakgroep Taalbeheersing, Faculteit der Letteren, UvA, Spuistraat 134, 1012 VB Amsterdam, email: mvdtolet@let.uva.nl

Joost Kircz

Vakgroep Experimentele Fysica, Faculteit der Wiskunde, Informatica, Natuur- en Sterrenkunde, UvA, Valckenierstraat 65, 1018 XE Amsterdam, email: kircz@phys.uva.nl, en Elsevier Science NL, Postbus 103, 1000 AC Amsterdam

URL: <http://www.phys.uva.nl/fnsis/onderzoek/comm/home.htm>

Samenvatting

In het project 'Communicatie in de Natuurkunde' wordt een model ontwikkeld voor modulair opgebouwde artikelen waarmee natuurwetenschappelijke informatie effectiever en efficiënter kan worden weergegeven en ontsloten in een elektronische omgeving. Een modulair artikel bestaat uit een samenhangende verzameling precies gekarakteriseerde informatie-eenheden met een interne structuur. Deze modules kunnen niet alleen met elkaar verbonden zijn binnen één artikel, maar ook met modules van andere modulaire artikelen. In ons model zal een nieuwe structuur worden gedefinieerd voor wetenschappelijke artikelen en zullen aanbevelingen worden gegeven aan auteurs en referenten. De constructie van het model berust op analyses van een corpus natuurkundige artikelen vanuit een vakinhoudelijk en een argumentatietheoretisch perspectief. We karakteriseren de informatie op basis van de bibliografische gegevens, de natuurkundige inhoud, de pragmatische functie en de reikwijdte. In de eerste aanzet tot het model baseren we de modules op de pragmatische functie van de informatie: Meta-informatie, Doel, Methoden, Resultaten, Discussie en Conclusies, en een aantal submodules daarbinnen. Om niet al bij voorbaat voorbij te gaan aan de verworvenheden van de bestaande publicatievorm is deze functionele karakterisering nog sterk gebaseerd op de archetypische paragraafindeling van het natuurwetenschappelijke artikel. We zullen in deze bijdrage het onderzoeksproject introduceren en enkele tussenresultaten bespreken.

1. Inleiding

Elektronische publicatie begint een steeds dominantere rol te spelen in de natuurwetenschappelijke communicatie [1-3]. Algemeen wordt aangenomen dat elektronische publicatie mogelijkheden biedt om problemen op het gebied van opslag en ontsluiting het hoofd te bieden. Uit de eerste experimenten met elektronische wetenschappelijke tijdschriften wordt echter ook duidelijk dat elektronisch publiceren niet *zonder meer* leidt tot een betere communicatie. Bij sommige elektronisch artikelen, zoals de PostScript- of PDF-bestanden die veel wetenschappelijke verenigingen en uitgevers ter beschikking stellen, wordt het nieuwe medium slechts gebruikt als een soort kopieerapparaat op afstand: de lezer wordt geacht het elektronische artikel meteen af te drukken, waarna het resultaat niet meer te onderscheiden is van een authentiek papieren artikel. Dit heeft tot gevolg dat ook bepaalde beperkingen van papieren publicatie volkomen onnodig worden mee-gekopieerd [4].

De effectiviteit en efficiëntie van wetenschappelijke communicatie kan worden vergroot als er meer gebruik wordt gemaakt van de intrinsieke mogelijkheden van de elektronische publicatietechnologie. Zo is het een wezenlijk voordeel van elektronische publicatie dat er een completere

* Dit onderzoek wordt financieel ondersteund door de Stichting Physica, Shell Research and Technology Centre Amsterdam, KNAW, KB en Elsevier Science.

rapportage mogelijk is. Dankzij de efficiënte opslag van informatie hoeven aan de omvang van het artikel geen beperkingen meer te worden opgelegd, zodat bijvoorbeeld omvangrijke wiskundige afleidingen en ruwe meetgegevens kunnen worden opgenomen. Verder bestaat de mogelijkheid om niet-afdrukbare informatie (geluid, bewegende beelden) over te dragen, wat essentieel kan zijn voor een compleet beeld van het onderzoek. Een elektronische publicatie-omgeving kan in principe ook de toegankelijkheid van wetenschappelijke informatie vergroten. Het is bijvoorbeeld mogelijk om complexe zoekoperaties uit te voeren en op een gemakkelijke manier in het informatie-landschap te grasduinen. Als er wordt gepubliceerd in een on-line-bestand, komt daar nog het voordeel bij dat de tijdspanne tussen de voltooiing van het artikel en de publicatie aanzienlijk kan worden bekort.

De genoemde voordelen kunnen echter gemakkelijk omslaan in nadelen als de welhaast oneindige opslagcapaciteit wordt benut voor redundante informatie en de bestaande zoekinstrumenten overbelast raken: hiermee wordt het zogenaamde 'informatie-infarct' [5], dat zich al voordoet bij de informatie-overdracht via papieren wetenschappelijke tijdschriften, alleen maar in de hand gewerkt. Wij gaan ervan uit dat elektronische publicatie kan leiden tot een verbetering van de wetenschappelijke communicatie als de huidige structuur van het artikel, zoals deze in een *papieren* traditie tot stand is gekomen, wordt toegesneden op een elektronische publicatie-omgeving. Op basis van de eisen voor wetenschappelijke communicatie kan de structuur van het artikel worden vernieuwd, waarbij de mogelijkheden van het medium beter worden benut [6,7].

Een essentiële eis voor effectieve en efficiënte communicatie is een adequate structurering en karakterisering van de informatie. Omdat thans de auteur van een wetenschappelijk artikel uiteenlopende soorten informatie tot één geheel samensmeedt, worden ook uiteenlopende classificatiecodes en trefwoorden toegekend aan het document als geheel, en niet aan de specifieke informatie daarbinnen. Veelal zijn vakgenoten echter juist geïnteresseerd in delen van het artikel, bijvoorbeeld in de methode of alleen in de resultaten van een bepaalde meting [8,9,10]. Met dit gevarieerde zoek- en leesgedrag wordt rekening gehouden als een wetenschappelijk artikel gepresenteerd wordt als een verzameling zelfstandig te raadplegen *modules* die elk een eigen, precieze karakterisering hebben. De samenhang van de informatie kan worden weergegeven door verbanden tussen deze modules te expliciteren. Hierdoor kunnen lezers met een minder specifieke interesse zich ook verdiepen in het artikel als geheel.

De verbanden tussen dergelijke informatie-eenheden kunnen met hypertext op een eenvoudige manier worden geïmplementeerd. In een elektronische omgeving kunnen eveneens verbindingen worden gelegd tussen modules uit verschillende artikelen. Zo kan een netwerk van informatie-eenheden gecreëerd worden, waarin de lezer vanuit een gekozen module gerelateerde informatie kan raadplegen in verschillende volgorden.

In het onderzoeksproject 'Communicatie in de Natuurkunde' onderzoeken we hoe een *modulair* artikel zodanig gestructureerd kan worden dat het, in vergelijking met een *lineair* artikel, daadwerkelijk een praktische verbetering oplevert voor de wetenschappelijke (i.c. natuurkundige) gemeenschap. We construeren een model waarin wordt aangegeven hoe de modules en de onderlinge verbanden eruit zouden kunnen zien. Hierin zullen ook aanbevelingen worden gegeven aan auteurs die wetenschappelijke informatie ter publicatie aanbieden en aan referenten die de aangeboden informatie moeten beoordelen. In dit stadium concretiseren we het model nog niet in een feitelijke implementatie. Later zal een prototype worden ontwikkeld waarmee in gebruikersonderzoek wordt nagegaan hoe de praktische bruikbaarheid van het model verder kan worden vergroot.

Het modulaire model zal worden gebaseerd op een analyse van op papier gepubliceerde artikelen. We bestuderen zowel de argumentatie als de wetenschappelijke inhoud van een samenhangend corpus op het gebied van molecuulfysica. We zullen onze analysemethoden en uitgangspunten in paragraaf 2 bespreken. Daarna geven we in paragraaf 3 de eerste aanzet tot het model, in de vorm waarin het zich in deze fase van de analyse bevindt.

2. Analysemethoden

Om bij de herstructurering van het wetenschappelijke artikel niet voorbij te gaan aan de verworvenheden van de papieren publicatievorm gaan we eerst na hoe de informatie in het lineaire, op papier gepubliceerde wetenschappelijke artikel gestructureerd, gekarakteriseerd en geformuleerd is. Daarna proberen we vast te stellen hoe het wetenschappelijke artikel op een verantwoorde manier in een modulaire structuur kan worden omgezet.

Hiertoe analyseren we een corpus artikelen waarin voor het eerst verslag wordt gedaan van nieuw onderzoek. We laten vanwege hun duidelijk andere vorm en functie overzichtsartikelen en populariserende teksten buiten beschouwing. Het onderzoekscorpus bestaat uit 38 natuurkundige artikelen die de weerslag vormen van een afgerond onderzoeksproject op het terrein van de molecuulfysica.¹ Deze artikelen rapporteren over een typisch experimenteel onderzoek dat ook een theoretische component heeft. Ze zijn gepubliceerd in internationale gerenommeerde tijdschriften. Later worden bij het onderzoek ook artikelen buiten het corpus betrokken om na te gaan waar zich bijvoorbeeld relevante verschillen voordoen tussen de disciplines, en tussen theoretische en experimentele artikelen.

We analyseren deze artikelen zowel vanuit een argumentatietheoretisch perspectief als vanuit een vakinhoudelijk perspectief. Zo proberen we te achterhalen wat de systematische kenmerken zijn van de argumentatieve stappen die in deze artikelen worden gezet en welke typen wetenschappelijke informatie in welke samenhang worden overgedragen. De resultaten van deze analyses worden gebruikt bij de ontwikkeling van het modulaire model.

2.1 De argumentatietheoretische analyse

Een wetenschappelijk artikel is niet alleen een samenbundeling van wetenschappelijke informatie. Deze informatie wordt overgedragen met een bepaald communicatief doel dat verbonden is aan het tekstgenre van het wetenschappelijke artikel, en lezers gaan er ook van uit dat de informatie met dit doel wordt overgedragen. Wetenschappelijke artikelen kunnen niet alleen als *rapporterende*, maar ook als *betogende* teksten worden beschouwd: teksten die als doel hebben standpunten over wetenschappelijk onderzoek te verdedigen tegenover een kritisch publiek van vakgenoten. In de eerste plaats wordt er altijd een standpunt ingenomen over de mate waarin het probleem is opgelost dat in het artikel centraal wordt gesteld, maar ook op 'lokaal' niveau kan bijvoorbeeld worden verdedigd waarom voor de gebruikte onderzoeksmethode is gekozen, waarom een meetapparaat op een bepaalde manier is afgesteld en waarom de meetresultaten juist op deze manier worden geïnterpreteerd.

In elk voorstel om het wetenschappelijke artikel te herstructureren zal rekenschap moeten worden gegeven van deze betogende functie. Een modulair artikel moet een lezer niet alleen kunnen informeren, maar ook in staat stellen de argumentatie kritisch te toetsen, wat bijvoorbeeld een moeilijke opgave wordt wanneer onderdelen van een redenering onherkenbaar over meerdere modules verspreid zijn. In het model zal daarom worden aangegeven hoe de argumentatie in een modulair artikel op een aanvaardbare manier kan worden gepresenteerd.

Voor het formuleren van dergelijke aanbevelingen moet er een theoretisch kader zijn waarbinnen argumentatie geanalyseerd en beoordeeld kan worden. Het argumentatietheoretische kader in dit project is dat van de *pragma-dialectiek* [11-13]. Deze benadering is pragmatisch omdat argumentatie wordt beschouwd als een vorm van doelgericht taalgebruik dat net als alle andere taalhandelingen gebonden is aan bepaalde communicatieve spelregels [14]. Iemand die een betoog analyseert en beoordeelt, moet er met andere woorden rekening mee houden dat argumentatie méér is dan een geheel van logische relaties tussen premissen en conclusies. Deze benadering is ook dialectisch omdat argumentatie wordt beschouwd als onderdeel van een discussie: er is een partij die een standpunt inneemt en daarvoor argumenten naar voren brengt en er is een partij die dit standpunt niet bij voorbaat aanvaardt.

In de pragma-dialectische theorie zijn regels geformuleerd waaraan taalgebruikers zich moeten houden wanneer zij op redelijke wijze een discussie willen voeren. Een redelijke discussie wordt in deze benadering opgevat als een discussie waarin de partijen het standpunt dat ter discussie staat kritisch toetsen. Het pragma-dialectische model fungeert als analyse- en beoordelingsinstrument waarmee systematisch kan worden vastgesteld waar en waarom een discussie eventueel 'ontspoot'.

In de pragma-dialectische analyse van het corpusmateriaal worden de artikelen gereconstrueerd als bijdragen aan een redelijke discussie. Kircz [9] en Sillince [15] hebben aanzetten gegeven tot een argumentatieve typering van de informatie in wetenschappelijke artikelen, om zo gecontextualiseerde zoekopdrachten mogelijk te maken. In deze publicaties blijft de analyse echter op het niveau van het klassieke artikel. Wij proberen niet alleen na te gaan hoe de verschillende bestanddelen van een wetenschappelijke discussie gerepresenteerd zijn in het artikel, maar ook welke implicaties dat heeft

¹ Dit onderzoek is in de jaren '70 uitgevoerd aan het FOM-Instituut voor Atoom- en Molecuulfysica (AMOLF) onder leiding van prof.dr. J. Los. Dankzij zijn medewerking beschikken we bij deze analyse over meer informatie dan in de artikelen weergegeven wordt.

voor de omzetting naar een modulaire structuur. Daarna kunnen, op basis van de pragma-dialectische regels, per module aanbevelingen worden geformuleerd voor aanvaardbare argumentatie.

2.2 De vakinhoudelijke analyse

Zoals in de vorige paragraaf naar voren is gebracht wordt wetenschappelijke informatie in artikelen gerepresenteerd in een taal om met een bepaald communicatief doel te worden overgedragen aan anderen. Naast de analyse van deze talige representatie op het *tekstuele* niveau bestuderen we ook de informatie zelf op het *conceptuele* niveau. Hiertoe analyseren we hetzelfde corpus vanuit een vakinhoudelijk perspectief, waarna we proberen de wetenschappelijke informatie opnieuw te representeren in een modulaire vorm. In dit stadium van het onderzoek reconstrueren we de informatie niet in de vorm van bewegende beelden of geluid [16], maar concentreren we ons vooralsnog op een tekstuele representatie.

In de vakinhoudelijke analyse identificeren we zowel de typen informatie die relevant zijn voor de lezer, als de samenhang van deze informatie. We brengen tevens in kaart welke informatie in de artikelen ontbreekt en welke verbanden impliciet zijn gebleven. Deze analyse is er dus op gericht om inzicht te krijgen in de aard, de compleetheid en de coherentie van de wetenschappelijke informatie die via natuurkundige artikelen wordt overgedragen. Bij het analyseren van de artikelen verplaatsen we ons als het ware in de natuurkundige als lezer, en bij het herschrijven van de artikelen in modulaire vorm in de natuurkundige als auteur. Door deze dubbelrol aan te nemen confronteren we ons met de praktische problemen en vragen die kunnen rijzen bij het gebruik van de modulaire vorm in vergelijking met de lineaire vorm.

De vakinhoudelijke analyse richt zich niet alleen op de afzonderlijke artikelen van het corpus, maar ook op de verbanden tussen deze artikelen en op de inbedding van het gehele corpus in de context van het vakgebied. Daarom voeren we analyses uit met verschillende reikwijdtes. We onderscheiden hierbij het niveau van het vakgebied (het *macroscopische niveau*), het niveau van het onderzoeksproject (het *mesoscopische niveau*) en het niveau van het afzonderlijke (deel)onderzoek waarover gerapporteerd is in één artikel (het *microscopische niveau*). Deze meerledige analyse van het corpus is mogelijk doordat het een coherent onderzoeksproject omvat.

2.3 Het startpunt van de analyse

Om te komen tot een zinvolle analyse hebben we eerst een startpunt bepaald. Daarbij stonden de volgende drie vragen centraal:

- 1) Welke classificatietypen moeten ten grondslag liggen aan de karakterisering van informatie in wetenschappelijke artikelen?
- 2) Hoe kan de informatie op grond van deze classificatietypen concreet worden getypeerd?
- 3) Hoe kan op basis van deze karakterisering een modulaire structuur worden geconstrueerd?

Bij het beantwoorden van de eerste vraag willen we benadrukken dat de classificaties die bij het indexeren van lineaire artikelen gangbaar zijn nog steeds een belangrijke rol spelen. We zullen in elk geval gebruik maken van zowel een *natuurkundige* karakterisering, als een karakterisering op basis van *bibliografische gegevens*. Verder maakt een modulaire structuur het mogelijk om de informatie expliciet te classificeren op grond van de specifieke rol die deze informatie speelt binnen het artikel, waarmee wordt gespecificeerd in welke context de informatie aan de orde komt. We noemen dit classificatietype de *pragmatische functie* van de informatie, omdat de modules die we op deze karakterisering baseren praktisch bruikbare informatie-eenheden moeten vormen. Aangezien een modulaire structuur ook de mogelijkheid biedt om hiërarchische verbanden tussen informatie-eenheden te expliciteren is een vierde en laatste classificatietype dat wij onderscheiden de *reikwijdte* van de informatie.

Vervolgens komt de vraag aan de orde hoe de informatie op grond van deze soorten classificaties concreet kan worden getypeerd. We zullen weinig woorden wijden aan de natuurkundige karakterisering van de informatie. We gaan ervan uit dat we bestaande, door anderen ontwikkelde natuurkundige classificaties kunnen gebruiken. Belangrijk is in elk geval dat de natuurkundige indexering, die aan iedere module kan worden toegekend in plaats van aan het artikel als geheel, in een modulaire omgeving

preciezer kan zijn dan in een lineaire omgeving. Bij de indeling op grond van de bibliografische gegevens zullen we evenzeer de lijnen volgen die in het vakgebied Information Retrieval zijn uitgezet.

Een veel minder ontgonnen terrein is de pragmatische functie van de informatie. Om de analyse richting te geven zijn we in eerste instantie uitgegaan van de archetypische paragraaf-indeling van het natuurkundige artikel, die in de wetenschappelijke communicatie haar waarde bewezen heeft: Inleiding, Methoden, Resultaten, Discussie, Conclusies [17]. In deze indeling komen zowel conceptuele typeringen voor (Methoden, Resultaten), als talige (Inleiding, Discussie, Conclusies). Een eerder voorstel voor een conceptuele karakterisering wordt beschreven door Penn en McCauley [18]. Zij typeren de rol van trefwoorden in metallurgische artikelen als bijvoorbeeld 'input', 'output' of 'impurity'. Auteurs die aanzetten hebben gegeven tot een talige karakterisering van de informatie in wetenschappelijke artikelen zijn Sillince [15] en Paice [19]. Wij proberen zowel conceptuele als talige functies te verwerken in een geïntegreerde karakterisering die leidt tot een praktisch bruikbare modulaire structuur. Het uitwerken van de pragmatische functie is het hoofddoel van onze analyse.

Het vierde classificatietype, de reikwijdte van de informatie, wordt geconcretiseerd in het onderscheid tussen het microscopische, mesoscopische en macroscopische niveau van onderzoek. Als in meerdere artikelen over hetzelfde onderzoeksproject wordt gerapporteerd maakt bijvoorbeeld de beschrijving van de meetapparatuur deel uit van al deze artikelen. Deze herhaalde, gemeenschappelijke informatie is mesoscopisch van aard en kan deel uitmaken van een aparte, als mesoscopisch gekarakteriseerde module. Het voordeel hiervan is dat auteurs in latere artikelen voor de apparaatbeschrijving kunnen volstaan met een verwijzing naar deze mesoscopische module, eventueel aangevuld met voor deze artikelen specifieke microscopische informatie.

Het is nog niet duidelijk in hoeverre de auteur daadwerkelijk in staat zal zijn om van tevoren de reikwijdte van de informatie al te bepalen. Bij de analyse van het corpus hebben we het voordeel dat we met terugwerkende kracht kunnen vaststellen welke informatie boven het artikelniveau blijkt uit te stijgen. We vermoeden echter dat bij grotere experimentele projecten veelal wel te voorspellen is dat informatie over de belangrijkste methoden, de uitgangspunten en de bredere context van het onderzoek mesoscopisch is. In elk geval kunnen in een elektronische publicatie-omgeving *retrograde* karakterisering, verbanden en mesoscopische modules worden toegevoegd, die om de authenticiteit van het artikel te behouden wel als zodanig herkenbaar zouden moeten zijn.

Voor de beantwoording van vraag 3 moet het begrip 'module' als bouwsteen voor ons model nader worden gedefinieerd. Een module vatten wij op als een conceptuele informatie-eenheid die gedefinieerd wordt in termen van de karakterisering. Daarbij maken we een onderscheid tussen *elementaire* en *samengestelde* modules. Een elementaire module is de kleinste eenheid van een artikel met een precieze karakterisering op grond van elk classificatietype (vakinhoudelijk, bibliografisch, en op basis van de pragmatische functie en de reikwijdte). Een samengestelde module is een grotere eenheid die opgebouwd is uit een verzameling, elementaire of ook weer samengestelde, submodules. Zo'n samengestelde module kan ten eerste ontstaan wanneer we binnen een classificatietype de karakterisering verder verfijnen. Een voorbeeld hiervan is het onderscheid tussen de submodules Experimentele en Theoretische Methoden binnen de module Methoden. Ten tweede kan er een samengestelde module ontstaan door de interactie van twee classificatietypen. Een pragmatische module kan bijvoorbeeld submodules bevatten met verschillende natuurkundige karakterisering.

De opzet van het modulaire en het traditionele artikel blijven tot op zekere hoogte vergelijkbaar doordat de indeling op grond van de pragmatische functie van de informatie als leidend principe wordt genomen. Hierdoor nemen de modules de plaats in van de paragrafen. Vervolgens worden deze 'hoofdmodules' mede op grond van de andere classificatietypen nader gestructureerd. In de volgende paragraaf zullen we voorbeelden geven van op deze manier geconstrueerde modules.

Een concrete leidraad voor de constructie van een module - en daarmee voor het apart karakteriseren van informatie - is dat de module een 'echte informatie-eenheid' moet behelzen. Dat wil zeggen dat zo'n eenheid een samenhangend geheel moet zijn dat ook betekenis heeft als het geïsoleerd is van de rest van het artikel; praktisch gezien houdt dat in dat het aannemelijk is dat lezers de eenheid ook apart willen raadplegen. De modulaire structuur is overigens nadrukkelijk conceptueel van aard en niet lay-outtechnisch. Grote modules kunnen worden gepresenteerd op verschillende schermen als dat bijvoorbeeld de leesbaarheid of de laadsnelheid ten goede komt. Naar de leesbaarheid van hypertext-documenten zijn de laatste jaren al veel studies verricht [20].

Het modulaire model komt tot stand door middel van een *iteratieve analyse*. Eerst worden enkele artikelen microscopisch geanalyseerd. Deze analyse houdt in dat op basis een gekozen classificatieprincipe wordt nagegaan welke informatie deze artikelen bevatten. Daarna worden ze in een eerste, rudimentaire modulaire vorm herschreven. Vervolgens vindt er een evaluatie plaats van het modulariseringsproces en de resulterende modulaire artikelen, zowel op microscopisch niveau per artikel, als op mesoscopisch niveau door de artikelen met elkaar te vergelijken. Na deze evaluatie wordt het model bijgesteld en weer gebruikt bij de analyse van de volgende artikelen. Deze iteratie zetten we voort totdat het model zover gestabiliseerd is dat de artikelen van het corpus herschreven kunnen worden volgens het model.

We hebben vijf artikelen uit het corpus geanalyseerd (zie appendix). Hierbij zijn we uitgegaan van hoofdmodules die vergelijkbaar zijn met de gestandaardiseerde paragrafen. Behalve deze vijf hoofdmodules voor de wetenschappelijke informatie hebben we ook een module 'Meta-informatie' onderscheiden voor bijvoorbeeld de bibliografische gegevens en het dankwoord. We hebben de informatie uit de artikelen geprobeerd volgens het pragmatische classificatietype te karakteriseren als Meta-informatie, Inleiding, Methoden, Resultaten, Discussie of Conclusies. Vervolgens hebben we geprobeerd deze informatie als een coherent geheel te groeperen in de desbetreffende hoofdmodules, waarbij ook rekening werd gehouden met het vierde classificatietype: de reikwijdte van de informatie.

De eerste modularisering was er ten eerste op gericht de karakterisering van de hoofdmodules te verbeteren. Ten tweede hebben we de hoofdmodules, die in eerste instantie nog elementaire modules waren, waar nodig onderverdeeld in submodules. Ten derde hebben we geprobeerd in grote lijnen vast te stellen welke bestanddelen van modules in het artikel vast en welke optioneel kunnen zijn. We concentreren ons in deze bijdrage op de constructie van de modules zelf en nog niet op de inventarisatie van de verbanden daartussen, noch op mogelijke aanbevelingen voor auteurs en referees.

3. De eerste modularisering

3.1 Belangrijkste bevindingen

Tijdens de analyse bleek dat de hoofdmodules weliswaar tot op zekere hoogte vergelijkbaar zijn met de paragrafen, maar daar zeker niet mee samenvallen. Ondanks de transparante opzet van de geanalyseerde artikelen dekten de paragraaftitels niet altijd de lading. Hoewel een deel van de informatie inderdaad conceptueel gekarakteriseerd kan worden door de bijbehorende titel, is een ander deel gepresenteerd in een andere paragraaf, impliciet gebleven, relevant voor meer dan één paragraaf, of helemaal niet opgenomen in het artikel.

De discrepantie tussen de modules en de paragrafen kan voornamelijk worden verklaard doordat de artikelen *lineair* geschreven zijn, en op papier gepubliceerd. Door de lineariteit worden conceptueel verschillende onderwerpen in één paragraaf ondergebracht omdat de lezer anders te veel moet zoeken binnen het artikel, of zelfs daarbuiten. Verder is het bijvoorbeeld niet altijd mogelijk om in een lineair artikel op papier alle stappen te expliciteren van een omvangrijke berekening: in een papieren tijdschrift wordt slechts een beperkte ruimte beschikbaar gesteld en de hoofdlijn van een lineair artikel zou te veel worden verstoord als de berekening wordt opgenomen in de lopende tekst (zie voorbeeld 1). Dit laatste probleem kan ook in een papieren omgeving worden ondervangen met niet-lineaire hulpmiddelen, zoals appendices.

Tijdens de eerste analyse bleek ook dat er geen vastomlijnd algoritme is om vast te stellen wat de pragmatische functie is van bepaalde informatie. Niet iedere passage waarin wordt geargumenteed kan bijvoorbeeld 'automatisch' worden ondergebracht in de hoofdmodule Discussie. Een voorbeeld hiervan is een specifieke discussie over detectiemethode die gebruikt is in het onderzoek. Het argumentatiethoretische kader kan bij de analyse van dit soort gevallen een structurerende rol spelen. Als we ervan uitgaan dat niet alleen het artikel als geheel een betogende tekst is, maar de afzonderlijke modules ook betogende teksten kunnen bevatten, dan is het duidelijk dat de genoemde specifieke discussie een bestanddeel is van de module Methoden. Dit (sub)betog is dus een onderdeel van de *interne* structuur van de module, waarbij het de functie heeft om aannemelijk te maken dat de gekozen detectiemethode geschikt is om een bepaald verschijnsel te onderzoeken (zie voorbeeld 2). Het herkennen van deze interne structuur kan dus de beslissing om de informatie onder te brengen in een bepaalde hoofdmodule (Methoden of Discussie bijvoorbeeld) vereenvoudigen.

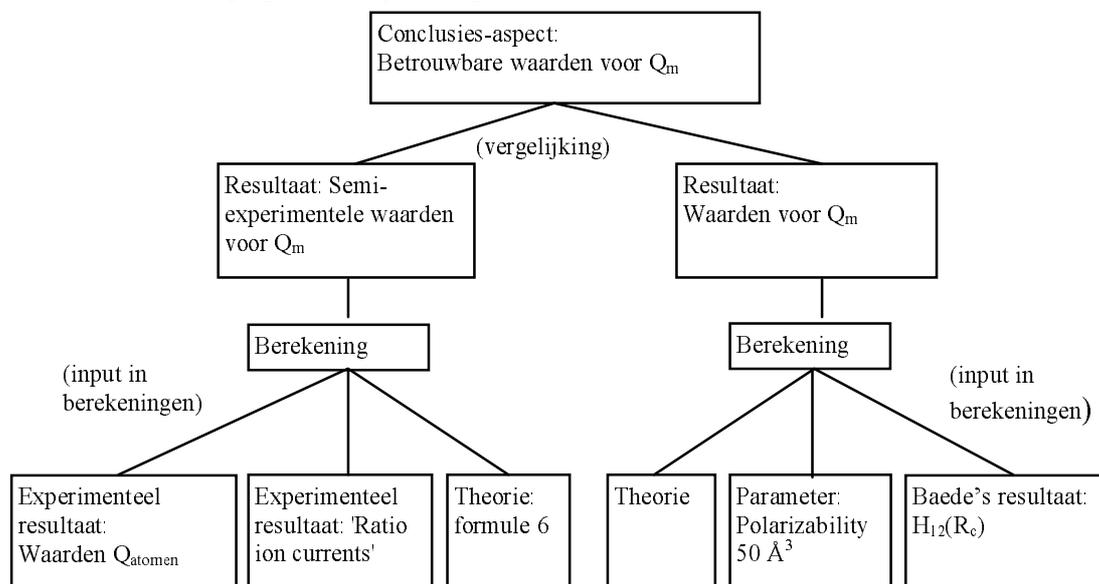
Voorbeeld 1: Vakinhoudelijke analyse

Uit: artikel 4, paragraaf 4.3: 'Discussion - Chemi-ionization by collisions with iodine atoms' (pp.487-488). In het onderstaande citaat zijn de karakterisering op basis van de pragmatische functie typografisch verschillend weergegeven: *Methoden*, **Resultaten**, Discussie.

'[...] Fig.11 gives the cross sections of Na + I and Li + I as functions of the relative velocity of the colliding particles. Using the measured ratio of the ion currents at room temperature and at 1000K, as displayed in Fig.9, and applying again formula 6, the chemi-ionization cross sections for Na + I₂ and Li + I₂ in the plateau region can be obtained. They are found to be $Q_m(\text{Na} + \text{I}_2) = 48 \text{ \AA}^2$ and $Q_m(\text{Li} + \text{I}_2) = 35 \text{ \AA}^2$.

Using Baede's⁶ estimated values for $H_{12}(R_c)$ and the polarizability 50 \AA^3 a direct calculation gives respectively $Q_m(\text{Na} + \text{I}_2) = 34 \text{ \AA}^2$ and $Q_m(\text{Li} + \text{I}_2) = 32 \text{ \AA}^2$, which numbers show a reasonable agreement with the semi-experimental values we report.

In dit fragment uit de paragraaf 'Discussion' komt informatie voor die gekarakteriseerd kan worden als Methoden, Resultaten en Discussie. In deze alinea's wordt over berekeningen gerapporteerd, die deel zullen uitmaken van de submodule Berekeningen, onderdeel van Theoretische Methoden. De eerste is een berekening op basis van eigen resultaten en de theorie, die leidt tot een nieuw resultaat. De tweede berekening is gebaseerd op gegevens van anderen. Vervolgens worden de resultaten van deze twee berekeningen met elkaar vergeleken in een discussie over de betrouwbaarheid van het nieuwe resultaat. We kunnen deze overwegingen ruwweg als volgt schematiseren:



In een elektronische publicatie-omgeving kunnen in het artikel ook de volledige meetgegevens beschikbaar gesteld worden, net als de eigenlijke berekeningen. Daarnaast wordt nu slechts verwezen met "direct calculation gives...". Verder kunnen in de Resultaten-module van dit artikel twee submodules worden onderscheiden met verschillende *natuurkundige* typeringen, namelijk met resultaten voor jodium-atomen en voor jodium-moleculen. In de paragraaftitel wordt alleen verwezen naar atomen. De moleculen worden eerder in het artikel uitvoerig besproken, maar vanwege de lineaire structuur van de rapportage kunnen deze 'molecuul-resultaten' pas na de 'atoom-resultaten' gegeven worden.

Om de interne structuur in kaart te brengen onderscheiden we in eerste instantie pragmatische 'aspecten' in de modules, die dezelfde pragmatische karakterisering hebben als de modules zelf, waarbij de module Meta-informatie buiten beschouwing wordt gelaten (zie figuur 1). Het nut van deze aspecten is dat ze bij de analyse houvast bieden in het geval van een keuzeprobleem, bijvoorbeeld wanneer de informatie als Discussie of als Methoden gekarakteriseerd kan worden. Dan kan, zoals in het bovengenoemde voorbeeld, de informatie als een discussie-aspect van de module Methoden worden opgevat, of andersom. Als het keuzeprobleem nog niet eenduidig opgelost is, kan een verwijzing naar deze Methoden-module worden opgenomen in de Discussie-module, of de informatie alsnog worden gedupliceerd.

Waarschijnlijk zullen niet alle aspecten ook werkelijk in alle modules voorkomen, en zullen tijdens vervolganalyses aspecten worden verwijderd en toegevoegd. De verschillende aspecten zullen vermoedelijk niet zichtbaar worden gemaakt voor de lezer van het modulaire artikel. Er zullen wel aanbevelingen op gebaseerd worden. Bij de bespreking van de eerste aanzet tot het model zullen enkele voorbeelden van aspecten van modules aan de orde komen.

DOEL-MODULE doel-aspect	METHODEN-MODULE doel-aspect	RESULTATEN-MODULE doel-aspect	DISCUSSIE-MODULE doel-aspect	CONCLUSIES-MODULE doel-aspect
DOEL-MODULE methoden-aspect	METHODEN-MODULE methoden-aspect	RESULTATEN-MODULE methoden-aspect	DISCUSSIE-MODULE methoden-aspect	CONCLUSIES-MODULE methoden-aspect
DOEL-MODULE resultaten-aspect	METHODEN-MODULE resultaten-aspect	RESULTATEN-MODULE resultaten-aspect	DISCUSSIE-MODULE resultaten-aspect	CONCLUSIES-MODULE resultaten-aspect
DOEL-MODULE discussie-aspect	METHODEN-MODULE discussie-aspect	RESULTATEN-MODULE discussie-aspect	DISCUSSIE-MODULE discussie-aspect	CONCLUSIES-MODULE discussie-aspect
DOEL-MODULE conclusies-aspect	METHODEN-MODULE conclusies-aspect	RESULTATEN-MODULE conclusies-aspect	DISCUSSIE-MODULE conclusies-aspect	CONCLUSIES-MODULE conclusies-aspect

Figuur 1: De aspecten van de modules

3.2 De eerste aanzet tot het modulaire model

Na de eerste analyse hebben we de hoofdmodules die we hadden gekozen als uitgangspunt bijgesteld en gedeeltelijk ingevuld. De aanzet tot het model voor modulaire artikelen is schematisch weergegeven in figuur 2. We zullen bij de verschillende modules een korte toelichting geven.

(1) *Meta-informatie*

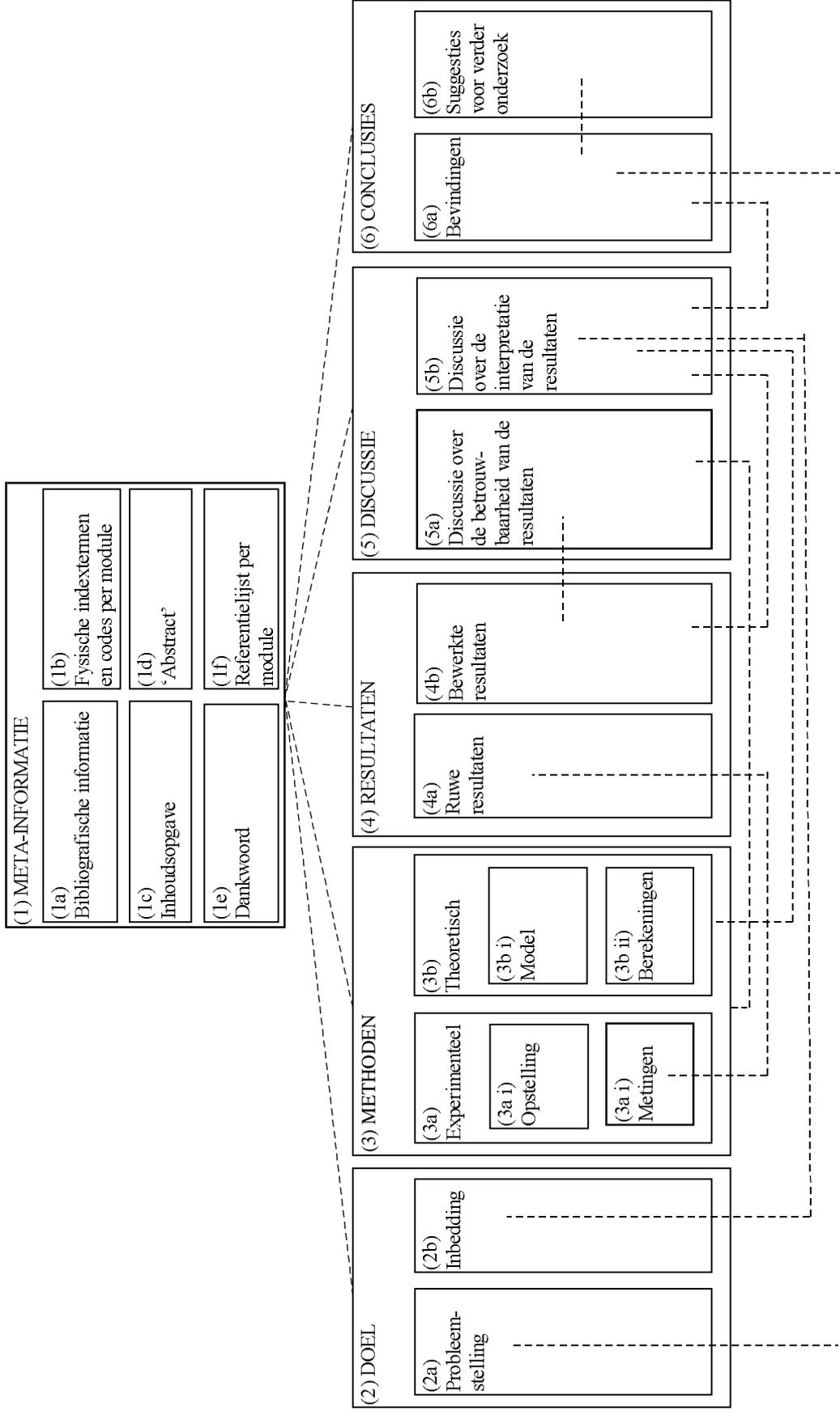
- (1a) *Bibliografische informatie*
- (1b) *Fysische indextermen en classificatiecodes per module*
- (1c) *Inhoudsopgave*
- (1d) *'Abstract'*
- (1e) *Dankwoord*
- (1f) *Referentielijst per module*

Deze 'meta-module' neemt een bijzondere plaats in aangezien hij geen wetenschappelijke informatie, maar informatie *over* het artikel bevat. Hij speelt een centrale rol bij het zoeken en het 'scannen' van het artikel. De submodules met de fysische indexering en de bibliografische informatie, zoals de namen van de auteurs en hun affiliaties, zijn noodzakelijk. Daarnaast bevat elk artikel een inhoudsopgave die de structuur van het artikel weergeeft en een samenvatting van het hele betoog, die beide (via hyperlinks) verwijzen naar alle hoofdmodules van het artikel. Een lijst met referenties naar andere artikelen, uitgesplitst per module, en een dankwoord kunnen een beeld geven van de context van het gerapporteerde onderzoek. Deze twee submodules zijn optioneel.

(2) *Doel*

- (2a) *Probleemstelling*
- (2b) *Inbedding*

De Inleiding bleek verschillende typen informatie te bevatten, namelijk informatie over de probleemstelling (hoewel deze niet altijd expliciet werd vermeld) en informatie over de context van het gerapporteerde onderzoek. Om een expliciete en scherpe formulering van de centrale probleemstelling te garanderen wordt deze ondergebracht in een vaste submodule. De Probleemstelling is zelf geen hoofdmodule, maar vormt samen met de andere submodule Inbedding de hoofdmodule Doel, want een probleemstelling heeft geen informatieve waarde als ze geheel geïsoleerd is van haar context. De Inbedding bevat een beschrijving van de (historische)



Figuur 2: De eerste aanzet tot de modulaire structuur van een artikel, met enkele geëxpliciteerde verbanden tussen de modules

context en de motivering van de relevantie van het gerapporteerde onderzoek. Als het artikel deel uitmaakt van een breder onderzoeksproject kunnen de probleemstelling en de context van het onderzoeksproject als geheel worden gepresenteerd in een mesoscopische module, waaraan afzonderlijke artikelen gerelateerd kunnen worden. De microscopische submodule Probleemstelling bevat de specifieke vraag die in het betreffende artikel centraal staat.

(3) *Methoden*

(3a) *Experimentele methoden*

(3a i) *Opstelling*

(3a ii) *Metingen*

(3b) *Theoretische methoden*

(3b i) *Model*

(3b ii) *Berekeningen*

De hoofdmodule Methoden omvat alle gebruikte methoden, technieken en gereedschappen en is daardoor zo omvangrijk dat nadere onderverdeling in submodules onvermijdelijk is. Het onderscheid tussen experimentele en theoretische (en later wellicht ook computationele) methoden is redelijk eenduidig aan te brengen en kan worden gemotiveerd doordat vermoedelijk veel lezers hoofdzakelijk in een van beide geïnteresseerd zijn. Het onderscheid tussen de beschrijving van het gereedschap zelf en de beschrijving van het *gebruik* van dit gereedschap is ingegeven door de reikwijdte van de informatie: de beschrijving van het gereedschap is vaak mesoscopische informatie die al elders gegeven kan zijn, terwijl een beschrijving van het gebruik van het gereedschap vaak microscopisch is.

Een vast onderdeel van alle elementaire submodules in (3) zijn de restricties van de beschreven methoden en technieken. Voorbeelden zijn de technische grenzen van het apparaat wat betreft het bereik en de nauwkeurigheid, en de aannames omtrent de simplificaties in het model en de verwaarlozing van hogere-ordetermen in de berekening. Deze restricties houden direct verband met de geldigheid en de precisie van de resultaten die met de gebruikte methoden en technieken zijn verkregen.

De belangrijkste complicatie bij de analyse hebben we in verband met deze module al genoemd: de keuze voor de pragmatische karakterisering van de informatie is niet altijd eenvoudig. Er is een duidelijk discussie-aspect in deze module, waarin geargumenteed wordt over de geldigheid van bijvoorbeeld bepaalde aannames. Tussenresultaten in berekeningen en metingen kunnen we als resultaten-aspect beschouwen. Behalve het identificeren, kan soms ook het scheiden van in het lineaire artikel nauw verweven informatie specifieke moeilijkheden opleveren: in de lopende tekst worden vaak onderdelen van een theorie besproken en berekeningen gedaan, in de context van een vergelijking met de experimentele resultaten, gevolgd door een interpretatie van die resultaten (zie voorbeeld 3).

(4) *Resultaten*

(4a) *Ruwe resultaten*

(4b) *Bewerkte resultaten*

Deze hoofdmodule bevat de resultaten van de metingen en berekeningen die moeten leiden tot het antwoord op de vraagstelling die centraal staat in het gerapporteerde onderzoek. In deze module wordt een compleet overzicht gegeven van de 'ruwe resultaten' (4a), en een bewerking van deze resultaten in bijvoorbeeld een tabel of een grafische representatie (4b).

De praktische waarde van het onderscheiden van een hoofdmodule Resultaten is dat de lezer zo in staat wordt gesteld de onderzoeksresultaten op elk niveau van bewerking te vergelijken met zijn eigen werk en de interpretatie daarvan in de module Discussie kritisch te toetsen. De lezer kan met de ruwe resultaten manipuleren, bijvoorbeeld om ze in een zelfgekozen model te interpreteren. Een heldere, bijvoorbeeld grafische, presentatie van bewerkte resultaten maakt het de lezer mogelijk de resultaten eenvoudig te doorzien.

Een fundamentele complicatie bij de invulling van deze module is dat we grenzen moeten definiëren in het continuüm dat zich uitstrekt van de werkelijk ruwe resultaten, via steeds meer bewerkte resultaten tot de uiteindelijke interpretatie die thuishoort in de Discussie-module [21]. De vraag op welk niveau van bewerking ruwe resultaten nog relevant zijn, hangt af van de onderzoekscontext. In de geanalyseerde artikelen zijn geen ruwe, maar wel grondig bewerkte resultaten opgenomen, met name in de vorm van grafieken. De meetgegevens die ten grondslag liggen aan deze grafieken zouden onderdeel kunnen uitmaken van een submodule Ruwe resultaten. Het onderscheid tussen de bewerking en de uiteindelijke interpretatie van de resultaten leverde bij de analyse meer problemen op.

Voorbeeld 2: Argumentatie-analyse

Uit: artikel 4, paragraaf 2: 'Apparatus' (p. 473)

'The primary beam of the alkali-atoms is monitored by two surface-ionization detectors, before and after crossing the secondary beam. We have used iridium-wires. The work function of this element is quite high, about 5.4 eV. The iridium wires were kept at 1325 K and periodically flashed in oxygen at 10⁻⁵ torr. This treatment guarantees a stable high work function which is essential for the detection of sodium and lithium.'

Dit fragment is een onderdeel van het betoog dat de juiste detector op de juiste manier gebruikt is om een bepaald type reactie te bestuderen. De lezer van dit artikel kan uit de laatste zin het standpunt reconstrueren dat er sprake is van een 'stable high work function', wat essentieel wordt genoemd voor detectie van natrium en lithium. Dat de werkfunctie *hoog* is, wordt ondersteund in de tweede en de derde zin: er zijn iridiumdraden gebruikt en de werkfunctie van dit element is zeer hoog, namelijk 5.4 eV. Dat de werkfunctie tijdens de detectie *stabiel* wordt gehouden, wordt ondersteund door de beschrijving van de gevolgde procedure en de garantie dat deze procedure leidt tot een stabiele werkfunctie.

Elders wordt verwezen naar een eerder gepubliceerd artikel, waarin dezelfde procedure wordt besproken om de werkfunctie van iridium stabiel te houden:

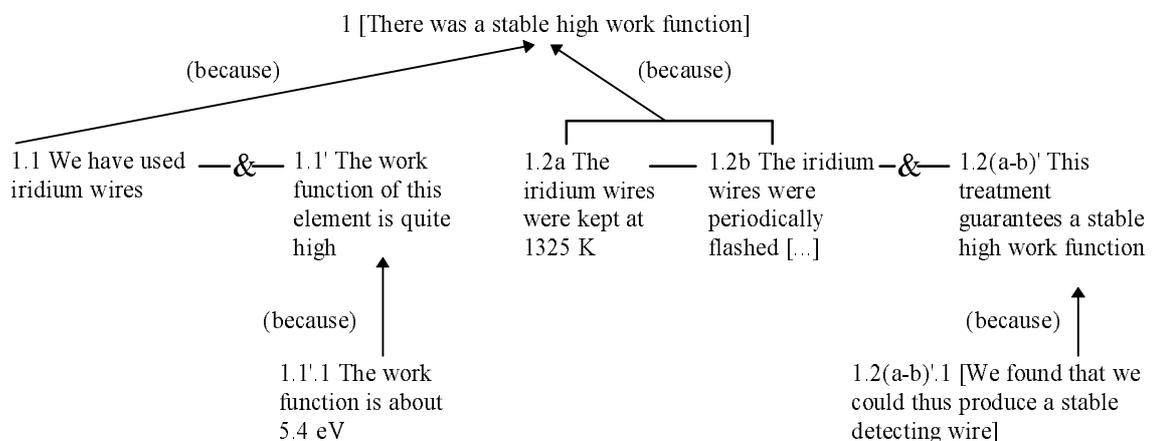
Uit: artikel 3, paragraaf 2: 'Experimental' (p. 426)

'Although Ir has a high work function [...] we sometimes had problems detecting Na and Li, especially in the low-energy range. We found that we could overcome these problems by a procedure, similar to that used by Touw and Trischka 19). We treated the wire at a working temperature of 1325 K in an oxygen bath of 1 x 10⁻⁵ torr for 5 minutes, thus producing a stable detecting wire for at least one day.'

De lezer kan de informatie die hij verkregen heeft uit artikel 4 aanvullen met de informatie uit dit fragment. Zo wordt hier vermeld dat de tijdsduur van de procedure vijf minuten bedroeg en dat de verkregen stabiliteit ten minste één dag standhield. Bovendien wordt er verwezen naar literatuur waarin een vergelijkbare procedure wordt beschreven. In dit fragment wordt niet alleen meer informatie gegeven over de stabiliseringsprocedure, maar wordt deze procedure ook experimenteel gerechtvaardigd ('we found that...'). In artikel 4 staat echter ook informatie over deze detector die in artikel 3 ontbreekt. De lezer die slechts een van de twee artikelen raadpleegt, beschikt dus niet over alle informatie.

In een modulaire publicatieomgeving zouden deze fragmenten kunnen worden vervangen door een verwijzing in beide artikelen naar een mesoscopische module Methoden waarin de beschrijving en de rechtvaardiging van de procedure compleet zijn ondergebracht. De lezer van de afzonderlijke artikelen heeft dan een afgeronde tekst over dit onderdeel van de detector tot zijn beschikking.

De argumentatiestructuur:



1.1, 1.1', 1.2a, 1.2b en 1.2(a-b)' zijn *hoofdargumenten*, waarvan 1.2a en 1.2b *nevenschikkend samengesteld* zijn. 1.1'.1 is een subargument bij 1.1'. Een argument voor 1.2(a-b)' is (1.2a-b)'.1.

Voor de verschillende typen van complexe argumentatie en de notatieconventies, zie [11].

De methoden die gebruikt worden om de resultaten te bewerken en op een verhelderende manier te presenteren vatten we op als een methoden-aspect van deze Resultaten-module. De foutendiscussie waarin de foutenmarge van de resultaten duidelijk wordt gemaakt, kan worden gezien als een discussie-aspect van de Resultaten-module.

In de Resultaten-module kunnen verschillende fysische submodules voorkomen. In artikel 4 uit het corpus komen bijvoorbeeld resultaten voor die aan de hand van de bestaande INSPEC-indexering van dit artikel kunnen worden gekarakteriseerd als 'total cross section' voor 'atomic collision processes' en voor 'molecular collision processes' (zie voorbeeld 1).

(5) Discussie

(5a) Discussie over de betrouwbaarheid van de resultaten

(5b) Discussie over de interpretatie van de resultaten

In de module Discussie kunnen twee soorten discussies worden onderscheiden: een discussie over de betrouwbaarheid van de verkregen resultaten en een discussie over de aanvaardbaarheid van de interpretatie van deze resultaten door vergelijking van de gegevens met (uitkomsten uit berekeningen op basis van) het gehanteerde theoretische model.

De verwoording van de resultaten, waarbij interessante kenmerken benadrukt worden en de resultaten van eventuele verschillende deexperimenten vergeleken worden, vatten wij op als een resultaten-aspect van de module Discussie (zie voorbeeld 3). De keuze van aandachtspunten impliceert namelijk een interpretatie. Een praktisch overweging is dat bovendien de evaluatie van

Voorbeeld 3: Vakinhoudelijke analyse

Uit: artikel 4, paragraaf 4.3: 'Discussion - Chemi-ionization by collisions with iodine atoms' (p.488). In het onderstaande citaat zijn de karakterisering op basis van de pragmatische functie typografisch verschillend weergegeven: *Methoden*, **Resultaten**, Discussie, Conclusies.

An interesting feature is the small oscillation observed in the **lithium cross section (Figs.4 and 11)**. Also in the case of Na there is a hint that the cross section is not rising monotonously. *The two-state Stueckelberg--Landau--Zener^{18,19} theory shows that oscillations can be superimposed upon the Landau--Zener cross section if the difference $H_{11}-H_{22}$ possesses an extremum within the crossing distance R_{∞} . The maxima appear at equal spacings **when the cross section is plotted as a function of v^{-1} , as has been derived by Olson¹⁹**. The amplitude is proportional to $v^{1/2} P(1-P)$ where P is the Landau--Zener probability. As for Li the experimental interval on the v/K scale is rather small, and for Na the oscillations are not well developed, it is difficult to draw any conclusions. It would be interesting to extend the measurements to higher energies, at the same time improving the energy resolution.*

In dit fragment staat de vraag centraal of bij het experiment zogenaamde Stueckelberg-Landau-Zener-oscillaties waargenomen zijn. In het begin van de deze alinea wordt er een kleine oscillatie als "interesting feature" in de grafiek genoemd. In een nadere omschrijving van deze oscillatie wordt iets verteld over de afstand tussen de maxima en de amplitude. Deze beschrijving van de resultaten beschouwen we als een resultaten-aspect van de Discussie-module. De methode waarmee de resultaten op een verhelderende manier worden gepresenteerd vallen we op als een methoden-aspect van de Resultaten-module: hier worden oscillaties inzichtelijk gemaakt door de werkzame doorsneden in een grafiek uit te zetten tegen v^{-1} .

De auteurs proberen de waargenomen oscillatie te interpreteren in termen van de Stueckelberg-Landau-Zener-theorie. De gegeven beschrijving van de Stueckelberg-Landau-Zener-theorie is zeer beknopt en gaat gepaard met referenties naar artikelen waarin meer details worden gegeven. Een dergelijke theorie zou bij uitstek het onderwerp kunnen zijn van een *mesoscopische* module, waarnaar hier kan worden verwezen. Op grond van een evaluatie van de betrouwbaarheid van de experimentele resultaten concluderen de auteurs dat hier geen antwoord kan worden gegeven op de vraag of er Stueckelberg-Landau-Zener-oscillaties zijn waargenomen. We karakteriseren deze informatie als een conclusies-aspect van de module Discussie, omdat deze conclusies niet in een rechtstreeks verband staat met de centrale vraag van het artikel als geheel. In de Conclusies-module kan naar dit conclusies-aspect van de module Discussie wel een verwijzing worden opgenomen. Omdat er over de oscillatie geen bevredigende conclusie kan worden getrokken, wordt een suggestie gedaan voor verder onderzoek. Alle nieuwe vragen die voortvloeien uit het gerapporteerde onderzoek brengen we samen in de submodule 'Suggesties voor verder onderzoek' van de Conclusies-module.

deze bijzondere kenmerken en de betrouwbaarheid van de resultaten zo overzichtelijker gepresenteerd kunnen worden. Een argumentatietheoretische analyse van de (meestal impliciete) stappen die worden gezet bij de verwerking van de gegevens kan over het grensgebied tussen Resultaten en Discussie verheldering brengen. Zodra zich in de rapportage indicatoren voordoen van argumentatie is dat een sterke aanwijzing dat de gezette stap onder vakgenoten nog niet gemeenschappelijk aanvaard is en daarom thuishoort in de module Discussie. Dit probleem zal in de argumentatietheoretische analyse verder worden onderzocht.

De vergelijking van de resultaten met het werk van anderen is een onderdeel van deze module, omdat deze een component is van de discussies over de betrouwbaarheid en/of de interpretatie van de verkregen resultaten. Omdat hiermee de resultaten in een bredere context worden geplaatst, zou deze vergelijking kunnen worden beschouwd als een 'inbedding-aspect', gezien het raakvlak met de submodule (2b).

(6) *Conclusies*

(6a) *Bevindingen*

(6b) *Suggesties voor verder onderzoek*

De Conclusies-module bevat in een heldere en beknopte submodule (6a) de belangrijkste bevindingen, met ofwel een (gedeeltelijk) antwoord op de centrale vraag die gesteld is in de submodule Probleemstelling (2a), ofwel de reden waarom de vraag niet beantwoord kan worden op basis van het verrichte onderzoek. Het hoofdstandpunt (de claim) van het artikel wordt dus in deze vaste submodule weergegeven. In de optionele submodule (6b) kunnen nieuwe vragen worden opgeworpen en suggesties worden gedaan voor een verdere ontwikkeling van het onderzoek.

4. Slotopmerkingen

In deze eerste aanzet tot het model hebben we een raamwerk voorgesteld voor modulaire natuurwetenschappelijke artikelen. We hebben vier classificatietypen geformuleerd die ten grondslag kunnen liggen aan de modulair structuur, namelijk op basis van de bibliografische gegevens, de natuurkundige inhoud, de pragmatische functie en de reikwijdte. Bij een analyse van vijf artikelen uit het onderzoekscorpus hebben we ons geconcentreerd op de pragmatische functie en de reikwijdte van de informatie. De analyse en de modularisering van empirisch materiaal blijken een waardevolle sturende werking te hebben bij de ontwikkeling van deze classificaties, met name omdat het resultaat praktisch toepasbaar moet zijn op informatie-overdracht via reële artikelen.

Uitgaande van de pragmatische functie van de informatie hebben we een globale indeling gegeven van samengestelde en elementaire modules. Deze indeling lijkt bruikbare informatie-eenheden te kunnen opleveren, ofschoon het begrip pragmatische functie nog onvoldoende gedefinieerd is om de informatie eenduidig te kunnen karakteriseren. Om de samenhang van de informatie weer te geven moeten tussen de aldus gedefinieerde modules verbanden geëxpliciteerd worden, waarvan er al een aantal zijn geschetst in figuur 2. We zullen de benodigde verbanden nader inventariseren om een completer beeld te krijgen van de mogelijke structuur van het modulaire artikel als geheel.

Het identificeren van een interne structuur van de modules geeft zowel houvast bij de karakterisering van de informatie, als meer samenhang in de eenmaal gevormde modules. Daarom hebben we verschillende *aspecten* in de modules onderscheiden. De aspecten zijn eveneens gebaseerd op de pragmatische functie en kunnen daarom pas uitgewerkt worden als deze karakterisering goed gedefinieerd is. We verwachten dat vervolganalyses vanuit een vakinhoudelijk en een argumentatietheoretisch perspectief een nadere verheldering kunnen geven van de conceptuele en de talige facetten van dit type karakterisering.

Bij de vervolganalyses zullen we ook verder stilstaan bij de karakterisering van modules op basis van de reikwijdte. Bij het herschrijven van de artikelen uit het corpus in modulaire vorm blijkt met name het onderscheiden van micro- en mesoscopische modules efficiënt. Er zal nader onderzocht moeten worden in welke mate deze indeling daadwerkelijk gehanteerd kan worden bij het schrijven van nieuwe artikelen.

Zoals gezegd zullen we in het model ook aanbevelingen geven voor het schrijven en het beoordelen van een modulair artikel. Enkele aanbevelingen hebben we al geschetst, bijvoorbeeld dat het wenselijk is om alle restricties die relevant zijn voor de methoden op te nemen in de Methoden-module. Verder zien we de zes hoofdmodules als vaste componenten van een modulair artikel, omwille van de

volledigheid van het betoog. Het is wel mogelijk dat enkele daarvan slechts verwijzingen naar andere (mesoscopische) modules bevatten, technisch gesproken dus pointers naar informatie elders in het gegevensbestand. Submodules kunnen ook optioneel zijn, of zelfs niet aan de orde, zoals de submodule Opstelling in een theoretisch artikel.

Na deze eerste modularisering van natuurkundige artikelen is duidelijk dat de omschakeling van lineariteit naar modulariteit niet betekent dat het gestandaardiseerde lineaire artikel eenvoudigweg kan worden doorgeknipt op de verschillende paragraafgrenzen, zoals momenteel wel gedaan wordt met automatische hypertext-converters. We denken echter dat het schrijven van een *origineel* modulaire artikel uiteindelijk efficiënter zal verlopen dan het vertalen van een intrinsiek lineair artikel in een modulaire vorm. De modulaire structuur is met name ontworpen om het communicatieproces efficiënter en effectiever te maken voor de lezer die geïnteresseerd is in specifieke informatie. Nu de contouren van het model zijn geschetst lijkt een modulaire structurering van wetenschappelijke artikelen inderdaad te kunnen leiden tot een praktische verbetering van de communicatie.

We danken F.H. van Eemeren, R. Grootendorst, W. van der Hoek en H.E. Roosendaal voor het waardevolle commentaar dat zij geleverd hebben op eerdere versies van dit artikel.

Literatuur

- [1] Schauder, D. (1994). Electronic publishing of professional articles: Attitudes of academics and implications for the scholarly communication industry. *JASIS*, 42(2), pp. 73-100.
- [2] Hitchcock, S., Carr, L. & Hall, W. (1995). A survey of STM on-line electronic journals 1990-95: The calm before the storm. URL: <http://journals.ecs.soton.ac.uk/survey/survey.html>
- [3] Boyce, P. & Dalterio, H. (jan. 1996). Electronic publishing of scientific journals. *Physics Today*, pp.42-47.
- [4] Ginsparg, P. (1996). Winners and losers in the global research village. In D. Shaw & H. Moore (Eds.), *Electronic publishing in science. Proceedings of the joint ICSU Press-UNESCO expert Conference*. (Paris, February, 1996) (pp. 83-88). Paris: ICSU Press.
- [5] Waard, A. de (1993). Informatie-infarct in de natuurkunde. *Intermediair*, 29(10), pp. 43-45.
- [6] Kircz, J.G. & Roosendaal, H.E. (1996). Understanding and shaping scientific information transfer. In D. Shaw & H. Moore (Eds.), *Electronic publishing in science. Proceedings of the joint ICSU Press-UNESCO expert Conference* (Paris, February, 1996) (pp. 106-116). Paris: ICSU Press.
- [7] Roosendaal, H.E. (1996, in press). Scientific communication in the science process. (Contribution to the International Summer School on the Digital Library, 4-16 augustus 1996, Tilburg University). URL: <http://kubax6.kub.nl:2080/summer/>
- [8] Bazerman, C. (1985). Physicists reading physics. Schema-laden purposes and purpose-laden schema. *Written Communication*, 2(1), pp. 3-23.
- [9] Kircz, J.G. (1991). Rhetorical structure of scientific articles: The case for argumentational analysis in information retrieval. *Journal of Documentation*, 47(4), pp. 354-372.
- [10] Line, M.B. (1988). Redesigning journal articles for on-line viewing. In L.J. Anthony (Ed.), *Lines of thought. Selected papers of Maurice B. Line*. (pp 281-295). London: Clive Bingley.
- [11] Eemeren, F.H. van & Grootendorst, R. (1992). *Argumentation, communication, and fallacies. A pragma-dialectical perspective*. Hillsdale, NJ: Erlbaum.
- [12] Eemeren, F.H. van & Grootendorst, R. (Eds.) (1994). *Studies in pragma-dialectics*. Amsterdam: International Centre for the Study of Argumentation (Sic Sat).
- [13] Eemeren, F.H. et al. (Eds.) (1996). *Fundamentals of argumentation theory*. Hillsdale, NJ: Erlbaum.
- [14] Eemeren, F.H. van & Koning, W.K.B. (Red.) (1981). *Studies over taalhandelingen*. Meppel: Boom.
- [15] Sillince, J.A.A. (1992). Argumentation-based indexing for information retrieval from learned articles. *Journal of Documentation*, 48(4), pp. 387-405.
- [16] Kircz, J.G. (1996). Naar een nieuwe modulariteit van wetenschappelijke informatie. J.J.W. Lisman, G. Goris & J.G. van Soest (Red.), *Van Kennis naar informatie, Van informatie naar kennis. Symposium ter gelegenheid van de officiële opening van de Walaeus Bibliotheek, Leiden* (pp. 71-81).

- [17] Swales, J.M. (1990). *Genre analysis. English in academic research settings*. Cambridge: Cambridge University Press. (The Cambridge Applied Linguistics Series).
- [18] Penn, C.A. & McCauley, M.J. (1965). How to use abstracts and key words. *Journal of Metals* 17(1), pp. 10-13.
- [19] Paice, C.D. (1991). The rhetorical structure of expository text. In K.P. Jones (Ed.), *The structuring of information* (Proceedings of the informatics 11 Conference, University of York, March 20-22, 1991) (pp. 1-25). London: Aslib.
- [20] Geest, Th. van der (1993). Hypertekst: Schrijven en lezen in een niet-lineair medium. *Tijdschrift voor Taalbeheersing*, 15(1), pp. 55-67.
- [21] Pinch, T.J. (1985). Towards an analysis of scientific observation: The externality and evidential significance of observational reports in physics. *Social studies of science*, 15(1), pp. 3-36.

Appendix

Geanalyseerde artikelen uit het corpus, die rapporteren over hetzelfde deelproject:

- (1) Baede, A.P.M., Moutinho, A.M.C., De Vries, A.E. & Los, J. (1969). Total cross sections for charge transfer between alkali atoms and halogen molecules. *Chemical Physics Letters*, 3(7), pp. 530-531.
- (2) Moutinho, A.M.C., Baede, A.P.M. & Los, J. (1971). Charge transfer between alkali atoms and oxygen molecules. *Physica*, (51), pp. 432-444.
- (3) Baede, A.P.M. & Los, J. (1971). Total cross sections for charge transfer and production of free electrons by collisions between alkali atoms and some molecules. *Physica*, (42), pp. 422-440.
- (4) Moutinho, A.M.C., Aten, J.A. & Los, J. (1971). Temperature dependence of the total cross section for chemi-ionization in alkali-halogen collisions. *Physica*, (53), pp. 471-492.
- (5) Baede, A.P.M., Auerbach, D.J. & Los, J. (1973). Fragmentation of negative ions formed in collisions of alkali atoms and halogen molecules. *Physica*, (64), pp. 134-148.