

# Het meten van spreekvaardigheid

WIL KNIBBELER

*Instituut voor Toegepaste Taalkunde, Katholieke Universiteit, Nijmegen*

## Samenvatting

*De auteur geeft de beschrijving van een methode om de spreekvaardigheid in een vreemde taal (Frans) te meten. Het is een toepassing van de 'integrated approach', de testsituatie is een gesimuleerde communicatiesituatie. De beoordeling van de testresultaten is echter sterk analytisch. De verkregen taalproducten werden aan de hand van uitvoerige instructies getranscribeerd en langs machinale weg verwerkt. De aldus door de computer verwerkte testresultaten werden in kwantitatieve termen geïnterpreteerd. De analytische beoordelingsmethode wordt getoetst door de kwantitatieve gegevens te vergelijken met scores toegekend via een globale impressionistische methode. De beschreven methode maakt deel uit van een onderzoeksproject getiteld Evaluatieonderzoek van een Multimediale cursus Frans (SVO project 0225).*

## 1. Theoretisch Kader

Bij het onderzoek van vreemde-taalverwerving speelt het meten van het leereffect een belangrijke rol. Of het nu gaat om de toetsing van een leertheorie, een onderwijsmethode of een instructiesysteem, dan wel om de evaluatie van een programma of een cursus, conclusies kunnen pas getrokken worden, als men erin slaagt het resultaat in objectieve termen te beschrijven. Zo het in z'n algemeenheid een moeilijke opgave is het niveau van taalbeheersing te beschrijven, dan geldt dit meer in het bijzonder voor de productieve mondelinge vaardigheid. Deze specifieke moeilijkheid bestaat niet alleen in de keuze van een adequate

methode om authentiek spreekgedrag uit te lokken, maar ook in de beoordeling van verkregen spraakproducten. De problematiek komt voort uit het feit dat spreekvaardigheid een creatieve en daarom complexe vaardigheid is en dat bij de beoordeling van taalproducten de linguïstische intuïtie een belangrijke rol speelt.<sup>1</sup> Op talrijke manieren werd en wordt dan ook getracht greep te krijgen op het probleem van het testen van spreekvaardigheid.<sup>2</sup> De pogingen om spreekvaardigheid te toetsen die in de zestigerjaren en het begin van de zeventigerjaren ondernomen zijn, kan men in de volgende classificaties onderbrengen. Voor wat de testconstructie betreft constateert men enerzijds de benadering van de totaalvaardigheid (integrated approach) en anderzijds de benadering via deelaspecten (discrete-point approach).<sup>3</sup> Voor wat de beoordeling betreft kan men een onderscheid maken tussen de globaal impressionistische methode en de analytische methode.

Hieronder volgt het verslag van een poging de spreekvaardigheid in de Franse taal van Nederlandse cursisten te meten. Deze methode moet voor wat de testconstructie betreft beschouwd worden als een toepassing van de 'integrated approach' en voor wat de beoordeling betreft als een analytische methode. De analytische werkwijze wordt hier zo ver doorgevoerd dat men zou kunnen spreken van een atomistisch-analytische methode. We hebben gekozen voor de 'integrated approach', omdat deze opzet een taalgedrag uitlokt dat het reële taalgebruik het dichtst benadert en voor de analytische beoordelingsmethode, omdat deze aanpak het meest geschikt is om spraakproducten objectief te kwantificeren en nader te analyseren, speciaal

voor wat betreft de foutieve uitingen.

## 2. Onderzoekskader

De hieronder beschreven methode van spreekvaardigheidsmeting maakt deel uit van het *Evaluatieonderzoek van een Multimediale Cursus Frans* (SVO project 0225). Over de opzet van dit onderzoeksproject, de toegepaste meetprocedures en de thans bekende resultaten is gerapporteerd tijdens de Onderwijsresearchdag op 29 maart te Amsterdam onder de titel: *Rendement van een Taalprogramma Frans verwerkt door volwassenen zonder hulp van een leraar*.<sup>4</sup> Hier zal enkel nader worden ingegaan op de opzet en evaluatie van de Vrije Spreektoetsen. Deze Vrije Spreektoets vormt namelijk de kern van genoemd onderzoek, omdat het project in eerste instantie is opgezet om na te gaan of het mogelijk is spreekvaardigheid als totaalvaardigheid tot ontwikkeling te brengen, zonder dat de cursist in een reële communicatiesituatie getraind wordt, zonder tussenkomst van een 'levende leraar'. Onze hypothese hieromtrent luidde:

De totaaltraining in spreken is geen conditio sine qua non voor verbetering van spreekvaardigheid. Oefening in deelvaardigheden heeft een grotere automatisering tot gevolg en een grotere automatisering brengt van zelf meer gemak in de totaalvaardigheid met zich mee, op voorwaarde althans dat er een juiste dosering en afwisseling plaats vindt tussen training in deelvaardigheden en training in totaalvaardigheid. Deze hypothese werd onderzocht aan de hand van een proeve met de Franse Teleac-Cursus *En Français*. Bij deze cursus is de tussenkomst van een leraar niet voorzien. Er is dan ook geen sprake van oefening in een reële communicatiesituatie. De enige oefentypen die toegepast worden, zijn: luisteren, en deelhandelingen zoals het nazeggen van zinnen en het vormen van zinnen volgens vaste patronen.

Om de verschillende variabelen onder controle te houden, is de cursus, nadat hij was uitgezonden via het open net, opnieuw aangeboden en verwerkt op het Instituut voor Toegepaste Taalkunde te Nijmegen. De televisielessen werden 4 ×

per week gepresenteerd in het videopracticum, de audio-oefeningen werden verwerkt in het talenpracticum, iedere cursist kon hier de hele dag terecht. De cursus werd 2 × aangeboden: 1e van januari 1972 tot juni 1972: 66 cursisten; 2e van oktober 1972 tot maart 1973: 157 cursisten.

De proefgroep was a-select. 120 cursisten hebben zowel aan de pretesten als aan de posttesten deelgenomen, althans voor wat de Vrije Spreektoets betreft.

## 3. Testconstructie en Testafname

De Vrije Spreektoets werd afgenomen voor de aanvang van de cursus en na afloop van de cursus. De opgaven en afnamecondities waren in beide gevallen identiek. Het mogelijk leereffect van de eerste afname is onderzocht aan de hand van de resultaten van 4 controlepersonen, die aan de pre- en posttesten deelgenomen hebben, maar de cursus niet gevolgd hebben. Om de resultaten van onze ppn. te kunnen normeren aan die van 'native speakers', zijn de toetsen ook afgenomen op 10 in Frankrijk woonachtige Fransen van middelbaar intellectueel niveau (administratief personeel bij Air France).

### 3.1 Vrije opdracht

Uitgangspunt bij de testconstructie van de Vrije Spreektoets was dat de opgave de pp. in staat moet stellen over onderwerpen die hem aanspreken iets in het Frans te zeggen zonder enige dwang omtrent de keuze van woorden of structuren die hij hiervoor nodig acht. Hierin onderscheidt zich de Vrije Spreektoets van de eveneens in dit project gebruikte Gebonden Spreektoets, waar de taakstelling zodanig was dat de testpersoon een bepaald woord of grammaticale vorm moest produceren, hiertoe gedwongen door een gegeven zinskader. Bij de Vrije Spreektoets kreeg de testpersoon in geschreven vorm 8 vragen of opdrachten naar aanleiding waarvan hij 5 minuten iets moest zeggen. Om zijn gedachten te ver-

zamelen, kreeg hij 5 minuten tijd om alle vragen/opdrachten in te zien. Hij was gebonden aan de gegeven volgorde, maar vrij in de tijd die hij aan ieder onderwerp zou besteden. Deze vrijheidsmarge leek ons een noodzakelijke voorwaarde voor een optimale taalgebruikssituatie. 'Different people will speak differently about the same subjects or in response to the same stimuli' (Perren 1968, p. 115). Upshur (1971) heeft gewezen op het postulaat van de subjectiviteit van de kant van de testpersoon: 'Test taking is and should be subjective. That is, there should *not* be an 'infallible' rule which each examinee applies in order that all come up with the same answer. If such were the case, there would be no individual differences, and hence nothing to test. The best test, therefore, provides the widest latitude for the student to be subjective in taking the test'. (p. 54) Dat tengevolge van deze subjectiviteit de testpersonen onderling moeilijk vergelijkbaar zijn, is nauwelijks een bezwaar te noemen. Immers het doel van ons onderzoek was niet pp A te vergelijken met pp B, maar pp A van voor de cursus met pp A van na de cursus.

### 3.2 Monoloogsituatie

Na de voorbereidingstijd nam de testpersoon plaats in een cabine van het talenpracticum, waar hij gedurende 5 minuten sprak in een microfoon zonder gesprekspartner. Ofschoon de meest reële spreeksituatie het interview is, hebben we deze situatie niet gekozen voor onze spreektest en wel om de redenen die Ingram (1968) aanhaalt: 'Oral interviews . . . have been shown to be highly unreliable. Their results depend on so many things over and above the candidate's command of spoken language' (p. 77). 'The speech situation is therefore inherently dependent on at least two persons, and a candidate is judged on a performance which is only a part of the total situation' (p. 96). Overigens kan men het aandeel van een gesprekspartner aan een conversatie beschouwen als een reeks min of meer korte monologen. Bovendien is het niet noodzakelijk dat de testsituatie identiek is aan de reële taalge-

bruikssituatie. Immers uit het gedrag dat zich manifesteert in een gereduceerde situatie, is het gerechtvaardigd verwachtingen af te leiden omtrent het gedrag in de totaal situatie.

### 3.3. Verbale stimulus

We hebben niet gekozen voor de visuele stimulus, omdat meerdere experimenten hebben aangetoond dat plaatjes een ongeschikt middel zijn om iemand tot spreken te stimuleren.<sup>5</sup>

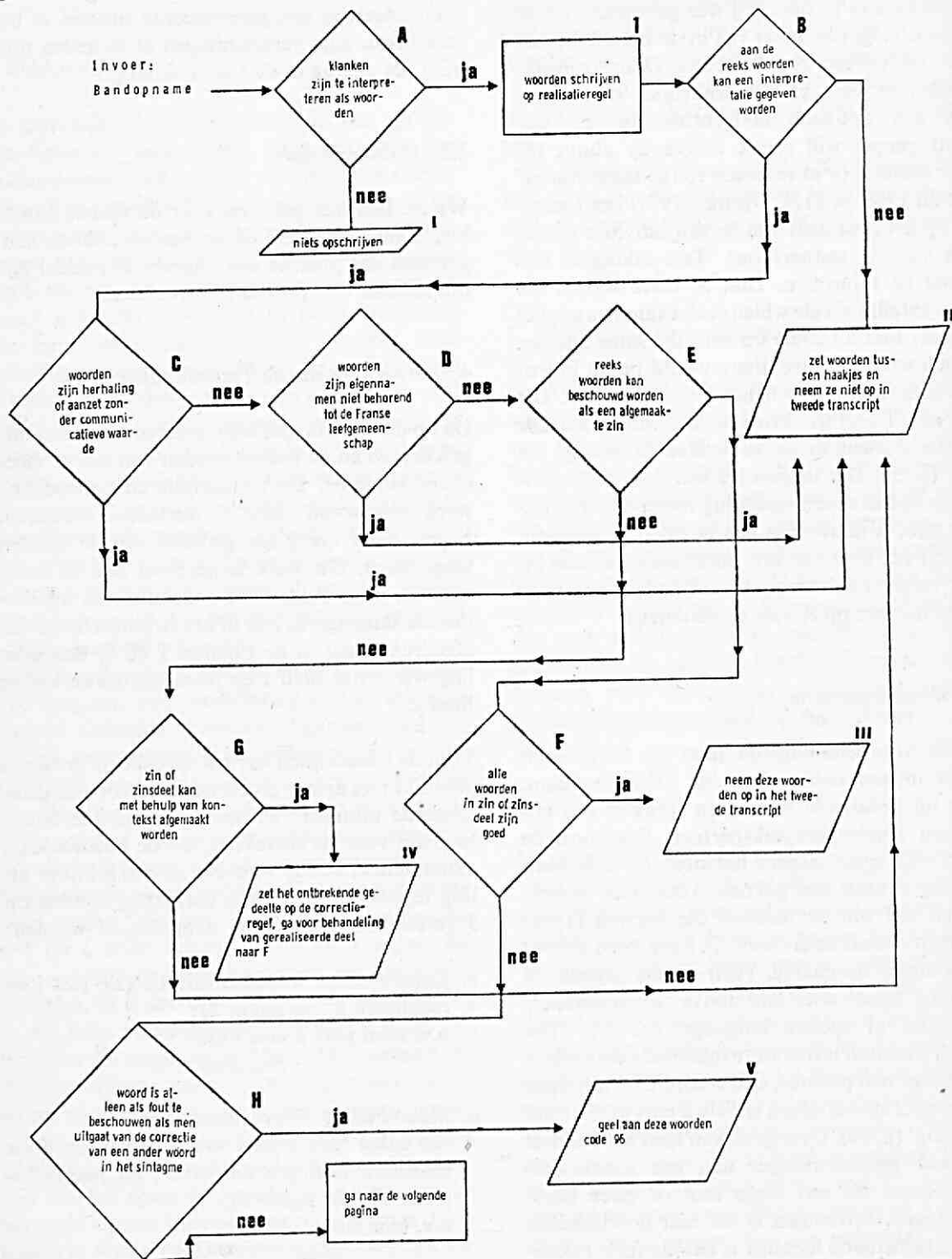
### 4. Verwerking van de Testresultaten

De op de band opgenomen prestaties werden uitgeschreven en de fouten werden van een of meer codes voorzien. De transcriptie en beoordeling werd uitgevoerd door 5 doctoraal studenten Frans. Ieder kreeg een gedeelte van de banden toegewezen. Dit werk is geschied aan de hand van een uitgebreide serie gedetailleerde instructies die samengevat zijn in het stroomschema dat afgedrukt staat in de Figuren 1 en 2, Beoordelingswerkwijze Blad 1 en Beoordelingswerkwijze Blad 2.

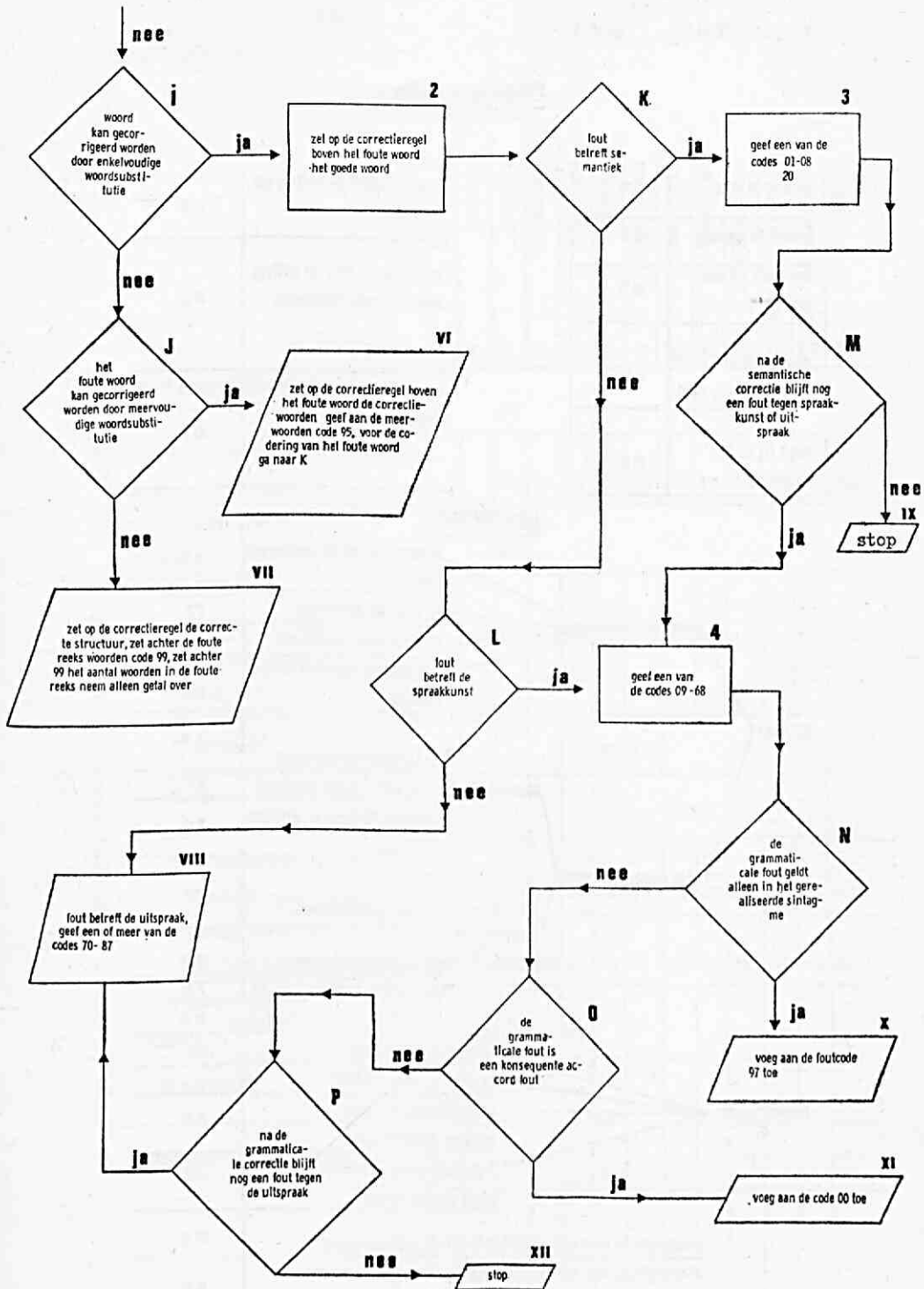
Voor de transcripten werden speciale formulieren gebruikt met aparte regels bestemd voor de generaliseerde uitingen<sup>6</sup> en aparte regels daarboven bestemd voor de correcties, die de beoordelaars aanbrachten, steeds wanneer ze een foutieve uiting tegenkwamen.<sup>7</sup> Deze correcties konden op 3 verschillende manieren aangebracht worden:

- a. Enkelvoudige woordsubstitutie (zie ruit i en rechthoek 2 van figuur 2);  
b.v. mon père a *cinq* ans.  
*cinquante*
- b. Meervoudige woordsubstitutie:  
één enkel fout woord wordt vervangen door meerdere woorden (zie ruit j en parallelogram VI van figuur 2);  
b.v. hier j' *avais* 20 ans.  
*ai eu*

Figuur 1 Beoordelingswerkwijze Blad 1



Figuur 2 Beoordelingswerkwijze Blad 2





Figuur 3 Foutcategorieën Blad 1 Codes

FOUTCATEGORIEËN Blad 1

CODES

Semantische fouten

woord te veel	20
woord te weinig	08
behoort tot een andere taal	02
fout in niet-morfologisch deel	03
fout in suffix of prefix	04

woord maakt de uiting tot nonsens	05
woord geeft aan de uiting een vreemde betekenis	06
woord in de combinatie ongebruikelijk	07

Uitspraakfouten

klinker	geaccentueerd	accent ligt op de verkeerde klinker	70	
		te veel gerealiseerd	71	
		te weinig gerealiseerd	72	
		het timbre van een andere klinker	73	
			74	
	niet geaccentueerd	toonloze ə	te veel gerealiseerd	75
			te weinig gerealiseerd	76
			krijgt timbre v. andere klinker	77
		andere klinker	te veel gerealiseerd	78
			te weinig gerealiseerd	79
	met ander timbre	andere	80	
medeklinker	te veel gerealiseerd	81		
	te weinig gerealiseerd	82		
	gerealiseerd als een andere medeklinker	83		
	2 medeklinkers gerealiseerd als 1 medeklinker	84		
	afwezigheid van een verplichte elisie na een pauze	85		
afwezigheid van een verplichte liaison na een pauze.	86			

Figuur 4 Foutcategoriën Blad 2 Codes

CODES

FOUTCATEGORIËN Blad 2

Grammaticale fouten

	tussenwerpsel	onbepaald v. n. w.	betrekkelijk v. n. w.	vragend v. n. w.	bezittelijk v. n. w.	aanwizend v. n. w.	persoonlijk v. n. w.	lidwoord	voegwoord	voorzetsel	bijwoord	telwoord	werkwoord	bijvoeglijk n. w.	zelfstandig n. w.	
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	
A. woord grammaticaal te veel	21															
B. woord grammaticaal noodzakelijk, te weinig.			55		09	10	11	12	13	14	15	16	17	18	19	
C. woord op verkeerde plaats	22															
D woord grammaticaal onjuist gekozen	a. hulpw. w. avoir être			23												
	b. persoon, getal, tijd, wijze			24												
	c. geslacht	25		26				27	28					27		
	d. getal		29											30		
	e. ander woord van dezelfde woordsoort			58	68	31	32	33	34	35	36	37	38	39	40	41
	f. ander woord van andere woordsoort	50	51	52	53	57	58	59	60	61	62	63	64	65	66	67
E. woord fout in zijn uitgang	42	43	44	45	46	47	48									
F. foute vorming				54												
G. afwezigheid van contractie								49								

- c. Vervanging van een hele structuur:  
de fout is zodanig dat correctie door woordsubstitutie niet mogelijk is (zie parallelogram VII van figuur 2);  
b.v. tu peux mieux acheter.  
il vaut mieux que tu achètes.

De foutcodes werden aangebracht aan de hand van een foutentypologie die weergegeven is in de figuren 3 en 4: Foutcategorieën Blad 1 en Foutcategorieën Blad 2.

Hierbij dient het volgende te worden opgemerkt:

*Ten eerste:* de behandeling van de fouten geschiedt in de volgorde van semantiek (betekenis-leer), grammatica, uitspraak. (Zie in dit verband de volgorde van de ruiten K en L en het parallelogram van figuur 2). Dit betekent dat in eerste instantie de vraag aan de orde komt of in de uiting woorden voorkomen die tot gevolg hebben dat de zin een andere betekenis heeft dan die welke de beoordelaar meent dat de spreker bedoeld heeft. Deze beslissing neemt de beoordelaar op grond van de kontekst dan wel op grond van buitentekstuele informatie. Pas wanneer deze vraag beantwoord is en dienovereenkomstig de correctie eventueel is aangebracht, komt de vraag aan de orde of de afwijking een inbreuk betekent op een grammaticale regel. Pas als deze vraag ontkennend beantwoord wordt, werd de afwijking als een uitspraakfout beschouwd. Deze opvolging van vraagstellingen is van cruciaal belang bij het categoriseren van fouten.

*Ten tweede:* het kan voorkomen dat een woord meerdere fouten bevat. Wanneer een testpersoon zegt:

'je veux condamne un menu'

en daarbij in de werkwoordvorm de medeklinkers m en n realiseert, maakt hij in dit woord 3 fouten:

1. condamner i.p.v. commander = semantische fout;
2. persoonsvorm i.p.v. infinitif = grammaticale fout;
3. medeklinker m te veel = uitspraakfout.

*Ten derde:* het in de figuren 3 en 4 weergegeven foutenschema pretendeert niets anders te zijn dan een frame waarin alle fouten voorlopig opgeborgen kunnen worden. Het is slechts een eerste stap in het systematiseren van alle voorkomende afwijkingen. Het ligt in de bedoeling om van hieruit nadere analyses uit te voeren die ons in staat stellen leertheoretische opvattingen te toetsen. Dit geldt voornamelijk voor het schema van de Grammaticale fouten, zie figuur 4. Dit schema heeft 2 ingangen: de woordsoorten (horizontaal) en de aard van de afwijkingen (verticaal). Dit maakt het mogelijk bij verdere analyses rubrieken samen te voegen zowel in horizontale als verticale richting.

De vraag kan nu gesteld worden of de 5 transcribenten-beoordelaars het hierboven geschetste correctiesysteem consequent hebben toegepast. Voorop zij gesteld dat wekelijks langdurige besprekingen plaats gevonden hebben, waarbij alle concrete problemen gezamenlijk opgelost werden en dat de beoordelaars elkaars werk verbeterd hebben. Om nu te toetsen of de beoordelaars overeenkomstig de gegeven instructies te werk zijn gegaan, werden 3 testbanden door ieder van de 5 beoordelaars verwerkt en heeft iedere beoordelaar 3 dezelfde banden getranscribeerd met een tussenpoos van 2 maanden.

De in de transcripten vastgelegde gegevens bleken zo omvangrijk en gedetailleerd te zijn, dat het wenselijk leek, ze langs machinale weg te analyseren. Hiertoe werden de transcripten door 5 andere student-assistenten, eveneens doctoraal studenten Frans, omgezet op ponsformulieren.<sup>8</sup> Met behulp van de computer konden nu exacte antwoorden verkregen worden op de volgende vragen:

1. Welke vraag of opdracht bleek de testpersonen het meest tot spreken te stimuleren?
2. Hoe vaak bleek het onmogelijk fouten door enkelvoudige woordsubstitutie te corrigeren?
3. Welke verschillen in beoordeling komen voor bij de 5 beoordelaars?
4. Welke verschillen in beoordelingswijze komen voor bij dezelfde beoordelaar tussen de eerste beoordeling en de tweede beoordeling (interval van 2 maanden)?



5. Om de testopdrachten uit te voeren, maakten de proefpersonen hoe vaak gebruik van woorden die in de cursus voorkomen, in pretest en in posttest?
6. Wat is de frequentie van de foutcategorieën in pretest en posttest?
7. In welke kontekst komen de fouten voor?
8. Welke zijn de linguïstische eigenschappen van iedere test?

### 5. Interpretatie van de Testresultaten

Hier geven we alleen een antwoord op de in de vorige paragraaf sub 8 gestelde vraag.

Van iedere toets werden 6 gegevens berekend:

1. totaal aantal woorden (tokens) (TW),
2. totaal aantal goede woorden (tokens) (TGW),
3. totaal aantal fouten (TF)  
(totaal aantal foute woorden, omdat een woord meerdere foutcodes kan hebben),
4. percentage goede woorden op het totaal aantal woorden (PGW),
5. aantal verschillende goede woorden (types) (VGW),
6. type-token relatie van de goede woorden (TT) = VGW/TGW.

Om inzicht te krijgen in de mogelijke vorderingen van de totale proefgroep en om eventuele vorderingen te kunnen normeren aan de prestaties van 'native speakers', werden van deze 6 gegevens gemiddelden berekend in de pretesten en in de posttesten en werden eveneens de gemiddelden vastgesteld van de 10 Franstalige normpersonen. Omdat bij voorafgaande analyses gebleken was dat de 5 beoordelaars op bepaalde punten verschillen vertoonden in hun beoordelingswijze, werden de gemiddelden alleen berekend van die ppn. wier pretest en posttest door dezelfde beoordelaar getranscribeerd en beoordeeld waren. Het lag aanvankelijk in de bedoeling om uit deze 6 gegevens een maat af te leiden voor 3 eigenschappen van het spreekgedrag, n.l.: correctheid, spreekgemak (fluency), gevarieerdheid in taalgebruik.

Voor de correctheid werd als index gekozen

het percentage goede woorden op het totaal aantal woorden (PGW). Voor de gevarieerdheid in taalgebruik bleek de oorspronkelijk overwogen type-token relatie (TT) geen juiste maat. Dit houdt verband met het feit dat de TT geïnterpreteerd moet worden in functie van het totaal aantal woorden en dat dit alleen zinvol is bij teksten van grote omvang. Dat de type-token relatie een moeilijk interpreteerbaar gegeven is, blijkt in nog sterkere mate, als men van dit gegeven de gemiddelde waarde berekent.

Tabel 1. Gemiddelde Type-token relatie.

TT Pretest	TT Posttest	TT Native Speakers
$\bar{M}=0.57(N=108)$	$\bar{M}=0.58(N=108)$	$\bar{M}=0.58(N=10)$

Uit Tabel 1 (Gemiddelde Type-token relatie) mag geenszins afgeleid worden dat de gevarieerdheid in taalgebruik bij de posttest gelijk zou zijn aan die van de 'native speakers'.

Het leek ons daarom juist om van iedere proefpersoon de type-token relatie af te zetten tegen het totaal aantal goede woorden (tokens) (TGW). Het verband tussen beide gegevens vindt men uitgedrukt in figuur 5 (Type-token relatie).

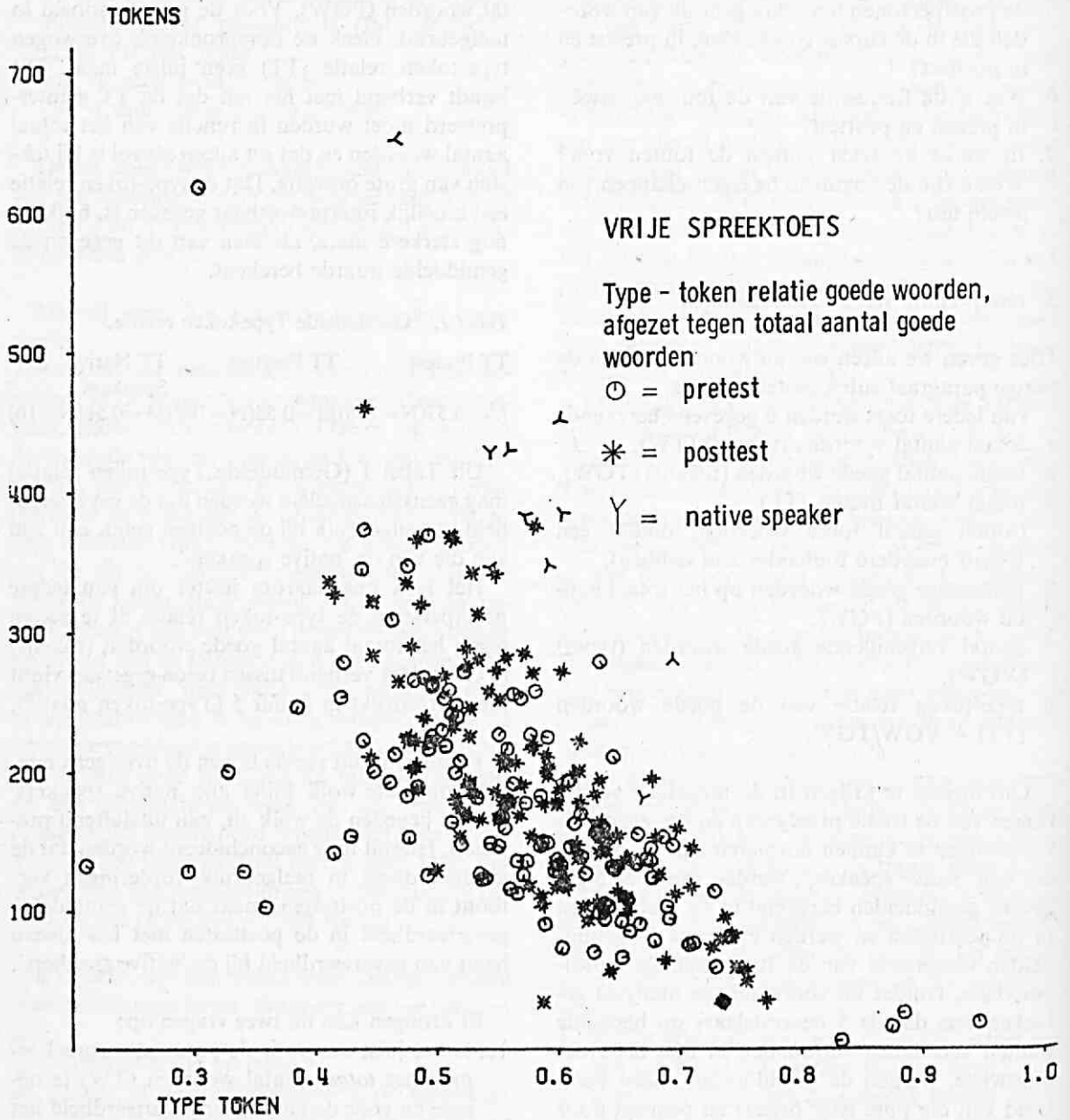
Figuur 5 toont rechts boven de overigens relatief compacte wolk bijna alle 'native speakers' en wat beneden de wolk zit, zijn uitsluitend pretesten. Hieruit mag geconcludeerd worden dat de gevarieerdheid in taalgebruik vorderingen vertoont in de posttesten, maar dat de gemiddelde gevarieerdheid in de posttesten niet het niveau haalt van gevarieerdheid bij de 'native speakers'.

Er drongen zich nu twee vragen op:

- 1e. Is het juist om voor de parameter spreekgemak het *totaal* aantal woorden (TW) te nemen en voor de parameter gevarieerdheid het aantal *verschillende* goede woorden (VGW)?
- 2e. Zoja, hoe dient dan het totaal aantal goede woorden (TGW) geïnterpreteerd te worden?

Alvorens deze twee vragen te beantwoorden, werden de correlaties berekend tussen deze 3 gegevens zowel in de pretesten als in de posttesten.

Figuur 5 Type-token relatie



Tabel 2. Correlaties TW, TGW, VGW.

N=108	TGW	VGW	TGW	VGW
	Pre	Pre	Post	Post
TW Pre	0.99	0.90		
TW Post			0.99	0.94

De hoge correlaties die uit Tabel 2 blijken, hebben ons doen besluiten, geen onderscheid te maken tussen 'fluency' en gevarieerdheid in taalgebruik. Bij de keuze van een van de 3 gegevens is de voorkeur uitgegaan naar de Verschillende

Goede Woorden (VGW), omdat hierin zowel de component gevarieerdheid als de component hoeveelheid ligt opgesloten. Dit gegeven zou men nu kunnen beschouwen als de algemene maat voor spreekvaardigheid naast spreekcorrectheid. Deze maat noemen we in het vervolg spreekgemak.

We hebben nu de kwantitatieve gegevens, die wij via minutieuze analyses verkregen hebben, omgezet in 2 maten voor spreekvaardigheid. Welke waarden deze 2 indices hebben bij onze proefpersonen voor en na de cursus en hoe deze waarden zich verhouden tot die van de 'native speakers', moge blijken uit Tabel 3 Vorderingen Vrije Spreektoets.

Tabel 3 Vorderingen Vrije Spreektoets

	Native Norm		Pretest		Posttest		
	(N = 10)		(N = 108)		(N = 108)		
	$\bar{M}$ =	s =	$\bar{M}$ =	s =	$\bar{M}$ =	s =	
Spreekgemak	220.90	45.00	92.30	35.41	107.47	38.05	t = 3.03** df = 214
Spreekcorrectheid	98.83	0.92	81.28	7.23	84.63	3.35	t = 4.00** df = 214

Tabel 3 Vorderingen Vrije Spreektoets toont aan dat, ofschoon de cursisten bij de posttest nog ver verwijderd blijven van de prestaties van de 'native speakers', ze bij een taakstelling die de reële communicatiesituatie benadert, significante vorderingen maken zowel in gemak van spreken als in correctheid van spreken. Deze conclusie mag beschouwd worden als een bevestiging van onze hypothese die luidde dat het mogelijk is spreekvaardigheid te ontwikkelen zonder training in reële communicatiesituaties.

6. *Toetsing van de analytische beoordelingsmethode door vergelijking met de globaal impressionistische methode.*

Uit bovenstaande beschrijving blijkt dat wij voor het meten van spreekvaardigheid een methode toegepast hebben die zo analytisch is dat de kwa-

lifikatie van atomistisch niet ongerechtvaardigd lijkt. De vraag is echter of de langs deze weg verkregen bevindingen overeenstemmen met de scores die men verkrijgt via de doorgaans toegepaste globale beoordelingsmethode. Om deze vraag te beantwoorden worden momenteel alle testbanden door 3 Fransen beoordeeld, die als opdracht kregen de testen eenmaal te beluisteren en hun globale impressie omtrent correctheid en spreekgemak uit te drukken in een cijfer tussen 1-10 of 1-20. Deze 3 beoordelaars zijn geselecteerd uit 10 in Frankrijk woonachtige Fransen op basis van objectiviteit, constantheid en nuance-gevoeligheid. Over de correlaties die we zullen vinden tussen onze getallen en de toegekende cijfers, zal later gerapporteerd worden.

Noten

1. Voor de problematiek van toetsing van spreek-

- vaardigheid zie Freudenstein (1973), Grayshon (1968), Perren (1968), Spolsky (1968) en Upshur (1969).
2. Voor de verschillende pogingen die op dit gebied ondernomen zijn, zie Carroll (1961), Perren (1967), Raash (1972), Robinson (1969), Roy (1967), Schrand (1973), Valette (1967) en Valette (1969).
  3. Voor een beschrijving van beide benaderingen zie Carroll (1961).
  4. Van deze inleiding zal een samenvatting verschijnen in het Verslag van de Onderwijsresearchdag. Overigens is over het Evaluatieonderzoek door Wil Knibbeler gerapporteerd in *Mededelingen* van het Instituut voor Toegepaste Taalkunde te Nijmegen no. 1, pp. 38-41; no. 2, pp. 45-56 en no. 3, pp. 21-28.
  5. Voor bedoelde experimenten zie Brière (1968), Creed (1969) en Baetens Beardsmore en Renkin (1971).
  6. Dit is wat de realisatieregels genoemd wordt in rechthoek 1 van Figuur 1 Beoordelingswerkwijze Blad 1.
  7. Dit is wat de correctieregels genoemd wordt in Figuur 1 Beoordelingswerkwijze Blad 1 parallellogram IV en in Figuur 2 Beoordelingswerkwijze Blad 2 rechthoek 2 en de parallellogrammen VI en VII.
  8. Dit is wat het tweede transcript genoemd wordt in de parallellogrammen II en III van Figuur 1 Beoordelingswerkwijze Blad 1.

#### Bibliographie

- Baetens Beardsmore, H., en A. Renkin, 'A Test of Spoken English', *IRAL* IX/1 (1971), 1-11.
- Brière, Eugène J., 'Testing ESL among Navajo Children', in: John A. Upshur en Julia Fata (eds.), *Problems in Foreign Language Testing*. Language Learning Special Issue 3 (1968), 11-21.
- Carroll, Joh. B., 'Fundamental Considerations in Testing for English Language Proficiency of Foreign Students' in: *Testing the English Proficiency of Foreign Students*, Washington, D.C.: Center of Applied Linguistics 1961, 30-40. Herdrukt in Mildred R. Donoghue (ed.), *Foreign Languages and the Schools: A Book of Readings*, Dubuque (Iowa) 1967, 276-285.
- Creed, Terence, 'Picture Composition and oral-response tests', in: Heinrich Schrand (ed.), *Leistungsmessung im Sprachunterricht*, Marburg 1969, 131-148.
- Freudenstein, R., 'Testen im Sprachunterricht - Notwendigkeit und Gefahren', in: H. Schrand (ed.), *Testen Probleme der Objektiven Leistungsmessung im neusprachlichen Unterricht*, Berlin 1973, 11-22.
- Grayshon, M. C., *The Examination of Spoken English*. A discussion of some fundamental weaknesses of examinations in spoken English for students whose mother tongue is not English, Nottingham, University of Nottingham, Institute of Education 1968.
- Ingram, Elisabeth, 'Attainment and Diagnostic Testing', in: Alan Davies (ed.), *Language Testing Symposium. A Psycholinguistic Approach*. (Language and Language Learning. nr. 21), London 1968, 70-97.
- Perren, George E., 'Testing ability in English as a second language: 3 spoken language', *English Language Teaching* XXII, 1 (1967), 22-29.
- Perren, George E., 'Testing Spoken Language: Some Unsolved Problems', in: Alan Davies (ed.), *Language Testing Symposium*, London 1968, 107-116.
- Raash, Albert, 'Tests im Sprachenunterricht', *Die Neueren Sprachen* 12 (1972), 696-706.
- Robinson, Peter, *Oral Expressions Tests*, Based on a paper presented at the Canadian Linguistics Association Congress, Toronto, June 10-13, 1969.
- Roy, Robert R., *Oral French Proficiency. Identification and Evaluation*, Alberta univ., Edmonton 1967.
- Schrand, Heinrich, 'Zur Kontrolle und Bewertung mündlicher Leistungen in der Zertifikats-Prüfung Englisch', in: Heinrich Schrand (ed.), *Testen Probleme der objektiven Leistungsmessung im neusprachlichen Unterricht*, Berlin 1973, 93-98.
- Spolsky, Bernard, *What does it mean to know a language or how do you get someone to perform his competence*, Communication presented at the Second conference on Problems in Foreign Language Testing, Idyllwild, Calif., 7-9, November 1968.

Upshur, John A., 'Measurement of Oral Communication', in: Heinrich Schrand (ed.), *Leistungsmessung im Sprachunterricht*, IFS Marburg/Lahn 1969, 53-80.

Upshur, John A., 'Objective Evaluation of Oral Proficiency in the ESOL Classroom', *TESOL Quarterly*, V 4 (1971), 47-59.

Valette, Rebecca M., *Modern Language Testing. A Handbook*, New York 1967.

Valette, Rebecca M., *Directions in Foreign Language Testing*, New York 1969.

*Curriculum vitae*

Drs. W. Knibbeler (geb. 1924) studeerde Romaanse taal- en letterkunde; was van 1954-1967 werkzaam als docent Frans bij het middelbaar onderwijs; van 1967-1969 medewerker van het Algemeen Pedagogisch Studiecentrum; vanaf 1969 als wetenschappelijk medewerker verbonden aan het Instituut voor Toegepaste Taalkunde van de Katholieke Universiteit te Nijmegen; publiceerde een aantal artikelen, voornamelijk in *Levende Talen*.

Adres: I.T.T., Erasmuslaan 40, Nijmegen.