



Universiteit Utrecht

De kansen van spreiding

Hugo Quené

ORATIE

Oratie uitgesproken bij de aanvaarding van het ambt van hoogleraar Kwantitatieve Methoden van Empirisch Onderzoek in de Geesteswetenschappen aan de Universiteit Utrecht op maandag 13 maart 2017.

*Mijnheer de Rector Magnificus,
geachte toehoorders,*

INLEIDING

Voor sommigen van U is dit wellicht de eerste keer dat U een inaugurele rede bijwoont en aanhoort. Anderen van U, daarentegen, hebben zo'n academische plechtigheid al veel vaker meegemaakt. En weer anderen van U hebben een tussenliggende score. Er is dus in deze steekproef van vanmiddag een aanzienlijke *spreiding* in uw oratie-ervaring. Anderzijds zijn er ook grote overeenkomsten: de meesten van U hebben twee werkende oren om mijn spraakklanken op te vangen, en er is grote *kans* dat U beschikt over de cognitieve vermogens om mijn Nederlandse woorden en zinnen te begrijpen. Er is dus een interessante tegenstelling tussen enerzijds spreiding en variatie, en anderzijds uniformiteit. Mijn uitgangspunt daarbij is dat we iets te weten kunnen komen over die uniformiteit, over wat ons verenigt, juist door die variatie mede te onderzoeken.

De leerstoel die ik sinds enige tijd bezet heeft een lange titel, namelijk “Kwantitatieve methoden van empirisch onderzoek in de geesteswetenschappen”. Ik wil graag beginnen met die laatste term, de geesteswetenschappen. De geesteswetenschappen, in het Engels “humanities” genaamd, houden zich bezig met de vraag “wat maakt ons tot mensen” (in het Engels, “what makes us human”)? Naar mijn idee vormt *taal* het belangrijkste element van een antwoord op die vraag. Taal onderscheidt ons mensen van andere levensvormen op deze planeet, en taal vormt het fundament van de studieobjecten van de geesteswetenschappen: geschiedenis, literatuur, cultuur, religie. Dank zij ons taalvermogen kunnen we deze diverse onderwerpen ook weer bestuderen. Maar ook in dat taalvermogen vertonen mensen grote variatie. In de eerste plaats natuurlijk omdat er duizenden verschillende talen zijn die mensen als zuigeling leren of geleerd hebben. En zelfs onder de sprekers van *dezelfde* taal, bijvoorbeeld het Nederlands of

bijvoorbeeld het Engels, horen we al opvallende verschillen in uitspraak, woordkeus, en zinsbouw. Maar ondanks die spreiding en variatie, of misschien juist daardoor, is er ook een gemeenschappelijke uniforme kern in de taal. Juist door te kijken naar de spreiding tussen talen, tussen gebruikers van die talen, en tussen talige uitingen die die taalgebruikers uitwisselen, kunnen we komen tot zinnige antwoorden op de vraag “wat maakt ons tot mensen”.

Dat geldt niet alleen voor het onderzoek naar taal en communicatie, maar ook voor andere terreinen van geesteswetenschappelijk onderzoek: literatuur en cultuur, geschiedenis, en wellicht ook filosofie. Natuurlijk moeten we op al deze onderzoeksterreinen zoeken naar onderliggende gemeenschappelijke patronen en verbanden. Maar daarbij moeten we de rijke en informatieve *variatie* in onze observaties wel in het oor en oog houden. Die variatie of spreiding biedt immers, zo zal ik betogen, ook nieuwe kansen op wetenschappelijke vooruitgang. Ik wil daarom vandaag de volgende stelling verdedigen:

Kwantitatieve methoden van empirisch onderzoek (en de bijbehorende computationele en statistische technieken) zijn onontbeerlijk voor verdere wetenschappelijke vooruitgang in de geesteswetenschappen.

Anders gezegd, aandacht voor spreiding in de data biedt ons dus nieuwe kansen, op robuuste, valide, empirisch goed gefundeerde inzichten in onze vakgebieden.

Over enkele ogenblikken zal ik terugkeren naar mijn stelling. Maar eerst wil ik kort met U terugkijken naar de geschiedenis van mijn oorspronkelijke vakgebied, de discipline die zich bezighoudt met de menselijke spraak, de *fonetiek*.

FONETIEK

Menselijke spraak is een uniek verschijnsel, dat zich afspeelt op het grensvlak van het menselijk lichaam en de menselijke geest. Om met elkaar te spreken en om elkaar te verstaan hebben we onze lichamen nodig (zoals longen, strottehoofd en tong, maar ook oren, en zenuwbanen). Maar net zo hard hebben we onze geesten nodig, om de bedoelde woorden en zinnen en teksten te bedenken, te produceren, en te begrijpen. De fonetiek is daardoor uit de aard der zaak multidisciplinair.

- Vanuit de taalwetenschap zijn we gericht op het *uniforme* menselijke taalvermogen dat in spraak tot uiting komt.
- Vanuit de logopedie, en ook vanuit forensisch perspectief, zijn we daarentegen juist gericht op de individuele *variatie* tussen en zelfs binnen sprekers.
- Vanuit de informatica zijn we gericht op computationele *modellering* van de processen van spreken en verstaan, voor wetenschappelijke doeleinden maar ook met het oog op praktische toepassingen (spraaksynthese, automatische spraakherkenning, en automatische sprekerherkenning).

De fonetiek heeft ruwweg drie stadia doorlopen. In het eerste stadium, ongeveer tot de Tweede Wereldoorlog, waren *impressionistische* methoden het meest belangrijk. Spraakklanken werden primair beschreven en geanalyseerd op het gehoor, dat dan ook goed getraind moest worden om de subtiele verschillen tussen spraakklanken te beschrijven. Deze impressionistische en introspectieve methode werd wel ondersteund met ingenieuze mechanische instrumenten waarmee men spraak kon visualiseren en observeren. Dit stadium van de fonetiek wordt fraai verbeeld in het toneelstuk *Pygmalion* van George Bernard Shaw¹ (1912) en de verfilming daarvan als *My Fair Lady* (1964).

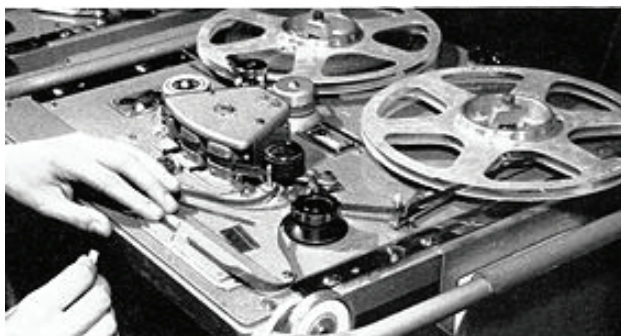
¹ G.B. Shaw (1912). *Pygmalion*. Beschikbaar via <http://www.gutenberg.org/files/3825/>. Zie ook Collins & Mees (1998).

Vanaf 1945 kwam er een grote verandering in de methoden van fonetisch onderzoek; ik beschouw dat als het tweede stadium. Fonetici bestudeerden vooral de *variatie* in fonetische realisaties, in allerlei talen en in allerlei contexten. Daaruit probeerde men af te leiden welke onderliggende, invariante patronen ten grondslag liggen aan al die variatie. De sleutelwoorden hierbij waren “variatie” en “invariantie”.

In het derde stadium, vanaf ongeveer 1980, worden *digitale* methoden en technieken gebruikt om spraak te registreren, te analyseren, en te modelleren. Ik kom daar later op terug. Eerst wil ik met U terugkijken naar die fonetische revolutie in het midden van de 20e eeuw.

BANDRECORDER

Die revolutie hangt namelijk nauw samen met de opkomst van de *bandrecorder*.



Figuur 1: EMI TR90 bandrecorder, waarbij de magnetische band geknipt en geplakt wordt in de daartoe bestemde mal (foto: <http://rfwilmut.net/broadcast/images/tr90.jpg>).

Al vanaf het midden van de 19e eeuw werkten onderzoekers aan apparatuur om spraak langs elektrische weg te transporteren, de zgn. telefoon. In de jaren '30 van de 20e eeuw ontwikkelde men in Duitsland het apparaat dat we kennen als de bandrecorder. Een bandrecorder

registreert geluiden door die geluiden om te zetten naar elektrische signalen, die weer worden vastgelegd als magnetische patronen, op een dunne band bekleed met minuscule metaaldeeltjes. Ik zal U de technische details besparen.

Wat waren nu de gevolgen van die bandrecorder? De gevolgen waren in ieder geval revolutionair voor de muzikale praktijk. Je kon er meerdere muziekinstrumenten en stemmen *afzonderlijk* mee opnemen, en vervolgens in de studio al die geluiden bewerken en vermengen tot nieuwe klankcomposities --- denk aan de Beatles², of Kraftwerk. De bandrecorder was ook de voorloper van de walkman en mp3-speler.

De bandrecorder leidde ook tot een revolutie in het onderzoek naar spraakklanken. Ineens werd het makkelijk om het geluid van spraakklanken, woorden en zinnen op te slaan voor nauwkeurige akoestische en perceptieve analyses. Dat kon bovendien betrouwbaar en relatief goedkoop. De nieuwe technische ontwikkeling maakte allerlei innovatief onderzoek mogelijk. Ik zal dat illustreren.

Met de bandrecorder kan het spraakgeluid ontleed worden tot in fracties van seconden, door de geluidsband te bewegen langs de weergavekop van de bandrecorder. Op die manier kon men ineens afzonderlijke spraakklanken afbakenen op geluidsband, de band uitknippen, en deze knipsels opnieuw combineren (U ziet dat op de foto). Zo experimenteerde men welke fonetische informatie van belang is voor de waarneming van gesproken woorden. We kunnen bijv. de [t] van het woord “thee” combineren met de [a] uit “laat” en de [l] uit “viel”. Het resultaat van deze combinatie van [t]+[a]+[l] blijkt dan echter heel anders te klinken dan een normaal uitgesproken versie van het woord “taal”. Het knip-en-plak-werk is zelfs onverstaanbaar! Het normale, vloeiende verloop tussen de spraakklanken is verstoord, en die vloeiende overgangen zijn blijkbaar essentieel om de spraak te verstaan (Harris, 1953). De fonetische aanwijzingen voor een spraakklank

2 The Beatles (1967). *Sgt. Pepper's Lonely Hearts Club Band*.

blijken niet alleen besloten te liggen in die klank zelf, maar ook in de omringende klanken. Normale spraak is wezenlijk anders dan een reeks van afzonderlijke spraakklanken: natuurlijke articulatie verloopt *vloeiend* van de ene klank naar de andere, en luisteraars houden daarmee ook rekening³. We kunnen die vloeiende bewegingen wel “geleden” naar afzonderlijke klanken⁴, maar dat is een taalkundige abstractie die tamelijk verwijderd staat van de fonetische werkelijkheid (Nooteboom, 2004). Normale spraak is dus niet te vergelijken met een gedrukte tekst, opgebouwd uit eenvormige losse letters. Spraak heeft nog het meeste weg van een snel en moeilijk leesbaar handschrift, dat vol zit met persoonlijke en contextuele eigenaardigheden.

Mijn punt bij dit voorbeeld is dat de bandrecorder het mogelijk maakte voor taalonderzoekers om nieuwe, nog niet eerder bedachte, en voorheen onmogelijk beantwoordbare vragen te stellen, en om zo tot geheel nieuwe inzichten te geraken.

- Welke klanken, woorden en zinsstructuren gebruiken sprekers van een dialect? Dialect-onderzoekers gingen daartoe met bandrecorders op pad in het land.
- In welke opzichten verandert de uitspraak van onze taal door de jaren heen? Daarvoor kunnen we oudere en recentere opnames met elkaar vergelijken (bijv. Quené, 2013).
- Hoe komen taalkundige structuren (zoals klemtonen, of woordgrenzen) tot uiting in verbonden spraak? Daarvoor kunnen we de geluidsregistraties vergelijken van spraak met contrastieve klemtonen, zoals “HARnas” vs. “komPAS”, of van spraak met contrastieve structuren, zoals “komPAS” zonder woordgrens vs. “kom # pas” met woordgrens (Quené, 1992).

3 Voor recent onderzoek naar perceptieve “compensation for coarticulation” zie bijv. Kang, Johnson & Finley (2016) en verwijzingen aldaar.

4 Het woord *articulatie* betekent “geleding”; het woord *artikel* betekent “lid” of “woord”.

In het spoor van de bandrecorder, nu zo'n 70 jaar geleden, volgden dus ook andere, meer kwantitatieve methoden van onderzoek. De eerder gebruikte methoden van onderzoek uit het eerste stadium (transcriptie, interpretatie, introspectie) raakten mettertijd op de achtergrond. De verspreiding van nieuwe kwantitatieve methoden en technieken bood nieuwe kansen voor wetenschappelijke vooruitgang in het onderzoek naar gesproken woorden en zinnen. Daarvoor moesten spraakonderzoekers trouwens wel iets leren over magnetische geluidsregistratie (bijvoorbeeld over bandsnelheid, magnetisering, en uitsturing) en over kwantitatieve en statistische methoden en technieken (bijvoorbeeld over kans en spreiding).

WAAROM

Verrijkt door dit voorbeeld wil ik met U terugkeren naar mijn stelling. Een belangrijke eerste vraag bij die stelling is natuurlijk: *waarom* hebben we die kwantitatieve methoden dan nodig? Het voorbeeld van de bandrecorder geeft aan, naar ik hoop, wat daarop mijn eerste antwoord is: een groter arsenaal van onderzoeksmethoden biedt ook grotere kansen op nieuwe en waardevolle inzichten.

Een tweede antwoord op de waarom-vraag is gelegen in de duizelingwekkende toename in het volume van data. Ik zal die toename met enkele citaten en cijfers illustreren (deels ontleend aan Kitchin, 2014, pp.69-74).

- “Between the dawn of civilisation and 2003, we only created five exabytes⁵ of information; now we’re creating that amount every two days” (Hal Varian, ca. 2012, geciteerd door Kitchin, 2014, p.xv).
- “The average person today processes more data in a single day than a person in the 1500’s did in an entire lifetime” (Smolan

5 1 exabyte is 10^{18} bytes.

& Erwitt, 2012).

- Meer dan 90% van alle data in de wereld zijn gegenereerd in de laatste 2 jaar (Helbing et al., 2017).
- Google ontvangt meer dan 40 000 zoekopdrachten, per seconde, oftewel 3.5 miljard per dag⁶; deze worden allemaal opgeslagen.
- De hoeveelheid opgeslagen⁷ data neemt toe met ongeveer 40% per jaar.
- Naar schatting is slechts 0.5% van alle data geanalyseerd.

De hoeveelheid data neemt dus explosief toe. De term “big data” is hierbij nog een “understatement”: de hoeveelheid van gegevens is niet gewoon groot, maar bijna *onvoorstelbaar* groot.

Bovendien zijn allerlei soorten van data op allerlei manieren aan elkaar verbonden (“linked data”), bijvoorbeeld in de vorm van *koppelingen* tussen een persoon, zijn of haar apparaat, locatie, koopgedrag, zoekopdrachten, enzovoort, enzovoort. Men zou permanent kunnen registreren wat U zegt, en tegen wie, wanneer, en waar, met behulp van data uit uw slimme telefoon en uit die van uw gesprekspartners. En wellicht gebeurt dat ook al. Het resultaat is een verregaande digitalisering van ons leven, en dus ook van de geestesvruchten (tekst, beeld, geluid) die bestudeerd worden in mijn Faculteit Geesteswetenschappen. We hebben kwantitatieve methoden hard nodig, om die onvoorstelbaar grote kwantiteit aan gegevens te destilleren tot kennis, begrip en inzicht.

De huidige digitale revolutie vertoont overeenkomsten met de elektromagnetische revolutie van de jaren '50. Maar er zijn ook verschillen. Een eerste verschil is dat een magnetische geluidsopname

6 Volgens <http://www.internetlivestats.com/google-search-statistics/> en https://www.domo.com/blog/wp-content/uploads/2014/04/DataNeverSleeps_2.0_v2.jpg.

7 Digitale data worden merendeels opgeslagen op harde magnetische schijven, met technologie ontleend aan de magnetische bandrecorder. De harde schijf is goed te beschouwen als een nazaat van de magnetische geluidsband. De hoeveelheid data groeit overigens sneller dan de opslagcapaciteit.

in de jaren '50 t/m '80 nog dermate bewerkelijk was dat je daar niet lichtzinnig aan begon. De eerste bandrecorders wogen zo'n 100 kilo, om maar een praktisch bezwaar te noemen. Tegenwoordig echter is een digitale registratie (in beeld en geluid) zeer makkelijk te maken en te verspreiden, ook als de geregistreerde personen dat niet wensen of niet weten. Een tweede verschil is de intensieve koppeling tussen digitale gegevens. Dat vergroot de mogelijkheden om verbanden te ontdekken. Anderzijds is er ook het reële gevaar dat die gekoppelde digitale gegevens een bedreiging vormen voor onze privacy, onze vrijheid van meningsuiting, onze keuzevrijheid (Wetenschappelijke Raad voor het Regeringsbeleid, 2016), en voor onze sociale omgangsvormen.

WAT

Naast de waarom-vraag zou U nog een tweede vraag kunnen stellen bij mijn stelling, nl. de vraag “*wat* zijn dan die kwantitatieve methoden”? Ik wil graag die vraag nog wat verbreden, en ook de aanpalende empirische methoden en statistische technieken in mijn antwoord betrekken. Deze empirische-kwantitatieve-statistische methoden-en-technieken hebben twee belangrijke kenmerken:

1. de gegevens worden uitgedrukt in numerieke vorm, als code of getal (kwantiteit);
2. de analyse van de gegevens is gebaseerd op principes van kansrekening, en maakt gebruik van statistische technieken.

Voor wat betreft het *eerste* kenmerk, de numerieke vorm van observaties: In het eerste stadium van de fonetiek werd een taal ontleed door te luisteren naar de woorden en klanken, en zo te expliciteren hoe sprekers die klanken produceerden, of zouden moeten produceren (zie voetnoot 1). In het elektromagnetische stadium van de fonetiek, daarentegen, werden vooral akoestisch-fonetische kenmerken gemeten in de geluidsopnamen, vanuit de gedachte dat “meten is weten”, en dat getallen objectiever zijn dan verbale beschrijvingen (Porter, 1995).

In het derde, digitale stadium geven we de computer een grote berg gekoppelde gegevens⁸ en we laten het aan een algoritme over om de onderliggende patronen op te sporen in die digitale gegevens.

Helaas ontbreekt vandaag de tijd om het *tweede* kenmerk, de statistische analyse, op gedegen wijze uit te leggen. We zullen ons behelpen met twee voorbeelden.

EERSTE VOORBEELD: UCU ENGLISH ACCENTS

Mijn eerste voorbeeld heeft betrekking op de spraakklank [s], in woorden als “student, semester, festival”. Het zal U niet verbazen dat er variatie of spreiding is in de uitspraak van die klank: soms wat scherper [s], soms wat doffer [s]. U denkt misschien, wat maakt zo’n [s] nou uit, we kunnen elkaar toch prima verstaan? Dat is zo, maar het accent waarmee iemand praat, en de wijze waarop een spreker [s] zegt, vormt desalniettemin een belangrijke sociale indicator. Denkt U maar aan het bijbelse verhaal over de Efraïmitische soldaten, die zichzelf bij hun terugtocht uit Kanaän verrieden door hun uitspraak. De Efraïmieten konden namelijk het woord [ʃ]ibbóleth niet goed uitspreken: zij spraken het fout uit als “[s]ibbóleth” (Rechteren 12:5-6). Het liep heel slecht af met deze soldaten. Ook nu nog worden sprekers sterk op hun accent beoordeeld⁹.

Zo vinden de meeste Nederlanders het heel erg, als een Nederlander Engels spreekt met een Nederlands accent (Van den Doel & Quené, 2013). Eén van de kenmerken van dat Nederlandse accent is de doffe uitspraak van de [s]. In de meeste varianten van het Engels is de [s] scherper van timbre dan de Nederlandse [s]. Dat verschil blijft meestal

8 Die gegevens kunnen bestaan uit opgenomen spraak, gekoppeld aan gevalideerde transcripties van diezelfde spraak, zoals in het *Corpus Gesproken Nederlands* (<http://tst-centrale.org/nl/tst-materialen/corpora/corpus-gesproken-nederlands-detail>).

9 Voor een illustratie hiervan, zie de video “Accenten en discriminatie” op de website van *Sprekend Nederland*, <http://www.dekennisvanu.nl/site/special/Sprekend-Nederland/17>.

onopgemerkt, en onverbeterd, omdat het inderdaad zelden leidt tot verwarring. Toch draagt het vermoedelijk wel bij aan de sociale indruk van een “ergerlijk” Nederlands accent in het Engels (of omgekeerd, aan de indruk van een “interessant” Engels accent in het Nederlands).

De gegevens die ik U wil presenteren zijn afkomstig uit een langdurig onderzoek, dat werd opgezet en uitgevoerd door wijlen mijn collega Rosemary Orr en mijzelf. We vragen ons af hoe het accent van een spreker verandert, als gevolg van aanpassing aan andere sprekers. Wij hebben daarvoor in Utrecht een vrijwel perfecte onderzoeksomgeving, nl. het Engelstalige University College Utrecht. Sinds 2010 hebben we bijna 1100 interviews afgenomen, en digitaal opgeslagen, van 283 studenten als insprekers. In elke opname-sessie hielden de studenten spontane monologen over informele onderwerpen, zoals reiservaringen, sport, campus life, enzovoort. Die monologen hielden ze zowel in hun moedertaal (indien dat niet Engels was) als in het Engels.

Voor dit voorbeeld wil ik me beperken tot een steekproef van 25 sprekers met alleen Nederlands als moedertaal, en Engels als tweede taal. Deze sprekers hebben heel vaak een [s] uitgesproken, in hun informele monologen, in het Nederlands en in het Engels. Al deze [s]’en zijn opgespoord met behulp van een automatische spraakherkenner¹⁰, en met de hand gecontroleerd --- mijn collega Rosemary Orr heeft zo ruim 5000 [s]-klanken gevalideerd. Van al deze klanken is het spectrale zwaartepunt gemeten, met behulp van scripts voor automatische digitale dataverwerking. Dat spectrale zwaartepunt (Centre of Gravity, afgekort COG) is een akoestische maat voor de scherpte of dofheid van een ruisklank als de [s]. Hoe scherper de [s] klinkt, hoe hoger het spectrale zwaartepunt ligt. We hebben dus metingen van dezelfde sprekers, uit meerdere opeenvolgende interviews, en steeds in twee talen.

10 Met dank aan David van Leeuwen voor het uitvoeren van deze automatische spraakherkenning m.b.v Kaldi. Voor meer informatie over de Kaldi herkenner, zie <http://kaldi-asr.org>.

Nu wordt het even wat ingewikkeld. We kunnen sprekers herkennen aan hun stem en aan hun uitspraak. Dat moet wel betekenen, dat er aanzienlijke spreiding moet zijn *tussen* sprekers, en dat er aanzienlijke *correlatie* moet zijn *binnen* sprekers, in de wijze waarop zij de spraakklanken uitspreken. Dat geldt ook voor de [s]. We moeten in de statistische analyse rekening houden met die correlatie binnen sprekers. Bovendien zijn er sterke effecten van de fonetische context (de naburige klanken), en ook daarmee moeten we rekening houden. Voor deze analyse heb ik me daarom beperkt tot [s]-klanken afkomstig uit 24 identieke woorden die voorkwamen in *zowel* de Nederlandse *als* de Engelse monologen: woorden als “student”, “semester”, of “festival”¹¹. Ik kijk daarbij naar de effecten van taal en tijdstip op de realisatie van de [s], *gegeven* een bepaalde spreker en *gegeven* een bepaald woord als fonetische context¹².

De kwantitatieve, statistische analyses leveren o.a. de volgende uitvoer op:

```
Note: Intercept: round=1, lang=en
(Intercept) b0= 5778  s.e.=310  95%CI=( 5202, 6392)  t(16)=18.641  p<.0001
lang.nl      b1=-1373  s.e.=310  95%CI=(-1957, -733)  t(16)=-4.429  p=.0002
```

Wat betekenen nu deze uitkomsten?

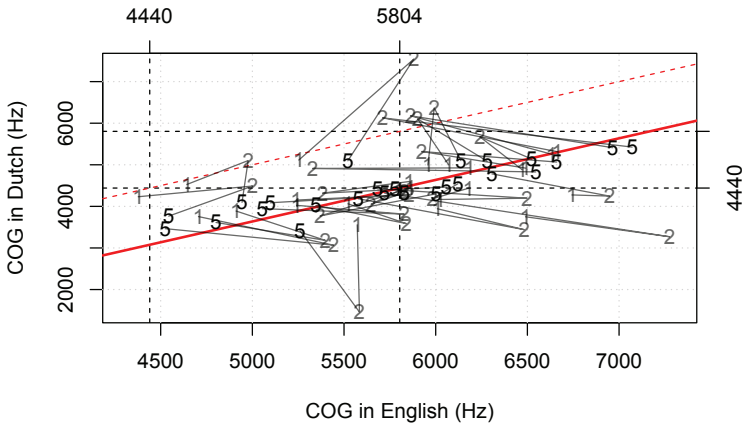
In het eerste interview is het spectrale zwaartepunt van de [s] in het Nederlands naar schatting zo'n 1373 Hz lager dan in het Engels (rekening houdend met spreker en woord). Die schatting van 1373 Hz heeft een foutmarge van 310 Hz: het werkelijke verschil zou ook wat groter of kleiner kunnen zijn, vanwege toevallige spreiding. Maar, het 95% betrouwbaarheidsinterval¹³ loopt hier van -1957 tot -733 Hz, en

11 Het betrof de volgende 24 woorden: “accent, basis, best, bus, campus, extra, festival, foto’s/photos, is, lustrum, plus, post, race, rest, school, semester, sport, stage, status, stop, student, test, was, west”.

12 Hiertoe heb ik een multilevel of mixed-effects model gebruikt, met sprekers en met woorden als twee gekruiste random effecten, en met random slopes over sprekers (Quené & Van den Bergh, 2008).

13 Het 95%-betrouwbaarheidsinterval is het interval dat statistisch gezien het ware verschil zou omvatten voor 95% van de herhaalde steekproeven.

dat omvat *niet* nul. Het ware verschil is dus vermoedelijk kleiner dan nul (Cumming, 2012). De kans om dit verschil te vinden, of een nog groter absoluut verschil, als er in werkelijkheid géén verschil is, bedraagt slechts $p=0.0002$ (of ruwweg 1:4500; Hox, 2010). We spreken hier dan van een *significant* verschil. Voor de andere relevante effecten, nl. van tijdstip (interviewnummer) en van de combinatie van tijdstip en taal, kunnen we soortgelijke cijfers in de statistische uitvoer inspecteren. Ik zal de gevonden patronen nu echter bespreken in visuele vorm, aan de hand van een figuur.



Figuur 2: Spreidingsdiagram van het gemiddelde spectrale zwaartepunt (COG) in Hz, in het Engels en in het Nederlands. Verbonden waarden (uit interviews 1, 2 en 5) zijn afkomstig van eenzelfde spreker.

Langs de horizontale as zien we het gemiddelde spectrale zwaartepunt van de [s] van een spreker die Engels praat (als tweede taal), en langs de verticale as zien we het corresponderende zwaartepunt van diezelfde spreker als die Nederlands praat (als moedertaal). De cijfers geven het volgnummer van het interview aan (1 of 2 of 5). Als sprekers géén verschil zouden maken tussen hun Engelse en Nederlandse [s], dan

zouden de observaties rond de bovenste diagonale stippellijn moeten liggen. In het eerste interview, aan het begin van hun eerste studiejaar, is dat voor drie van de 25 sprekers inderdaad het geval: die 3 sprekers maken dus geen contrast tussen hun Engelse en Nederlandse [s]. (Een opmerking terzijde: ik ben uiteraard van mening dat bij ieder meetpunt ook bijbehorende spreiding aangegeven dient te worden. Toch heb ik deze spreiding hier niet gevisualiseerd, om de figuur niet nog complexer te maken.). Op tijdstip 1 maken de meeste sprekers wél een contrast, aangegeven met de dikke rode lijn. Zoals gezegd, het spectrale zwaartepunt in het Nederlands (verticale as) is 1373 Hz lager dan in het Engels. De meeste sprekers beheersen dus het verschil tussen de Engelse en Nederlandse [s] al goed op het moment dat ze aankomen op UCU.

Dat patroon wordt echter erg rommelig als we kijken naar het tweede interview, aan het einde van het eerste studiejaar. Sommige sprekers vergroten het verschil, anderen verkleinen het juist. Sommigen gaan een overdreven scherpe [s] in het Engels maken, en anderen gebruiken hun scherpe Engelse [s] ineens ook als ze Nederlands praten. Kortom, grote variatie tussen en binnen sprekers.

Op het laatste tijdstip, aan het einde van het laatste studiejaar, zien we juist weer een sterke convergentie. Bijna alle sprekers maken een systematisch onderscheid tussen hun [s]’en in het Engels en die in het Nederlands. De spreiding tussen de sprekers in de steekproef is afgenomen. Bijna alle sprekers gebruiken in beide talen de passende variant van de [s]. In de toekomst zullen we nader onderzoeken, met behulp van onze opnames, wat er in de tussenliggende twee jaren is gebeurd met de uitspraak van het Engels en van het Nederlands, en welke factoren daarbij van invloed zijn geweest.

Wat hebben we nu geleerd van dit voorbeeld van de “UCU English Accents”?

- Ten eerste zien we hoe je op grond van numerieke observaties, en met gebruik van principes van kansrekening, kunt komen

tot empirisch onderbouwde conclusies over accent en uitspraak.

- Ten tweede: De koppeling van data, in dit voorbeeld over meerdere interviews en over meerdere talen per interview, die koppeling verschaft inzicht in de longitudinale ontwikkeling van de accenten per persoon, en in de (hier niet besproken) *individuele factoren* die in die ontwikkeling een rol spelen.
- Ten derde, we hebben expliciet rekening gehouden met de spreiding tussen proefpersonen en tussen woorden, en met de correlaties binnen proefpersonen en binnen woorden. Daardoor is er een grotere kans dat onze conclusies valide en robuust zijn, en dat deze conclusies ook generaliseerbaar zijn naar andere proefpersonen en andere woorden. Ik wil dit leerpunt graag verbinden met de titel van mijn rede, en met mijn eerdere stelling: rekening houden met spreiding, biedt betere kansen op valide en robuuste inzichten.

EERSTE TEGENWERPING

Dames en heren, tijd voor een korte tussenstand. Ik heb geprobeerd U ervan te overtuigen dat kwantitatieve methoden onontbeerlijk zijn in de geesteswetenschappen. Daarbij ben ik ingegaan op de vraag *waarom* die methoden onontbeerlijk zijn, en op de vraag wat die methoden dan behelzen.

We zijn nu aangekomen bij het onderdeel van mijn rede dat in de klassieke retorica bekend staat als *refutatio*, het weerleggen van de tegenargumenten.

Het eerste tegenargument is samen te vatten als dit zoekpatroon:

Wij doen dat al (goed)?(genoeg)?

Dit zoekpatroon beschrijft vier varianten van dit tegenargument:

Wij doen dat al.
Wij doen dat al goed.
Wij doen dat al genoeg.
Wij doen dat al goed genoeg.

Het is zeker waar dat er in diverse plaatsen in onze Faculteit gewerkt wordt met digitale gegevens, met statistische technieken, met kwantitatieve methoden, enzovoort. Dat geldt onder andere voor sommig onderzoek naar taal en spraak, naar sociaal-economische geschiedenis, naar gebruik van nieuwe digitale media, naar kunstmatige intelligentie. Toch denk ik dat het vaak nog niet genoeg is, en niet goed genoeg is.

Eén van de problemen die ik constateer in veel empirisch kwantitatief onderzoek, is het alarmerend lage onderscheidend vermogen, d.w.z., onvoldoende “power”. *Power* is een statistisch concept: het is de kans om de nulhypothese te verwerpen, als de nulhypothese inderdaad niet waar is. Het is dus de kans om als onderzoeker gelijk te *krijgen* (een significant resultaat te vinden) als je ook gelijk hebt. In het vorige voorbeeld is dat dus de kans om een verschil in spectraal zwaartepunt op *te sporen*, als dat verschil er ook *is* in de populatie. De power hangt samen met de spreiding in observaties. Naarmate de [s]’en meer van elkaar verschillen, wordt de power van mijn analyse lager. Veel onderzoekers *onderschatten* echter die toevallige spreiding. De kans op *toevallige spreiding* in de observaties is doorgaans veel groter dan ingeschat door de onderzoeker. We komen daarmee bij de tweede interpretatie van de titel van deze rede: De kansen van spreiding zijn doorgaans groter dan gedacht. Dat leidt dan tot een empirische, kwantitatieve studie die eigenlijk te klein is: te weinig proefpersonen, te weinig observaties per proefpersoon, te weinig power. Dat soort onderzoek heeft weinig wetenschappelijke waarde. Als we wèl een positief resultaat vinden, zou dat ook spurieus kunnen zijn (een Type-I-fout). En als we niet een positief resultaat vinden, dan zou dat ook kunnen komen omdat de studie niet gevoelig genoeg was (een Type-II-fout). Dit alles wordt nog verergerd doordat we ons vaak bezig houden met relatief zwakke effecten. Deze problemen zijn op heldere

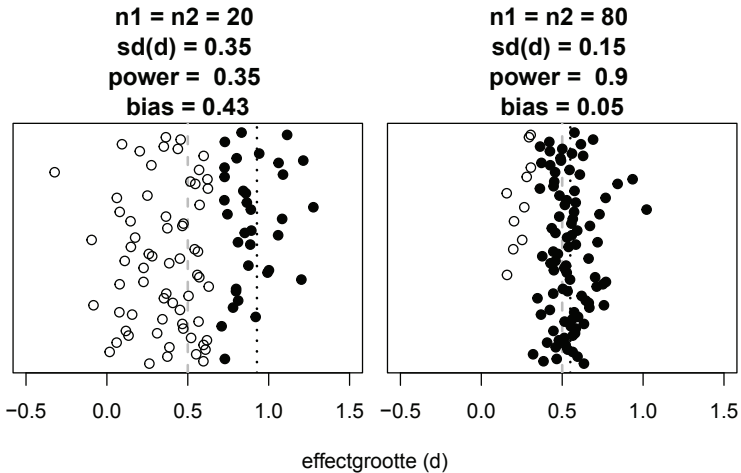
wijze verwoord in een artikel getiteld “Why most published research findings are false” (Ioannidis, 2005; i.h.b. Corrolaries 1 and 2), dat ik mijn kwantitatief werkende collega’s zeer wil aanbevelen.

Wat zijn nu de nadelen van een te geringe power? Is dat zo ernstig? Ik meen van wel. Ten eerste wordt onze wetenschappelijke vooruitgang belemmerd, omdat we falen in het opsporen van daadwerkelijk bestaande effecten, of verbanden, of verschillen.

Ten tweede geven herhaalde studies, bij onvoldoende power, steeds wisselende uitkomsten te zien. Gepubliceerde bevindingen kunnen niet goed herhaald worden in replicatie-onderzoek. Enkele jaren geleden zijn 100 toonaangevende psychologische experimenten zo goed mogelijk herhaald door een groot team van onderzoekers (Open Science Collaboration, 2015). In theorie zou je bij ongeveer 89 van de herhaalde studies ongeveer dezelfde uitkomst verwachten. Echter, in werkelijkheid werd bij slechts 47 van de herhaalde studies *ongeveer* hetzelfde resultaat gevonden. Van deze 100 toonaangevende studies kon dus slechts de *helft* gerepliceerd worden, als we de term “gerepliceerd” ruimhartig interpreteren. Er is helaas geen reden om te vermoeden dat deze “replication crisis” aan de geesteswetenschappen voorbij zou gaan.

De wisselende uitkomsten in herhaalde experimenten leiden typisch tot vervolg-onderzoek om de discrepantie te verklaren. Team A vindt wel een significant resultaat, en team B niet. Sommigen van ons hebben daar ervaring mee. We gaan dan nader onderzoeken waarom het effect wel optrad in de studies van team A, en niet in die van team B. Zou dat komen door verschil in proefpersonen? in stimuli? taken? instrumentatie? Al dat vervolg-onderzoek is echter overbodig: de verschillen in uitkomsten van teams A en B zijn meestal goed te verklaren door de onvoldoende power in hun onderzoeken. Al dat vervolg-onderzoek kost ondertussen veel tijd, geld, moeite, plus allerlei immateriële kosten. Eén goed ontworpen studie met ruim voldoende power had al dat nodeloze vervolg-onderzoek kunnen voorkomen.

Het derde nadeel is wederom wat ingewikkeld om uit te leggen; het hangt samen met mijn vorige punt. Studies met onvoldoende power leiden tot *overschatting* van de onderzochte effecten. In een studie met onvoldoende power, moet een effect immers tamelijk groot zijn om statistisch significant te zijn. Ik zal dat proberen te illustreren in deze figuur, die ik heb gebaseerd op 100 computersimulaties per paneel.



Figuur 3: Gevonden effectgroottes (langs horizontale as) in 100 gesimuleerde experimenten (tweezijdige t-toets voor onafhankelijke waarnemingen, $\alpha=.05$), uitgesplitst naar steekproefgrootte (links $n=20$, rechts $n=80$) en naar toetsingsresultaat (donkere symbolen: wel significant; lichte symbolen: niet significant).

In alle simulaties is het ware effect gelijk aan $d=0.50$ (op een willekeurige schaal), aangegeven met de grijze stippellijn. In het linker paneel zien we 100 replicaties van onderzoek met te kleine steekproeven en onvoldoende power. In dat linker paneel, met slechts $n=20$ proefpersonen per groep, wordt het ware effect slechts $35\times$ opgespoord, en $65\times$ gemist (de power is dus slechts 35%). De gemiddelde grootte van die 35 significante uitkomsten is 0.93 (zwarte stippellijn), dat wil zeggen dat de significante resultaten een forse vertekening of ‘bias’ geven van 0.43 ten opzichte

van het ware effect van 0.50 ¹⁴. Geen wonder, dus, dat Team B dat vertekende resultaat vervolgens niet kan repliceren!

In het rechter paneel zien we de situatie met voldoende grote steekproeven ($n=80$ proefpersonen), en voldoende power. Het ware effect wordt nu $90\times$ opgespoord, en wordt slechts $10\times$ gemist. De gemiddelde effectgrootte van deze 90 significante uitkomsten (0.55) wijkt weinig af van het ware effect. Kortom, we moeten terdege rekening houden met de aanzienlijke toevallige spreiding in een steekproef, om valide conclusies te kunnen trekken uit de kwantitatieve analyses (zie ook Button et al., 2013; Loken & Gelman, 2017).

Er is nòg een vierde, maatschappelijke reden waarom onderzoekers moeten zorgen voor sterk empirisch onderzoek met voldoende statistische power. Als onderzoekers steeds wisselende resultaten vinden, met soms wel en soms niet significante uitkomsten, en met grote variatie in de gerapporteerde effecten, dan leidt dat tot maatschappelijke *onverschilligheid* jegens wetenschappelijke kennis: ik citeer, “wetenschap is ook maar een mening”, einde citaat. De samenleving meent dan, en niet ten onrechte, dat de wetenschappelijke evidentie voor het onderzochte effect niet sterk is, dat de onderzoekers het er nog niet over eens zijn, enzovoort. Dat belemmert de maatschappelijke toepassing van, en waardering voor wetenschappelijke inzichten, en het schaadt het aanzien van de wetenschap in het algemeen. Onderzoekers moeten dus zorgen voor methodologisch sterke studies, met ruim voldoende statistische power, in overeenstemming met de academische beginselen van zorgvuldigheid en verantwoordelijkheid (VSNU, 2014). Sterke kwantitatieve methoden zijn onontbeerlijk.

Soortgelijke problemen doen zich overigens ook voor bij exploratief of toetsend onderzoek dat gebaseerd is op correlaties. Ook in correlatieve onderzoek zijn de steekproeven vaak te klein, en/of het aantal variabelen

14 De experimenteel gevonden effecten vertonen bovendien grote spreiding, en zijn soms zelfs negatief in plaats van positief.

is te groot. Een gevonden correlatie zou dan ook *spurious* kunnen zijn, d.w.z., er is grote kans dat de gevonden correlatie niet reëel is, maar een artefact is van toevallige spreiding. Omwille van de tijd moet ik dit punt echter verder laten rusten.

TWEEDE TEGENWERPING

Het is immers hoogste tijd voor de tweede tegenwerping:

In mijn vakgebied hebben we die kwantitatieve methoden helemaal niet nodig.

Het is waar dat veel van mijn collega's hun empirische materiaal benaderen op een niet-kwantitatieve manier: zij *interpreteren* hun empirisch materiaal, in plaats van dat zij daaraan meten, tellen en rekenen. We zien deze kwalitatieve, interpretatieve methode bij de disciplines die zich bezighouden met de kunsten, met geschiedenis, en met filosofie en religie. Ik denk echter, dat de ervaringen uit het verleden geen garantie bieden dat dit in de toekomst zo zal blijven. De digitale revolutie, die nu gaande is, zal ook deze hermeneutisch georiënteerde vakgebieden raken. De primaire empirische materialen, de teksten, boeken, foto's, en beelden, worden immers allemaal digitaal. De 'dataficatie' of 'informatisering' of 'digitalisering' zal alle disciplines van de geesteswetenschappen sterk beïnvloeden¹⁵. Ik ben ervan overtuigd dat deze revolutie grote kansen biedt voor nog onvermoede vragen en voor nieuwe inzichten, ook of misschien zelfs juist in deze meer hermeneutisch georiënteerde disciplines. Het is dan wel van belang om de computationele en kwantitatieve methoden en technieken goed en zinnig in te zetten, met een kritisch oog voor hun mogelijkheden en beperkingen.

Er is nog een tweede weerlegging voor deze tegenwerping. Ook wie

15 Zie voor soortgelijke analyses ook Clariah, www.clariah.nl.

kwantitatieve methoden gebruikt, moet de gevonden statistische patronen wél kunnen duiden en interpreteren. Dat speelt vooral een rol bij niet-experimenteel onderzoek, waarbij frequent gebruik gemaakt wordt van correlatie, regressie, en daarop gebaseerde technieken. We vinden dan bijvoorbeeld statistische verbanden, die vervolgens geïnterpreteerd moeten worden. Een goede interpretatie vereist uiteraard expertise in het vakgebied waarop de gegevens betrekking hebben. Maar een goede interpretatie vereist ook hermeneutische en interpretatieve vaardigheden van een deskundige. De “harde” kwantitatieve analyses gaan uiteindelijk hand in hand met de “zachte” interpretaties van de uitkomsten, en deze twee zullen elkaar in de toekomst zeer nodig hebben.

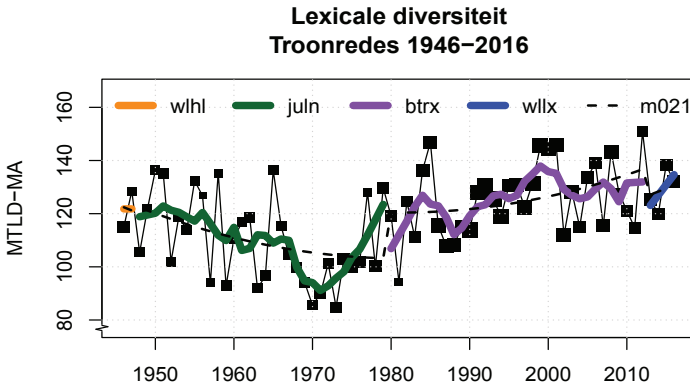
TWEEDE VOORBEELD

Ik wil mijn weerleggingen graag illustreren met een ander voorbeeld, dat betrekking heeft op een ander soort redes, nl. de Troonredes uitgesproken door ons staatshoofd. Het empirisch materiaal voor dit voorbeeld bestaat uit alle naoorlogse Troonredes,¹⁶ van 1946 tot en met 2016. Dit voorbeeld kunt U beschouwen als een wat losse illustratie van de wijze waarop kwantitatieve methoden een rol kunnen spelen in andere domeinen dan de taalwetenschap.

Laten we eens kijken naar de woordenschat van de Troonredes. De rijkdom van de woordenschat van een tekst kunnen we uitdrukken als de zogeheten “type/token-ratio” (TTR). Dat is de verhouding tussen unieke woorden (types) en alle woorden (tokens) in een tekst. Deze maat is echter gevoelig voor de lengte van de tekst, en die verschilt nogal tussen de Troonredes. Ik heb daarom een afgeleide maat gebruikt, afgekort als MTLDMA (dat staat voor Measure of Textual Lexical Diversity with Moving Average, hier berekend met factor 0.69; McCarthy & Jarvis, 2010).

16 De teksten van de Troonredes zijn ontleend aan www.troonredes.nl, <https://www.parlement.com/> en <https://nl.wikipedia.org/wiki/Troonrede>.

Nu ben ik geïnteresseerd in de historische ontwikkelingen in die Troonrede. Werden de Troonredes gaandeweg minder gevarieerd in woordkeus, of juist meer gevarieerd? Zijn er verschillen tussen staatshoofden? We kunnen die vragen beantwoorden door statistische analyse van de gemeten lexicale diversiteit van de Troonredes.



Figuur 4: Lexicale diversiteit (MTLD-MA, berekend met factor 0.69) van de Troonredes van 1946 tot en met 2016, met rechte (zwart) en versmeerde (gekleurd per staatshoofd) verbindingslijnen, en met gemodelleerde waarden (onderbroken lijn).

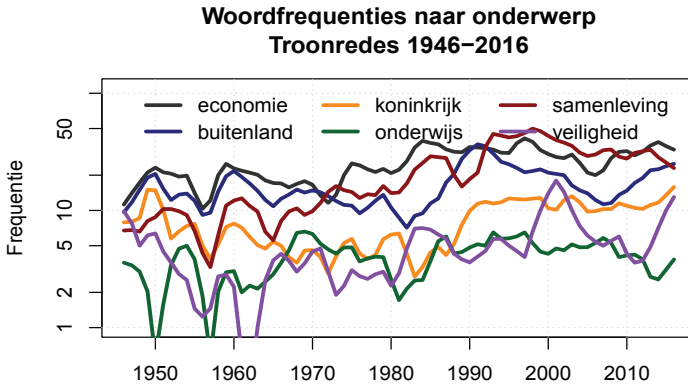
De resultaten laten twee trends zien in het gekozen tijdvak. Ten eerste is er een concaaf (d.w.z. hol) patroon: de lexicale diversiteit neemt af in de jaren tot ca 1980; daarna neemt de lexicale diversiteit weer toe¹⁷. Ten tweede is er een significante toename in lexicale diversiteit in de Troonredes voorgelezen door Koningin Beatrix¹⁸.

Beide trends vragen om nadere uitleg. Welke woorden of onderwerpen

17 Statistisch gezien is het concaaf patroon te beschrijven met een kwadratische regressiecoëfficiënt (met jaren gecentreerd rond 1981), $\beta = +54.7$ (s.e. 17.1), $t(15) = 3.19$, $p = .0022$.

18 ten opzichte van Koningin Juliana als referentie, $\beta = +17.3$ (s.e. 6.8), $t(15) = 2.53$, $p = .0138$.

kwamen “er bij” in de Troonredes vanaf 1980? Laten we daarvoor eens kijken naar de frequenties van woorden, ingedeeld naar onderwerp.



Figuur 5: Woordfrequenties voor 6 belangrijkste onderwerpen in de Troonredes van 1946 tot en met 2016.

We zien dat de hoeveelheid woorden besteed aan onderwijs (groen) nauwelijks verandert door de decennia. Maar vanaf de jaren '80 is er wel een groeiende aandacht voor de economie (grijs) en voor het buitenland (blauw, inclusief Europese eenwording). Een gestage en significante stijging zien we echter vooral bij het onderwerp “samenleving” (roodbruin, met woorden als “burgers, samenleving, maatschappij”, enz). In de jaren '90 werd de samenleving zelfs belangrijker dan de economie, althans in de Troonredes. Opvallend is ook de afname van het onderwerp “veiligheid” in de jaren '50 en '60. Deze statistische trends vragen om verdere duiding en interpretatie, die verder ontwikkeld moet worden in overleg met historici, politicologen, en andere deskundigen. Ook verdere nieuwe onderzoeksvragen, die weer worden opgeroepen door deze statistische analyses, vereisen zulke interdisciplinaire samenwerking.

Met dit simpele voorbeeld van “text mining” en statistische modellering hoop ik aangetoond te hebben, zij het schetsmatig, dat kwantitatieve methoden kunnen leiden tot nieuwe inzichten en tot nieuwe onderzoeksvragen, die ons nopen tot interdisciplinaire samenwerking.

HOE

Tenslotte wil ik kort aangeven, *hoe* we deze kwantitatieve methoden en statistische technieken kunnen inzetten voor onderzoek en onderwijs. Er zijn daarvoor nog wel belemmeringen, die we eerst zullen moeten wegnemen.

- Ten eerste worden de digitale en kwantitatieve inspanningen in onze Faculteit geremd door de versnippering over diverse teams en departementen. Ik ben dan ook verheugd dat de Faculteit heeft besloten om al deze inspanningen te bundelen in een nieuw op te richten *Utrecht Centre for Digital Humanities*. De taak om leiding te geven aan dit nieuwe centrum is aan mij toebedeeld. Ook hier is de titel van mijn rede van toepassing: spreiding van methoden en van perspectieven biedt ons meer kans op succes.
- Ten tweede: Docenten en onderzoekers melden een grote behoefte aan bijscholing op het gebied van mijn leerstoel. Ik wil hierin voorzien door middel van trainingen. Collega-onderzoekers uit alle geesteswetenschappelijke domeinen kunnen zich dan verdiepen in computationele en statistische technieken, zowel theoretisch als toegepast op hun eigen onderzoekspraktijk.
- Ten derde: Niet alleen onderzoekers, maar *al* onze afgestudeerden zouden getraind moeten zijn in kwantitatieve methoden. Waarom? Omdat *kwantitatieve geletterdheid* deel moet gaan uitmaken van de “transferrable skills” van onze afgestudeerden. Kwantitatieve en digitale geletterdheid zullen mijns inziens essentiële vaardigheden zijn voor de

arbeidsmarkt van 2020 en daarna, voor al onze alumni. Voor dit doel zullen we per september 2017 twee bachelor cursussen aanbieden, *Methoden en Statistiek 1* (vanuit Taalwetenschap) en 2 (een nieuwe cursus vanuit Dept Talen, Literatuur en Communicatie). Deze cursussen zullen toegankelijk zijn voor alle studenten Geesteswetenschappen. Uiteraard blijf ik daarnaast streven naar verdere versterking van het methoden-onderwijs binnen alle GW-opleidingen, zowel bachelor als master.

CONCLUSIE

Dames en heren, wat maakt ons tot mensen? En wat blijft daar nog van over, na de digitale revolutie? De geschiedenis van de bandrecorder heeft ons getoond dat nieuwe technieken leiden tot nieuwe onderzoeksvragen en nieuwe methoden. Met de digitale revolutie rijzen nu nog weer nieuwe soorten van onderzoeksvragen op in de geesteswetenschappen. Digitale en kwantitatieve methoden zijn nodig om die nieuwe vragen zinnig te beantwoorden. Ik hoop U ervan te hebben overtuigd dat we verschillende methoden nodig hebben om verder te komen: spreiding biedt immers grotere kansen op vooruitgang. De eerste stappen in deze richting zijn al lang geleden gezet, door volhardende statistische en computationele pioniers. Mijn leerstoel vormt een mijlpaal langs deze weg. Ik hoop dat die mijlpaal ook een kruispunt markeert, waar verschillende wegen tot inzicht samenkomen.

DANKWOORD

Dames en heren, graag wil ik deze inaugurele rede afsluiten met het uitspreken van woorden van dank. Mijn dank gaat vooreerst uit naar het College van Bestuur van de Universiteit en naar het Faculteitsbestuur van de Faculteit Geesteswetenschappen, voor het inrichten van deze nieuwe leerstoel, en voor het in mij gestelde vertrouwen om deze te

bekleden. Mijn dank betreft in het bijzonder het voormalige hoofd van mijn Departement Talen, Literatuur en Communicatie, Ted Sanders, en de onderzoeksdirecteur van het Utrechts instituut voor Linguïstiek OTS, Frank Wijnen, voor al jullie inspanningen in de afgelopen jaren.

Mijn dank gaat ook uit naar mijn leermeesters: naar mijn promotor Toon Cohen, en naar diens unieke opvolger Sieb Nooteboom. Hoewel Sieb en ik sterk verschillen van elkaar, zijn er ook overeenkomsten in onze achtergronden en interesses. We hebben de afgelopen jaren met succes samen onderzoek gedaan naar de vraag hoe mensen zich verspreken, en hoe ze die versprekingen weer repareren (of niet repareren), en naar de vraag in hoeverre sprekers daarin onderling verschillen. Juist *doordat* Sieb en ik ook onderling verschillen, heb ik al bijna 30 jaar veel geleerd van Siebs inzicht en ervaring. Ik ben je daarvoor zeer dankbaar, en ik zie uit naar een lange en vruchtbare verdere samenwerking en vriendschap. Een tweede collega-onderzoeker die ik graag dank is Rosemary Orr, die afgelopen november helaas veel te vroeg is overleden. Zonder haar enthousiasme en inzet zou ons project over de UCU English Accents, waarover ik U verteld heb, nooit van de grond gekomen zijn. Ook dank ik haar man, David van Leeuwen, voor zijn steun en assistentie bij ons onderzoek.

Ook spreek ik mijn dank uit naar mijn mede-onderzoeker Huub van den Bergh, voor zijn constante inspiratie bij het propageren van zuivere empirische methoden en geschikte kwantitatieve technieken, ook in het geesteswetenschappelijk onderzoek.

Graag wil ik ook mijn voormalige en huidige studenten danken, alsook mijn promoti (Noor, Saskia, Esther, Maya, en Hans Rutger), mijn huidige promovenda (Anne), en mijn mede-promotores (Sieb, Ted, Nivja, en Jos) voor jullie nieuwsgierigheid en voor de inspiratie die jullie mij boden en bieden als docent, als begeleider, en als onderzoeker. Het motto van de Universiteit Utrecht is “bright minds, better future”. De *nieuwsgierigheid* van jullie heldere geesten zie ik als de brandstof voor de wetenschappelijke motor, die onze wereld moet brengen naar die betere toekomst.

Ook mijn vele collega's ben ik zeer dankbaar. Die dank betreft niet alleen mede-docenten en mede-onderzoekers, maar ook de vele ondersteunende medewerkers in de onderwijs-organisatie, in onderzoekslaboratoria, in de bibliotheek, in beleidswerk, en op de secretariaten. Jullie werken allemaal met buitengewoon grote inzet, voor het beste onderwijs en het beste onderzoek, waarvoor veel dank.

Daarnaast wil ik dank uitspreken naar mijn mede-bestuurders in het Departement, in het bijzonder Kiene Brillenburgh Wurth en Els Stronks, alsook in de Faculteit, in het bijzonder Peter Schrijver en de mede-onderwijsdirecteuren. Ik verheug me erop om de komende tijd samen verder te werken aan goed, en waar mogelijk nòg beter onderwijs en onderzoek.

Van een andere aard is de dank die ik wil uitspreken aan wijlen mijn ouders: mijn vader Theo Quené, en mijn moeder Ineke Quené geboren Boterenbrood. Het doet me groot verdriet dat zij beiden te vroeg zijn overleden om mijn benoeming als hoogleraar te kunnen meemaken. Dat ik vandaag hier sta, heb ik in hoge mate te danken aan de opvoeding die ik van hen genoten heb: zowel genuanceerd als wetenschappelijk verantwoord.

Mijn kinderen, Jeanine, Tobias en Vincent, ben ik diep dankbaar voor wie ze zijn, en voor alles wat jullie iedere dag in mijn leven brengen: verfrissende ideeën, kritische reflectie, en onmetelijke vreugde.

De meeste dank komt toe aan mijn liefste Karin. Al ruim 35 jaar ben jij naast me en bij me. Als enige heb jij vandaag niets nieuws gehoord. Ik dank je van ganser harte voor alles wat ons verbindt, en ik verheug me op het vervolg.

Ik heb gezegd.

BIBLIOGRAFIE

- Button, K. S., Ioannidis, J. P. A., Mokrysz, C., Nosek, B. A., Flint, J., Robinson, E. S. J., & Munafò, M. R. (2013). Power failure: why small sample size undermines the reliability of neuroscience. *Nature Reviews: Neuroscience*, 14(5), 365–376. doi:10.1038/nrn3475
- Collins, B. S., & Mees, I. M. (1998). *The Real Professor Higgins: The life and career of Daniel Jones*. Berlin: De Gruyter.
- Cumming, G. (2012). *Understanding The New Statistics*. New York: Routledge.
- van den Doel, R., & Quené, H. (2013). The endonormative standards of European English: Emerging or elusive? *English World-Wide*, 34(1), 77–98. doi:10.1075/eww.34.1.04van
- Harris, C. M. (1953). A study of the building blocks in speech. *J. Acoustical Society of America*, 25(5). doi:10.1121/1.1907227
- Helbing, D., Frey, B.S., Gigerenzer, G., Hafen, E., Hagner, M., Hofstetter, Y., et al. (2017). Will democracy survive big data and artificial intelligence? *Scientific American*, 25 Feb 2017. Beschikbaar via <https://www.scientificamerican.com/article/will-democracy-survive-big-data-and-artificial-intelligence/>
- Hox, J. J. (2010). *Multilevel Analysis: Techniques and Applications* (2nd ed.). Mahwah, NJ: Lawrence Erlbaum.
- Ioannidis, J. (2005). Why most published research findings are false. *PLoS Medicine*, 2(8), e124. doi:10.1371/journal.pmed.0020124
- Kang, S., Johnson, K., & Finley, G. (2016). Effects of native language on compensation for coarticulation. *Speech Communication*, 77, 84–100. doi:10.1016/j.specom.2015.12.005
- Kitchin, R. (2014). *The Data Revolution: Big data, open data, data infrastructures & their consequences*. London: Sage.
- Loken, E., & Gelman, A. (2017). Measurement error and the replication crisis. *Science*, 355(6325), 584–585. doi:10.1126/science.aal3618
- McCarthy, P. M., & Jarvis, S. (2010). MTLID, vocd-D, and HD-D: A validation study of sophisticated approaches to lexical diversity assessment. *Behavior Research Methods*, 42(2), 381–392. doi:10.3758/brm.42.2.381
- Nooteboom, S. G. (2004). *Waar komen de letters van het alfabet vandaan? Een verlate reactie op het college “Alfabetiek”, gegeven door prof.dr. H. Mol in 1964/1965*. Afscheidscollege 23 april 2004, Universiteit Utrecht.
- Open Science Collaboration (2015). Estimating the reproducibility of psychological science. *Science*, 349(6251), aac4716. doi:10.1126/science.aac4716
- Porter, T. (1995). *Trust in Numbers: The pursuit of objectivity in science and public*

- life*. Princeton, NJ: Princeton University Press.
- Quené, H. (1992). Durational cues for word segmentation in Dutch. *J. Phonetics*, 20(3), 331-350.
- Quené, H. (2013). Longitudinal trends in speech tempo: The case of Queen Beatrix. *J. Acoustical Society of America*, 133(6), EL452-EL457. doi:10.1121/1.4802892
- Quené, H., & Van den Bergh, H. (2008). Examples of mixed-effects modeling with crossed random effects and with binomial data. *Journal of Memory and Language*, 59(4), 413-425. doi:10.1016/j.jml.2008.02.002
- Smolan, R., & Erwit, J. (2012). *The Human Face of Big Data*. New York: Sterling.
- VSNU. (2014). *Nederlandse Gedragcode Wetenschapsbeoefening*. Beschikbaar via <http://www.vsnu.nl/gedragcodes>
- Wetenschappelijke Raad voor het Regeringsbeleid (2016). *Big Data in een vrije en veilige samenleving* (Rapport nr. 95). Amsterdam: Amsterdam University Press. Beschikbaar via <https://www.wrr.nl/publicaties/rapporten/2016/04/28/big-data-in-een-vrije-en-veilige-samenleving>

CURRICULUM VITAE



Hugo Quené (1960) studeerde eerst Nederlands, en behaalde daarna zijn doctoraal in de Fonetiek aan de RU Utrecht. Voor zijn proefschrift uit 1989 onderzocht hij de productie en perceptie van woordgrenzen in verbonden spraak. Naast dit onderzoek ontwikkelde hij ook taalkundige analyses en prosodische regels voor Nederlandstalige spraaksynthese, en werkte hij als docent bij de RU Leiden. In 1989 werd hij aangesteld als universitair docent bij de opleiding Fonetiek aan de RU

Utrecht, tevens verbonden aan het programma Taal- en Cultuurstudies (1999-2006). In 2001-2002 werkte hij als Fulbright Visiting Senior Scholar aan de Indiana University, Bloomington, Indiana, U.S.A., aan onderzoek naar de productie en perceptie van ritme in spraak. Van 2004 tot 2016 was hij universitair hoofddocent bij de opleiding Taalwetenschap van de UU. Per 1 augustus 2016 is hij benoemd als hoogleraar Kwantitatieve Methoden van Empirisch Onderzoek in de Geesteswetenschappen.

In diverse projecten onderzoekt Hugo Quené de wisselwerkingen tussen spraakproductie en spraakperceptie: hoe beïnvloeden de Engelse accenten van moedertaalsprekers en niet-moedertaalsprekers elkaar (met Rosemary Orr †), hoe worden eigen versprekingen door een spreker opgespoord en eventueel gerepareerd (met Sieb Nooteboom), hoe dragen vloeiende beurtwisselingen in een dialoog bij aan affectieve oordelen over de sprekers (met Jos van Berkum en Anne van Leeuwen)? En hoe worden de wisselwerkingen tussen spraakproductie en spraakperceptie beïnvloed door te glimlachen tijdens het spreken? Daarnaast verricht hij onderzoek en verzorgt hij onderwijs op het gebied van kwantitatieve empirische methoden en statistische technieken (met Huub van den Bergh).

Sinds 2014 is hij tevens onderwijsdirecteur van de Undergraduate School voor Talen, Literatuur en Communicatie, en vanaf 2017 zal hij ook leiding geven aan het nieuwe Utrecht Centre for Digital Humanities.

COLOFON

Copyright: Hugo Quené, 2017

Vormgeving: Communicatie & Marketing, faculteit
Geesteswetenschappen, Universiteit Utrecht

Afbeelding titelpagina: Piet Mondriaan, *Compositie in kleur B* (1917)

© Collectie Kröller-Müller Museum, Otterlo

Fotograaf portret Hugo Quené: Ed van Rijswijk