De geheimen die ‘ik’ vertelt

PSYCHOLOGIE De onopvallende woorden die nauwelijks nodig zijn om te begrijpen waar iemand over praat, verraden veel over mensen.

Ellen de Bruin

De Texaanse psycholoog James Pennebaker is een voorstander van gelijkheid. Hij had nooit gedacht dat er een groot statusverschil zichtbaar zou zijn tussen de mailtjes die hij aan voormalige studenten stuurde en zijn mailtjes aan beroemde collega’s. Totdat hij zijn oude mailtjes door zijn eigen taalscanner haalde. Toen zag hij: in zijn mails aan vooraanstaande hoogleraren gebruikt hij voortdurend het woord ‘ik’, in zijn mails aan studenten komt het woord ‘ik’ vrijwel niet voor.

Niet dat Pennebaker dat bewust deed. Hij kon zich van de meeste mails amper herinneren dat hij ze geschreven had. En wat hij zo fascinerend vindt, vertelt hij op een bankje op de grotendeels verlaten universiteitscampus te Tilburg: die statusverschillen zijn zichtbaar in zijn mails terwijl hij niet neerbuigend is tegen de studenten. Hij citeert: “‘Beste student, dank je voor je mail, het klinkt alsof je hier echt iets te pakken hebt, houd me alsjeblieft op de hoogte’ – dat is een heel warme mail van iemand met een hogere status. Alleen: ik gebruik het woord ‘ik’ niet.” ‘Ik’-gebruik, ontdekte Pennebaker, is namelijk een signaal van lage status. Vergelijkbaar, beaamt hij, met een diepe Japanse beleefdheidsbuiging – maar dan *vrijwel* onmerkbaar. Evengoed: “Als je ‘ik’ zou mijden in mails aan iemand met een hogere status, denk ik dat hij dat op een heel subtiele manier zou opmerken.”

Het is zondag; in het Dantegebouw op de campus is net een congres over emoties begonnen waar Pennebaker te gast is. Oorspronkelijk is hij dan ook geen taalpsycholoog. Hij werd bekend, onder psychologen beroemd, als grondlegger van de schrijftherapie. Een kwartiertje per dag schrijven, een dag of vier achter elkaar, maakt mensen minder somber en verbetert het immuunsysteem. Om te onderzoeken hoe dat kan, ontwikkelden Pennebaker en collega’s een computerprogramma dat telt hoe vaak welke soorten woorden in een tekst voorkomen. Ze noemden het *Linguistic Inquiry and Word Count*, afgekort LIWC, liefkozend uitgesproken als ‘Luke’.

Aanvankelijk ging het Pennebaker vooral om inhoudsanalyse, om woorden die emoties of redeneringen weergaven: hij ontdekte dat mensen meer opknapten naarmate ze in de loop der tijd meer positieve en minder negatieve emotiewoorden gebruikten én als er meer structuur in hun verhaal kwam, dus als ze anders naar hun emoties leerden kijken.

Heel interessant, maar inmiddels bijzaak. Momenteel richt Pennebaker zich vooral op onopvallende, korte, schijnbaar betekenisloze woorden: hulpwerkwoorden, lidwoorden, persoonlijk voornaamwoorden, voegwoorden, ontkenningen, voorzetsels, bijwoorden. Woorden die je nauwelijks opmerkt. Waar je in een vreemde taal de meeste fouten mee maakt. Die je meestal nauwelijks nodig hebt om een betoog te begrijpen. Hij vertelt er meer dan enthousiast over, met zijn gezellige, Texaans knauwende accent. Genietend is het woord.

Juist in het gebruik van die functiewoorden, zoals Pennebaker ze noemt, verschillen mensen. Er valt dan ook van alles uit af te leiden over hun persoonlijkheid, gemoedstoestand of de situatie waarin ze verkeren. Neem het woord ‘ik’. Dat signaleert niet alleen onderdanigheid – veel ‘ik’ gebruiken is ook een teken van depressie. Zelfmoord komt meer voor onder dichters die vaker het woord ‘ik’ gebruiken. Vrouwen gebruiken ‘ik’ ook vaker dan mannen, jongere mensen vaker dan oudere mensen, waarheidssprekers vaker dan leugenaars, mensen die rustig zitten te eten vaker dan mensen die in groepsverband werken of sporten. En zulke patronen bestaan uiteraard niet alleen voor ‘ik’, maar voor alle woordsoorten, schrijft Pennebaker in zijn net verschenen populair-wetenschappelijke boek *The Secret Life of Pronouns*. Op basis van deze kennis, en met de LIWC, kun je bijvoorbeeld berekenen dat er 74 procent kans is dat de auteur van een tekst een vrouw is, of 60 procent kans dat het dezelfde schrijver is als van een andere tekst.

En nee, Pennebaker zit niet voortdurend te letten op het taalgebruik van andere mensen om hun persoonlijkheid te doorgronden. “Dat is bijna onmogelijk. Die functiewoorden vloeien zo snel van de tong, je kunt ze niet waarnemen. Je moet bijna je hersens uitschakelen om ernaar te kunnen luisteren.”

Maar een computer analyseert een geschreven tekst in seconden, minuten hooguit. Dus wat hij wel graag doet, is ’s avonds nog snel wat tekst door de LIWC heen jassen. “Ik heb wel een ziljoen tekstbestanden op mijn computer”, zegt hij opgewekt. Dus nadat hij op een congres met collega’s had gediscussieerd over de vraag of personages in fictie praten als de auteur of als ‘zichzelf’, analyseerde hij op zijn hotelkamer even het volledige werk van Shakespeare en het filmscript van *Sleepless in Seattle*, van Nora Ephron. “Ik ontdekte dat Romeo en Julia allebei praten als een man en de personages van Nora Ephron als vrouwen. Dat had ik nooit gedacht! Onze hersenen werken gewoon niet zo.”

U schrijft in uw boek ook dat u aanbevelingsbrieven voor sollicitanten heeft geanalyseerd op mogelijke misleiding. Doet u dat nu altijd?

In Shakespeares ‘Romeo en Julia’ praten Romeo én Julia als mannen

“Dat zou ik graag willen... Ze komen ook steeds vaker als pdf binnen en niet op papier, dus het zou kunnen.”

Zou het ethisch zijn?

“Ik zie er geen probleem mee. We kleden die brieven toch wel helemaal uit. Alleen: het zou heel dom zijn om onze beslissing daarop te baseren. De meeste van die brieven zijn niet heel betrouwbaar. Ze zijn allemaal positief. Je moet in feite ontdekken wat iemand *niet* zegt. Als iemand niet zo slim is, staat er: charmant, vriendelijk, werkt hard, werkt goed samen.

“Weet je, iedereen schrijft wel een béétje misleidende aanbevelingsbrieven. Ik zelf ook, hoewel ik dat niet in de gaten had totdat ik mijn eigen brieven analyseerde.”

We hebben op dit moment een groot fraudegeval in de Nederlandse sociale psychologie...

“Ja, hier op deze plek.” Hij wijst geestdriftig naar de grond. “We zitten hier precies op Ground Zero!”

Begin september werd bekend dat Diederik Stapel, hoogleraar sociale psychologie te Tilburg, op grote schaal onderzoeksgegevens had verzonnen. Maandag presenteert de commissie die de fraude onderzoekt naar verwachting haar rapport.

Is die fraude bekend onder sociaal psychologen in de Verenigde Staten?

“O, *it’s huge*. Zelf had ik nog nooit van deze man gehoord, maar ik ben geen traditionele sociaal psycholoog. Ik was op een klein congres op Harvard toen het nieuws naar buiten kwam. Er waren een paar Nederlanders bij en die waren er kapot van. De helft van de mensen op dat congres kenden de man of zijn werk. Het was heel indrukwekkend.

“Ik heb twee andere fraudegevallen meegemaakt in mijn carrière en het is zo wreed. Het maakt iedereen achterdochtig tegenover iedereen. Een paar jaar geleden gebeurde iets dergelijks aan mijn universiteit en plotseling merkte ik dat ik de data van mijn studenten nauwkeuriger begon te checken – en zij begonnen mijn data waarschijnlijk ook nauwkeuriger te checken. En dat had niets positiefs. In de wetenschap moet je mensen kunnen vertrouwen.”

Zou u Stapels artikelen kunnen analyseren om te zien in welke er gefraudeerd is en in welke niet?

“Ja, daar heb ik ook over nagedacht. Het probleem is: dan hebben we onderzoek nodig waarvan we zeker weten dat hij eerlijk was toen hij het beschreef. En dat is heel moeilijk.

“Iemand als deze jongen is interessant omdat hij duidelijk al jaren aan het frauderen was. Ik ben door nationale veiligheidsinstanties in de Verenigde Staten benaderd om zulke gevallen te onderzoeken. Hoe weet je of een werknemer bezig is de FBI of een particulier beveiligingsbedrijf te verraden? Dat is hetzelfde type probleem.

“Ik ben al een tijd bezig met een groot project waarbij we de e-mails van mensen bekijken die ze naar hun vrienden en kennissen sturen vóórdat ze een enorm geheim hebben, en ten tijde van dat geheim. In de meeste gevallen gaat het waarschijnlijk om geheime relaties, maar sommige geheimen hebben met werk te maken, of met gezondheid, of iets anders. Ik probeer uit te zoeken: hoe veranderen de communicatiepatronen van deze mensen in het algemeen? En er treden interessante verschuivingen op. Voordat je een geheim hebt, ben je geneigd om eerlijk te zijn tegen je vrienden en meer afstandelijk tegen mensen die je niet zo goed kent. En dan heb je ineens dat enorme geheim – en wat er gebeurt is: je wordt afstandelijk tegen je vrienden en attenter en vriendelijker tegen mensen die je niet zo goed kent.”

U analyseert dus ook e-mails van mensen die bij beveiligingsdiensten werken?

“Ja. Als je bij een openbare instantie werkt, of bij een instituut dat overheidsgeld krijgt, dan kunnen al je e-mails geanalyseerd worden. Dus je kunt mensen volgen in zoverre als ze e-mails versturen. Je komt dan wel op wat *creepy* terrein, je kunt je afvragen of we al onze e-mails de hele tijd geanalyseerd willen hebben – maar aan de andere kant, het wordt toch wel gedaan, door Google.”

Maar Google kijkt toch alleen naar de inhoud, naar steekwoorden om advertenties bij te zetten, en niet naar persoonlijke stijl?

“O jawel, Google begint ook al op stijl te letten. Ze zouden bij Google idioot zijn als ze niet op persoonlijk voornaamwoorden zouden letten. Als mensen depressief zijn, gebruiken ze het woord ‘ik’ veel vaker. Dus als iemands gebruik van het woord ‘ik’ sterk stijgt, zou ik reclame voor antidepressiva naast hun inbox zetten, als ik Google was. Die correlatie is klein, maar wat maakt het uit! Als je tien miljoen mensen hebt, dan kan de correlatie 0,01 zijn, en dan heb je nog steeds een enorm verkooppotentieel.”

Terug naar die beveiligingsinstanties... Kunt u hun al iets vertellen waar ze wat aan hebben?

“Ik denk dat dat begint te komen, ja. Maar ook hier geldt: het is allemaal probabilistisch.”

U kunt dus zien, bijvoorbeeld, dat er een grote kans is dat iemand een verschuiving heeft doorgemaakt van ‘geen geheim hebben’ naar ‘wel een geheim hebben’.

Maar u weet dan niet waar het geheim over gaat. Iemand kan een geheime relatie hebben...

“Klopt. Misschien kunnen we trouwens wel meer zien, maar daar hebben we nog niet goed naar gekeken.”

Misschien weet u binnenkort of een geheim privé is of werkgerelateerd, maar nu nog niet?

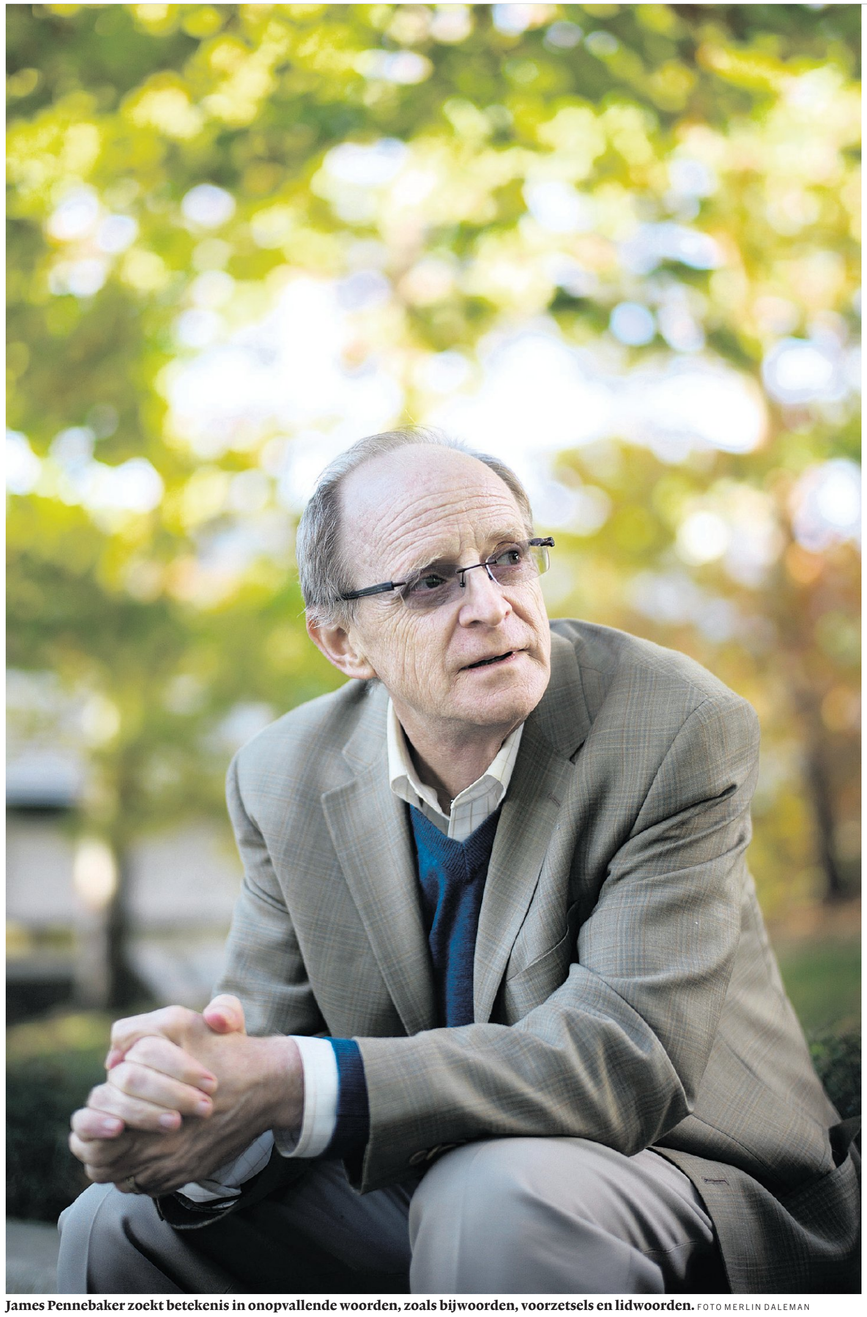
“Dat klopt, ja. Een ander voorbeeld dat eigenlijk ook over hetzelfde gaat, en dit is een heel *cool* voorbeeld, is het manifest van Anders Breivik.” De Noorse extreemrechtse terrorist die in juli 69 mensen doodschoot op het eilandje Utøya. “Nu weet ik wel dat hij ook dingen in dat manifest gestolen heeft, maar als je aanneemt dat hij de rest van het begin naar het eind heeft geschreven, dan is het taalkundig gezien een zeer interessant document. In mijn vak geldt: hoe meer woorden hoe beter – en dit manifest bevat heel veel woorden [778.230 om precies te zijn, red.]. Ik heb het opgebroken in blokken van 2.500 woorden, ruim 300 blokken, en de LIWC-analyse gedaan.

“Nu weten we, hoewel ik er nog niet heel veel data over heb: als mensen op het punt staan om een bedreiging uit te voeren, gaat hun gebruik van het woord ‘ik’ sterk omlaag. Dat was zo bij Bush voordat hij Irak binnenviel, bij Hitler voordat hij Polen binnenviel, bij Truman voordat hij de atoombommen op Japan afwierp. En wat er interessant is aan Breivik: zijn gebruik van het woord ‘ik’ is ongeveer anderhalf, twee procent, in bijna het hele boek. Maar in de laatste twintig procent gaat zijn gebruik van het woord ‘ik’ enorm omhoog, tot wel zeven procent.”

Heeft u naar de structuur van van het manifest gekeken? Het grootste deel is een soort maatschappijanalyse, maar het laatste deel is in dagboekvorm.

“Dus dat is het deel waar het ‘ik’-gebruik hoog is – ik heb het niet gelezen, alleen het begin. Het interessante is: in de laatste twee, drie procent van het boek daalt het ‘ik’-gebruik weer, naar vrijwel nul.”

En dat is, inderdaad, vlak voordat Breivik zijn aanslagen heeft gepleegd.



James Pennebaker zoekt betekenis in onopvallende woorden, zoals bijwoorden, voorzetsels en lidwoorden. FOTO MERLIN DALEMAN

‘Als mensen een bedreiging gaan uitvoeren, gebruiken ze ineens minder “ik”’

Wat zouden we over tien of vijftien jaar met deze onderzoekslijn kunnen?

Pennebaker denkt even na. “Om te beginnen, en terzijde: alles waar ik het nu over heb, daar zullen ze over tien, vijftien jaar om lachen. Gewoon omdat het zo simplistisch is. De technologie gaat zo snel vooruit. Neem het Google Books project, dat is echt spectaculair – ik kan niet beginnen te vertellen hoe opwindend ik dat vind. Het artikel dat er afgelopen december over in *Science* stond, zou verplichte literatuur moeten zijn in alle geschiedenis-, psychologie- en antropologiefaculteiten. Maar de realiteit is dat heel weinig mensen ervan af weten of het belangrijk vinden.”

In dat artikel – deze bijlage besteedde er op 15 januari een pagina aan – beschreven enkele tientallen onderzoekers en technici van onder meer Harvard, MIT, Google Books en de Encyclopaedia Britannica hoe je kwantitatief, objectief cultuuronderzoek kunt doen door slim zoeken in Google Books, gebaseerd op (op dat moment) 12 procent van alle boeken die ooit gepubliceerd zijn.

Pennebaker struikelt bijna over zijn woorden van enthousiasme: “Ga naar ngrams.googlelabs.com en typ iets in, wat dan ook.” Het resultaat is een grafiek waarin te zien is hoe vaak die term in een bepaalde periode in boeken voorkwam. “Je kunt zien hoe taal zich heeft ontwikkeld, hoe onze belangstelling voor bepaalde onderwerpen zich heeft ontwikkeld... Je kunt bijvoorbeeld Shakespeare en Milton intypen. Milton was heel populair in 1800, Shakespeare was toen niet zo populair, en” – Pennebaker doet een ingewikkelde autorace na met zijn wijsvingers – “toen ging het een tijdje om en om, en Shakespeare heeft uiteindelijk gewonnen, althans de laatste tien, vijftien jaar. Zo kun je elk idee testen. Het is zo’n andere manier van denken! En de methoden die ontwikkeld worden, die zijn zó creatief. Je kunt naar de inhoud kijken, een stilistische benadering hanteren, of een mengvorm.

“Dus wat gebeurt er over tien, vijftien jaar? We zullen in staat zijn om taalgebruik echt heel goed te ontginnen. We zullen in staat zijn de manier waarop de cultuur denkt af te tappen. Een beetje zoals veiligheidscamera’s nu gebruikt worden: we kunnen erachter komen hoe mensen in Oost-Tilburg op dit moment denken, omdat er zó veel woorden tegelijk worden gegenereerd...”

Je zou kunnen analyseren hoe de schok van slecht nieuws, of van een fraudegeval, zich door de gemeente en het land verspreidt.

“Precies, precies. En dan kun je ook nog iets concluderen als: deze mensen zouden op dit moment echt baat hebben bij Coca Cola!” En hij lacht weer zijn knauwende Texaanse lach.

James W. Pennebaker – The Secret Life of Pronouns: What Our Words Say About Us. (Bloomsbury, 352 blz., €24,99).

James W. Pennebaker

‘Jamie’ Pennebaker (1950) studeerde in de jaren zeventig psychologie aan de universiteit van Texas (Austin, Texas). Aan diezelfde universiteit is hij sinds 1997 hoogleraar. Hij begon als gezondheidspsycholoog. In het begin van de jaren ‘80 onderzocht hij hoe mensen hun fysieke symptomen en klachten waarnemen (we voelen vaak wat we verwachten te voelen). Daarna deed hij onderzoek naar traumatische ervaringen, geheimen en mensen in vertrouwen nemen.

Begin jaren ‘90 ontdekte hij het gunstige effect van schrijven over nare emotionele ervaringen op de geestelijke en lichamelijke gezondheid, óók voor wie niets aan anderen laat lezen. Om de inhoud van het geschrevene te onderzoeken ontwikkelde hij met collega’s een computerprogramma dat telt hoe vaak welke soorten woorden in een tekst voorkomen: Linguistic Inquiry and Word Count (LIWC). Daarmee tellen en analyseren psychologen nu woorden in alle mogelijke teksten: van Beatles-liedjes en speeches van presidenten tot gedichten, filmscripts, romans, tweets en blogs.

De LIWC is in meer dan tien talen vertaald, waaronder het Nederlands, door Utrechtse psychologen. Dat viel nog wel mee, vindt Pennebaker, maar talen als Arabisch en Chinees waren “*a pain in the ass*”, zegt hij. “Het Chinees kent geen spaties tussen woorden, het Arabisch loopt van rechts naar links... Makkelijk aan te passen? Dat denk jij! En functiewoorden zijn in andere talen vaak geen aparte woorden maar opgenomen in andere woorden. Het is gewoon zó veel werk.”