

De Nederlandse versie van de 'Linguistic Inquiry and Word Count' (LIWC)

Een gecomputeriseerd tekstanalyseprogramma

Hanna Zijlstra, Tanja van Meerveld, Henriët van Middendorp,
James W. Pennebaker en Rinie Geenen*

Summary

Dutch version of the Linguistic Inquiry and Word Count (LIWC), a computerized text analysis program

Language, whether spoken or written, is an important window into people's emotional, social, and cognitive worlds. Text analysis of narratives is a means to examine the underlying processes, such as processes that are considered crucial to achieve a desired health outcome after expression of emotions. The Linguistic Inquiry and Word Count (LIWC) was developed to analyse text on a word-by-word basis. This article describes the development of a Dutch version of the LIWC, its manual, and results of equivalence analyses. The Dutch LIWC dictionary contains 6568 words in 66 word categories. Eighty-nine percent of the word categories showed a high to very high correlation between the Dutch and English LIWC-results. This reflects considerable equivalence. The main word categories 'pronouns', 'positive emotions', 'negative emotions' and 'cognitive processes' showed satisfactory equivalence. The evaluation of intra-individual processes and interindividual differences in language use according to the principles of the LIWC program is from now on possible in Dutch texts. This will be of interest for research in psychology, but also in law, marketing, communication, medicine and many other areas.

Inleiding

Expressie van emoties wordt gezien als een belangrijke werkzame factor in psychotherapie en wordt ook in het dagelijkse leven als gezond beschouwd (Van Middendorp & Geenen, 2004). Bijna twintig jaar geleden startte experimenteel onderzoek naar het effect van expressie van emoties. In het 'emotional disclosure' paradigma krijgen personen de opdracht om 15 tot 20 minuten per dag op 3 á 4 opeenvolgende dagen te schrijven of te praten over de meest traumatische gebeurtenis die ze ooit hebben meegemaakt (Pennebaker & Beall, 1986). De op-

* Hanna Zijlstra, psycholoog en diëtist en Tanja van Meerveld, psycholoog, hebben onderzoek gedaan naar emotieregulatie bij de Capaciteitsgroep Gezondheidspsychologie van de Universiteit Utrecht. Henriët van Middendorp en Rinie Geenen, beiden psycholoog, zijn respectievelijk als onderzoeker en universitair docent verbonden aan de Capaciteitsgroep Gezondheidspsychologie van de Universiteit Utrecht. James W. Pennebaker is hoogleraar psychologie bij het Department of Psychology of the University of Texas, Austin, Texas. Correspondentieadres: Hanna Zijlstra, L. van Tulderstraat 61, 7558 JM Hengelo. E-mail: hanna.zijlstra@wxs.nl. Dit artikel verschaft de handleiding van de Nederlandse LIWC. Het Nederlandse woordenboek is bij de auteurs te verkrijgen.

dracht is anoniem en er wordt geen feedback gegeven. Mensen in de controleconditie schrijven over een relatief oppervlakkig niet-emotioneel onderwerp. Deze 'eenvoudige' schrijfo opdracht blijkt in veel studies gunstige gevolgen te hebben voor de gezondheid, het emotioneel welbevinden, de fysiologische gesteldheid en het algemeen functioneren (Smyth, 1998).

Op dit moment zijn de voorwaarden voor het bereiken van gunstige effecten nog onduidelijk. Sommige mensen hebben meer baat bij de opdracht dan anderen. Bijvoorbeeld, mensen die hoog scoren op de persoonlijkheidseigenschap 'vijandigheid' en mensen die moeilijk bespreekbare traumatische ervaringen hebben meegemaakt, lijken meer te profiteren van de interventie (Pennebaker & Graybeal, 2001; Pennebaker & Susman, 1988). Expressie van emoties lijkt minder gunstig voor alexithyme personen, als het gaat om verliesverwerking en bij boosheid (Bushman, 2002; Lumley, Tojek & Macklem, 2002; Stroebe, Stroebe, Schut, Zech & Van den Bout, 2002). Het is nog grotendeels onduidelijk welk mechanisme verantwoordelijk is voor de gerapporteerde effecten. Schrijven over emotionele gebeurtenissen lijkt niet zozeer voordeel op te leveren vanwege het vrijkomen van opgekropte spanningen (catharsis), maar vooral omdat schrijven leidt tot de omzetting van traumatische ervaringen in woorden, waardoor verwerking en begrip van de situatie toenemen en negatieve emoties die samenhangen met de traumatische gebeurtenis verminderen (Lutgendorf & Ullrich, 2002; Pennebaker, Mayne & Francis, 1997).

Tekstanalyse lijkt een veelbelovende methode om inzicht te krijgen in het proces dat in gang gezet dient te worden om te kunnen profiteren van een dergelijke emotionele interventie. Het computerprogramma 'Linguistic Inquiry and Word Count' (LIWC) is ontwikkeld om geschreven of gesproken teksten te analyseren (Pennebaker, Francis & Booth, 2001). De LIWC is een kwantitatief onderzoeksinstrument, dat op objectieve en geautomatiseerde wijze het aantal woorden per woordcategorie telt. De LIWC bevat 66 woordcategorieën, uitgesplitst in vijf dimensies: standaard linguïstische dimensies, psychologische processen, relativiteit/betrekkelijkheid, persoonlijke aangelegenheden en experimentele dimensies. De dimensies zijn onderverdeeld in categorieën, waaronder voornaamwoorden (ik, wij, zij), emotionele processen (afschuw, angstig, amuseren) en cognitieve processen (analyseren, bedenken, realiseren). Enkele van deze categorieën zijn nog verder onderverdeeld in subcategorieën, zoals 'positieve emoties' en 'negatieve emoties' binnen de categorie 'emotionele processen'.

De mogelijkheden tot tekstanalyse met de LIWC beperken zich niet tot de analyse van teksten betreffende expressie van emoties. Wel werd de LIWC voor dat doel ontwikkeld. De eerste onderzoeken rapporteerden dat personen die het meest profiteren van het schrijven of praten een hoog percentage positieve emotiewoorden, een gemiddeld aantal negatieve emotiewoorden en, het meest belangrijk, een toename in het aantal cognitieve woorden over de sessies heen gebruiken (Pennebaker & Francis, 1996; Pennebaker et al., 1997). Uit recenter onderzoek komt naar voren dat individuen hun eigen specifieke woordgebruik hebben, dat correleert met gezondheidskenmerken, sociaal gedrag en cognitieve mechanismen. Als dit woordgebruik over de sessies heen verandert, heeft dit een gunstige invloed op de gezondheid en het sociale gedrag (Pennebaker & Graybeal, 2001). Met name een flexibel gebruik van voornaamwoorden lijkt een goede graadmeter voor verbetering (Campbell & Pennebaker, 2003).

Het LIWC-woordenboek is in het Engels opgesteld. Het doel van ons onderzoek was het ontwikkelen van een Nederlandse versie van het standaard Engelse LIWC-woordenboek, om het mogelijk te maken ook Nederlandse teksten met het LIWC-computerprogramma te analyseren. In dit artikel bespreken we de ontwikkeling van de Nederlandse LIWC, presenteren we de handleiding en to-

nen we de resultaten van de equivalentie-analyses van de Nederlandse en Engelse LIWC.

Procedure van de vertaling

Het Engelse LIWC-woordenboek is uitgesplitst in de 66 woordcategorieën, waarbij dezelfde woorden in meerdere categorieën kunnen voorkomen. Bijvoorbeeld 'think' behoort tot de woordcategorieën 'cognitieve processen', 'inzicht' en 'werkwoorden in de tegenwoordige tijd'.

Twee vertalers (HZ en TvM) hebben onafhankelijk van elkaar alle woorden vertaald met behulp van twee woordenboeken Engels-Nederlands (Coenders, 1996a; Coenders, 1996b). Indien het woordenboek bij een Engels woord meerdere Nederlandse vertalingen gaf, zijn al deze vertalingen meegenomen, mits ze pasten in de woordcategorieën van de Engelse LIWC.

In de Engelse LIWC is een beperkt aantal werkwoorden vervoegd. In de Nederlandse versie is er voor gekozen om alle veelgebruikte werkwoorden te vervoegen, zodat de woordcategorieën 'werkwoorden in de verleden tijd', 'werkwoorden in de tegenwoordige tijd' en 'werkwoorden in de toekomstige tijd' een beter beeld geven van het gebruik van werkwoorden dan de Engelse versie.

In de Engelse versie wordt gebruik gemaakt van woordstammen. Dit wordt aangegeven met een sterretje, bijvoorbeeld 'hungr*', waarbij alle woorden die beginnen met 'hungr' op dezelfde wijze gescoord worden. Dit is in de Nederlandse versie, afgezien van de woordcategorie 'getallen' niet overgenomen, om de kans op het ten onrechte tellen van woorden die niet passen in de woordcategorie te verkleinen. Een voorbeeld van een woordstam waarbij in het Engels een verkeerde telling kan ontstaan is 'cash*'. Niet alleen cash-desk wordt meegenomen, maar ook 'cashew' dat in de woordcategorie 'eten, drinken en dieet' hoort en niet in de categorie 'geld en financiële zaken'.

Na het vertaalwerk zijn de resultaten van beide vertalers vergeleken. De woorden die overeenkwamen zijn beoordeeld op relevantie voor het Nederlandse LIWC-woordenboek. Het criterium was dat de woorden in algemene spreek- en schrijftaal gebruikt worden. Woorden die zelden gebruikt worden in de Nederlandse taal, hoogdravende woorden en vakjargon, zijn verwijderd. Afwijkende vertalingen werden door de vertalers besproken. Indien dit leidde tot overeenstemming werd het woord toegevoegd aan het Nederlandse LIWC-woordenboek en was dit niet het geval dan werd het woord geschrapt.

Van alle in het Nederlandse LIWC-woordenboek opgenomen woorden werd vervolgens gecontroleerd of de aangegeven woordcategorieën correspondeerden met de betekenis van het woord. Zo nodig zijn woorden aan andere of aan minder woordcategorieën gekoppeld. Een voorbeeld is het Engelse woord 'damn', waarvan de vertaling 'verdomme' wel en 'afkraken' niet is gekoppeld aan de woordcategorie 'vloekwoorden' aangezien het Nederlandse woord 'afkraken' geen vloekwoord is.

Achtereenvolgens zijn alle Nederlandse woorden uit de 66 woordcategorieën samengevoegd, de dubbele woorden verwijderd en de woorden op alfabetische volgorde gezet. Het Nederlands LIWC-woordenboek werd voorzien van de definities van de woordcategorieën en als tekstbestand (.txt) opgeslagen, zodat het binnen het gecomputeerde LIWC-programma van Pennebaker et al. (2001) gebruikt kan worden (zie www.erlbaum.com, search: LIWC).

Tabel 1 Voorbeelden en aantal woorden per woordcategorie van de Nederlandse LIWC.

Nr	Woordcategorieën	Voorbeelden	Aantal
I Standaard linguïstische dimensies			
1	Totaal aantal voornaamwoorden	ik, jij, onze	38
2	1 ^{ste} persoon enkelvoud	ik, mij, mijn	7
3	1 ^{ste} persoon meervoud	wij, onze, ons	6
4	Totaal eerste persoon	ik, wij, mij	13
5	Totaal tweede persoon	jij, jullie, jouw	7
6	Totaal derde persoon	zij, hij, hen	12
7	Ontkenningen	nee, nooit, niet	36
8	Toestemming/instemming/goedkeuring	ja, oké, goed	50
9	Lidwoorden	de, het, een	3
10	Voorzetsels	op, naar, van	48
11	Getallen	één, dertig, miljoen	107
II Psychologische processen			
12	Emotionele processen	blij, verdrietig, somber	2048
13	Positieve emoties	gelukkig, dankbaar, dapper	690
14	Positieve gevoelens	plezier, liefde, glimlachen	129
15	Optimisme en energie	trots, vurig, wilskracht	128
16	Negatieve emoties	bedroefd, vijandig, wanhoop	1347
17	Angst en vrees	zenuwachtig, bang, gespannen	207
18	Boosheid	boos, dreigen, ergeren	439
19	Treurigheid of depressie	huilen, somber, teleurstelling	238
20	Cognitieve processen	oorzaak, weten, denken	1035
21	Veroorzaken	omdat, waarom, vandaar	125
22	Inzicht	overwegen, ophelderen, realiseren	459
23	Tegenstrijdigheid	zou, behoren, conflict	62
24	Remming/geremdheid	blokkeren, hinderen, inhouden	199
25	Tentatief/voorzichtig/aarzelend	misschien, waarschijnlijk, voorlopig	171
26	Zekerheid	volstrekt, absoluut, vastbesloten	75
27	Zintuigen en Perceptuele processen	zien, voelen, horen	347
28	Zien	kijken, zicht, blik	104
29	Horen	geluid, luisteren, klank	127
30	Voelen	aanraken, aftasten, vasthouden	93
31	Sociale processen	communiceren, delen, helpen	819
32	Communicatie	interviewen, gesprek, gerucht	585
33	Verwijzingen naar andere personen	wij, zij, zichzelf	28
34	Vrienden	vriend, vriendschap, vriendin	30
35	Familie	moeder, ouders, neef	80
36	Mensen	meisje, mens, volwassenen	53
III Relativiteit/ betrekkelijkheid			
37	Tijd	zomer, vroeger, zodra	269
38	Werkwoorden in de verleden tijd	deelde, discussieerde, domineerde	1773
39	Werkwoorden in de tegenwoordige tijd	afleren, bewaren, gebeuren	1886
40	Werkwoorden in de toekomstige tijd	zal, wens, wil	19
41	Ruimte	vlakbij, plaats, noord	121
42	Omhoog/ de hoogte in	boven, hoger, top	15
43	Beneden / neer	dieper, beneden, laag	14
44	Inclusief	en, inbegrepen, ook	31
45	Exclusief	buiten, tenzij, uitgezonderd	23
46	Beweging	benaderen, lopen, beklimmen	390

Tabel 1 (vervolg) Voorbeelden en aantal woorden per woordcategorie van de Nederlandse LIWC.

Nr	Woordcategorieën	Voorbeelden	Aantal
IV Persoonlijke aangelegenheden			
47	<i>Beroep/bezigheden</i>	bereiken, bevorderen, chef	500
48	School	educatief, eerstejaars, eindexamen	171
49	Werk	carrière, collega, concurreren	221
50	Prestatie	behalen, bekroning, beloning	139
51	<i>Vrijtijdsbesteding</i>	fietsen, fitness, trainen	226
52	In en om het huis	keuken, huis, tuin	50
53	Sporten	wedstrijd, lichaamsbeweging, conditietraining	93
54	Televisie en films	film, video, televisie	21
55	Muziek	zingen, lied, gitaar	60
56	<i>Geld en financiële zaken</i>	winst, factuur, failliet	190
57	<i>Godsdienst / Zingeving</i>	dopen, preek, katholiek	199
58	Religie	bidden, vereren, zegen	139
59	Dood en sterven	dood, treuren, sterfbed	59
60	<i>Lichamelijke toestand</i>	eetlust, flauwvallen, doezelig	653
61	Lichamelijk functioneren	diarree, vitaal, dorst	403
62	Seks en seksualiteit	flirten, zoen, beminnen	96
63	Eten, drinken en dieet	drinken, honger, voeding	124
64	Slapen en dromen	dromen, slaperig, wakken	64
65	Lichamelijke verzorging	douche, wassen, make-up	50
V Experimentele dimensies			
66	Vloekwoorden	verdorie, goddomme, shit	37

Het Nederlandse LIWC-woordenboek

Het Nederlandse LIWC-woordenboek telt 6.568 woorden en bevat 66 woordcategorieën. In Tabel 1 wordt per woordcategorie een aantal voorbeelden en het aantal woorden per categorie gegeven.

De hoofdcategorieën (bijvoorbeeld emotionele processen) bevatten zowel alle woorden uit de subcategorieën (bijvoorbeeld positieve emoties) als woorden die alleen in de hoofdcategorie voorkomen.

Voor het gebruik van de Nederlandse LIWC kan de beknopte Nederlandse handleiding worden gebruikt, zoals weergegeven in Tabel 2. De uitgebreide Engelstalige handleiding is te vinden op: website <http://homepage.psy.utexas.edu/homepage/class/Psy394V/Pennebaker/LIWC2001.pdf>.

Equivalentie onderzoek

Voor toetsing van de equivalentie van de Nederlandse en de Engelse LIWC-woordenboeken zijn 75 teksten verzameld die zowel in het Engels als in het Nederlands in digitale vorm bestaan. De teksten zijn grotendeels via het Internet verkregen en zijn onder meer afkomstig van filmscripts, boeken, verhalen, verslagen, columns, liedjes, gedichten, persberichten en websites. Teksten die slechts in één van beide talen beschikbaar waren, zijn vertaald met het vertaalprogramma 'Worldlingo' (zie website www.worldlingo.com/wl/translate).

Tabel 2 *Beknopte handleiding van de LIWC.*

Het LIWC-programma is te bestellen via de website www.erlbaum.com, search: LIWC. Start het LIWC-programma door op het LIWC-icoon te klikken of het LIWC2001.EXE bestand te openen. Via het menu kunnen de diverse mogelijkheden worden bekeken.

Laden van het gewenste woordenboek

Laad het gewenste LIWC-woordenboek. Het programma geeft het laatst gebruikte woordenboek aan. Bijvoorbeeld 'dictionary currently loaded is: Dutch using its own category definitions'. Klik op 'OK', ook als een ander woordenboek dient te worden gebruikt. Kies eventueel via 'dictionary' in het menu een ander woordenboek.

Analyseren van tekstbestanden

Het programma kan één of meerdere teksten tegelijk analyseren. Teksten dienen opgeslagen te zijn als txt-bestand (tekst zonder opmaak). Word-bestanden worden wel geanalyseerd, maar geven onjuiste aantallen. Het programma geeft een foutmelding als tekstbestanden in te veel submappen staan. Ga voor het analyseren van teksten in het menu naar 'file' en vervolgens naar 'process text'. Selecteer de te analyseren teksten en druk op 'select'. Het programma vraagt naar de 'bestandsnaam' voor de LIWC resultaten en naar de 'directory' waar het bestand opgeslagen moet worden. Klik na dit gedaan te hebben op 'opslaan/save'. Binnen enkele seconden zijn de teksten geanalyseerd en verschijnt het bericht 'Results are in file ...'. Door op 'OK' te klikken kunnen de resultaten worden bekeken.

LIWC-resultaten

Op de bovenste regel van het resultatenbestand staan de woordcategorieën vermeld met daaronder per regel de resultaten van iedere tekst. In de eerste kolom worden de bestandsnamen vermeld. De resultaten van de woordcategorieën 'segment', 'WC' (word count) en 'WPS' (words per sentence) geven aantallen weer; alle andere woordcategorieën zijn percentages (het aantal woorden binnen die woordcategorie gedeeld door het totale aantal woorden). De LIWC-resultaten zijn in te lezen en statistisch te bewerken in Excel en SPSS. Voor gebruik in SPSS dienen de decimaalteken (punten) in het resultatenbestand vervangen te worden door komma's.

Methode

De Engelse en Nederlandse versies van alle teksten zijn geanalyseerd met respectievelijk het Engelse en het Nederlandse LIWC-woordenboek. Voor alle woordcategorieën zijn correlaties berekend tussen de Engelse en Nederlandse LIWC-resultaten over alle parallelle teksten samen. Bij verdelingen met een scheefheid tussen de -1,5 en 1,5 is de parametrische Pearson correlatiecoëfficiënt berekend en bij de overige verdelingen de non-parametrische Spearman's rho correlatiecoëfficiënt. Daarnaast zijn voor alle woordcategorieën de mediaan en de spreiding (minimum - maximum) berekend, evenals de effectgrootte van het verschil tussen het Nederlandse en Engelse gemiddelde van de score op de woordcategorie. De effectgrootte verschaft globaal inzicht in het bestaan van systematische verschillen tussen de twee versies van de LIWC. Voor de scheef verdeelde categorieën moet de effectgrootte met enige voorzichtigheid worden geïnterpreteerd. De effectgrootte (d) werd berekend door het verschil in gemiddelde score op een woordcategorie van de Nederlandse (M_{NL}) en Engelse (M_{EN}) versie te berekenen en deze te delen door de gemiddelde standaarddeviatie van de beide versies: $d = (M_{NL} - M_{EN}) / [(SD_{NL} + SD_{EN}) / 2]$. Een effectgrootte tussen 0,2 en 0,5 wordt als klein beschouwd, een effectgrootte tussen 0,5 en 0,8 als middelmatig en een effectgrootte vanaf 0,8 als groot (Cohen, 1977).

Resultaten

In Tabel 3 wordt van alle woordcategorieën de mediaan, de spreiding (min-max) en de correlaties vermeld. In 89% van de gevallen zijn de correlaties tussen de Engelse en Nederlandse LIWC-scores op de woordcategorieën hoog (tussen 0,50 en 0,70) tot zeer hoog (tussen 0,70 en 0,99). Voor zes woordcategorieën is sprake van een middelmatige correlatie: ruimte ($r = 0,34$), beneden/ neer ($r = 0,42$), sporten ($r = 0,40$), televisie en films ($r = 0,23$), muziek ($r = 0,40$) en vloekwoorden ($r = 0,38$). Voor de woordcategorie '% afkortingen' is de correlatie laag ($r = 0,07$).

Tabel 3 *Equivalentie van de Engelse en Nederlandse LIWC gebaseerd op 75 teksten.*

Nr.	Woordcategorieën	Engels		Nederlands		<i>r</i>
		Mediaan	Min - Max	Mediaan	Min - Max	
	Aantal woorden per tekst	794,0	348 - 13645	799,0	295 - 13972	0,99
	Aantal woorden per zin	17,9	4,5 - 44,9	17,4	4,4 - 46,0	0,83 ^a
	Aantal zinnen eindigend op ?	5,6	0,0 - 26,4	5,3	0,0 - 6,6	0,83 ^a
	Aantal unieke woorden	39,7	13,8 - 62,6	41,9	16,2 - 1,6	0,91 ^a
	% woorden uit woordenboek	73,2	61,0 - 83,3	70,6	57,4 - 78,0	0,85 ^a
	% afkortingen	0,0	0,0 - 0,2	0,0	0,0 - 1,4	0,07
	% emoticons (emotionele symbolen)	0,0	0,0 - 0,7	0,0	0,0 - 0,7	0,73
I Standaard Linguïstische dimensies						
1	Totaal aantal voornaamwoorden	11,8	2,2 - 21,1	9	1,1 - 20,8	0,93 ^a
2	1 ^{ste} persoon enkelvoud	1,3	0,0 - 8,7	1,2	0,0 - 11,5	0,94
3	1 ^{ste} persoon meervoud	0,7	0,0 - 5,1	0,4	0,0 - 3,1	0,92
4	Totaal eerste persoon	2,4	0,0 - 9,1	2,0	0,0 - 11,5	0,87
5	Totaal tweede persoon	1,1	0,0 - 7,7	1,1	0,0 - 7,1	0,98
6	Totaal derde persoon	3,9	0,2 - 16,7	3,6	0,4 - 12,8	0,96 ^a
7	Ontkenning	1,8	0,1 - 3,5	1,7	0,0 - 3,7	0,91 ^a
8	Toestemming/instemming/ goedkeuring	0,1	0,0 - 1,3	0,2	0,0 - 1,2	0,55
9	Lidwoorden	7,9	3,8 - 15,4	9,6	6,0 - 17,2	0,73 ^a
10	Voorzetsels	13,6	7,1 - 19,2	12,6	4,5 - 20,2	0,79 ^a
11	Getallen	1,2	0,3 - 9,6	0,9	0,1 - 7,7	0,88
II Psychologische processen						
12	Emotionele processen	3,4	1,1 - 8,1	2,9	0,6 - 7,1	0,85 ^a
13	Positieve emoties	2,0	0,6 - 4,6	1,4	0,1 - 5,4	0,75
14	Positieve gevoelens	0,3	0,0 - 1,1	0,1	0,0 - 1,2	0,52
15	Optimisme en energie	0,4	0,0 - 1,5	0,3	0,0 - 1,4	0,967 ^a
16	Negatieve emoties	1,3	0,0 - 4,0	1,3	0,0 - 3,2	0,72 ^a
17	Angst en vrees	0,2	0,0 - 0,9	0,1	0,0 - 0,7	0,59 ^a
18	Boosheid	0,3	0,0 - 1,2	0,1	0,0 - 0,9	0,57 ^a
19	Treurigheid of depressie	0,3	0,0 - 2,0	0,4	0,0 - 1,7	0,67
20	Cognitieve processen	6,1	1,6 - 8,8	5,4	1,7 - 8,1	0,70 ^a
21	Veroorzaken	0,8	0,0 - 2,2	0,6	0,0 - 2,5	0,62
22	Inzicht	1,8	0,2 - 4,0	1,8	0,5 - 3,8	0,59 ^a
23	Tegenstrijdigheid	1,8	0,0 - 4,1	2,0	0,0 - 3,8	0,57 ^a
24	Remming/geremdheid	0,3	0,0 - 1,0	0,1	0,0 - 0,9	0,53
25	Tentatief/voorzichtig/ aarzelend	1,7	0,1 - 3,6	1,1	0,2 - 2,9	0,57 ^a
26	Zekerheid	0,9	0,0 - 1,9	0,8	0,0 - 1,9	0,56 ^a
27	Zintuigen en perceptuele processen	2,2	0,2 - 5,6	1,9	0,0 - 5,4	0,86 ^a
28	Zien	0,6	0,0 - 3,4	0,5	0,0 - 3,4	0,77
29	Horen	0,9	0,0 - 3,0	0,8	0,0 - 3,2	0,87 ^a
30	Voelen	0,2	0,0 - 1,1	0,2	0,0 - 1,1	0,80 ^a

Tabel 3 (vervolg) *Equivalentie van de Engelse en Nederlandse LIWC gebaseerd op 75 teksten.*

Nr.	Woordcategorieën	Engels		Nederlands		r
		Mediaan	Min - Max	Mediaan	Min - Max	
31	<i>Sociale processen</i>	11,7	2,6 - 20,0	9,5	2,6 - 17,3	0,95 ^a
32	Communicatie	1,5	0,3 - 4,0	1,4	0,3 - 4,1	0,85 ^a
33	Verwijzingen naar andere personen	7,9	1,0 - 17,9	6,7	0,8 - 13,9	0,97 ^a
34	Vrienden	0,0	0,0 - 0,4	0,1	0,0 - 0,4	0,58 ^a
35	Familie	0,3	0,0 - 3,9	0,4	0,0 - 3,5	0,78
36	Mensen	0,9	0,0 - 2,4	0,5	0,0 - 2,8	0,72 ^a
III Relativiteit/ betrekkelijkheid						
37	<i>Tijd</i>	3,1	1,4 - 6,0	3,0	1,4 - 5,5	0,83 ^a
38	Werkwoorden in de verleden tijd	4,5	0,6 - 10,1	5,5	0,7 - 11,8	0,96 ^a
39	Werkwoorden in de tegenwoordige tijd	7,3	1,2 - 16,7	9,9	4,9 - 18,6	0,88 ^a
40	Werkwoorden in de toekomstige tijd	1,3	0,0 - 3,9	0,9	0,0 - 3,1	0,55 ^a
41	<i>Ruimte</i>	2,2	1,0 - 4,5	1,7	0,5 - 3,3	0,34 ^a
42	Omhoog/ de hoogte in	1,0	0,4 - 3,2	0,9	0,3 - 2,2	0,59 ^a
43	Beneden / neer	0,2	0,0 - 1,1	0,1	0,0 - 0,6	0,42
44	Inclusief	6,6	3,4 - 9,9	7,1	2,8 - 11,1	0,91 ^a
45	Exclusief	3,2	0,8 - 5,9	3,7	1,4 - 6,8	0,81 ^a
46	<i>Beweging</i>	1,1	0,0 - 2,9	1,4	0,3 - 3,8	0,63 ^a
IV Persoonlijke aangelegenheden						
47	<i>Beroep/bezigheden</i>	1,4	0,0 - 7,5	0,8	0,0 - 5,4	0,82
48	School	0,3	0,0 - 3,2	0,1	0,0 - 2,3	0,64
49	Werk	0,3	0,0 - 3,6	0,3	0,0 - 3,1	0,61
50	Prestatie	0,7	0,0 - 2,0	0,2	0,0 - 1,3	0,65 ^a
51	<i>Vrijtijdsbesteding</i>	0,6	0,0 - 1,9	0,3	0,0 - 2,1	0,67
52	In en om het huis	0,3	0,0 - 1,5	0,1	0,0 - 2,1	0,68
53	Sporten	0,0	0,0 - 0,9	0,0	0,0 - 0,7	0,40
54	Televisie en films	0,0	0,0 - 0,5	0,0	0,0 - 0,1	0,23
55	Muziek	0,1	0,0 - 1,3	0,0	0,0 - 1,0	0,40
56	<i>Geld en financiële zaken</i>	0,2	0,0 - 3,7	0,1	0,0 - 2,3	0,65
57	<i>Godsdienst / zingeving</i>	0,6	0,0 - 5,2	0,4	0,0 - 5,0	0,91
58	Religie	0,3	0,0 - 5,1	0,3	0,0 - 4,9	0,89
59	Dood en sterven	0,1	0,0 - 1,4	0,0	0,0 - 1,4	0,93
60	<i>Lichamelijke toestand</i>	1,3	0,0 - 6,3	1,1	0,0 - 5,0	0,78
61	Lichamelijke functioneren	0,8	0,0 - 5,7	0,7	0,0 - 4,3	0,79
62	Seks en seksualiteit	0,1	0,0 - 1,9	0,0	0,0 - 0,9	0,60
63	Eten, drinken en dieet	0,1	0,0 - 1,2	0,1	0,0 - 1,4	0,82
64	Slapen en dromen	0,1	0,0 - 1,2	0,1	0,0 - 1,2	0,79
65	Lichamelijke verzorging	0,0	0,0 - 0,7	0,0	0,0 - 0,2	0,50
V Experimentele dimensie						
66	Vloekwoorden	0,0	0,0 - 0,9	0,0	0,0 - 0,6	0,38

a = Pearson correlatie; de overige correlaties zijn Spearman rho correlaties

De volgende verklaringen kunnen gegeven worden voor de middelmatige en lage correlaties. Het '% afkortingen' is een telling die buiten het LIWC-woordenboek om gaat. Verschil in frequentie van gebruik van afkortingen in beide talen verklaart de lage correlatie. Bijvoorbeeld 'a.u.b.' in het Nederlands wordt in het Engels vertaald als 'please'. De woordcategorie 'ruimte' bevat woorden die in letterlijke zin een plaats aanduiden, bijvoorbeeld 'close' heeft de vertalingen 'aanengesloten', 'dichtbij', 'nabij' en 'vlakbij'. Als woorden uit deze woordcategorie in overdrachtelijke zin gebruikt worden blijken de vertalingen niet meer overeen te stemmen met de woorden uit deze woordcategorie, waardoor tellingen niet

meer parallel lopen. Bijvoorbeeld 'Close your eyes' heeft de vertaling 'doe je ogen dicht', waarbij het woord 'close' wel in de categorie 'ruimte' valt maar het woord 'dicht' niet. Woorden uit de woordcategorieën 'beneden/ neer', 'sporten', 'televisie en films', 'muziek' en 'vloekwoorden' komen weinig voor in de geanalyseerde teksten. Incidentele scores leiden snel tot lage correlaties.

Voor de woordcategorieën die het meest relevant zijn voor onderzoek naar expressie van emoties en het bereiken van gunstige cognitieve, sociale en gezondheidseffecten zijn de correlaties hoog: voornaamwoorden ($r = 0,93$), positieve emoties ($r = 0,75$), negatieve emoties ($r = 0,72$) en cognitieve processen ($r = 0,70$).

De effectgrootte is voor 92% van de woordcategorieën middelmatig of klein ($< 0,8$). Dit betekent dat het verschil tussen het gemiddelde aantal woorden in een woordcategorie van de Engelse en Nederlandse tekst niet groot was. Voor vijf woordcategorieën werd wel een groot verschil tussen de Nederlandse en Engelse versie ($d > 0,8$) gevonden. Dit waren de woordcategorieën lidwoorden ($d = 0,87$), remming / geremdheid ($d = 0,99$), ruimte ($d = 1,13$), werkwoorden in de tegenwoordige tijd ($d = 0,89$), en prestatie ($d = 1,23$).

De grote effectgrootten kunnen op de volgende manier verklaard worden. De woordcategorie 'lidwoorden' bestaat in het Engels uit 'a, an, the' en in het Nederlands uit 'een, de, het'. In het Engels wordt een zelfstandig naamwoord minder vaak voorafgegaan door een lidwoord. Daarnaast is het Nederlandse lidwoord 'het' ook een persoonlijk voornaamwoord ('het is moeilijk'), een onbepaald voornaamwoord ('het regent') en een zelfstandig naamwoord ('dat is het'), waardoor het vaker geteld wordt in teksten. Uit de woordcategorie 'remming/ geremdheid' zijn een aantal woorden weggelaten omdat zij in de Nederlandse vertalingen niet goed in deze woordcategorie pasten. De woordcategorie 'werkwoorden in de tegenwoordige tijd' bevat in de Nederlandse versie veel meer vervoegde werkwoorden dan in de Engelse versie. De woorden uit de woordcategorie 'ruimte' worden vooral in het Engels ook vaak in overdrachtelijke zin gebruikt. De woordcategorie 'prestatie' bleek nogal typisch Amerikaans te zijn. In de Nederlandse cultuur worden veel van deze woorden wel gebruikt, maar niet geassocieerd met prestatie, waardoor vertalingen niet in het Nederlandse LIWC-woordenboek zijn opgenomen. Van het woord 'purpose' is bijvoorbeeld wel het woord 'doel' maar niet het woord 'bedoeling' in de woordcategorie 'prestatie' geplaatst.

De effectgrootten van vier belangrijke woordcategorieën variëren van geen verschil tussen de Nederlandse en Engelse versie tot een middelmatig verschil ($0,5 < d < 0,8$): voornaamwoorden ($d = 0,45$), positieve emoties ($d = 0,76$), negatieve emoties ($d = 0,03$) en cognitieve processen ($d = 0,32$).

Tot besluit

De Nederlandse LIWC is een onderzoeksinstrument dat is ontwikkeld voor toepassing in onderzoek naar effecten van expressie van emoties in Nederlandstalige teksten. Het Nederlandse LIWC-woordenboek telt 6.568 woorden in 66 woordcategorieën. Tussen het Nederlandse en Engelse LIWC-woordenboek zijn twee belangrijke verschillen. In de Nederlandse versie worden meer werkwoorden vervoegd dan in de Engelse versie, zodat de woordcategorieën 'werkwoorden in de verleden tijd', 'werkwoorden in de tegenwoordige tijd' en 'werkwoorden in de toekomstige tijd' een beter beeld geven van in de Nederlandse taal gebruikte woorden. Daarnaast bevat de Nederlandse versie, met uitzondering van de categorie 'getallen' geen woordstammen, omdat deze leiden tot veel ruis in de resultaten: woorden die worden gescoord in verkeerde woordcategorieën.

De equivalentie tussen het Nederlandse en het standaard Engelse LIWC-woordenboek is redelijk tot goed te noemen. Juist voor de woordcategorieën die het meest relevant lijken voor onderzoek naar uiting van emoties en het bereiken van gunstige cognitieve, sociale en gezondheidseffecten blijken de verschillen klein te zijn. Voor een gering aantal woordcategorieën is sprake van een slechte equivalentie. Dit is het gevolg van taalverschillen en een bewuste beslissing om minder ruis te creëren. Het betreft voornamelijk woordcategorieën die tot nu toe minder relevant zijn gebleken in onderzoek naar expressie van emotie. Momenteel wordt de validiteit en betrouwbaarheid van de Nederlandse LIWC onderzocht onder meer door woordgebruik van emotionele expressie te vergelijken met neutrale instructies. Ook zal de convergente validiteit worden bestudeerd door de LIWC-scores op de woordcategorieën 'positieve en negatieve emoties' te vergelijken met scores verkregen met het Nederlandstalige instrument dat is ontwikkeld om positieve en negatieve emoties in geschreven Nederlandse teksten te scoren (Vording, Weiss, Van Oeveren & Garssen, 1999). De Nederlandse LIWC kan van waarde zijn in divers onderzoek waarbij taalgebruik in algemene zin centraal staat zoals in taalonderzoek, onderzoek naar mediareacties na emotionele gebeurtenissen als een ramp, nationale rouw of een sportoverwinning, cognities van patiëntengroepen en therapie-evaluatieonderzoek. Tegenwoordig worden interventies gericht op emotionele expressie ook aangeboden via Internet met op het individu toegesneden feedback (Lange, Rierdijk, Hudcovicova, Van der Ven, Schrieken & Emmelkamp, 2003). Tekstanalyseprogramma's zouden misschien in de toekomst kunnen worden geïntegreerd in dergelijke interventies. Er zou dan geautomatiseerd enige controle kunnen zijn op de voortgang van het proces en eventuele risicovolle situaties zouden in een vroeg stadium kunnen worden ontdekt, zodat een professionele hulpverlener kan ingrijpen (Van Middendorp & Geenen, 2004). De tabel met medianen en correlaties van de Nederlandse en Engelse versie, zoals gepresenteerd in dit artikel kan worden gebruikt als referentietabel.

Een beperking van het LIWC-programma is dat het uitsluitend woorden telt. Het programma geeft geen informatie over het onderwerp waarover geschreven is en houdt geen rekening met de context waarin woorden gebruikt worden. Dit kan zorgen voor enige vertekening van de resultaten, met name bij woorden die meerdere betekenissen hebben. Een voorbeeld hiervan is het woord 'loslaten', dat zowel in letterlijke zin als in overdrachtelijke zin gebruikt kan worden.

Met het beschikbaar komen van de Nederlandse LIWC kunnen nu ook Nederlandse teksten op snelle en efficiënte wijze geanalyseerd worden. Het programma biedt de gebruiker de mogelijkheid om zelf aan te geven op welke woordcategorieën gescoord dient te worden. Verder zijn de LIWC-resultaten direct beschikbaar voor interpretatie en statistische bewerking. De equivalentie tussen het Nederlandse LIWC-woordenboek en het standaard Engelse LIWC-woordenboek is redelijk tot goed. Het woordenboek is geschikt voor gebruik in onderzoek naar expressie van emoties en ander onderzoek gericht op verbale expressie in de Nederlandse taal.

Literatuur

- Bushman, B.J. (2002). Does venting anger feed or extinguish the flame? Catharsis, rumination, distraction, anger, and aggressive responding. *Personality and Social Psychology Bulletin*, 28, 724-731.
- Campbell, R.S. & Pennebaker J.W. (2003). The secret life of pronouns: Flexibility in writing style and physical health. *Psychological Science*, 14, 60-64.
- Coenders, H. (1996a). *Kramers Handwoordenboek Engels-Nederlands Nederlands-Engels*. Zutphen: Koninklijke Wöhrmann.
- Coenders, H. (1996b). *Kramers Vertaalwoordenboek Engels-Nederlands. Nieuwe spelling*. Amsterdam: Bonaventura.
- Cohen, J. (1977). *Statistical power analysis for the behavioral sciences*. New York: Academic Press.
- Lange, A., Rietdijk, D., Hudcovicova, M., Ven, J. P. van de, Schrieken, B. & Emmelkamp, P. M. (2003) Interapy: a controlled randomized trial of the standardized treatment of posttraumatic stress through the internet. *Journal of Consulting and Clinical Psychology*, 71, 901-909.
- Lumley, M.A., Tojek, T.M. & Macklem, D.J. (2002). The effects of written emotional disclosure among repressive and alexithymic people. In S.J. Lepore & J.M. Smyth (Eds.). *The writing cure. How expressive writing promotes health and emotional well-being* (pp. 75-79). Washington, DC: American Psychological Association.
- Lutgendorf, S.K. & Ullrich, P. (2002). Cognitive processing, disclosure, and health: Psychological and physiological mechanisms. In S.J. Lepore & J.M. Smyth (Eds.). *The writing cure. How expressive writing promotes health and emotional well-being* (pp. 177-196). Washington, DC: American Psychological Association.
- Middendorp, H. van & Geenen R. (2004). Onze allerdiepste gedachten en gevoelens over emotionele expressie. *Gedrag & Gezondheid*, 32, 40-48.
- Pennebaker, J.W. & Beall, S.K. (1986). Confronting a traumatic event: Toward an understanding of inhibition and disease. *Journal of Abnormal Psychology*, 95, 274-281.
- Pennebaker, J.W. & Francis, M. (1996). Cognitive, emotional, and language processes in disclosure. *Cognition and Emotion*, 10, 601-626.
- Pennebaker, J.W., Francis, M.E. & Booth, R.J. (2001). *Linguistic Inquiry and Word Count, LIWC 2001*. Mahwah, NJ: Erlbaum.
- Pennebaker, J.W. & Graybeal, A. (2001). Patterns of natural language use: Disclosure, personality, and social interaction. *Current Directions in Psychological Science*, 10, 90-93.
- Pennebaker, J.W., Mayne, T. & Francis, M. (1997). Linguistic predictors of adaptive bereavement. *Journal of Personality and Social Psychology*, 72, 863-871.
- Pennebaker J.W. & Susman, J.R. (1988). Disclosure of traumas and psychosomatic processes. *Social Science and Medicine*, 26, 327-332.
- Smyth, J.M. (1998). Written emotional expression: Effect sizes, outcome types, and moderating variables. *Journal of Consulting and Clinical Psychology*, 66, 174-184.
- Stroebe, M., Stroebe W., Schut, H., Zech, E. & Bout, J. van den (2002). Does disclosure of emotions facilitate recovery from bereavement? Evidence from two prospective studies. *Journal of Consulting and Clinical Psychology*, 70, 169-178.
- Vording, J.R., Weiss, J.J., Oeveren, I. van & Garssen, B. (1999). Ontwikkeling van een methode voor het meten van emotionele expressie in Nederlandse geschreven teksten. *Gedrag & Gezondheid*, 27, 148-157.