



Beleefdheid op Twitter

(On)beleefde nachtbrakers

Mike de Wildt

Begeleider: Dr. J.J. Paijmans

Tweede lezer: Dr. M. M. van Zaanen

Faculteit Geesteswetenschappen aan de Universiteit van Tilburg

Communicatie- en Informatiewetenschappen

Bedrijfscommunicatie & Digitale Media

22 maart 2012

Voorwoord

Deze scriptie is geschreven ter afsluiting van de master Bedrijfscommunicatie en Digitale Media aan het Departement Communicatie- en Informatiewetenschappen van de faculteit Geesteswetenschappen, Tilburg University.

Direct na het afsluiten van mijn HBO studie (Communication, Advertising and Design Management) in 2008 heb ik in overleg met de docenten aldaar een vervolgstudie uitgekozen. Na een looptijd van drie studiejaren en zes maanden is dit vervolg dan eindelijk afgelopen. Tilburg University en de bijhorende studentenperiode hebben veel van mij gevraagd, en daarvoor heb ik dubbel zoveel teruggekregen. Mijn competenties, mijn kennis en mijn vaardigheden zijn enorm gegroeid. Daarnaast heb ik vrienden voor het leven leren kennen en een groot sociaal netwerk opgebouwd. Met tevredenheid kijk ik terug op deze periode.

Om een geschikt onderwerp te vinden voor mijn scriptie heb ik een andere keuze moeten maken dan in eerste instantie mijn idee was. Een onderzoek naar videogames (in het veld van cognitiwetenschappen) was niet haalbaar in Tilburg, althans niet op de manier zoals ik dat wilde. Het huidige onderwerp betreft een ander veld van interesse: online sociale media en taal. Geïnspireerd door de cursus Sociolinguïstiek, is taal voor mij een uitermate interessant onderwerp geworden. Om die reden heb ik besloten om naar dr. A.R. Vermeer te stappen met de vraag of hierin een interessant onderwerp te vinden was. Omdat grof taalgebruik mij al eerder was opgevallen als moderator op gamingfora, heb ik deze constatering gekoppeld aan de mogelijke aanwezigheid van onbeleefde taal op het populaire sociale microblog Twitter.

Terugkijkend op deze studie is het afstuderen mij het meest belastend gebleken. Er komt meer kijken bij een groots opgezette scriptie dan bij het voorbereiden van bijvoorbeeld een tentamen. Daarnaast zorgt een bestuursjaar tijdens je afstuderen ook niet voor een optimale focus op het werk. Ondanks de werkdruk van de scriptie, mijn bijbaan, mijn voorzitterschap bij mijn hockeyvereniging SHOT, mijn relatie en het onderhouden van mijn sociale contacten is dan toch de tijd aangebroken om af te studeren. Dit had ik echter niet kunnen doen zonder de toewijding en het enorme enthousiasme van Hans Paijmans, de steun en liefde van Lindsay, de ideeën en motivationele speeches van Björn en tot slot de support, in de breedste zin van het woord, van mijn familie. Mijn dank gaat uit naar al deze personen die ieder op hun eigen manier van groot belang zijn geweest voor de totstandkoming van deze scriptie.

Mike de Wildt, 22 maart 2012 te Tilburg

Michiel ‘Gilles’ Veenstra

13-8-1986 - 7-5-2010

Inhoudsopgave

Voorwoord	II
Samenvatting	6
1: Inleiding en probleemstelling.....	7
1.1 Korte introductie van Twitter	9
1.2 Van normen en waarden naar grof taalgebruik en #dreigtweets	11
1.3 Jongerencultuur en taal.....	12
2: Theoretisch kader	15
2.1 Inleiding.....	15
2.2 Theorieën aan de basis van dit onderzoek	15
2.2.1 Beleefdheidstheorie.....	15
2.2.2 Cues-Filtered-Out (CFO) theorie/perspectief	17
2.2.3 Social Information Processing (SIP) theorie.....	18
2.2.3.1 Hyperpersonal model	20
2.3 Aanzet tot het gebruik van scheldwoorden.....	20
2.4 Het construct beleefdheid	23
3: Onderzoeksopzet	24
3.1 Methode	24
3.2 Materiaal.....	24
3.3 Procedure	24
3.4 Verzameling en verwerking van de gegevens	24
3.4.1 Vier weken tweets uit heel Nederland	27
3.4.2 Geen toestemming bestanden uit SoNaR corpus	28
3.4.3 Toestemming bestanden uit SoNaR corpus	28
4: Opzet van het onderzoek	29
4.1 Vraagstelling en hypothesen.....	29
5: Resultaten	33
5.1 Resultaten grote dataset	33
5.1.1 Heel Nederland verdeeld over vier weken	33
5.1.2 Omgeving Tilburg.....	40

5.1.3 Nieuwjaarsdag 2012.....	42
5.1.4 Frequentie van type scheldwoorden.....	43
5.2 Resultaten SoNaR: STEVIN dataset	44
5.2.1 Geen toestemming groep	44
5.2.2 Met toestemming groep	45
5.3 Samenvatting van de resultaten	46
6: Conclusie en discussie	48
6.1 Inleiding.....	48
6.2 Conclusie	48
6.3 Discussie.....	50
6.3.1 Limitaties en vervolgstudies	50
6.3.2. Interpretatie van de resultaten	51
Referenties.....	53
Bijlage I – Neologismen en de verklarende woordenlijst voor Twitter	58
Bijlage II – Tabel uit “woorden van en voor emotie” Van Sterkenburg ‘07	60
Bijlage III - Veelgebruikte bestandstypen voor Twitteronderzoek	61
Bijlage IV – Scripts	63
Bijlage V – Scheldwoordenlijst	67
Bijlage VI – Meest frequent gebruikte scheldwoorden (absolute aantallen)	68
Bijlage VII – Frequentiegrafieken van alle tweets (per dag).....	69
Bijlage VIII – Frequentiegrafieken van alle onbeleefde tweets (per dag).....	71
Bijlage IX - Matrixen van alle twitterdata.....	73
Bijlage X – Zoekopdracht E.F. Tjong Kim Sang	82
Bijlage XI - SPSS Output.....	83
Dankwoorden.....	85

Samenvatting

Het taalgebruik van personen op verschillende online communicatieplatformen is onderhevig aan verandering, zo ook op Twitter. Daaraan toegevoegd maakt men steeds vaker gebruik van krachttermen, iets dat in deze scriptie wordt aangeduid als „onbeleefdheid”. Dit beleefdheidsonderzoek op Twitter analyseert de Nederlandstalige tweets van miljoenen unieke twitteraars op verschillende tijdstippen en weekdagen gedurende vier weken in januari 2012. Er wordt een verschil in de mate van beleefdheid verwacht gedurende de dag en gedurende de gehele week. Daarnaast wordt er verwacht of er een verschil bestaat tussen Tilburg en heel Nederland. Vervolgens wordt er verwacht dat twitteraars afkortingen en alternatieve schrijfwijzen voor scheldwoorden gebruiken in hun tweets. Tot slot wordt er verwacht dat twitteraars die tweets met toestemming afstaan, zodat die kunnen worden gebruikt voor wetenschappelijk onderzoek, minder onbeleefde uitingen bevatten dan in tweets van twitteraars die geen toestemming hebben gegeven om hun tweets te gebruiken.

Voor dit onderzoek zijn er bijna 65 miljoen tweets gedownload van ongeveer 2,2 miljoen twitteraars. Met behulp van awk scripts is er uit een UNIX *commandline* omgeving informatie onttrokken uit de tweets. Enkele belangrijke gegevens die hieruit voort zijn gekomen, zijn de aantallen (onbeleefde) tweets per uur en per dag van week, frequentieoverzichten van hoort soort scheldwoord dat is gebruikt, overzichten van twitteraars en geolocaties.

Er zijn statistisch significante resultaten gevonden met betrekking tot de mate van beleefdheid in Nederlandstalige tweets. Er wordt ’s nachts relatief veel gescholden, terwijl het absolute piekmoment in Nederland in de avonds ligt. Over het algemeen wordt er meer gescholden dan wanneer er wordt gecommuniceerd via andere kanalen zoals chat en sociaal netwerk MySpace en men scheldt vooral vaker in het weekend. De keuze voor scheldwoorden in de tweets is eenduidig te noemen. Het overgrote deel van de scheldwoorden komt overeen in alle vier de groepen. Verrassend was het resultaat op de vierde hypothese. Het zijn juist de groepen die geen toestemming geven om tweets af te staan waarbij de beleefdheid hoger is.

Er is aangetoond dat er inderdaad veel onbeleefd taalgebruik op Twitter in Nederland voorkomt, terwijl dat in de groepen van het SoNaR: STEVIN juist amper het geval was. De reden hiervoor kan waarschijnlijk gezocht worden in de wijze van dataverzameling (at random en niet at random). De meeste resultaten zijn dan ook onttrokken aan de grote dataset.

Keywords: scheldwoorden, grof taalgebruik, onbeleefdheid, Twitter

1: Inleiding en probleemstelling

Taalverruwing, oftewel het toenemende gebruik van onder andere scheldwoorden, vloeken en verwensingen, is aan de orde van de dag en het lijkt (volgens een deel van de bevolking) bij te dragen aan de *verhuftering* van Nederland. Zowel via de digitale- als via de analoge weg wordt er vaker en grover gescholden¹ in het dagelijks leven dan bijvoorbeeld een decennium geleden (de Jongh, 2007). Ook de woordkeuze is onderhevig aan verandering. Daar waar de Duitsers en Italianen het achterwerk gebruiken om iemand op zijn plek te zetten, daar benoemen de Turken liever de familieleden van anderen in hun beledigende boodschappen. De Grieken vloeken voornamelijk met handicaps terwijl de Vlamingen voornamelijk schelden met genitaliën en fecaliën. Bij Nederlanders vindt een verschuiving plaats van vloeken waarin God voorkomt naar de Angelsaksische gewoonte waarin men onder andere met *shit* en *fuck* scheldt (Sterkenburg, 2007 en zie voetnoot 2). Maar waar Nederland vooral om bekend staat is het schelden met (ernstige) ziektes om hevige emoties te uiten. Hierin zijn Nederlanders haast uniek, want enkel in het Hebreeuws komt dit nog voor². Bij de Nederlanders komt dit met name door de invloed van de calvinistische mentaliteit: vloeken uitspreken over datgene waar men zelf het bangst voor is (Sterkenburg, 2007). Een voorbeeld hiervan is „de kanker”. Het woord wordt gebruikt als een ernstige verwensing, maar tegenwoordig wordt het ook gebruikt als bijvoeglijk naamwoord.

De inspiratie van deze scriptie komt voornamelijk uit het werk van Prof. Dr. Van Sterkenburg (2007). In zijn afscheidsrede als hoogleraar in de Lexicologie benoemde hij scheldwoorden welk hij koppelde aan het aantal hits op Google en gerangschikt van hoog naar laag. De raakvlakken met het huidige idee van onbeleefd taalgebruik op Twitter zijn alom aanwezig. De uitkomsten van zijn werk maken het interessant om een masterstudie te doen naar het onderwerp van deze scriptie. Een passage uit Van Sterkenburg’ werk luidt als volgt: “Mensen gebruiken taal naast andere middelen omdat ze een grote behoefte hebben om hun emoties uit te drukken. Vloeken, verwensingen, scheldwoorden, taboewoorden enz. zijn daardoor hoogfrequent in onze communicatie, zoals de volgende tabel (zie bijlage II) wil aantonen³” (Sterkenburg, 2007).

Scheldwoorden en vloeken worden niet alleen geuit op straat of in de huiskamer, maar ook op sociale media websites. Daarin representeert een gebruiker zichzelf door middel van

¹ <http://www.trouw.nl/tr/nl/4324/nieuws/archief/article/detail/1662202/2007/05/08/Taalverruwing-op-televisie->

² <http://taalschrift.org/reportage/001784.html> We vloeken anders dan vroeger.

³ Zie bijlage II.

een gebruikersnaam en in de meeste gevallen een profielfoto. Facebook biedt gebruikers de mogelijkheid om bijnamen te gebruiken wanneer ze zich registreren op dit sociale netwerk. Maar het is haast een ongeschreven regel dat de ware identiteit wordt blootgegeven. Dit komt voort uit de herkomst van Facebook als een communicatiemiddel voor studenten tussen universiteiten. Het lag in de lijn der verwachting dat men zijn of haar „echte’ naam opgaf. „Geen pseudoniemen, maar meer integriteit’ zo zegt Mark Zuckerberg, oprichter en CEO van Facebook. Google+ gaat een stap verder en eist door middel van hun *policy* dat men gebruik maakt van echte namen – een keuze die zwaar onder vuur ligt bij de (potentiële) gebruikers van Google+.

Elk sociaal netwerk (i.e. Facebook, Google+, LinkedIn, MySpace, Twitter en vele anderen) wordt voor afzonderlijke zaken gebruikt en verschillende connecties worden wel (of juist niet) toegelaten door de eigenaar van het profiel. Dit heeft als gevolg dat het taalgebruik ook afgestemd is op diegenen die jouw berichten kunnen lezen, zoals dat ook in de echte wereld gebeurt. Een gebruiker van Facebook laat connecties toe tot zijn Facebookprofiel uit persoonlijke, offline, sociale kringen. Op LinkedIn worden bedrijven, medestudenten of collega’s (waaronder ook de baas) toegevoegd als connectie. Het verschil zit in het doel van de *Social Network Sites* (SNSs) en de content die daar op komt te staan. Gebruikers hebben dit zelf in de hand, maar verwacht wordt dat men zich gedraagt zoals ze dat ook in de echte wereld zouden doen.

Op fora gaat dat anders. Daar wordt men streng aangesproken door de dienstdoende *moderators*, een soort digitale politie, terwijl op de sociale media deze rol wordt vervuld door „vrienden’. Vrienden beschikken echter niet over dezelfde autoriteit als een *moderator*. Dit heeft als gevolg dat de controle zoals die op fora was (en is) eigenlijk niet bestaat binnen sociale media, waardoor men tegenwoordig zelf verantwoordelijk is voor het schetsen van een juist beeld van zichzelf. Tweets met daarin beledigende content rapporteren is niet mogelijk.

Wanneer men in tweets vormen van taalverruwing niet schuwt, verschuilen zij zich niet meer achter een (anonieme) avatar⁴ of verzonnen identiteiten, zoals dat op fora mogelijk was. Men speelt juist open kaart en staat veelal geregistreerd onder de eigen naam of anders wel met een profielfoto en daarmee wordt het Internet steeds meer een afspiegeling van de „echte’ wereld. Vanwege die constatering zal men tweemaal het taalgebruik moeten overdenken alvorens iets te plaatsen op de SNSs.

Het bestuderen van de verruwing van taal op Twitter heeft betrekking op drie zaken.

⁴ Een virtuele representatie van jezelf. Dit kan een afbeelding zijn of een personage (in bijvoorbeeld een virtuele wereld zoals een videogame). Avatar komt uit het Sanskriet (Avatara) hetgeen incarnatie betekent.

Deze drie punten worden vervolgens in voorbeelden toegelicht. Taalverruwing op Twitter staat in verband met:

1. taalontwikkeling;
2. het ontstaan van niet-wenselijke situaties;
3. gezichtsverlies.

Twitter, maar ook andere SNSs, dragen bij aan de ontwikkeling van nieuwe taalelementen die weer kunnen leiden tot taalverruwing. Veel gebruikte afkortingen (die via het Internet zijn intrede deden in de Nederlandse taal) krijgen het Engelse *„fuck(ing)“* met zich mee om sterker en stoerder over te komen en in Nederland is tegenwoordig alles „kk⁵ dom/gaaf/slecht/mooi“. Voorbeelden om de drie eerdergenoemde punten toe te lichten luiden als volgt:

- ad1. Continu ontstaan er afkortingen van woorden en compleet nieuwe woorden die gebruikt worden in tweets, status updates op Facebook en WWW's⁶ op Hyves. Doordat een restrictie in de hoeveelheid woorden die gebruikt mogen worden, wordt de gebruiker van het communicatiemedium gedwongen creatief te zijn.
- ad2. Het komt ook voor dat er dreigementen op Twitter worden geplaatst, zogenaamde dreigtweets. Twitter is de laatste jaren veel in het nieuws geweest, zowel in Nederland als in het buitenland, naar aanleiding van door twitteraars ingezonden dreigtweets. Een dreigtweet houdt in dat een persoon een tekst tweet om zodoende, bedoeld of niet, onrust te veroorzaken. Hiervoor zijn zelfs Twitteraccounts geopend waarop dreigementen worden geplaatst (meer over dreigtweets in hoofdstuk 1 par. 2).
- ad3. Door te twitteren creëert een twitteraar een beeldvorming voor andere twitteraars, deze kan positief en negatief zijn. Het type taalgebruik heeft invloed op zaken die buiten Twitter om afspelen. Hierover meer in paragraaf 2.2.1.

1.1 Korte introductie van Twitter

Twitter, het communicatie systeem dat werd gelanceerd in 2006 onder de vlag van oprichter Obvious Corp (San Fransisco) heeft als hoofddoel het in (maximaal) 140 karakters weergeven van een status- of nieuwsbericht. Met name dat laatste, verslaglegging van de actualiteiten, is waar Twitter zich goed voor leent en waarmee het zich meteen onderscheidt van de meeste andere SNSs.

⁵ De lettercombinatie „KK“ wordt veelal gebruikt als afkorting van het scheldwoord kanker.

⁶ Een WWW op Hyves is een afkorting voor Wie Wat Waar. Een manier om in het kort aan al je Hyvesvrienden te vertellen wat je op dat moment aan het doen bent.

Het microblog heeft in het verleden moeite gehad met het creëren van eigen onderscheidend vermogen. De kernfuncties van Twitter liggen volgens Kevin Thau, Twitter's *Vice President business & corporate development*, anders dan op social media websites zoals Facebook. Twitter is volgens hem voor „nieuws, content en informatie”. Desalniettemin, wanneer de volgende definitie van een social media website wordt aangehouden, dan blijkt Twitter op alle punten te scoren: “*Any website designed to allow multiple users to publish content themselves. The information may be on any subject and may be for consumption by (potential) friends, mates, employers, employees, etc. The sites typically allow users to create a ‘profile’ describing themselves and to exchange public or private messages and list other users or groups they are connected to in some way. There may be editorial content or the site may be entirely user-driven.*” – Dictionary.com⁷.

Hetgeen waar Twitter om draait, de statusupdates van de gebruikers, is in de loop der jaren veranderd. In de begindagen van Twitter was dit “Wat ben je aan het doen?”, terwijl sinds kort is dit veranderd in “Wat gebeurt er?”. Dit geeft de nieuwe *core business* van Twitter weer waarmee Twitter zich voornamelijk profileert als dienst die snel berichten over de actualiteiten laat uitwisselen.

Gegeven het bovenstaande verspreiden nieuwsberichten zich snel over de grote gemeenschap die actief is op Twitter. Deze online gemeenschap (ook wel twitterverse) bestaat uit gebruikers (tweeps) die berichten (tweets) kunnen plaatsen en die berichten kunnen doorsturen van anderen (re-tweeten). De tweets kunnen worden verstuurd vanaf de officiële website www.twitter.com en vanaf mobiele applicaties voor tablets en mobiele telefoons. Volgers van het account waarvan het bericht is verstuurd kunnen vervolgens dit bericht ook lezen. Wanneer de twitteraar de tweet interessant genoeg vindt om te re-tweeten, dan kunnen al zijn of haar volgers dat bericht ook lezen. Twitter heeft in het verleden aan de basis gestaan van nieuwswaardige gebeurtenissen zoals het oppakken van verdachten⁸, het wereldwijd beïnvloeden van de financiële beurzen⁹ en complete revoluties in het Midden-Oosten¹⁰.

Twitter kent nog een functionaliteit, de *hashtag* oftewel het „hekje” (#). De hashtag wordt door een twitteraar in een tweet geplaatst om aan te geven wat het onderwerp van de tweet is. Vervolgens verzamelt Twitter al de identieke hashtags die vervolgens als zogenaamde *trending topics* (i.e. lijst van de meest populaire onderwerpen op Twitter op dat moment) zichtbaar is op twitter.com. Handig voor specifieke marketingcampagnes,

⁷ <http://www.inc.com/tech-blog/is-twitter-a-social-network.html>.

⁸ <http://www.tctubantia.nl/regio/twente/9439548/Opgepakt-na-tweet-over-'bloedbad'-.ece>.

⁹ <http://www.wired.com/wiredscience/2010/10/twitter-crystal-ball/>.

¹⁰ http://www.foreignpolicy.com/articles/2011/01/14/the_first_twitter_revolution.

voorspellingen (zoals de beurs) en het starten van complete revoluties (e.g. Arabische lente).

Vijf jaar na de oprichting van Twitter is de SNS een populair communicatiemedium geworden. Per week worden er wereldwijd een miljard tweets verstuurd. Het belang van het delen van (lokaal) nieuws en de wijze waarop dit geschiedt is onderwerp van diverse studies geweest (Smith, 2011). Men twittert om hun dagelijkse activiteiten mede te delen en om informatie te zoeken en delen met andere twitteraars (Java, Song, Finin, & Tseng, 2007). Bedrijven rekruteren nieuwe werknemers en promoten hun producten via Twitter¹¹.

1.2 Van normen en waarden naar grof taalgebruik en #dreigtweets

In 2009 deed *verhuftering* zijn intrede in de Nederlandse taal¹². Het is een verwijzing naar de situatie waar de maatschappij volgens sommigen op het moment aan onderhevig is. De basis van deze observatie hiervan ligt in het normen- en waardendebat aan het einde van 2002, onder leiding van toenmalig minister-president Jan-Peter Balkenende (Balkenende I). Na een periode van ruim een jaar werden de normen en waarden van „de Nederlander” naar de achtergrond verdrongen, waarna in 2004 het begrip weer werd gebruikt tijdens een Europese conferentie onder leiding van Balkenende. Balkenende gebruikte hiervoor ter inspiratie de kennis van socioloog Amitai Etzioni¹³. Die stelt dat niet de overheid, maar de gemeenschap de norm van een samenleving bepaalt (Etzioni, 1996). TNS-NIPO deed vervolgens onderzoek naar de populariteit van het normen- en waardendebat in Nederland waaruit zou blijken dat Balkenende de samenleving goed aanvoelde. 87 procent van de participanten achtte „het een goede zaak dat Balkenende de discussie over normen en waarden stimuleert” en zelfs 90 procent geeft aan dat het debat voortgezet moet worden (TNS-NIPO, 2006).

Nu Twitter populairder wordt en de totale frequentie aan tweets toeneemt (wereldwijd worden er per week een miljard tweets verstuurd), merken de politici in de Tweede Kamer op dat er sinds 2010 ook een toename van het aantal dreigtweets is, waardoor er volgens hen ook online sprake is van *verhuftering*. Daarop diende demissionair Minister van Justitie E.H. Ballin op 28 juli 2010 een wetsontwerp in waarmee (zonder rechter) content van websites gehaald kon worden. Het Internet zou hiermee tevens een onderdeel van de grondwet worden. Tweets zouden in dat geval direct tot een straf kunnen leiden in de toekomst.

Sindsdien hebben enkele tweets het landelijke nieuws gehaald waaronder de volgende

¹¹ Jobvite's juni 2011 onderzoek geeft aan dat 47% van de ondervraagde HR-managers en recruitment professionals (N=800) via Twitter nieuwe werknemers zoekt.

¹² <http://www.grenswetenschap.nl/permalink.asp?grens=2952>.

¹³ The New Golden Rule, A. Etzioni (1996).

voorbeelden: een Amersfoortse scholiere (17) die haar school wilde opblazen¹⁴, een scholier (13) uit Tolbeek die zijn school dreigde op te blazen¹⁵, een jongen (15) die de terreurdaad van Karst T. wilde herhalen¹⁶ en zo zijn er inmiddels een behoorlijk aantal tweets die de landelijke media hebben bereikt. Maar het kan nog verder gaan, waarbij zelfs (preventieve) arrestaties werden verricht: Een Rotterdammer (17) die de schietpartij in Alphen aan den Rijn¹⁷ wilde overtreffen kreeg een celstraf¹⁸, een Twitteraccount¹⁹ met 1.634 volgers (22 maart 2012) en bijhorende website²⁰ die zijn gemaakt om landelijke doodsverwensingen weer te geven en tot slot de bedreiging en het terroriseren van een scholiere (13) waarbij de politie moest ingrijpen.

1.3 Jongerencultuur en taal

Het lijkt aannemelijk dat het overgrote deel van de tweets met onbeleefde taalvormen afkomstig zijn van jongeren en een kleiner deel afkomstig van volwassenen. Elke generatie jongeren kent volgens de historie afkeer tegen de gevestigde orde en de snelle, wereldwijde communicatie helpt deze generatie een stem te geven. Jongeren hanteren een eigen systeem van normen en waarden, waarin taal een belangrijke functie vervult en een veranderende maatschappij leidt tot lexicale veranderingen (Holmes, 2008) zoals bij de opkomst van een medium zoals Twitter. Straattaal, jongerentaal, turbotaal, *slang* of Internettaal. Het zijn allemaal benamingen van de door de jeugd gesproken en geschreven talen naast, dan wel gemixt, met het Standaard Nederlands.

Sinds het verschijnen van het Internet in zijn huidige vorm (consumenten Internet dateert in Nederland uit 1993, XS4ALL) is het taalgebruik op dit medium aan verandering onderhevig. Een verklaring daarvoor kan worden gevonden in het feit dat het Internet in de loop der tijd steeds meer is gebruikt als een netwerk van sociale contacten met aanvullende optie het opvragen van informatie, waardoor het persoonlijker wordt benaderd door de gebruikers. Dit wordt ook wel de overgang van Web 1.0 naar Web 2.0 genoemd.

Een voorbeeld van één van de eerste taalveranderingen door het gebruik van de computer is de (combinatie)taal „L33t”. L33t is een verbastering van elite en wordt ook weleens gespeld als 1337. In L33t worden veelal klinkers vervangen door getallen zodat de gebruikers, met name hackers en gamers, niet ontdekt konden worden door (overheids)

¹⁴ <http://www.scholieren.com/artikel/1183/leerling-opgepakt-om-dreiging-met-bomaanslag-op-twitter>.

¹⁵ <http://www.volkskrant.nl/vk/nl/2694/Internet-Media/article/detail/2933390/2011/09/28/Jongen-13-dreigt-op-Twitter-school-op-te-blazen.dhtml>.

¹⁶ <http://www.depers.nl/binnenland/564637/Aanhouding-om-dreigtweet.html>.

¹⁷ <http://nos.nl/artikel/231756-zes-doden-bij-schietpartij-alphen-aan-den-rijn.html>.

¹⁸ <http://www.depers.nl/binnenland/575289/Celstraf-na-dreigtweet.html>.

¹⁹ @Doodsbedreiging // <https://twitter.com/#!/doodsbedreiging>.

²⁰ <http://www.doodsbedreiging.nl/>.

instanties maar ook om een eigen identiteit te creëren. Een ander kenmerk van typisch Internettaalgebruik is het gebruik van afkortingen en acroniemen. Hieraan worden telkens weer krachttermen toegevoegd die in de afkorting of het acroniem zelf niet direct veel voorstellen, maar wanneer het acroniem volledig uitgesproken of uitgeschreven wordt grof overkomt (e.g. LMAO en LMFAO, respectievelijk „Laugh[ing] My Ass Off” en „Laugh[ing] My Fucking Ass Off”; RTFM of FML wat staat voor respectievelijk „Read the fucking manual’ en „Fuck my life”).

Het is geconstateerd en bewezen (Feldweg, Kibiger, & Thielen, 1995, p. 150) dat *posts* (bijdragen) van foragebruikers zowel elementen kent van geschreven als van gesproken taal, terwijl in chatprogramma’s de gebruikers praten (typen) zoals ze schrijven, waarbij meer wordt gelet op spelling conventies, interpunctie en grammaticale compleetheid (zie Lankshear, 1997; Abbott, 1998; Merchant, 2001, p. 296). Een tweet zit tussen een forumbijdrage en een chatbericht in en moet dus op zichzelf worden gezien. Tekstuele bijdragen op het Internet lezen alsof ze worden gesproken (Davis & Brewer, 1997) en voor Twitter geldt dit ook.

Dit oefent invloed uit op de wijze hoe gebruikers van het Internet communiceren, ongeacht of oudere doelgroepen ook gebruik maken van Internet. Inmiddels heeft 93% van de Nederlandse bevolking toegang tot het Internet, hebben tien miljoen Nederlanders een Hyvesaccount en hebben Twitter en Facebook sinds 2008 een hoge naamsbekendheid gekregen in Nederland. Hierdoor is het aantal Nederlandse Facebookgebruikers nu op 5.361.200²¹ komen te staan en het aantal Twitter gebruikers op een miljoen actieve gebruikers²².

Het taalgebruik op Computer-Mediated Communication (CMC) systemen is onderhevig aan veranderingen sinds de opkomst van het Internet. Waar voorheen de mogelijkheid bestond om anoniem berichten te plaatsen op een blog²³, daar moest later in Multi-User Domains (MUD’s), zoals bijvoorbeeld een Instant Messenger programma²⁴, een account worden aangemaakt. Hierdoor kreeg deze persoon een identiteit aangemeten. Dat wil

²¹ Cijfer afkomstig van <http://www.checkfacebook.com/> dat zich baseert op de officiële advertentie applicatie van Facebook. Terwijl op 24 maart 2011 het aantal nog op 4,8 miljoen lag volgens Arno Lubrun (Facebook Benelux) in De Pers.

²² Cijfer afkomstig van MarketingFacts

http://www.marketingfacts.nl/berichten/20110812_nederland_telt_1_miljoen_actieve_twitteraars_infographic/.

²³ Een blog, ook wel weblog, is een website waarop bijdragen van een persoon vergelijkbaar aan een dagboek, maar dan in een chronologisch omgekeerde volgorde.

²⁴ Een Instant Messenger is een chatprogramma waarmee geregistreeerde gebruikers onder hun eigen naam of pseudoniem berichten uitwisselen met andere gebruikers van dat programma die ook geregistreerd staan onder hun eigen naam dan wel hun pseudoniem. Een bekend voorbeeld is Windows Live Messenger (voorheen MSN Messenger).

echter niet zeggen dat op de MUD's alleen maar gebruik werd gemaakt van de eigen naam. De trend bleek echter dat over het algemeen de officiële naam wel werd gebruikt om herkenbaar te zijn voor de overige gebruikers van de MUD, vaak vrienden en kennissen. Men wordt hiertoe ook gestimuleerd door Facebook, LinkedIn, Google+ en Twitter, maar alleen in het geval van Google's dienst is het ook verplicht. Het feit of iemand anoniem een bericht kan plaatsen heeft ook bijgedragen in de ontwikkeling van taalverandering op het Internet.

2: Theoretisch kader

2.1 Inleiding

In dit hoofdstuk zal eerst een korte inleiding op het onderzoek worden gegeven. Hierna zullen er diverse theorieën worden besproken die belangrijk zijn geweest voor de vorming van een theoretische achtergrond van de scriptie. Wanneer de verschillende theorieën uitgelegd zijn zal er uitgebreid ingegaan worden op het construct „beleefdheid’, de rode draad door deze scriptie. In de daaropvolgende paragraaf zal de methodiek van het onderzoek worden uitgelegd en zal het hoofdstuk worden afgesloten met het verwerken van de verkregen data.

2.2 Theorieën aan de basis van dit onderzoek

Om het onderzoek te onderbouwen is er voor gekozen om een aantal lang bestaande theorieën te gebruiken en deze toe te passen op de huidige technologie. De beleefdheidstheorie van Brown & Levinson (1987) is daarvan de rode draad door deze scriptie. Hier uit voort komt ook de Social Information Processing theorie van Walther (1992) die in de loop van dit hoofdstuk zal worden uitgelegd en wat de relevantie van de theorie is voor dit onderzoek.

2.2.1 Beleefdheidstheorie

De beleefdheidstheorie van Brown & Levinson (1987) is de best toepasbare en meest uitgebreide benadering van het construct „beleefdheid’. Om de vragen met betrekking tot beleefdheid in deze scriptie te beantwoorden, biedt het perspectief van Brown en Levinson het beste uitgangspunt. In deze theorie komt naar voren wanneer sociale afstand tussen de spreker en de ontvanger zal toenemen, de mate van beleefdheid ook zal toenemen.

Gezichtsbedreigende factoren (*Face Threatening Acts*, kortweg FTA’s) zoals machtsverhouding, de sociale afstand en de plaats van het individu in de cultuur zijn volgens Brown & Levinson belangrijke factoren. Het plaatsen van een tweet kan er voor zorgen dat een persoon positiever, gelijk of negatiever overkomt op zijn of haar volgers.

Brown & Levinson (1987) ontleenden hun kijk op beleefdheid aan de literatuur van sociaalpsycholoog Erving Goffman (1955, 1967) waarin het concept „*face*’ (gezicht) in zijn figuurlijke (en dus niet letterlijke) zin werd geïntroduceerd wat op zijn beurt „beleefdheid’ nader verklaart. Goffman omschrijft beleefdheid als het uiten van een intentie van een spreker om zijn of haar gezichtsbedreiging(en) te verzachten wanneer deze door bepaalde FTA’s worden opgewekt. In het boek *Interaction Ritual* beschrijft hij twee typen „gezicht’ die Brown & Levinson een ruime plaats toebedelen in hun theorie. Het eerste is het positieve gezicht,

ook wel positieve beleefdheid, waarbij men hoopt te worden herinnerd als een begerlijk persoon. Het tweede type is het negatieve gezicht of negatieve beleefdheid. Hierbij willen mensen „onafhankelijk zijn, de vrijheid hebben om te handelen en willen niet dat hen een wil wordt opgelegd door anderen’ (Brown & Levinson, 1987).

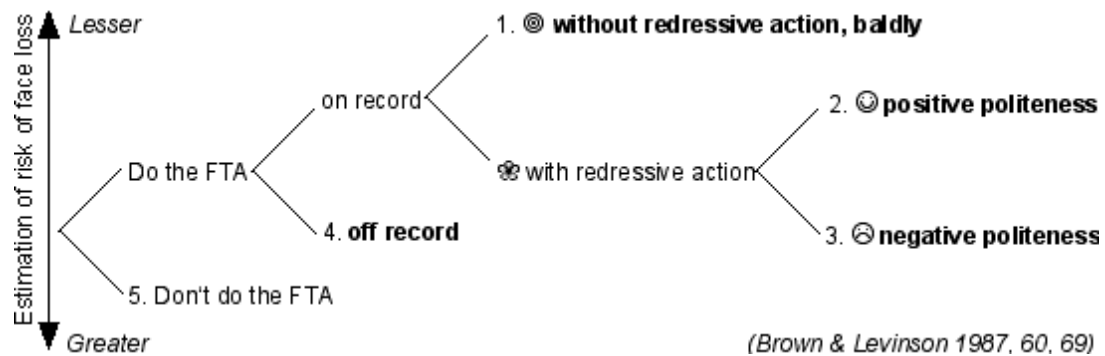
Brown & Levinson (1987) hebben op basis van deze twee typen gezichten de theorie verder ontwikkeld. De doorontwikkeling van de theorie leidde tot de totstandkoming van de volgende formule (Brown & Levinson, 1987, pp. 15 & 76-78): $W_x = R + D + P$, oftewel de zwaarte van de FTA (*Weight*) is afhankelijk van de volgende drie factoren; de lasten van de gezichtsbedreigende taalhandeling uitgedrukt in een numerieke waarde (*rate of imposition*), sociale afstand (*distance*) en macht (*power*). Een zware FTA wil dus zeggen dat er een grote kans is op gezichtsverlies en komt tot stand door bovengenoemde drie factoren en suggereert een bepaalde strategie (Bousfield, 2008, p. 88). Aansluitend hebben Brown & Levinson drie behoeften vastgesteld die een persoon moet afwegen om een FTA uit te voeren, te weten (Brown & Levinson, 1987):

1. Behoefte om de inhoud van de FTA te communiceren;
2. Behoefte om urgent dan wel efficiënt te zijn richting de gesprekspartner;
3. Behoefte om te proberen het gezicht van de gesprekspartner zoveel mogelijk te behouden.

Daarnaast hebben Brown & Levinson (1987) vier beleefdheidsstrategieën ontwikkeld die met FTA's om kunnen gaan en zodoende het „gezicht te redden’. De vier strategieën zijn:

1. Bald On-Record;
2. Positive Politeness;
3. Negative Politeness;
4. Off-Record indirect strategy

In figuur 1 staat schematisch weergegeven wanneer deze beleefdheidsstrategieën worden gebruikt.



Figuur 1 – Beleefdheidsstrategieën van Brown & Levinson (1987, 60-69)

De vier strategieën uit de voorgaande figuur kunnen als volgt uitgelegd worden:

- ad1.** De meest directe beleefdheidsstrategie, hierin is de kans op gezichtsverlies minimaal te noemen (e.g. “Open het raam!”).
- ad2.** Positieve beleefdheidsstrategie is niet zo direct als de Bald On-Record strategie. Door middel van de strategie probeert men het positieve gezicht te verbeteren (e.g. “Zou u het raam alstublieft willen openzetten?”).
- ad3.** Ook deze strategie is weer iets minder direct dan de voorgaande strategie. Hiermee wordt geprobeerd om te voorkomen dat men een negatief gezicht krijgt (e.g. “De temperatuur is hier wel erg hoog opgelopen”).
- ad4.** De laatste soort strategie is zeer indirect en zorgt dat de spreker niet aansprakelijk kan worden gehouden voor eventuele gezichtsbedreigingen, maar het eigen gezicht wordt hierbij vergeten waardoor er een kans is op gezichtsverlies van de spreker zelf (e.g. “Wat heb ik het warm”).

Door middel van controle uitoefenen op FTA’s (i.e. goed nadenken over wat je tweet) kan iemand zijn eigenwaarde behouden waardoor iemand anders zich gerespecteerd voelt. Wanneer een persoon geen aandacht schenkt aan datgene dat hij plaatst op Twitter kan diegene zijn „gezicht” verliezen ten opzichte van een grote groep volgers en in sommige gevallen (wanneer er geen privacy restricties zijn) voor iedereen die het wil lezen. Een tweet met onbeleefde content kan van grote invloed zijn op iemands „gezicht” zonder dat diegene daar van op de hoogte is.

2.2.2 Cues-Filtered-Out (CFO) theorie/perspectief

Een belangrijke theorie die veelal aangehaald wordt in onderzoeken naar sociale media is de Cues-Filtered-Out (CFO) theorie/perspectief van Culnan en Markus (Culnan & Markus, 1987). De theorie is gevormd om aan te tonen dat op tekst gebaseerde CMC een mindere

vorm van communicatie is door het gebrek aan fysieke- en sociale aanwijzingen. CMC wordt door Boudourides (1995) beschreven als “*the process of one-to-one, one-to-many, and many-to-many communicative discourse using a computer-based communication channel, taking place predominantly in a text-based environment*”. De theorie gaat uit van een situatie waarbij personen elkaar niet kunnen horen, zien of voelen (i.c. Twitter) wat er feitelijk op neerkomt dat er geen non-verbale signalen zijn. Door het gemis van *social context cues* in CMC is het volgens Boudourides van belang om deze theorie nader te bekijken en te gebruiken bij het voeren van gesprekken via CMC. Er zou sprake zijn van absentie van een regelmatige wijze om anderen van feedback te voorzien, echter door nieuwe technieken is dit juist wel mogelijk (e.g. Like buttons, porren²⁵, re-tweeten, delen van foto, video- en audiobestanden). Op basis van bovenstaande is het CFO perspectief niet geïmplementeerd in deze scriptie. Daarvoor dient het perspectief eerst te worden herzien.

Op tekst gebaseerde CMC leidt volgens de theorie tot minder socio-emotionele²⁶ (Santrock, 2007) banden bijvoorbeeld in een bedrijfshiërarchie. Men is eerder geneigd van zich te doen spreken door middel van CMC juist door de afwezigheid van de kenmerken die bij *Face-to-Face* (FtF) wel aanwezig zijn. De typische FtF kenmerken dienen zich bij CMC wel aan, maar de kenmerken hoeven niet te worden gebruikt, het zijn opties. Deze constatering is tevens de reden dat deze theorie niet geschikt is voor dit onderzoek. „Vele studies wijzen uit dat de effecten van de *Cues-Filtered-Out* benadering alleen gelden voor de beginperiode van een CMC relatie (e.g. een gesprek tussen twee twitteraars). Wanneer er langer wordt gemeten is er geconstateerd dat de onpersoonlijke aspecten verdwijnen, omdat men door de hoeveelheid berichten elkaar al „goed’ leert kennen. Er ontstaat kennelijk een positieve relatie zolang ze maar genoeg tijd krijgen om nader tot elkaar te komen’ (Walther, Burgoon, & Judee, 1992, pp. 76-77).

2.2.3 Social Information Processing (SIP) theorie

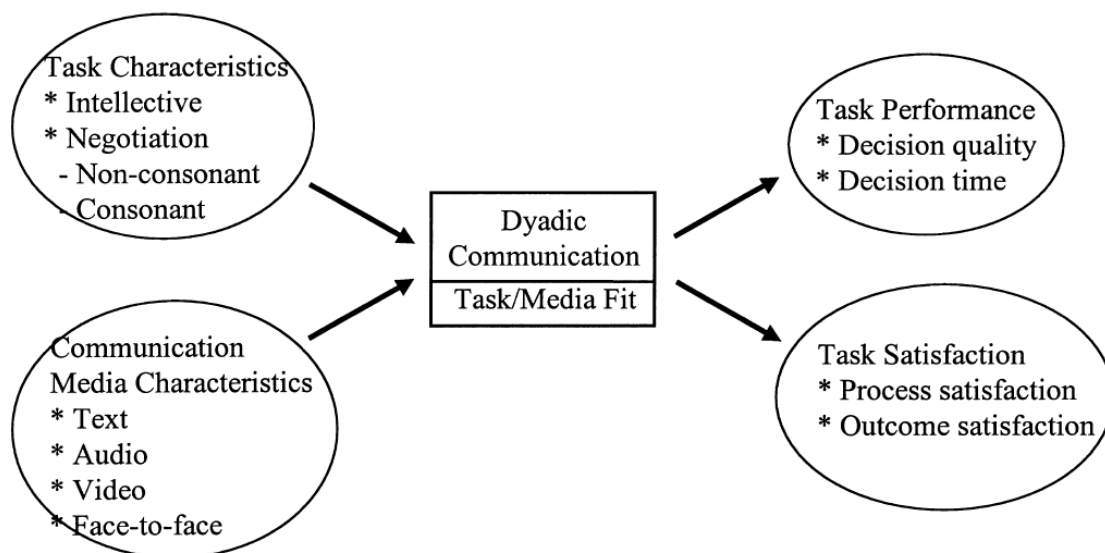
Een theorie die wel betrekking heeft tot dit onderzoek is de Social Information Processing theorie (Walther J. B., 1992), oftewel, het verwerken van sociale informatie. Deze theorie is ook wel de tegenhanger van het CFO perspectief en wordt daarom ook wel *Cues-Filtered-In* (CFI) perspectief genoemd (Walther & Parks, 2002). In de SIP theorie wordt beschreven dat individuen gemotiveerd raken om een impressie te vormen en een relatie te bewerkstellingen, ongeacht het medium dat zij daarvoor gebruiken.

²⁵ Een notificatie op Facebook dat iemand aan je denkt.

²⁶ “Socioemotional is a process that consists of variations that occur in an individual’s personality, emotions, and relationships with others during one’s lifetime” – Santrock (2007).

Sinds het ontstaan van online communicatie wordt er gezocht naar alternatieven voor non-verbale communicatie. De emoticon is hiervan een goed voorbeeld, gezien de essentie van dit fenomeen een weergave is van de blik op het gezicht van de gesprekspartner(s). De theorie van Joseph Walther is gebaseerd op CMC. Ondanks de afwezigheid van aanwijzingen hoe iemand er uitziet of hoe iemand zich gedraagt in CMC ten opzichte van FtF, is CMC volgens Walther en Parks (Walther & Parks, 2002) toch in staat om een goed alternatief te zijn voor FtF communicatie. Door de mogelijkheid te bieden om impressies en percepties van iemand te krijgen zonder dat daar aanwijzingen voor zijn die wel aanwezig zijn in FtF.

Een theorie om de effectiviteit van diverse media vehikels te testen is de Media Richness theorie (King & Xia, 1997). Deze theorie wijst uit dat Twitter in zijn huidige vorm in de buurt komt van FtF communicatie vanwege de „media rijkheid” van het communicatiesysteem. Door de evolutie van de afgelopen jaren is het medium, evenals andere CMC systemen, steeds dichterbij FtF communicatie komen te staan. Ook vanwege die reden is Twitter een medium waar op taal wordt gelet door de ontvangers van het bericht (c.q. waarschijnlijk veel meer dan in een FtF conversatie) en waar de verzender op afgerekend kan worden. Al deze factoren wegen mee in de resultaten van dit onderzoek waar correlaties worden gelegd tussen deze theorieën en de werkelijkheid.



Figuur 2 – Media richness model van Suh (1999)²⁷

²⁷ Dyad is groep van twee personen, dyadic is een bijvoeglijk naamwoord die een type communicatie beschrijft. Een dyad is in dit geval de kleinst mogelijke groep om communicatie mogelijk te maken.

2.2.3.1 Hyperpersonal model

Het „Hyperpersoonlijke model’ (Walther J. B., 1996) is een inter-persoonlijke communicatietheorie die verklaart dat CMC hyperpersoonlijk kan worden. Dit komt doordat CMC ook communicatieve voordelen heeft voor een persoon die een boodschap verzendt, ten opzichte van FtF interacties. De uitdrukking dient om een bepaald type communicatie te beschrijven die alleen plaatsvindt in tekstuele communicatie en is één van de drie typen communicatie die door Walther zijn voorgesteld. De andere twee zijn onpersoonlijke- en inter-persoonlijke communicatie.

In de beginnende jaren van het online communiceren, bijvoorbeeld door middel van e-mail, was de achtergrond van de communicatie-uiting vooral onpersoonlijk (Parks & Floyd, 1996). Het werd zakelijk ingezet, terwijl dat tegenwoordig het tegenovergestelde is waarbij zakelijke communicatie ook persoonlijk is. Zowel naar medewerkers van een bedrijf als naar klanten van een bedrijf wordt er op persoonlijke noot gecommuniceerd.

Het derde punt uit de Social Information Processing theorie, inter-persoonlijke communicatie, is een type communicatie die georiënteerd is op het sociale vlak (Turner, Grube, & Meyers, 2001). Dit is een stap vooruit naar de hedendaagse kijk op (online) communicatie waarbij persoonlijke informatie toegevoegd werd aan de communicatieve uiting. Doordat met deze ontwikkelingen de sociale relaties toenamen en men het FtF gesprek steeds minder nodig had om persoonlijke informatie te delen. De sociale kring waarin iemand zich bevond werd langzamerhand uitgebreid met personen die ze niet in het echt kenden, waardoor het Internet een plek werd waar inter-persoonlijke relaties ontstonden.

Uiteindelijk heeft dit geleid tot het Hyperpersoonlijke model van Walther (1996) waarin intensiteit en intimiteit snel tot een hoog niveau reiken. De snelle toename van intimiteit wordt verklaard aan de hand van het missen van visuele aanwijzingen. Hierdoor zal eerder in een gesprek een intiemere band ontstaan. Walther verklaart deze vorm van communicatie als volgt: *“(CMC is hyperpersonal) when users experience commonality, are self-aware, physically separated, and communicating via a limited-cues channel, they selectively self-present and edit; to construct and reciprocate representations of their partners and relations without the interference of environmental reality”* – Walther (1996, p. 33).

2.3 Aanzet tot het gebruik van scheldwoorden

Het is aangetoond en bewezen dat schelden een natuurlijk onderdeel van de mens is (Jay, 2009), er is in feite niets raars aan schelden. Volgens het onderzoek van Jay schelden we 0,3 procent (McEnery, 2006) tot 0,7 procent (Jay, 1980) van de tijd, ter vergelijking gebruikt men

gemiddeld in 1 procent van de gevallen een veelgebruikt persoonlijk voornaamwoord. Die hoeveelheid gebruikte scheldwoorden is dus een klein, maar wel een significant percentage van al onze mondeling communicatie. Uit Jay (2008) blijkt dat er op websites zoals MySpace het percentage scheldwoorden op 0,3 procent (Thelwall, 2008) ligt en drie procent van de zinnen die worden getypt in een chatroom bevat een scheldwoord (Subrahmanyam, Smahel and Greenfield, 2006). Ook onderzoeksbureau Reppler²⁸ heeft zich verdiept in schelden op sociale netwerken, in dit geval Facebook. Ook op dit medium komt het regelmatig voor dat een gebruik minsten één onbeleefde uiting op zijn tijdlijn heeft staan. Van de 30.000 participanten van het onderzoek, bleek 47 procent van hen een tijdlijn te bezitten met daarop één of meerdere onbeleefde uitingen. In 80 procent van de gevallen zijn die woorden afkomstig van een Facebookvriend. Het is dus niet vreemd dat we schelden, al is de acceptatie daarvan een ander verhaal.

Volgens het Sociaal en Cultureel Planbureau (SCP) vormen de gehanteerde normen en waarden in het land de grootste zorg voor Nederlanders²⁹. Nederlanders schatten zelfs de huidige economische crisis qua belangrijkheid lager in dan de omgangsvormen die tegenwoordig worden gehanteerd. De omgangsvormen worden ook belangrijker gevonden dan populaire onderwerpen zoals immigratie en criminaliteit. Motivaction, dat hiernaar onderzoek³⁰ heeft gedaan, ergert men zich met regelmaat aan vormen van openbaar onbeschoft gedrag. Maar vervolgens spreekt men elkaar niet aan op dit gedrag om uiteenlopende redenen als „heeft toch geen zin’ en „bang voor represailles’.

Het blijkt dus dat veel Nederlanders zich ergeren aan onbeschoft gedrag in het openbaar, zoals bumperkleven of afsnijden in het verkeer, het gooien van afval op straat, het gebruik van grove taal, uitgaansoverlast en hondenpoep op straat. Voor een persoon zijn er gedurende de dag dus veel prikkels om deze frustraties van zich af te schrijven. De hoge toegankelijkheid van communicatie middelen, maakt het mogelijk om het medium te gebruiken om deze emoties de wereld in te brengen. Om hevige emoties te uiten, maken we voornamelijk gebruik van vloeken (Sterkenburg, 2007).

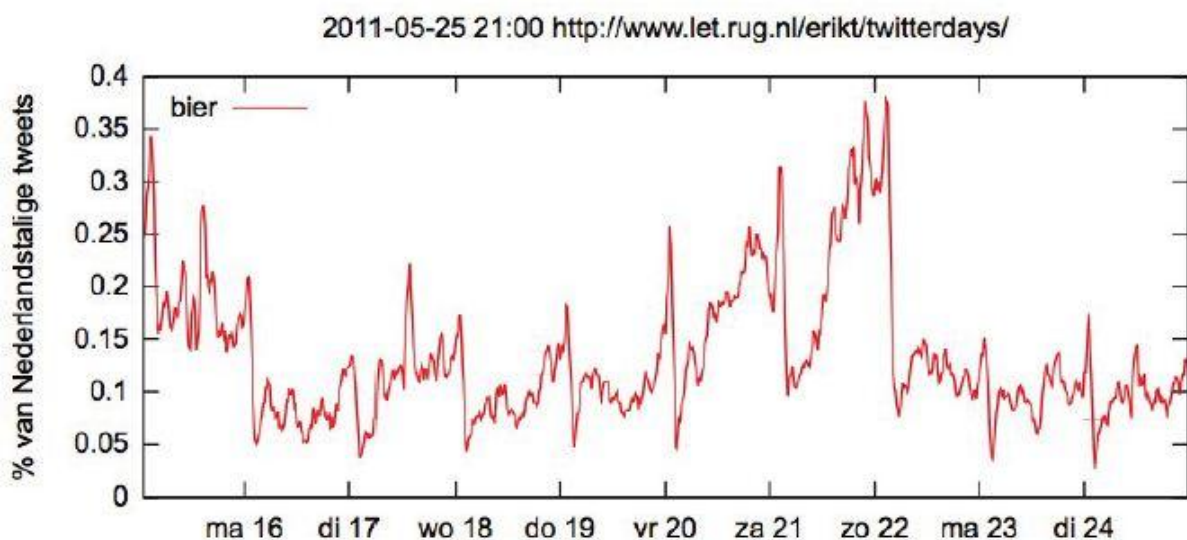
Wanneer deze frustraties opgedaan worden loopt uiteen, maar van enkele frustraties staat het praktisch vast wanneer de hoogst mogelijke kans daarop aanwezig is. Verkeersfrustraties zullen voornamelijk tijdens de spitsuren voorkomen en op de zogenaamde zwarte (zater)dagen. Voetbalwedstrijden (e.g. verlies favoriete team) en zelf sporten (e.g. zelf

²⁸ Reppler research “47% of Facebook users have profanity on their wall” N=30.000 <http://bit.ly/wUZ5Ed>.

²⁹ Trouw van 5 januari 2012 <http://bit.ly/w08rqO>.

³⁰ Onderzoek van Motivaction, N=619 <http://bit.ly/xeyQEM>.

verliezen van een wedstrijd), spelen zich respectievelijk vooral doordeweeks 's avonds of overdag in het weekend af. Het uiten van emoties zal onder invloed van alcohol voornamelijk van donderdag tot en met zondag in de avonduren plaatsvinden. Van de werkende stof in drank, alcohol, is bekend dat het onder andere remmingen wegneemt en zorgt voor een verhoogde impulsiviteit. Kleine irritaties kunnen hierdoor sneller veranderen in verbale (en fysieke) agressieve uitbarstingen. Daaraan toegevoegd zijn Nederlanders gemiddeld gezien grootgebruikers als het gaat om alcoholconsumptie. Een deel van de Nederlanders drinkt amper, maar de groep die drinkt neemt grote hoeveelheden alcohol tot zich³¹. Aan de hand van een klein testje met de software demo „Twitter search” van dr. ir. E.F. Kim Sang³² wordt duidelijk wanneer men twittert met het woord „bier” in de tweet. Hieruit valt op te maken dat dit voornamelijk 's avonds en 's nachts wordt geschreven, daarnaast is het door de weeks ook een aanzienlijk minder voorkomend woord. Hieruit lijkt naar voren te komen dat het bioritme van de mens van betekenis kan zijn bij de verdeling in frequentie van tweets. Het lijkt aannemelijk dat er 's nachts minder wordt getwitterd, maar of dat tijdens kantooruren ook zo is, kan dit onderzoek deels aantonen.



Figuur 3 – Het woord ‘bier’ is een week lang gemeten op Twitter in de periode van 16 mei 2011 tot en met 24 mei 2011 door dr. ir. E.F. Kim Sang. Hieruit wordt onder andere duidelijk dat er naarmate het weekend dichterbij komt men vaker het woord ‘bier’ gebruikt in hun tweets. Hoe later het wordt hoe vaker het woord ‘bier’ voorkomt.

³¹ Onderzoek van GFK (N=17.350 Europeanen, 17 landen).

³² <http://urd.let.rug.nl/erikt/bin/twitter>.

Studenten hebben volgens een onderzoek van NRC-Next³³ meer te besteden aan het eind van de maand (studiefinanciering) en dat geldt ook voor veel niet-studenten volgens softwarebedrijf Unit 4. Zij verzorgen iedere maand de loonstroken van 3,1 miljoen werknemers in Nederland. Hieruit kan worden opgemaakt dat er vanaf de 21^e tot de 25^e van de maand meer geld valt te besteden. Volgens taxichauffeurs en kroegeigenaren in respectievelijk Rotterdam en Amsterdam wordt het inderdaad aanmerkelijk drukker in de uitgaansgelegenheden vanaf die dagen. Er wordt ook nog eens meer geld verdiend aan de verkoop van drank.

2.4 Het construct beleefdheid

Het definiëren van het concept „beleefdheid’ is in de literatuur terug te vinden in onderzoeken naar onder andere emoties, gedragingen, psycholinguïstiek en onderhandelingsstrategieën. Om iets over de beleefdheid van een tweet te zeggen, moet er wel een grondslag voor het construct „beleefdheid’ zijn en die is aanwezig. Zo wordt linguïstische beleefdheid, hetgeen dat wordt onderzocht, binnen enkele theorieën vanuit een pragmatische standpunt vertaald als “het gebruik van communicatieve strategieën om sociale harmonie te bewaren of te promoten” (Leech, 1983; Brown & Levinson, 1987; Lakoff, 1989). Watts (2003: p. 9) beschrijft het als volgt “*(im)politeness is a term that is struggled over at present, has been struggled over in the past and will, in all probability, continue to be struggled over in the future*”, waarmee hij aangeeft dat het nooit geheel duidelijk zal worden wat (on)beleefdheid nu eenmaal is. Die opvatting lijkt nogal naïef. De Van Dale geeft de betekenis kort weer. Hierin wordt beleefdheid als volgt gedefinieerd: “Het beleefd-zijn, hoffelijkheid”. Dit is te vaag om een fundament te vormen voor een wetenschappelijk onderzoek. Deze wetenschappelijke benadering van Leech, Lakoff en Brown & Levinson biedt een basis om een onderzoek op te bouwen.

In dit onderzoek is ervoor gekozen om een tweet als beleefd te bestempelen wanneer er geen enkel woord uit de woordenlijst (zie bijlage V) voorkomt in de tweet. Wanneer er wel een woord uit deze lijst voorkomt in de tweet, dan zal deze worden gemarkeerd als onbeleefd. Ook wanneer er meerdere onbeleefde woorden uit de woordenlijst in een tweet staan, verandert dat niet de mate van beleefdheid van de tweet. Er wordt enkel onderscheid gemaakt tussen een beleefde- en een onbeleefde tweet.

³³ <http://www.nrcnext.nl/blog/2011/12/29/is-er-meer-cafebezoek-aan-eind-van-de-maand/>.

3: Onderzoeksopzet

3.1 Methode

Om de gehanteerde beleefdheid in tweets te meten werd er een reeks tests uitgevoerd via enkele speciaal daarvoor geschreven scripts. Deze scripts werken binnen een Linux omgeving in een *command line* programma. Elke tweet draagt metadata (i.e. in de Twitteromgeving niet-zichtbare gegevens die een gebruiker meestuurt met zijn tweets) met zich mee, waaruit relevante gegevens zijn gehaald voor dit onderzoek. Het gaat in dit geval om locatie, tijd en dag, maar ook de *hashtag* is in zekere zin metadata.

3.2 Materiaal

De tweets werden binnengehaald en geanalyseerd op een computer die was uitgerust met Linux Ubuntu 11.10. Het voordeel van het gebruiken van tweets voor onderzoek is in eerste instantie het feit dat het materiaal gratis is en in enorme hoeveelheden (gratis) te verkrijgen op het Internet. De actualiteit is makkelijk en nauwkeurig te meten door de constante updates van twitteraars. Met de juiste scripts (zie bijlage IV) zijn de mogelijkheden precies af te stemmen op hetgeen dat onderzocht moet worden.

Iedere tweet bevatte op zijn minst de volgende informatie:

- Tijd – op welk tijdstip is de tweet geschreven
- Dag – op welke dag is de tweet verstuurd
- Gebruikersnaam – wie heeft de tweet geschreven

3.3 Procedure

Via de Twitter API zijn de bestanden gedownload. De tests zijn deels uitgevoerd in het HAIT lab³⁴ van Tilburg University en grotendeels uitgevoerd op een thuis PC. Daar de middelen op beide locaties evenzogoed waren, heeft dit geen effect gehad op de uitkomsten van het onderzoek.

3.4 Verzameling en verwerking van de gegevens

De eerste drie hypothesen uit het onderzoek kunnen worden beantwoord aan de hand van een dataset die bestaat uit tweets die zijn gedownload door dr. ir. E.F. Tjong Kim Sang,

³⁴ Naar de studierichting van het departement CIW: Human Aspects of Information Technology (HAIT).

onderzoeker van de Alfa-informatica leerstoel-groep van de Universiteit Groningen³⁵. Tjong Kim Sang downloadt dagelijks tweets ten behoeve van zijn eigen onderzoek. Hierdoor heeft Tjong Kim Sang inmiddels al een zeer groot bestand opgebouwd waardoor het voor hem geen enkel probleem was de data te selecteren en over te dragen ten behoeve van deze scriptie. Tjong Kim Sang heeft er tevens voor gezorgd dat het gebied „Tilburg” uit deze data is gefilterd. In zijn paper „Het Gebruik van Twitter voor Taalkundig Onderzoek” (Tjong Kim Sang, 2011) wordt uitgelegd op welke wijze zijn data wordt gedownload. Met cURL³⁶ worden tweets automatisch gedownload van de Twitter website (zie bijlage X voor de zoekopdracht). Door “curl -d '@BESTAND' URL -u 'GEBRUIKER:WACHTWOORD'” in te voeren in een *command line* programma kan er gezocht worden naar tweets. De Nederlandse tweets zijn geselecteerd met behulp van twee technieken:

1. Selectie van tweets die een veelvoorkomend Nederlands woord bevatten of die een typisch Nederlandstalige *hashtag* bevatten;
2. De overige tweets worden door een automatische taalrader beoordeeld.

De data is binnengehaald vanaf 1 januari 2012 om 00:00 uur tot en met 29 januari 2012 om 23:59 uur en bevat 13 procent van de Nederlandstalige tweets die in die periode zijn geplaatst op Twitter. Het betreft hier dus exact vier weken en één dag aan tweets. Deze ene dag betreft nieuwjaarsdag en is later in het onderzoek losgekoppeld van de vier weken data.

Daarnaast wordt er voor de vierde hypothese gebruik gemaakt van tweets die afkomstig zijn uit een database gecreëerd door drs. E.P. Sanders, Radboud Universiteit Nijmegen. Sanders verzamelde de tweets voor het SoNaR: STEVIN Nederlandstalig Referentiecorpus³⁷ (Oostdijk, et al., 2008). De eerste tweets in deze database dateren uit 2007, maar het merendeel komt uit de periode 2010 en 2011. Dr. M.W.C. Reynaert en dr. M. M. van Zaanen, onderzoekers binnen Tilburg University hebben deze tweets vervolgens beschikbaar gesteld voor dit onderzoek.

De tweets uit de groep zonder toestemming zijn via de Twitter API³⁸ gedownload door drs. E.P. Sanders. Per specifiek persoon konden de afzonderlijke tweets (in series van twintig) in een loop worden gedownload. De `max_id` (zie onderstaande voorbeeld zoekopdracht) is bij elke loop telkens de oudste tweet zodat er steeds verder teruggegaan kan worden in de

³⁵ Zijn meest recente Twitteronderzoek dateert van 2011 <http://ifarm.nl/erikt/papers/tabu2011.pdf>.

³⁶ Gratis te downloaden op <http://curl.haxx.se>.

³⁷ <http://lands.let.ru.nl/projects/SoNaR/> en <http://www.sonarproject.nl/> Een corpus bestaande uit 500 miljoen Nederlandstalige woorden.

³⁸ Application Programming Interface. De Twitter API is te vinden op <https://dev.twitter.com/>.

*timeline*³⁹ van de twitteraar. Een voorbeeld van een dergelijke zoekopdracht luidt als volgt:
http://api.twitter.com/1/statuses/user_timeline.json?screen_name=mikedewildt&count=20&include_entities=1&max_id=29121684039

Om een duidelijk overzicht te geven welke dataset gebruikt maakt van hoeveel twitteraars en hoeveel (onbeleefde) tweets, en tevens om aan te geven welke groepen bij welke dataset/corpus horen is tabel 1 gerealiseerd:

Tabel 1

Overzicht van de data die is gebruikt per corpus. De grote dataset is verkregen via dr. ir. E.F. Tjong Kim Sang (onder andere gespecialiseerd in Twitter zoekopdrachten). Hij heeft eveneens gezorgd voor de data waar deel B uit bestaat, bewerkstelligd is. Deze data is gescheiden van deel A. De data in groep B is dus wel onderdeel van groep A. De data uit het SoNaR: STEVIN corpus zijn beschikbaar gesteld door drs. E.P. Sanders van de Radboud Universiteit Nijmegen en beschikbaar gemaakt voor dit onderzoek door dr. M.M. van Zaanen. Dit corpus bevat twee groepen die elk zijn opgebouwd uit verschillende data.

	Grote dataset		SoNaR: STEVIN corpus	
Paragraaf	3.4.1	3.4.1	3.4.2	3.4.3
Groep	1 t/m 29 jan 2012	1 t/m 29 jan 2012	Geen toestemming	Toestemming
	Nederland	Tilburg		
Subnaam	Deel A	Deel B	-	-
Verzamelwijze	At random	At random	Handmatig	At random
Twitteraars	2.162.670	1.152	115	368
Tweets	64.934.780	10.395	262.967	681.265
Onbeleefde tweets	1.665.181	186	574	5.819

In de volgende paragrafen wordt er per groep nader ingegaan op de informatie uit tabel 1. Er zal onder andere expliciet worden vermeld op welke corpora de groep is gebaseerd, hoe de data is verzameld, hoeveel twitteraars in de groep zitten en wat de absolute frequentie van de (onbeleefde) tweets is. Deze tweets worden geanalyseerd om een mogelijk verschil aan te tonen het wel of niet geven van toestemming om de tweets te gebruiken en de gehanteerde beleefdheid in de tweets.

³⁹ Tijdslijn, hetgeen waar alle tweets van iedere gebruiker die een persoon volgt, inclusief de gebruiker zelf, komt te staan.

3.4.1 Vier weken tweets uit heel Nederland

Deze dataset bestaat uit 13 procent van alle Nederlandstalige tweets verzonden vanaf zondag 1 januari 2012 om 00:00 uur tot en met 23:59 uur op 29 januari 2012. Hierdoor zijn er exact vier weken en één dag aan data verzameld. Nieuwjaarsdag is vervolgens uit de data gehaald, omdat dit een vertekend beeld zou kunnen opleveren. Vanwege deze constatering is er op het einde van dit hoofdstuk een vergelijking gemaakt tussen alle vier de zondagen uit deze groep. Zodoende kan er iets worden gezegd over deze bewering over nieuwjaarsdag en of het juist was om de zondag niet mee te nemen in de data.

Deel A van de dataset bevat alle tweets uit Nederland inclusief Tilburg en bestaat uit: $N=2.162.670$ unieke twitteraars. In totaal zaten er 64.934.780 tweets in deze groep, waarvan waren er 1.665.181 tweets onbeleefd.

Aansluitend, door een vierkant over Tilburg te leggen met een omtrek van tien vierkante kilometer, is er een gebied afgebakend dat de Tilburgse twitteraar in kaart kan brengen. Aan de hand van geocodes van de omgeving Tilburg kon worden bepaald welke metadata uit de dataset moest worden gefilterd om iets te kunnen zeggen over de tweets die zijn verzonden vanuit dit gebied. Niet elke tweet bevatte dergelijke metadata. Ongeveer uit één procent van de complete dataset konden er geocodes worden gehaald uit de metadata. Door middel van deze gegevens konden er observaties worden gedaan en conclusies worden getrokken naar aanleiding van het twittergedrag van de Tilburger ten opzichte van de rest van Nederland. De vergelijking is uiteindelijk gedaan door de tweets met geocode Tilburg te vergelijken met alle tweets (ook wanneer deze geen geocode bevatte). Ter verduidelijking, al deze „Tilburg” tweets komen ook voor in de deel A van de dataset.

Deel B van de dataset bevat alle tweets uit de omgeving Tilburg die ook in deel A van deze dataset zaten en bestaat uit: $N=1.152$ unieke twitteraars, waarvan er 98 twitteraars een onbeleefde tweet hebben geplaatst. In totaal zaten er in deze groep 10.395 tweets, waarvan waren er 186 tweets onbeleefd waarin 218 scheldwoorden zijn gebruikt.

De gegevens zijn verwerkt aan de hand van de volgende definitie voor nacht, ochtend, middag en avond. De nacht valt in om 00:00 en duurt tot en met 05:59, de ochtend begint om 06:00 en duurt tot en met 11:59, de middag begint om 12:00 en duurt tot en met 17:59 en tot slot de avond begint om 18:00 en duurt tot en met 23:59. Wanneer er gesproken wordt over weekend dan betreft het vrijdag 18:00 uur tot en met maandag 5:59 uur. Doordeweeks wil zeggen maandag 6:00 uur tot en met vrijdag 17:59 uur.

Tot slot kan het voorkomen in de data dat er bots⁴⁰ tussen zitten. Er zijn geen directe aanleidingen gevonden dat die daadwerkelijk aanwezig zijn, al wekt een specifiek geval wel de indruk. De tweets hiervan zijn wel meegenomen in de dataset, omdat er geen onbeleefde uitingen in zitten. Het betreft een „persoon” genaamd „Echt_WTF_Gewoon” die het meest frequent heeft getwitterd, namelijk 19.331 tweets. De onbeleefde afkorting „WTF” in zijn naam is ook niet als scheldwoord genoteerd, evenals bij andere twitteraars die een onbeleefde uiting in hun gebruikersnaam hebben staan.

3.4.2 Geen toestemming bestanden uit SoNaR corpus

Dit corpus beschikt over tweets die afkomstig zijn van een groep twitteraars die niet specifiek toestemming hebben gegeven om hun tweets te gebruiken voor onderzoek. Gezien het feit dat tweets voor dergelijke doeleinden vrijelijk te gebruiken zijn, levert dit geen problemen op.

Ongeveer een kwart van deze tweets zijn verzameld bij publieke figuren, de rest van de tweets komt bij „gewone” personen vandaan. De accountgegevens zijn gevonden op homepages en Wikipedia pagina’s van de publieke personen. Het specifiek zoeken naar twitteraccounts is gedaan om de werkelijke achtergrondgegevens te krijgen van deze personen. Op deze manier zijn beide groepen (wel en geen toestemming) gelijk aan elkaar en ontstaan er geen gaten in de dataverzameling voor het corpus. Op basis van het feit dat er specifiek gezocht is naar publieke figuren in deze database, zal hier rekening mee worden gehouden in de resultaten. De dataset bestaat uit: N=115 unieke twitteraars. In totaal zaten er 262.967 tweets in deze groep, daarvan waren er 574 tweets gemarkeerd als onbeleefd.

3.4.3 Toestemming bestanden uit SoNaR corpus

In de „met toestemming” groep hebben de twitteraars hun geslacht, leeftijd en woonplaats met de tweets meegestuurd via de website van SoNaR en zodoende impliciet toestemming gegeven om hun tweets te download en vervolgens te gebruiken voor wetenschappelijk onderzoek (in eerste instantie het SoNaR project). Deze mensen hebben vrijwillig hun data afgestaan aan de wetenschap. De dataset bestaat uit: N=368 unieke twitteraars. In totaal zaten er 681.265 tweets in deze groep, daarvan waren er 5.819 tweets gemarkeerd als onbeleefd.

⁴⁰ Citaat afkomstig van Dictionary.com uit 1999: “From “{robot}”) Any type of autonomous {software} that operates as an {agent} for a user or a {program} or simulates a human activity. On the {Internet}, the most popular bots are programs (called {spiders} or crawlers) used for searching. They access {web sites}, retrieve documents and follow all the {hyperlinks} in them; then they generate catalogs that are accessed by {search engines}. A {chatbot} converses with humans (or other bots). A {shopbot} searches the Web to find the best price for a product. Other bots (such as {OpenSesame}) observe a user's patterns in navigating a web site and customises the site for that user. {Knowbots} collect specific information from {web sites}.” Vertaald zou dit betekenen dat een bot, ook wel webrobot, een computerprogramma is dat in het geval van Twitter, automatisch berichten verstuurt en/of genereert in opdracht van degene die het computerprogramma heeft geschreven.

4: Opzet van het onderzoek

Het onderzoek waar deze scriptie op is gebaseerd gaat na of taalverruwing optreedt in Nederlandstalige tweets en op welk tijdstip onbeleefde tweets het meest frequent voorkomen. Vanaf dit punt in de scriptie zal er enkel worden gesproken over scheldwoorden of een onbeleefde uiting wanneer een vorm van taalverruwing wordt bedoeld (i.e. krachttermen, vloeken, taboewoorden, verwensingen en scheldwoorden). Deze woorden, die de meest onbeleefde vorm van Nederlandstalige communicatie vormen, zijn ondergebracht in een woordenlijst bestaand uit 120 woorden (zie bijlage V). Deze woordenlijst wordt gebruikt om de tweets uit de datasets te toetsen op deze woorden. Deze lijst met woorden kent een grote verzameling onbeleefde uitingen gebaseerd op populaire, veelgebruikte scheldwoorden. De keuze voor de woorden is deels gebaseerd op de tabel in bijlage II en deels op zoekacties via Twitter. Een kanttekening hierbij is het feit dat sommige woorden, die op meerdere manieren interpreteerbaar zijn, uit de lijst zijn gelaten. Sommige scheldwoorden zijn ook „gewone” woorden die veelal in dagelijkse communicatie terugkomen. Voorbeelden hiervan zijn „hond”, „homo”, „kanker”, „del”, en „aids”. De scripts konden wel onderscheid maken tussen het gebruik van „kanker” als voorvoegsel (e.g. „kankerlijer”), maar niet wanneer men de ziekte gebruikte als toewensing of als een boodschap van een dokter (e.g. „krijg de kanker!” en „er is bij de patiënt terminale kanker geconstateerd door de dokter.”).

Aan de hand van de informatie die beschikbaar komt na de analyse van de data zal worden bekeken of er een conclusie mag worden getrokken uit dat wat gehypothetiseerd is. Zodoende kan worden bekeken op welk tijdstip en op welke dag men het meest gebruikt maakt van onbeleefde uitingen in tweets. Ook zal er een overzicht komen van de meest frequent gebruikte scheldwoorden in alle groepen.

4.1 Vraagstelling en hypothesen

In dit onderzoek naar de beleefdheid die wordt gehanteerd op Twitter wordt er gekeken naar de frequentie van het plaatsen van onbeleefde tweets in verhouding tot het tijdstip en de dag van plaatsing. Daarnaast wordt er eveneens gekeken of de stad van waaruit de tweets zijn verzonden een verschil maakt in de mate van beleefdheid. De centrale vraagstelling in dit onderzoek luidt:

Is er een verschil waar te nemen in de beleefdheid van Nederlandstalige tweets op verschillende tijdstippen en op verschillende dagen van de week?

Op basis van de literatuur zijn er enkele hypothesen opgesteld die aansluiten bij de onderzoeksvraag. Aan de hand van de resultaten zullen de hypothesen aan het eind van hoofdstuk 5 van de scriptie worden aangenomen dan wel worden verworpen. Vervolgens zal de centrale vraagstelling van dit onderzoek worden beantwoord.

Door de aanwezigheid van stress op het werk, files, een verloren sportwedstrijd, het eindigen van het weekend en/of het gebruik van alcohol, is de kans hoger dat er vaker onbeleefde tweets op het Internet verschijnen dan wanneer deze situaties zich niet voor doen. Twitter lijkt onder andere te worden gebruikt als een communicatievehikel waarmee frustraties kunnen worden geuit en de rol van bijvoorbeeld alcohol zou hier een rol in kunnen spelen. Alcohol is een middel dat zorgt voor een lossere tong, het neemt remmingen weg. Deze remming zouden iemand in nuchtere toestand wellicht weerhouden berichten met een onbeleefd karakter te versturen via Twitter. Over het algemeen genomen wordt er voornamelijk gedronken op uitgaansavonden. Naarmate het later op de avond wordt zouden er mogelijk grotere aantallen onbeleefde tweets kunnen verschijnen. Het in kaart brengen van de activiteit per uur en een weergave van de frequentie van onbeleefde tweets op verschillende dagdelen, zal mogelijk verschillen laten zien. Op basis van deze gegevens is de volgende hypothese opgesteld:

H1. De beleefdheid in Nederlandstalige tweets varieert met de tijd van de dag.

Factoren zoals die zijn genoemd in de inleiding van de eerste hypothese komen in de regel vaker voor op vaste tijdstippen in de week. Filefrustraties en problemen in het openbaar vervoer treden bijvoorbeeld vooral op tijdens de spitsuren van maandag tot en met vrijdag. Woensdagmiddag rondom het middaguur komt @zaagmans langs om „de week door midden te zagen’ en aan het eind van de week is er de vrijdagmiddagborrel. Aansluitend neemt het cafébezoek vanaf donderdagavond toe, en zijn vanaf vrijdagavond de kantoorverplichtingen ook niet meer aanwezig waardoor er een grote deel van de bevolking de uitgaansgelegenheden bezoekt. Hierdoor zouden de tijdstippen waarop getwitterd wordt, in vergelijking met doordeweeks, enigszins opgerekt kunnen worden. Er wordt dan ook verwacht dat van donderdag tot en met zondag de onbeleefde tweets in aantallen zullen toenemen en dan in het bijzonder in de avond en in de nacht. De reden hiervoor ligt, zoals eerder gesteld, bij het alcoholgebruik van de twitteraars. De volgende hypothese is geformuleerd:

H2. Beleefdheid in Nederlandstalige tweets varieert met de dag van week.

Een veelvoorkomend verschijnsel is, mede door het gebruik van hashtags, het gebruik van afkortingen en acroniemen. Niet alleen om lange woorden af te korten, maar ook om snel en simpel, *to the point* een tweet te versturen. Door de laagdrempeligheid van Twitter (en andere social media) is een bericht zo verzonden. Snel reageren is immers inherent aan social media. Door het bericht zo kort mogelijk te maken, maar toch nog begrijpelijk voor een ieder die men wil bereiken, haalt men het uiterste uit een tweet. Een afkorting (tzt voor „te zijner tijd”) of acroniem (AFAIK voor „as far as I know”) kan soms de hele tweet beslaan, dat geldt ook voor onbeleefde tweets (FML voor „fuck my life”). Dit zijn kort en krachtig geformuleerde antwoorden. Gezien deze geringe hoeveelheid tekens die een twitteraar kan gebruiken in zijn tweets, wordt verwacht dat er ook meer afkortingen voor onbeleefde uitingen zullen worden gebruikt in tweets.

H3a. In Nederlandstalige tweets wordt met regelmaat gebruik gemaakt van afkortingen van onbeleefde uitingen.

Vervolgens wordt er in het tweede deel van de derde hypothese verwacht dat er veel woordvariaties voorkomen in tweets. Op het Internet komen vele (soms zelf verzonden) vormen van één enkel woord voor. Dit stamt af uit de tijd dat fora een populaire „sociaal” netwerk vormde en al het nodige moest worden gedaan om de „wetten” van de moderators van de website te omzeilen. Een voorbeeld uit de policy van een forum luidt als volgt:

“De administrator kan ervoor kiezen sommige woorden te censureren. Indien uw post dergelijke woorden bevat, zullen deze vervangen worden door andere woorden of door sterretjes ().Elke gebruikersgroep heeft de zelfde geblokkeerde woorden en de censurering wordt automatisch doorgevoerd. Het censuursysteem zoekt naar de woorden en vervangt deze.”* – Uit de veelgestelde vragen van www.podiumcc.nl

Online bijdragen worden opgemaakt door de twitteraars alsof het gesproken tekst betreft (Davis & Brewer, 1997) en dat was voorheen ook al zo op fora (Feldweg, Kibiger, & Thielen, 1995). Dit leidt dat tot een toename van woordverbasteringen, omdat iedereen zijn eigen interpretatie geeft aan woorden. Deze creatieve taalveranderingen komen voort uit de tijd dat het SMSen is geïntroduceerd (Locher & Watts, 2005). Voor scheldwoorden geldt over het algemeen dat deze worden geplaatst met spaties tussen de letters (e.g. k l o t e), klinkers en medeklinkers die worden vervangen door leestekens of algemene typografie (e.g. k*t, f#ck, sh!t) en er wordt gebruik gemaakt van veelvuldig toevoegen van klinkers (e.g. teeeeeering) om het woord kracht mee te geven. De hypothese luidt daarom als volgt:

H3b. In Nederlandstalige tweets wordt met regelmaat gebruik gemaakt van woordvariaties van onbeleefde uitingen.

Tot slot wordt er gekeken naar de toegang die werd verleend door de zenders van de tweets, zoals te lezen valt in paragraaf 3.4. Verwacht wordt een hogere mate van onbeleefdheid in de tweets die zijn verzameld zonder toestemming van de twitteraars, dan wanneer de tweets beschikbaar zijn gesteld door de twitteraars zelf. Wanneer men zelf zijn of haar tweets beschikbaar stelt dan zal er meer bewustzijn optreden wat men doet. Daarnaast stelt men de tweets beschikbaar voor de wetenschap, waardoor mogelijk, en vooral niet uit te sluiten, de tweets zijn opgeschoond om „zo goed mogelijk over te komen”. In de „zonder toestemming” groep zit een groep twitteraars die zorgen voor een vergelijkbare situatie. Er zijn „normale” twitteraars (75 procent) opgenomen in de dataset, maar ook publieke figuren (25 procent). De groep publieke figuren, waaronder politici, zullen een zeer gering aantal, zo niet geen, scheldwoorden gebruiken, terwijl verwacht wordt dat de overige groep twitteraars dat wel met enige regelmaat zal doen.

H4. Tweets uit de „met toestemming” groep bevatten minder onbeleefde uitingen dan tweets uit de „zonder toestemming” groep.

In hoofdstuk vijf worden de resultaten van het onderzoek weergegeven en tot slot de hypothesen worden aangenomen dan wel worden verworpen.

5: Resultaten

De belangrijkste resultaten in paragraaf 5.1 tonen aan of er een verschil bestaat tussen het versturen van tweets op bepaalde tijdstippen van de week in Nederland. Daarnaast wordt er gekeken of de dagen van de week van elkaar verschillen met betrekking tot de frequentie aan tweets. Behalve de totale frequentie zal de nadruk liggen op de verschillen in beleefdheid per dag. In paragraaf 5.2 zal vervolgens aan de hand van een weergave van de resultaten duidelijk worden of er op basis van de manier van data verzamelen, met of zonder toestemming van de twitteraars, een verschil bestaat in de gehanteerde beleefdheid in de tweets.

In tabel 1 aan het begin van paragraaf 3.4 is reeds duidelijk vermeld bij welke corpora de verschillende groepen horen, welke subgroepen er zijn en hoe de groepen qua omvang en hoeveelheid (onbeleefde) tweets zijn opgebouwd. In de paragrafen van dit hoofdstuk zal er bij elke groep nogmaals kort worden uitgelegd hoe de groep in elkaar zit, zodat het exact duidelijk is om welke groep (vier weken Nederland, Tilburg, met toestemming, zonder toestemming) het gaat.

In de volgende paragrafen zal met betrekking tot de tweets onder andere worden gekeken naar de tijd van verzenden, de dag van verzenden, of er een verschil bestaat tussen diverse groepen uit het onderzoek en of er een verband bestaat tussen tijdstippen en dagen van verzenden over een periode van vier weken. Aan de hand van deze resultaten zullen de hypothesen in paragraaf 5.3 worden aangenomen, dan wel worden verworpen. Vervolgens zullen de keuzes voor het aannemen of het verwerpen worden beredeneerd.

5.1 Resultaten grote dataset

De grote dataset is onderverdeeld in twee groepen, respectievelijk tweets uit heel Nederland (waaronder Tilburg) en de tweets die enkel afkomstig zijn uit Tilburg. De eerste groep (die heel Nederland omvat) zal worden behandeld in paragraaf 5.1.1. en de twee groep (die de omgeving Tilburg omvat) zal worden behandeld in paragraaf 5.1.2. In paragraaf 5.1.3 worden tot slot de resultaten van nieuwjaarsdag besproken en er zal worden uitgelegd waarom deze dag uit voorzorg uit het onderzoek is gehouden.

5.1.1 Heel Nederland verdeeld over vier weken

In totaal bestond deze groep uit 2.162.670 unieke twitteraars die in totaal 67.280.661 tweets verstuurd in 28 dagen. Gemiddeld verstuurden deze twitteraars 31 tweets per account in deze periode. In totaal bevat de groep 1.665.181 onbeleefde tweets (2,47 procent) die weer bestaat

uit een totaal van 1.890.567 scheldwoorden. Maandag en dinsdag waren de dagen waarin de meeste onbeleefde tweets werden verstuurd, terwijl vrijdag en zaterdag de dagen waren waarin het meest werd getwitterd.

Over de hele maand genomen wordt er voornamelijk tussen 8:00 uur 's ochtends en 2:00 uur 's nachts getwitterd. Het absolute piekmoment van alle tweets ligt rond 17:00 uur, terwijl het absolute dal twaalf uur later ligt rond 5:00 uur 's nachts. Wanneer er wordt gekeken naar de onbeleefde tweets, dan blijkt het piekmoment rond 22:00 uur 's avonds te liggen en om 5:00 uur 's nachts worden de minste onbeleefde tweets gestuurd. Opmerkelijk is dat 17:00 uur ver van het meest onbeleefde tijdstip ligt, hierdoor ligt de hoeveelheid onbeleefde tweets rond dit tijdstip op een relatief laag punt. Eveneens opmerkelijk is het gegeven dat 22:00 uur 's avonds de op zes na hoogste absolute frequentie aan tweets bevat. Hier geldt dus het omgekeerde, er worden in verhouding tot alle onbeleefde tweets relatief veel onbeleefde tweets verstuurd rond dit tijdstip.

Onbeleefde tweets worden voornamelijk 's nachts verstuurd. Verhoudingsgewijs liggen de aantallen tweets uitgesplitst per uur en totale hoeveelheid tweets in datzelfde uur het hoogst tussen 00:00 uur en 7:00 uur (tussen de 3,05 procent en 3,47 procent van de gevallen, laatstgenoemde piekmoment ligt op 3:00 uur). Het voorgenoemde piekmoment (3:00 uur 's nachts) scoort relatief zeer hoog terwijl dit tijdstip absoluut gezien een van de minst populaire momenten is om te twitteren (26.222 onbeleefde tweets uit 756.085 tweets). De bandbreedte in deze groep loopt van 2,29 procent tot en met 3,47 procent ($M=2.68$, $SD=.37$).

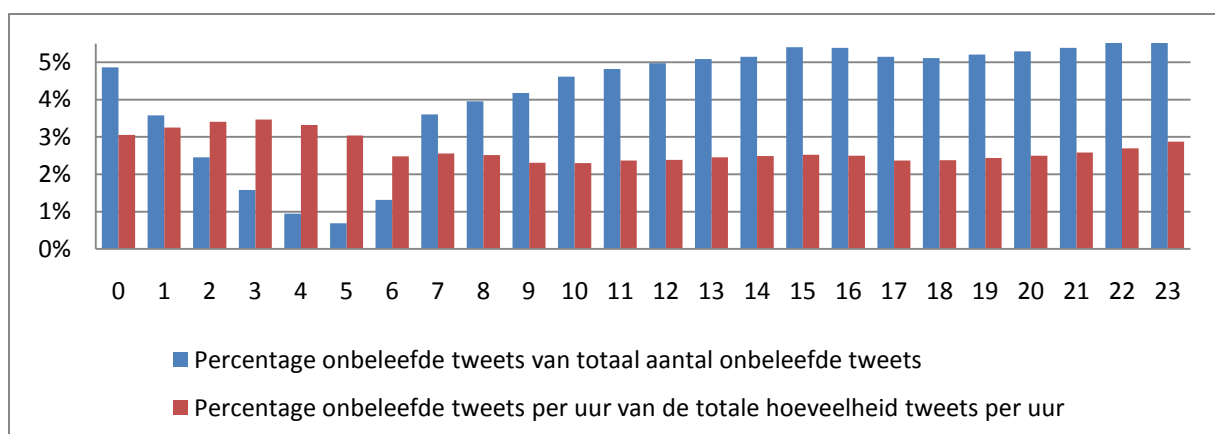
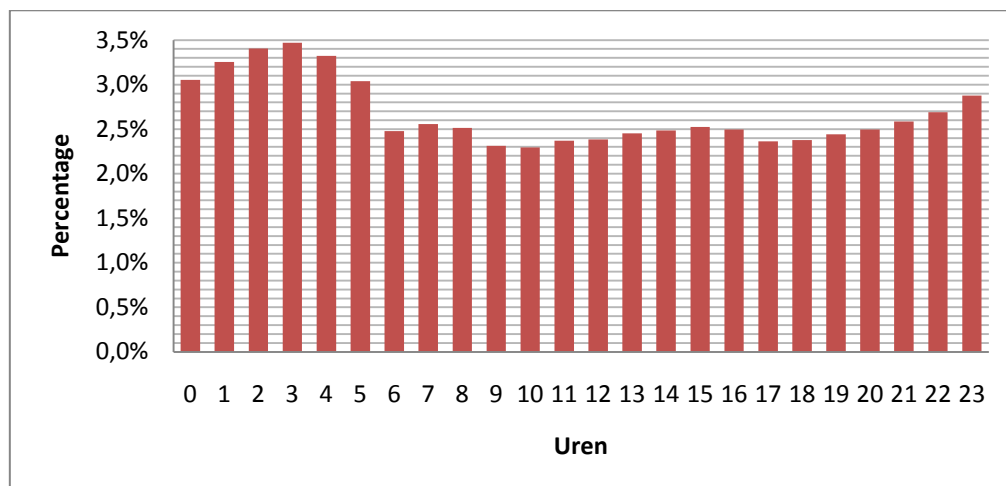
In de grafieken in bijlage VII en VIII zijn een aantal opmerkelijkheden ontdekt. Zo is er in de daggrafieken van alle tweets ontdekt dat er op enkele dagen dipjes ontstaan, zonder dat daar direct aanleiding voor is vanuit de reeds beschreven theorie. Het gaat hierbij om maandag 14:00 uur, woensdag 23:00 uur tot en met donderdagnacht, vrijdag 21:00 uur en tot slot ook zaterdag en zondag de overgang van nacht naar ochtend. Deze laatste constatering laat een veel minder abrupte overgang van nacht naar ochtend zien.

In de daggrafieken in bijlage VIII (onbeleefde tweets) verandert de week naarmate de week vordert steeds verder qua vorm tot het uiteindelijk op het „schoolvoorbeeld” lijkt. In dit geval is dat de zondag, deze weergeeft de exacte vorm zoals verwacht werd van te voren. Op dinsdag bestaat er na 15:00 uur tot 21:00 uur een dip, op donderdag vindt deze dip ook plaats maar dan al vanaf 14:00 uur. Op vrijdag vindt, evenals in bijlage VII, een daling in de hoeveelheid onbeleefde tweets. De abrupte overgang van nacht naar ochtend geldt ook hier doordeweeks en is verdwenen bij het begin van het weekend op zaterdagochtend.

Het effect van de onafhankelijke variabelen (tijdstip en weekday) op de twee

afhankelijke variabelen (onbeleefde tweets en totaal aantal tweets) is door middel van een One-Way ANOVA (F-toets) getoetst. Er is inderdaad een statistisch significant verschil ontdekt tussen beleefde tweets en onbeleefde tweets ($F(23,144) = 29.511$; $p < .001$). Alle paarsgewijze vergelijkingen lieten een significant scoreverschil zien (Tukey's HSD-procedure). De tweets die rond 5:00 uur 's nachts zijn verstuurd, scoorde het laagst; de tweets die om 22:00 uur 's avonds zijn verstuurd scoorde het hoogst.

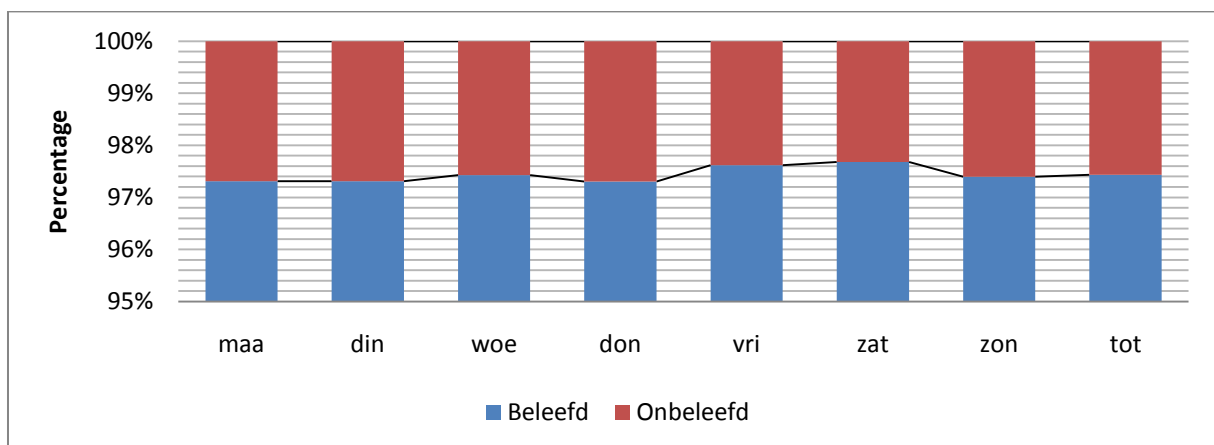
Deze percentages zijn ontstaan uit een berekening van het aantal onbeleefde tweets in een bepaald uur die vervolgens zijn afgezet tegenover het totaal aantal tweets in datzelfde uur. Overdag tussen 9:00 uur en 11:00 uur (tussen de 2,29 procent en 2,31 procent) wordt er relatief weinig gescholden, terwijl er in absolute aantallen rond 9:00 uur (69.568 onbeleefde tweets) vrijwel gelijk is aan het gemiddelde aantal onbeleefde tweets (69.382) ligt. Figuur vier weergeeft een duidelijk overzicht van hetgeen dat hierboven is verteld.



Figuur 4 – De blauwe staven geven het percentage onbeleefde tweets weer van het totaal aantal onbeleefde tweets in de ‘heel Nederland’ groep. De rode staven geven de onbeleefde tweets aan afgezet tegenover alle tweets op datzelfde tijdstip.

Wanneer de frequentie van het schelden het hoogst is, rond 22:00 uur 's avonds, bereikt de hoeveelheid scheldwoorden een aantal van 94.551 (5,68 procent van de onbeleefde tweets en 0,15 procent van de gehele groep).

Om een beeld te schetsen op welke dagen er relatief de meeste taalverruwing in tweets voorkomt, is figuur vijf gerealiseerd. Hieruit wordt duidelijk dat zaterdag de dag is waarop het minst wordt gescholden, terwijl dat op maandag een stuk vaker gebeurt. Wanneer er wordt gekeken naar de percentages aan onbeleefde tweets ten opzichte van het totaal aantal tweets, dan valt daaruit op te maken dat de verschillen tussen klein zijn.



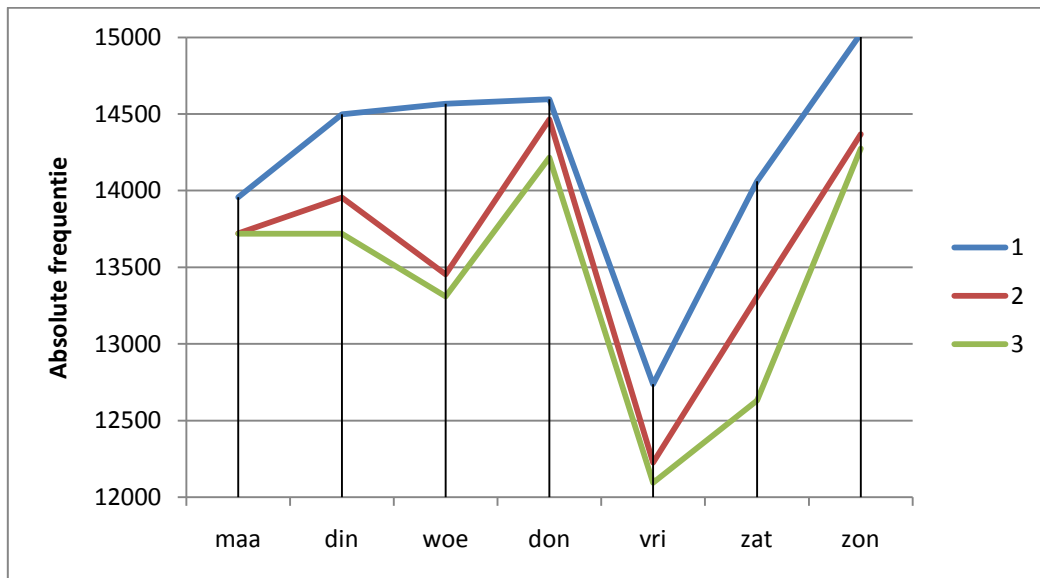
Figuur 5 – De verhouding tussen de beleefde tweets en onbeleefde tweets per dag in de ‘heel Nederland’ groep. De totale hoeveelheid wordt pas weergegeven vanaf 96 procent vanwege het feit dat de onbeleefde tweets minder dan drie procent van het totaal in beslag nemen.

Figuur zes is gerealiseerd om een beeld te schetsen van absolute frequenties van de onbeleefde tweets op de piekmomenten, weergegeven per dag. Aanvullend zijn het op één na populairste en op twee na populairste moment ook toegevoegd in de grafiek.

Eerder in de resultaten werd al bekend dat 22:00 uur het absolute piekmoment was om onbeleefde tweets te versturen. Over het algemeen blijkt dat gedurende de week de drie piekmomenten (in de figuur aangeduid met de getallen 1, 2 en 3) ook op of rond 22:00 uur liggen, een uitzondering hierop is donderdag waar de drie meetmomenten allen in de middag liggen.

Vanuit figuur zes kan er worden opgemaakt dat behalve op woensdag, de piekmomenten in gelijke maten stijgen of dalen ten opzichte van de voorgaande dag. Er heerst dus een cohesie in het twittergedrag gedurende de week. Vervolgens valt er uit de grafiek af te lezen dat er van dinsdag op woensdag een stijgende hoeveelheid onbeleefde tweets wordt verzonden op de piekmomenten, maar dat het tweede en derde piekmoment juist een

aanzienlijk kleinere hoeveelheid onbeleefde tweets bevatten. In navolging op deze daling, stijgen alle piekmomenten weer van woensdag op donderdag. Wanneer er wordt gekeken naar de tabel onder de grafiek in figuur 6, dan blijkt dat er op vrijdag in vergelijking met zaterdag de piekmoment ver van elkaar liggen, terwijl er op vrijdag in zijn totaliteit vaker onbeleefd wordt getwitterd (225.502 om 218.674 tweets).



	maa	din	woe	don	vri	zat	zon
--- 1	22:00	15:00	22:00	14:00	23:00	00:00	23:00
--- 2	00:00	16:00	21:00	16:00	00:00	23:00	22:00
--- 3	23:00	22:00	20:00	15:00	16:00	22:00	00:00

Figuur 6 – De drie momenten per dag in heel Nederland waarop de hoogste frequentie verzonden onbeleefde Nederlandstalige tweets wordt behaald. De blauwe lijn (1) weergeeft de hoogste frequentie per dag, de rode lijn weergeeft de op een na hoogste frequentie per dag en tot slot de groene lijn (3) weergeeft de op twee na hoogste frequentie per dag. Nota bene: de grafiek begint bij de ondergrens van 12.000 tweets. Een piekmoment representeert een tijdstip waarop die dag in absolute aantallen de hoogste frequentie aan onbeleefde tweets is geplaatst. De tabel onder de grafiek weergeeft de tijdstippen waarop het piekmoment werd bereikt.

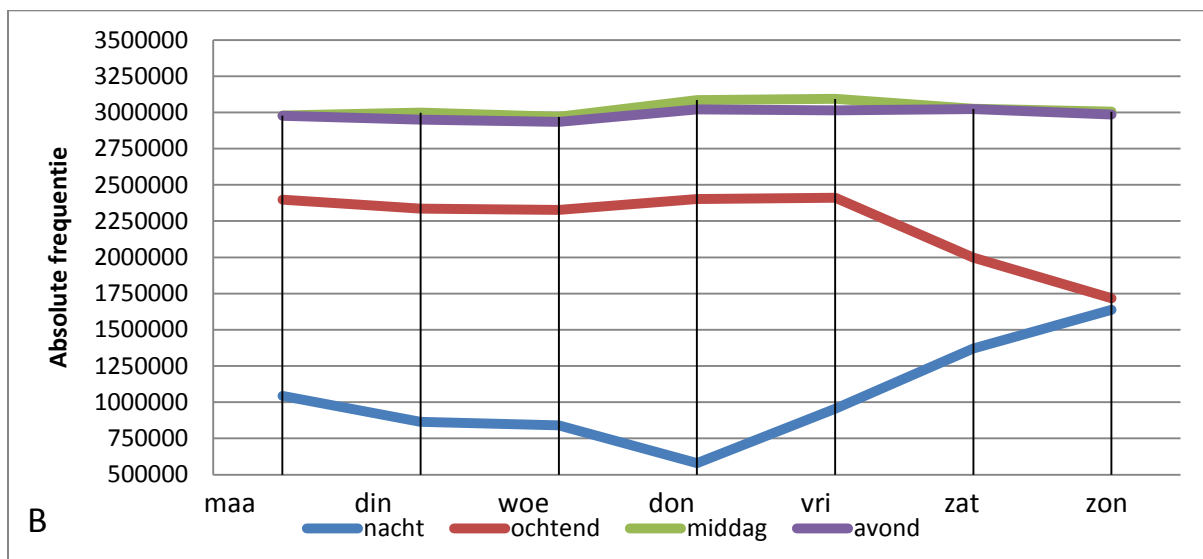
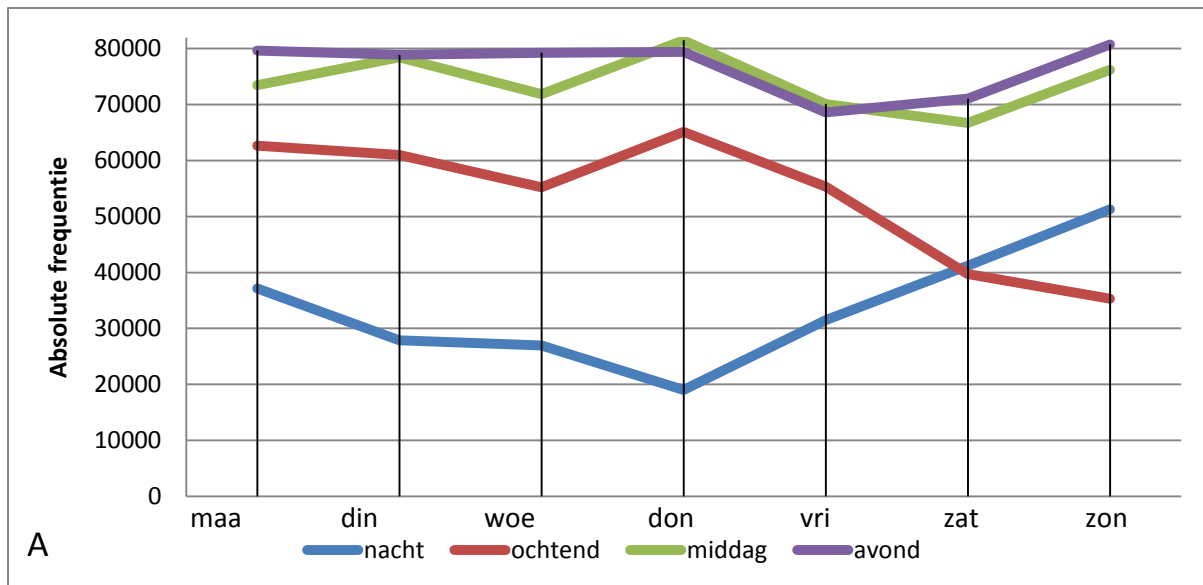
Wanneer er nader wordt gekeken naar patronen in het verzenden van tweets (beleefd en onbeleefd) dan blijkt er een algeheel patroon te bestaan. Wanneer de nacht invalt dalen de hoeveelheden tweets en vervolgens nemen de aantallen weer toe bij het aanbreken van de dag. Naarmate de dag vordert bereiken de frequenties een piek rond 17:00 uur waarna de aantallen weer langzaam afnemen. Op enkele punten in de data komen uitschieters voor die het patroon

verstoren. In dit geval gaat het om maandag 14:00 uur, dinsdag 20:00 uur en vrijdag 21:00 uur. Daarnaast verbreken zaterdag en zondag het patroon enigszins. In de weekenden behouden de absolute frequenties gedurende langere tijd eenzelfde volume in plaats van een golfend patroon.

Over het algemeen dalen de hoeveelheden van onbeleefde tweets wanneer de nacht invalt en beginnen de aantallen weer toe te nemen bij het aanbreken van de dag. De rest van de dag schommelen de frequenties heen en weer waardoor een eenduidig patroon niet aangewezen kan worden. Vervolgens worden in figuur zes de verschillen per dagdeel tegen elkaar afgezet voor zowel de complete groep als de onbeleefde tweets.

Uit figuur 7A worden afgelezen dat de absolute frequenties van het versturen van onbeleefde tweets 's middags en 's avonds het hoogst zijn op nagenoeg alle dagen van de week. Alleen op vrijdag wordt er 's middags meer gescholden, dan in de avond. Wanneer de aantallen onbeleefde tweets in deze dagdelen afneemt, stijgen juist de hoeveelheid onbeleefde tweets in de nacht tot het moment dat er op zondag een piek wordt bereikt. Op dit punt is de absolute frequentie aan onbeleefde tweets meer dan de nacht van woensdag en donderdag samengenomen.

Het overzicht van de gehele groep (fig. 7B) geeft een rustiger beeld dan wanneer er enkel naar de onbeleefde tweets worden gekeken. Overeenkomend met de voorgaande figuur stijgt 's nachts de absolute frequentie van de tweets gedurende de week en zijn de middag en avond de populairste momenten om te twitteren. De nachten vertonen een overeenkomend beeld met de onbeleefde tweets evenals de ochtenden. Er zit meer regelmaat in de ochtenden over de gehele week gezien. De middagen vertonen een ander beeld dan de onbeleefde tweets. Bij de onbeleefde tweets kwamen nog uitschieters voor, maar in de gehele groep zijn de scores eenduidig met minieme verschillen tussen de dagen. In de avonden geldt dat ook. De daling die bij de onbeleefde tweets voorkomt, komt niet meer voor in de gehele groep. Daar komt zelfs een lichte stijging van de hoeveelheid tweets voor.



Figuur 7 – In figuur 7A zijn de onbeleefde tweets die zijn verstuurd vanuit heel Nederland weergegeven per dagdeel. Hoe (on)beleefd zijn de Nederlandstalige twitteraars op een bepaalde dag in de week op een bepaald dagdeel? Op de verticale as staan de absolute frequenties van het aantal onbeleefde Nederlandstalige tweets die zijn verstuurd in de periode van 2 januari tot en met 29 januari. In figuur 7B is de totale frequentie van tweets weergegeven per dagdeel. Dagdelen: Nacht (00:00 tot 05:59), ochtend (06:00 tot 11:59), middag (12:00 tot 17:59) en avond (18:00 tot 23:59)

Wanneer er wordt gekeken naar de verdeling van de tweets over het weekend en doordeweeks⁴¹, dan blijken beide elkaar nauwelijks te ontlopen. In totaal worden er in het weekend (60 uur) 567.912 onbeleefde tweets verstuurd, terwijl dat aantal doordeweeks (108

⁴¹ Een definitie van weekend en doordeweeks wordt gegeven in paragraaf 3.4.1.

uur) 1.097.269 onbeleefde tweets bevat. Dat wil zeggen dat, wanneer je de absolute frequentie deelt door het aantal uur, dat er gemiddeld per uur 9.465,2 onbeleefde tweets in het weekend worden verstuurd en 10.159,9 onbeleefde tweets doordeweeks. Dit verschil, 694,7 tweets, is 0,1 procent van alle onbeleefde tweets en is tevens gaf een t-toets voor gepaarde metingen aan dat deze score significant is $t(1)=28,289$, $p < .025$.

5.1.2 Omgeving Tilburg

Vanuit de omgeving Tilburg werden in dezelfde 28 dagen 10.395 tweets verstuurd waaronder 186 onbeleefde tweets (1,79 procent) met daarin 218 onbeleefde uitingen. De groep bestond uit 1.152 unieke twitteraars. In totaal werd er vanaf 98 Twitteraccounts (8,51 procent) een of meerdere onbeleefde tweets verstuurd.

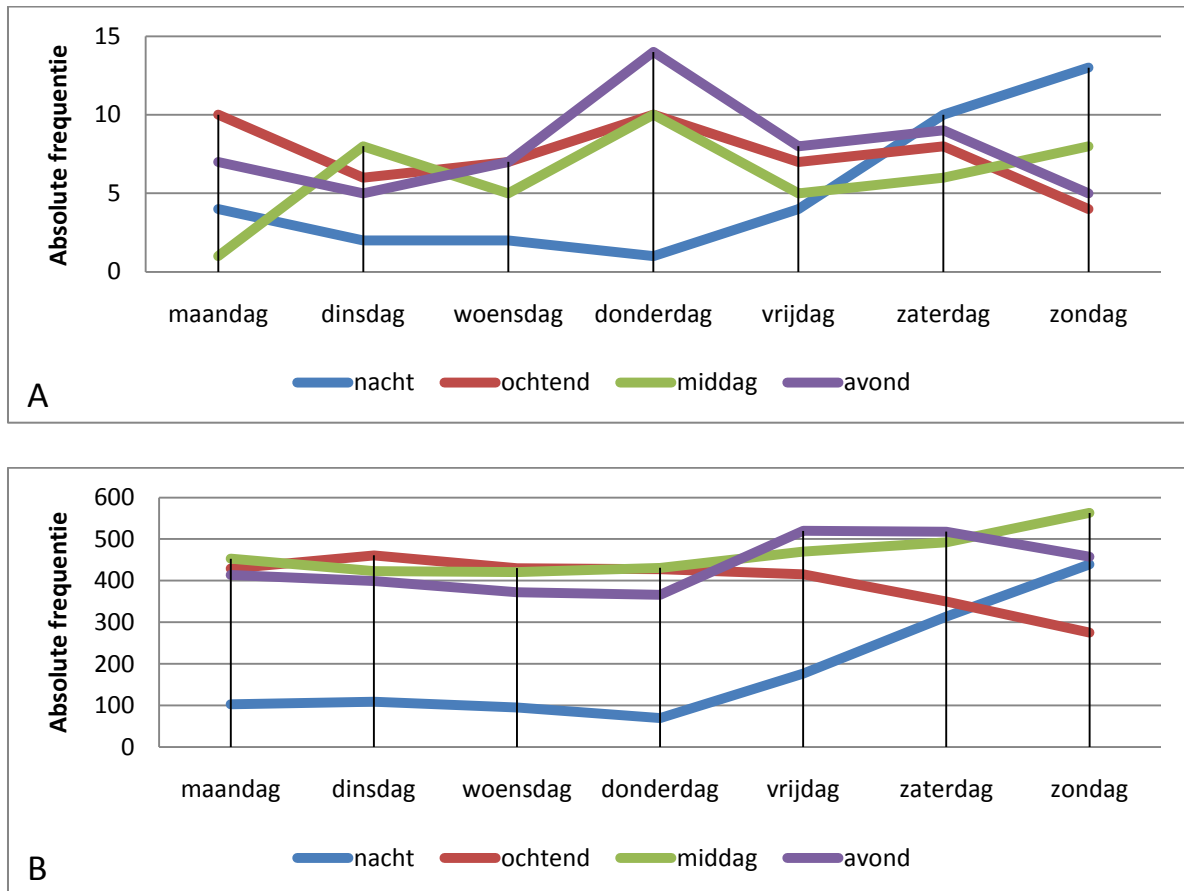
Vanuit Tilburg wordt er voornamelijk in het weekend getwitterd. Er worden op zaterdag en zondag gemiddeld 300 tweets meer verstuurd dan doordeweeks. Zondag had in absolute aantallen de hoogste frequentie, namelijk 1.735 tweets gevolgd door zaterdag met 1.674 tweets. Donderdag was voor de twitteraars de minst drukke dag met 1.295 tweets. Voor de onbeleefde tweets in deze groep liggen de verhoudingen vrijwel gelijk met de totale verdeling van tweets er is echter één uitschieter: donderdag. De dag met in totaal de laagste frequentie aan tweets scoort het hoogst qua onbeleefde tweets (35 tweets met één of meer scheldwoorden, 18,82 procent). Het laagst scoorde zowel de dinsdag als de woensdag met 21 tweets (11,29 procent).

Zowel bij alle verzonden tweets als bij de onbeleefde tweets liggen de tijdstippen waarop de absolute frequentie van het versturen van tweets in Tilburg het hoogst is, behoorlijk uiteen. Het valt op dat er op uiteenlopende tijdstippen onbeleefde tweets worden verstuurd. Wanneer er naar alle tweets wordt gekeken zit er meer regelmaat in het twittergedrag.

Tot slot worden de verschillen per dagdeel tegen elkaar afgezet voor zowel alle tweets als de onbeleefde tweets. De onderstaande twee grafieken geven weer hoe de verdeling exact is. Bij het aflezen van figuur 8A wordt duidelijk dat in de ochtend en in de avond de meeste onbeleefde uitingen worden gebruikt (respectievelijk 52 en 55 tweets). De verdeling van de frequentie schommelt heen en weer gedurende de week met een absolute piek op donderdagavond.

Met betrekking tot figuur 8B wordt het duidelijk dat er 's nachts sprake is van een schommelende frequentie aan tweets naarmate de week vordert. Maandag tot en met donderdag gaan de frequenties op en neer, maar de aantallen blijven dicht bij elkaar. In het

weekend stijgt het aantal explosief. In de ochtend valt er een patroon af te lezen in de vorm van een dalende frequentie naar mate de week vordert. Aan het begin van de week is er veel activiteit en dat bouwt zich weer af naarmate het einde van de week in zicht komt. In de middag nemen de hoeveelheden tweets toe, zodat ze in de avond weer afnemen. Ook hier, evenals 's nachts en 's middags, stijgen de aantallen vanaf vrijdag. De activiteit is een stuk hoger op donderdag- tot en met zaterdagavond.



Figuur 8 – In figuur 8A zijn de onbeleeftde tweets vanuit Tilburg weergegeven per dagdeel. Hoe (on)beleeft zijn de Tilburgse twitteraars op een bepaalde dag op een bepaald dagdeel? Op de verticale as staan de absolute frequenties van het aantal onbeleeftde tweets die zijn verstuurd in de periode van 2 januari tot en met 29 januari vanuit Tilburg. In figuur 8B is de totale frequentie tweets die verstuurd zijn vanuit Tilburg, weergegeven per dagdeel.

Dagdelen: Nacht (00:00 tot 05:59), ochtend (06:00 tot 11:59), middag (12:00 tot 17:59) en avond (18:00 tot 23:59)

Wanneer er wordt gekeken naar de verdeling van de tweets over het weekend en doordeweeks, dan blijken beide elkaar amper te ontlopen. In totaal worden er in het weekend

(60 uur) 75 onbeleefde tweets verstuurd, terwijl dat aantal doordeweeks (108 uur) 111 onbeleefde tweets omvat. Dat wil zeggen dat, wanneer je de absolute frequentie deelt door het aantal uur, dat er gemiddeld per uur 1,25 onbeleefde tweets per uur in het weekend worden verstuurd en 1,03 onbeleefde tweets per uur doordeweeks. Het verschil bestaat uit slechts 0,22 tweets. Dit is slechts 0,13 procent van alle onbeleefde tweets. Na het uitvoeren een t-toets voor ongepaarde metingen blijkt dat dit verschil niet significant is ($t(1)=1,049$, $p=.49$).

5.1.3 Nieuwjaarsdag 2012

Vanwege het feit dat nieuwjaarsdag de allereerste dag van deze dataset was, is er besloten om deze dag uit de data te halen (dat is op dit punt van het onderzoek dan ook reeds gedaan).

Aangezien er in totaal 29 complete dagen aan data zijn binnengehaald, bleef het hoofdonderzoek intact (vier keer zeven dagen). Verwacht werd dat 1 januari een bijzondere dag was (officiële nationale feestdag) en dat deze dag invloed zou uitoefenen op de resultaten. In deze paragraaf wordt bekeken of 1 januari inderdaad een bijzondere dag is geweest.

In totaal werden er op vier zondagen in de complete dataset (1, 8, 15 en 22 januari 2012) 11.689.902 tweets verstuurd, waarvan er 313.846 onbeleefd waren. Op 1 januari werden er 2.345.886 tweets verstuurd, waarvan 60.322 onbeleefde tweets (19,22 procent), terwijl er gemiddeld 3.114.672 tweets werden verstuurd op de overige zondagen. Hiervan bevatte 84.508 tweets (26,93 procent) één of meerdere scheldwoorden uit de woordenlijst.

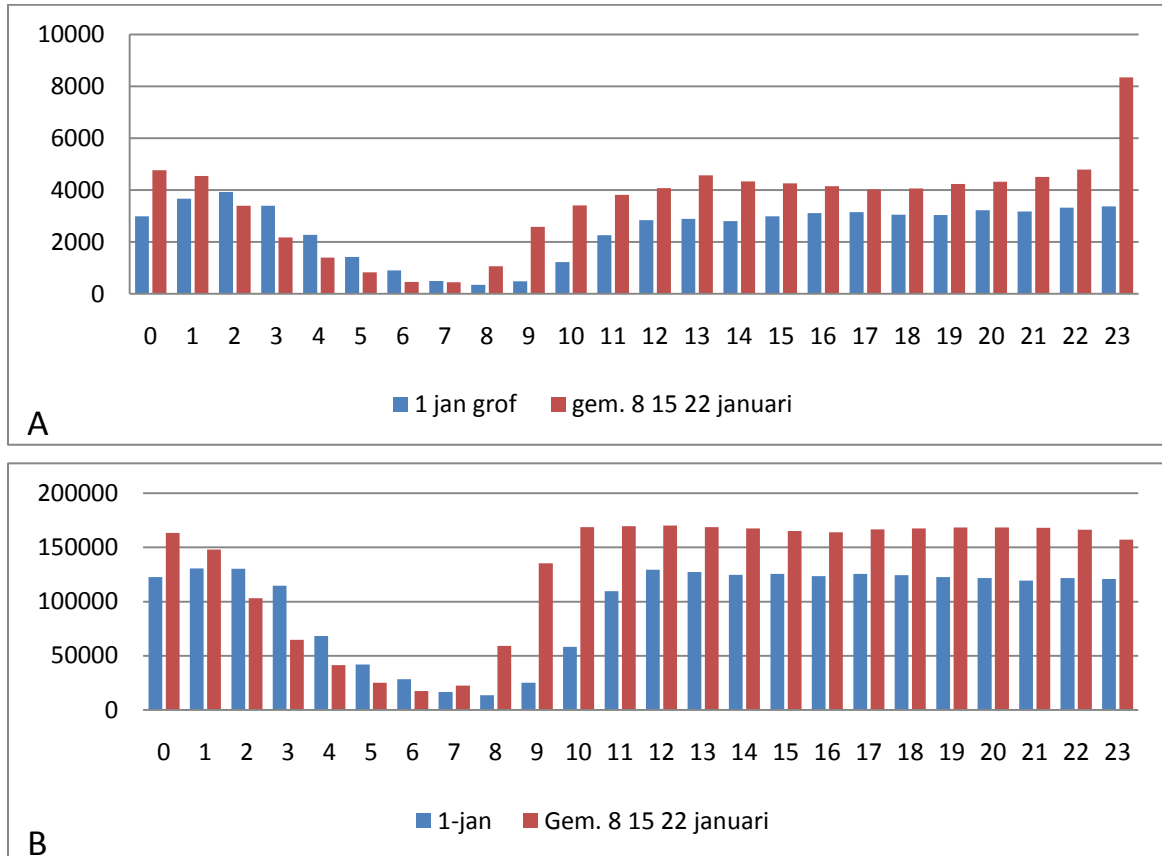
Tabel 2

Overzicht van de resultaten op respectievelijk zondag 1 januari 2012 en het gemiddelde van zondag 8, 15 en 22 januari 2012 (beide over de gehele dataset genomen). De aantallen van de overige zondagen zijn gesommeerd en vervolgens gedeeld door het aantal zondagen om tot een gemiddelde van die week te komen. De gegevens zijn afkomstig uit de 'heel Nederland' groep.

	1 januari	Gemiddelde overige zondagen	Totaal
Onbeleefde tweets	60.322	84.508	313.846
Totaal	2.345.886	3.114.672	11.689.902
Percentage van totaal	2,57	2,71	2,68

In figuur 9 wordt duidelijk dat er op 1 januari inderdaad verschillen optreden ten opzichte van 8, 15 en 22 januari. Gemiddeld scoren twitteraars op de drie zondagen na 1 januari hoger op onbeleefdheid dan op nieuwjaarsdag ($t(23)=3.881$, $p<.005$). Om te beginnen

met figuur 9A valt het op dat over het algemeen meer wordt gescholden op de drie zondagen na 1 januari. Enkel tussen 2:00 uur 's nachts en 8:00 uur 's ochtends wordt er op die dagen meer gescholden. Vanaf 2:00 uur 's nachts tot 7:00 uur 's ochtends ligt de absolute frequentie van alle tweets ook hoger dan op de andere zondagen. De rest van de dag scoren de andere zondagen gemiddeld (ruim) hoger.



Figuur 9 – In afbeelding 9 A worden onbeleefde tweets op 1 januari afgezet tegen de gemiddelde frequentie van onbeleefde Nederlandstalige tweets op de overige drie zondagen op respectievelijk 8, 15 en 22 januari.

In afbeelding 9 B wordt de totale frequentie van zowel beleefde als onbeleefde Nederlandstalige tweets op 1 januari afgezet tegen het gemiddelde aantal tweets op de overige drie zondagen. De gegevens uit zowel A als B zijn afkomstig uit de 'heel Nederland' groep.

5.1.4 Frequentie van type scheldwoorden

In groep 1 (heel Nederland) worden er in de top vijf van meest frequent voorkomende scheldwoorden voornamelijk Angelsaksische scheldwoorden gebruikt. Opgevolgd door scheldwoorden met geslachtsorganen en ziektes. Een overzicht van de meest gebruikte

scheldwoorden voor alle onderzoeksgroepen is te vinden in bijlage V. Afkortingen en woordvariaties komen in verhouding niet vaak voor, maar wel vaker dan bij de Tilburgers het geval is. De afkorting die het vaakst voorkomt staat op een tweede plaats, namelijk *kk* gevolgd door *gvd* op de tiende plaats. Aansluitend volgen *f*ck*, *f*cking* en *k*t* op respectievelijk de 20^e, 22^e en 23^e plaats. De eerste woordvariatie komt pas voor op de 30^e plaats, *fokking*.

In groep 2 (de Tilburgers) is dit als volgt. Van de in totaal 218 gebruikte scheldwoorden komen de meest voorkomende uit verschillende categorieën, te weten Angelsaksische scheldwoorden, godslasterende vloeken, ziektes en de scheldwoorden die een geslachtsorgaan benoemen. De Tilburgers gebruiken, evenals de gehele populatie, scheldwoorden uit diverse categorieën. Daarnaast worden er ook afkortingen door de Tilburgers gebruikt om een scheldwoord te uiten, daartegenover staat dat woordvariaties amper voorkomen. Er zijn twee afkortingen die hoog scoren, namelijk *kk* op de vierde plek en *gvd* op de vijfde plek. Samen zijn deze afkortingen goed voor 21,19 procent van alle scheldwoorden. Vervolgens komt pas op de negentiende plaats het zelfgecensureerde woord *k*t* voor. In de lijst komen enkel *f*ck* en *f*cking* nog voor op respectievelijk plek 21 en 22. Deze drie tezamen zijn goed voor slechts 3,39 procent van alle scheldwoorden.

Tot slot is er ook gekeken naar de meest frequent gebruikte scheldwoorden op 1 januari 2012 en de opvolgende drie zondagen in 2012. Er komen slechts kleine verschillen voor. Wanneer er een top tien van meest frequent gebruikt scheldwoorden wordt opgemaakt dan wordt duidelijk dat op zondag 1 januari dezelfde scheldwoorden gebruikt worden als op de andere zondagen. Het enige verschil is de volgorde, al blijft de top drie gelijk. Deze bestaat respectievelijk uit *kut*, *kk* en *fuck*. Ook in de periode van 2 januari tot en met 29 januari is dat de top drie van meest frequent gebruikte scheldwoorden.

5.2 Resultaten SoNaR: STEVIN dataset

5.2.1 Geen toestemming groep

De ‘geen toestemming’ groep bestaat uit 115 twitteraars en bevat in totaal 262.967 tweets. Gemiddeld werden er per twitteraar 2.286,67 tweets verstuurd in deze groep. Dit zijn zowel onbeleefde tweets als tweets die compleet vrij van scheldwoorden zijn. In totaal waren er 574 onbeleefde tweets (0,22 procent) verzonden door 87 unieke twitteraars met een of meerdere scheldwoorden in hun tweets. Deze 87 twitteraars (75,65 procent) gebruikte één of meerdere scheldwoorden in zijn of haar tweets. Gemiddeld versturen de onbeleefde twitteraars 6,6

onbeleefde tweets.

De persoon die de meeste tweets verzond met daarin een of meer scheldwoorden, verstuurde in totaal 3.613 tweets waaronder 64 onbeleefde tweets. Deze persoon scoorde naar verhouding van totaal aantal tweets en onbeleefde tweets ook het hoogst (1,77 procent). Doordat er geen *outlier* in de groep voorkomt kan er worden vastgesteld dat er geen enkel persoon verantwoordelijk is voor extreme waarden in de gemiddelden.

Wanneer er wordt gekeken naar de tijdstippen waarop het meest frequent wordt getwitterd dan loopt deze gedurende de week behoorlijk uiteen. Dit in tegenstelling tot de „met toestemming” groep waar een duidelijker patroon valt te ontdekken in de piekmomenten van het twitteren. Kijkend naar alle tweets uit de „geen toestemming” groep dan kan worden geconstateerd dat de avonduren het populairst zijn om te twitteren, met name tussen 20:00 uur en 22:00 uur.

Wanneer er enkel naar de onbeleefde tweets wordt gekeken dan blijkt dat er van dit patroon niets over is. De tijden waarop de meeste onbeleefde tweets worden verstuurd verschillen per dag van elkaar. Hierdoor kan het voorkomen dat de piekmomenten verdeeld over de week op zes verschillende tijdstippen plaatsvinden.

5.2.2 Met toestemming groep

De „met toestemming” groep bestaat uit 368 unieke twitteraars en bevat in totaal 681.265 tweets. Gemiddeld werden er per twitteraar 1.851 tweets verstuurd in deze groep. Dit zijn zowel onbeleefde tweets als tweets die compleet vrij van scheldwoorden zijn. In totaal waren er 5.819 onbeleefde tweets (0,85 procent) verzonden door 310 unieke twitteraars (84,24 procent) met een of meerdere scheldwoorden in hun tweets.

In deze groep is een duidelijk patroon te zien op welk tijdstip men tweets verstuurt. In zes van de zeven gevallen gebeurt dit tussen 20:00 en 21:00 uur. Enkel op de vrijdag ligt dit punt op 9:00 uur ’s ochtends. Ook het tijdstip met de op één na hoogste frequentie van verzonden tweets is eveneens constant. De frequenties liggen op alle dagen tussen 19:00 uur en 20:00 uur, behalve op de toch al uitzonderlijke vrijdag. Vrijwel alle piekmomenten liggen die dag op een eerder tijdstip dan de rest van de week. Dit geldt voor alle tweets.

De persoon die de meeste tweets verzond met daarin een of meer scheldwoorden, verstuurde in totaal 3.187 tweets waaronder 185 onbeleefde tweets. Deze persoon scoorde naar verhouding van het totaal aantal tweets en onbeleefde tweets ook het hoogst (5,8 procent). Doordat er geen *outlier* in de groep voorkomt kan er worden vastgesteld dat er geen enkel persoon verantwoordelijk is voor extreme waarden in de gemiddelden.

Wanneer er alleen tweets met één of meerdere onbeleefde uitingen in acht worden genomen, dan treden er meer schommelingen op in de piekmomenten, evenals in de „geen toestemming’ groep. Naarmate de week vordert loopt het piekmoment af van 22:00 uur naar 19:00 uur op zaterdag om vervolgens weer te stijgen tot 21:00 op zondag.

5.3 Samenvatting van de resultaten

De onderzoeksresultaten worden nu besproken aan de hand van de in paragraaf 4.1 opgestelde hypothesen. In deze paragraaf zullen de hypothesen worden aangenomen dan wel worden verworpen.

H1. De beleefdheid in Nederlandstalige tweets varieert met de tijd van de dag.

De eerste hypothese van dit onderzoek kan worden aangenomen. De verhouding van de frequentie van tweets en de frequentie van onbeleefde tweets neemt in de nachten geleidelijk af tot aan het begin van de ochtend waarna er weer een toename van de hoeveelheid onbeleefde woorden worden geconstateerd tot aan de nacht. Het is duidelijk geworden dat in absolute aantallen de hoogste frequentie aan onbeleefde tweets worden verstuurd tussen 22:00 en 00:00 uur terwijl er tussen 2:00 en 5:00 uur relatief gezien weinig onbeleefde tweets werden verstuurd.

Op basis van het totale aantal (beleefde en onbeleefde) tweets blijkt dat ’s nachts juist veel onbeleefde tweets worden geplaatst. 3:00 uur ’s nachts is relatief gezien het tijdstip waarop in verhouding met alle tweets de hoogste frequentie onbeleefde tweets wordt verstuurd.

H2. De beleefdheid in Nederlandstalige tweets varieert met de dag van week.

De tweede hypothese van dit onderzoek kan worden aangenomen. De verschillen tussen de dagen lopen uiteen tot een maximum verschil van 4,94 procentpunten. Dit aantal representeert een totaal van 85.239 tweets met daarin onbeleefde uitingen. De verschillen zijn als zodanig hoog bevonden dat er gesproken mag worden van een variatie met de dag van de week in de beleefdheid van Nederlandstalige tweets.

H3a. In Nederlandstalige tweets wordt met regelmaat gebruik gemaakt van afkortingen van onbeleefde uitingen.

Het eerste deel van de derde hypothese kan worden verworpen. In zowel de groepen wel toestemming en geen toestemming, als in de groepen Nederland en Tilburg is ontdekt dat er niet met regelmaat met afkortingen wordt gescholden.

H3b. In Nederlandstalige tweets wordt met regelmaat gebruik gemaakt van woordvariaties van onbeleefde uitingen.

Het tweede deel van de derde hypothese kan worden verworpen. Ondanks dat vele populaire samenvoegingen en misspellingen van onbeleefde uitingen in de woordenlijst zaten, scoorden deze woorden verhoudingsgewijs zeer laag in alle vier de groepen.

H4. Tweets uit de „met toestemming” groep bevatten minder onbeleefde uitingen dan tweets uit de „zonder toestemming” groep.

De vierde hypothese van dit onderzoek kan worden verworpen. Het is juist de „met toestemming” groep die de meeste onbeleefde tweets bevat ten opzicht van de „zonder toestemming” groep.

Nu de vier hypothesen van dit onderzoek zijn beantwoord, kan er gekeken worden naar de beantwoording van de centrale vraagstelling van het onderzoek. Dit zal in het volgende hoofdstuk gaan plaatsvinden, evenals het interpreteren en bediscussiëren van de onderzoeksresultaten.

6: Conclusie en discussie

6.1 Inleiding

In paragraaf 6.2 wordt een antwoord gegeven op de centrale vraagstelling van dit onderzoek. Allereerst wordt er ingegaan op de resultaten uit de grote dataset om vervolgens de resultaten uit het SoNaR: STEVIN corpus te behandelen. De conclusie wordt vervolgens in paragraaf 6.3 verder bediscussieerd met aansluitend de limitaties van dit onderzoek en de mogelijkheden voor toekomstig onderzoek aan de hand van dit onderzoek. Tot slot zal er een interpretatie worden gegeven van de resultaten waarmee deze masterscriptie zal worden besloten.

6.2 Conclusie

Taalverruwing op het Internet is volgens een deel van de Nederlandse bevolking een probleem (in wording). Men wordt niet alleen grover, maar gaat ook vaker over op het gebruik van grove taal. Daarbij speelt ook mee dat er gebruik wordt gemaakt van andere typen scheldwoorden dan een decennium geleden. Waar vroeger werd gevloekt met hogere machten, hanteert men tegenwoordig steeds vaker (ernstige en actuele) ziektes en Angelsaksische krachttermen. In zijn algemeenheid stapt men volgens prof. dr. Van Sterkenburg over op het gebruik van scheldwoorden, vanwege de hoge nood om onze emoties te uiten en op Twitter bleek dit niet anders.

Na het beantwoorden van de hypothesen in paragraaf 5.3, kan er inmiddels een antwoord worden gegeven op de in paragraaf 4.1 opgestelde centrale vraagstelling van dit onderzoek. Hieronder staat de centrale vraagstelling nogmaals geformuleerd:

Is er een verschil waar te nemen in de beleefdheid van Nederlandstalige tweets op verschillende tijdstippen en op verschillende dagen van de week?

De beleefdheid in de grote dataset bleek voor zowel Tilburg als heel Nederland tegen te vallen. Veel tweets in de grote dataset bevatten één of meerdere scheldwoorden. Gezien vanuit het standpunt dat onbehoorlijk taalgebruik doorgaans niet *comme il faut* is, en zo is dat ook op Twitter, valt het relatieve aantal tweets hoog te noemen. Zo werd er bijvoorbeeld relatief meer gescholden op Twitter dan in normale gesprekken of op MySpace. Daarnaast gaf de absolute frequentie per uur aan dat er vaker overdag werd gescholden dan dat er 's nachts werd gescholden, maar relatief gezien was dit juist andersom. Daaruit bleek dat er 's nachts in

hogere mate onbeleefde tweets in de data voorkwamen, hetgeen er op kan duiden dat er in (bijvoorbeeld) het uitgaansleven meer aanleiding is om een onbeleefde tweet te versturen. Daarnaast maakte het ook nog uit welke dag van de week het was. Zo bleek maandag de meest onbeleefde dag en zaterdag de minst onbeleefde dag. Daarnaast was vrijdag de drukste dag om op te twitteren en zondag net als in het echte leven, een rustdag voor de twitteraars.

Op enkele dagen ontstonden dipjes in de lijngrafieken zoals deze te zien zijn in bijlage VII en VIII. De momenten hiervoor zijn niet perse vreemd. Zo is een dip rond het middaguur ongetwijfeld het gevolg van een hogere activiteit tijdens de lunch en de minder abrupte overgang van nacht naar ochtend in het weekend heeft te maken met het feit dat er meer mensen 's nachts actief zijn en men op minder reguliere tijdstippen wakker wordt in het weekend. Waarom er vanaf woensdagavond 23:00 tot donderdagochtend een veel mindere activiteit op Twitter is en er op vrijdag om 21:00 uur een dip optreedt, valt niet te verklaren aan de hand van de theorie of voorspelbare situaties.

Het verwachte verschil tussen twitteraars die toestemming gaven om hun tweets te gebruiken en twitteraars waarvan de tweets zonder toestemming werden gebruikt, bleef uit. Ondanks het feit dat de verdeling in de 'zonder toestemming' groep bestaat uit 75 procent doorsnee Nederlanders en 25 procent bekende Nederlanders, heeft deze laatstgenoemde groep toch invloed gehad op de resultaten. Bekende Nederlanders, waarvan het merendeel politici betrof, gebruiken, zo werd aangenomen, over het algemeen zeer correct taalgebruik. Zoals het nu blijkt komen de woorden uit de woordenlijst vrijwel nooit voor in hun tweets. De hypothese is dus verworpen.

Men scheldt in alle vier datasets met nagenoeg dezelfde scheldwoorden. De top-15 van meest gebruikte scheldwoorden (op een totaal van 120 woorden) is slechts in de volgorde van meest frequent gebruikt tot minst frequent gebruikt, afwijkend. Dit valt echter te wijten aan de hoeveelheid onbeleefde tweets in de datasets van SoNaR: STEVIN. Afkortingen komen nauwelijks voor en het gebruik van leestekens of bijzondere wijze van spelling zijn amper voorgekomen in de hele dataset. Mede dankzij deze constatering kan er worden vastgesteld dat de creativiteit laag ligt bij de onbeleefde Twitteraars, men scheldt vrijwel alleen met die vijftien woorden.

Om tot een antwoord op de centrale vraagstelling te komen, kan er worden gerefereerd aan de resultaten in hoofdstuk 5.1.1 tot en met 5.1.3. Hieruit kwam naar voren dat zowel in heel Nederland als in Tilburg de beleefdheid varieert op zowel tijd als dag. Vooraf werd verwacht dat het bioritme van de mens en wederkende dagactiviteiten een indicatie geven van de hoeveelheid tweets. Dit blijkt achteraf ook grotendeels overeen te komen met de resultaten.

De verschillen die optreden in de data, waardoor de centrale vraagstelling met „ja’ kan worden beantwoord, waren vooral merkbaar in relatieve zin. Hieruit bleek dat er, zoals verwacht, verschillen waarneembaar zijn doordat er ’s nachts meer taalverruwing optreedt op Twitter. Ook naarmate de week vorderde werden er meer grove tweets verstuurd en werd het piekmoment verschoven naar een later tijdstip. Interessant om te zien is het feit dat maandag de minst beleefde dag is als het gaat om tweets. Men lijkt echt last te hebben van een zware start van de week, wellicht dat het piekmoment op maandagnacht daarmee te maken heeft (de laatste ronde in de kroeg) en de daarop volgende blauwe maandag.

6.3 Discussie

6.3.1 Limitaties en vervolgstudies

Het onderzoek heeft te maken met enkele limitaties die in deze paragraaf worden besproken. Een limitatie van het onderzoek is het feit dat er niet met zekerheid iets kan worden gezegd over „de Tilburger’. Het gaat immers over een persoon die vanuit het gebied van Tilburg tweet. De nauwkeurigheid van de geolocaties⁴² kloppen, daarover geen twijfel, maar het feit of de eigenaar van de tweet ook daadwerkelijk in dat gebied woont is nog maar de vraag. Per slot van rekening, als persoon A in Tilburg woont maar werkt in Amsterdam en op beide locaties twittert wordt vanaf zijn account meerdere locaties aangeroepen. Daarnaast moet men handmatig aangeven of ze een locatieaanduiding bij hun tweets willen of niet. Standaard staat dit uit, waardoor soms nuttige metadata mist in de onderzoeksdata. Aanvullend op het voorgaande punt, maar niet onbelangrijk, is het feit dat men achteraf alle locatieaanduidingen kan verwijderen. Daarnaast heeft een twitteraar ook nog de keuze om een stad in te vullen waar diegene woont. Echter is dat niet verplicht en is het makkelijk te vervalsen. Locaties van twitteraars is eigenlijk geen valide meetinstrument, daar het geen verplicht onderdeel van een account binnen de Twitter omgeving is.

In dit onderzoek is vrijwel geen rekening gehouden met de invloed van actuele gebeurtenissen op het taalgebruik van twitteraars. Wanneer er bijvoorbeeld een voetbalwedstrijd is geweest, wanneer er examens zijn of wanneer er bijvoorbeeld vakantie is. Vermoedelijk spelen deze en vele anderen factoren een rol in mate van onbeleefd taalgebruik op Twitter. Twitter is namelijk een goede weergave van de actualiteiten.

Tot slot moet er worden opgemerkt dat de tweets van zondag 1 januari 2012 uit de data zijn gehaald. De invloed van zondag 1 januari was te nadrukkelijk aanwezig waardoor de

⁴² <http://nl.wikipedia.org/wiki/Geolocatie>.

data enigszins gemanipuleerd is. Door deze dag af te zetten tegenover drie andere zondagen blijkt dat er het een invloedrijke dag in de data zou zijn. Dat wil zeggen dat er op die datum langer en vaker 's nachts is getwitterd. Over het algemeen gaat men later naar bed en wil men elkaar nog de beste wensen doorgeven en vervolgens slaapt men langer uit. Dit zorgt voor een scheve verdeling van de tweets al wordt dit gecompenseerd door het feit dat er nog drie zondagen in de database zitten. Voor een goede weergave van de realiteit in vervolg onderzoek zullen er tweets moeten worden gedownload in een maand zonder vakanties, feestdagen of bijzonderheden of deze dagen moeten op een dergelijke wijze worden gefilterd.

Voor vervolgstudies op deze masterscriptie zijn diverse onderwerpen interessant om nader te bekijken. Allereerst is het voor sommige bedrijven fijn om te weten hoe er over hen wordt geschreven op Twitter. Een voorbeeld hiervan zijn de landelijke vervoersbedrijven. Wanneer er treinen uitvallen dan zal er vermoedelijk een tendens ontstaan van negatieve uitlatingen over de Nederlandse Spoorwegen. Hoe verhoudt het scheldgedrag zich tot het gebruik van merknamen die de actualiteiten bezig houden op de tijdlijnen van twitteraars.

Wanneer een persoon zich onbeleefd uitlaat op Twitter, dan heeft dat wellicht ook consequenties voor hem of haar. Het is echter niet onderzocht wat die consequenties dan zijn en op welke schaal dit invloed kan hebben. Heeft dit invloed op hoe mensen (volgers of Facebook vrienden) tegen je aankijken, invloed op je werk(relaties) en sollicitaties of op de onderlinge verstandshouding tussen de twitteraar en contacten uit het dagelijks leven. Kan iemands online gedrag zijn of haar offline gedragingen doen veranderen?

Omdat er in deze scriptie voor een periode direct na het nieuwe jaar is gekozen, is het interessant om in de zomertijd te kijken naar de beleefdheid. Daarnaast zou er ook naar tweets rond het terugzetten of voorruit zetten van de tijd kunnen worden gekeken. Op deze manier zou er dus nader naar het bioritme van de mens worden gekeken. Tot slot is het wellicht interessant om nader te kijken naar kantoortijden en het verband van scheldwoorden. De data uit deze scriptie kan in dat geval worden gebruikt worden als referentiepunt.

Dit onderzoek is grotendeels gebaseerd op literatuur over beleefdheidstheorieën en sociale media onderzoeken. Er zou ook kunnen worden overwogen om op grote schaal, mogelijk via Twitter, enquêtes af te nemen waarin men aangeeft wat zij van hun twittergedrag vinden of van twittergedrag in het algemeen (bijvoorbeeld van diegene die zij volgen).

6.3.2. Interpretatie van de resultaten

De vermoedens dat er op Twitter veel gescholden wordt, zijn redelijk gegrond zo blijkt uit dit onderzoek. Omdat „veel’ een begrip is dat door een ieder zelfinterpreteerbaar is, zal er onder de

bevolking een verschil van mening zijn of het ook daadwerkelijk „veel’ is. In dit onderzoek is gesteld, dat wanneer een tweet een scheldwoord bevat, deze onbeleefd is. Wanneer dus meerdere Nederlandstalige tweets een of meer scheldwoorden bevatten, dan kan er worden gesteld dat er met enige regelmaat onbeleefd wordt getwitterd. Naar aanleiding van onderzoek van Jay (2009) is er gekozen voor een grens om te bepalen wat „veel’ nu eigenlijk is. Naar aanleiding van deze gegevens is de centrale vraagstelling van dit onderzoek ook beantwoord met „ja, tweets zijn onbeleefd’. Naar aanleiding van diverse toetsen is er een significant waarneembaar verschil gevonden op verschillende tijdstippen en op verschillende weekdays.

Een simpele zoektocht op het Internet geeft duizenden meningen over wat wel en wat absoluut niet geschreven mag worden. Men is in dit land (en met name op het Internet) vooral vrij om te zeggen wat hij of zij wil, mits het binnen de grenzen van het strafrechtelijke valt. De één noemt het vrijheid van meningsuiting wanneer er tweets online verschijnen met daarin onbeleefde uitingen, de ander spreekt van een gebrek aan opvoeding, en weer een ander merkt op dat er verloedering (veelal bij de jeugd) optreedt. Het mogen uitten van emoties is ook een populair weerwoord op de vraag of het toelaatbaar is dat men scheldt op Twitter.

Het feit dat er applicaties worden ontwikkeld om men in de nachtelijk uren te behoeden voor dergelijk ongewenst twittergedrag⁴³ lijkt niet voor niets te zijn. Dit onderzoek toont aan dat men tegen zichzelf in bescherming dient worden genomen om te zorgen voor een minimaal gezichtsverlies. Het eerder aangekaarte gebruik van alcohol bevattende dranken lijkt door de applicatie Social Media Sobriety test⁴⁴ te worden bevestigd als „een’ probleem van ongewenste uitingen op sociale media. Deze applicatie, die een persoon test op zijn of haar nuchterheid, kan iemands toegang tot allerlei sociale media blokkeren die via de mobiele telefoon benaderbaar zijn.

⁴³ In 2008 ontwikkelde Google al een applicatie wat ervoor zorgde dat je vanaf vrijdagavond enkele rekensommen moest oplossen alvorens een e-mail verstuurd kon worden. Maar ook mobiele applicaties zoals „Last Night Never Happend’ zijn ontwikkeld om „een te snel verstuurd bericht tegen te houden’. De ontwikkelaars spreken van een „*morning after app*’. (<http://lastnightapp.com/>).

⁴⁴ http://www.webroot.com/En_US/sites/sobrietytest/index.html.

Referenties

- Abbott, C. (1998). Making Connections: Young people and internet. In J. Sefton-Green, *Digital Diversions: Youth culture in the age of multimedia* (pp. 84-105). London, UK: UCL Press.
- Boon den, T. (2007). Ongezien de tiefus. De taal van 'shocklog' GeenStijl. *Onze Taal* , 76 (12), 328-331.
- Boudourides, M. A. (1995). Social and Psychological effects in Computer-mediated Communication. *Neties '95*. Piraeus.
- Bousfield, D. (2008). *Impoliteness in Interaction*. Amsterdam, Noord-Holland, Nederland: John Benjamins Publishing Company .
- Brown, P., & Levinson, S. C. (1987). *Politeness: Some universals in language usage*. Cambridge, UK: Cambridge University Press.
- Crystal, D. (2001). *Language and the Internet*. Cambridge: Cambridge University Press.
- Culnan, M. J., & Markus, M. L. (1987). Information technologies. In F. M. Jablin, *Handbook of organizational communication: An interdisciplinary pererspective* (pp. 420-443). Newbury, CA: Sage.
- Davis, B. H., & Brewer, J. P. (1997). *Electronic discourse: linguistic individuals in virtual space*. Albany, NY, US: State University of New York Press.
- de Jongh, M. (2007). *Vloeken in Nederland: trends sinds 1995*. Veenendaal: TNS-NIPO.
- Etzioni, A. (1996). *The New Golden Rule: Community And Morality In A Democratic Society*. New York: Basic Books.
- Feldweg, H., Kibiger, R., & Thielen, C. (1995). Zum Sprachgebrauch in deutschen Newsgruppen. In M. Beisswenger, L. Hoffmann, & A. Storrer, *Osnabrücker Beiträge zur Sprachtheorie* (pp. 143-154). Duitsland: Redaktion OBST (2004).
- Fulk, J., Steinfield, C. W., Schmitz, J., & Power, J. G. (1987). A Social Information Processing Model of Media Use in Organizations. *Communication Research* , 14 (5), 529-552.
- Goffman, E. (1967). *Interaction ritual;; Essays on face-to-face behavior*. Garden City, NY, US: Doubleday Books.
- Graham, P. (2006). *Hypercapatilism: New media, Language, and Social Perceptions of Value*. New York, New York: Peter Lang Publishing Inc.
- Holmes, J. (2008). *Introduction to Sociolinguistics* (3e editie ed.). London and New York: Pearson Education.

- Huffaker, D. A., & Calvert, S. L. (2005). Gender, Identity, and Language Use in Teenage Blogs. *Journal of Computer-mediated Communication* , 10 (2).
- Huls, E. (2002). Beleefdheid als communicatief principe. In T. Janssen, *Taal in gebruik; een inleiding in de taalwetenschap* (pp. 143-161). Den Haag: SDU.
- Java, A., Song, X., Finin, T., & Tseng, B. (2007). Why We Twitter: Understanding Microblogging Usage and Communities. *Joint 9th WEBKDD and 1st SNA-KDD Workshop '07* (pp. 1-10). Baltimore County: University of Maryland.
- Jay, T. (2009). The utility and ubiquity of taboo words. *Perspectives on Psychological Science* , 4 (2), 153-161.
- Joinson, A. N. (2002). *Understanding the Psychology of Internet Behaviour: Virtual Worlds, Real Lives*. Basingstoke, Hampshire: Palgrave Macmillan.
- Kapitein, J. (2006, Februari 28). *Aanspreekvorm en Blikrichting (master thesis)*. Retrieved Januari 9, 2012, from <http://igitur-archive.library.uu.nl/student-theses/2006-0324-083753/UUindex.html>
- Kasper, G. (1990). Linguistic politeness: Current research issues. *Journal of Pragmatics* (14), 193-218.
- Kim Sang, E. T. (2011). Het gebruik van Twitter voor Taalkundig Onderzoek. *TABU: Bulletin voor Taalwetenschap* , 39 (1/2), 62-72.
- King, R. C., & Xia, W. (1997). Media Appropriateness: Effects of Experience on Communication Media Choice. *Decision Sciences* , 28 (4), 877-910.
- Lakoff, R. T. (1989). The limits of politeness: Therapeutic and courtroom discourse. 101-129.
- Lankshear, C. (1997). *Changing literacies*. Buckingham, UK: Open University Press.
- Leech, G. N. (1983). *Principles of Pragmatics*. London en New York: Longman.
- Locher, M. A., & Watts, R. J. (2005). Politeness theory and relational work. *Journal of Politeness Research* , 1 (1), 9-33.
- Merchant, G. (2001). Teenagers in cyberspace: an investigation of language use and language change in internet chatrooms. *Journal of Research in Reading* , 24 (3), 293-306.
- Oostdijk, N., Reynaert, M., Monachesi, P., Noord, G. J., Ordelman, R., Schuurman, I., et al. (2008). From D-Coi to SoNaR: A reference corpus for Dutch. *Proceedings of the sixth international language resources and evaluation (LREC 2008)*, (pp. 1437-1444). Marrakech.
- Ouden, d. H., & Wijk, v. C. (2007). Om vet gaaf op te kicken: overjongerentaal en het gebruik ervan in productadvertenties. *Tijdschrift voor Communicatiewetenschap* , 35 (3), 232-248.

- Parks, M. R., & Floyd, K. (1996). Making Friends in Cyberspace. *Journal of Communication* , 46 (1), 80-96.
- Petrovic, S., Osborne, M., & Lavrenko, V. (2010). The Edinburgh Twitter Corpus. *Proceedings of the NAACL HLT 2010 Workshop on Computational Linguistics in a World of Social Media* (pp. 25-26). Los Angeles, California: Association for Computational Linguistics.
- Santrock, J. W. (2007). *Adolescence* (11e editie ed.). Boston, MI, US: McGraw-Hill.
- Smith, A. (2011). *Why Americans use social media*. Washington, D.C.: Pew Research Center.
- Sterkenburg, v. P. (1997). *Vloeken. Een cultuurbepaalde reactie op woede, irritatie en frustratie*. Leiden: Sdu/Standaard.
- Sterkenburg, v. P. (2007). *Woorden van en voor emotie*. Retrieved 1 15, 2012, from Leiden Univ:http://www.leidenuniv.nl/tekstboekjes/content_docs/afscheidsrede_van_sterkenburg.pdf
- Subrahmanyam, K., Smahel, D., & Greenfield, P. (2006). Connecting development constructions on the Internet: Identity presentation and sexual exploration in online teen chat rooms. *Development Psychology* (42), 395-406.
- Thelwall, M. (2008). Fk yea I swear: Cursing and gender in a corpus of MySpace pages. *Corpora* (3), 83-107.
- Tjong Kim Sang, E. F. (2011). Het gebruik van Twitter voor Taalkundig Onderzoek. *TABU: Bulletin voor Taalwetenschap* , 39 (1/2), 62-72.
- TNS-NIPO. (2006, oktober 5). *Normen- en waardendebat Balkenende moet doorgaan*. Retrieved december 1, 2011, from TNS-NIPO: <http://www.tns-nipo.com/tns-nipo/nieuws/van/normen--en-waardendebat-balkenende-moet-doorgaan/>
- Turner, J., Grube, J., & Meyers, J. (2001). Developing an optimal match with in online communities: an exploration of CMC support communities and traditional support. *Journal of communication* , 51 (2), 231-251.
- Walther, J. B. (1996). Computer-mediated communication: Impersonal, interpersonal and hyperpersonal interaction. *Communication Research* , 23 (1), 3-43.
- Walther, J. B. (1992). Interpersonal effects in computer-mediated interaction: A relational perspective. *Communication Research* , 19 (1), 52-90.
- Walther, J. B. (2008). Social Information Processing Theory. In L. A. Baxter, & D. O. Braithewaite, *Engaging Theories in Interpersonal Communication: Multiple Perspectives* (pp. 391-404). Thousand Oaks, CA, US: Sage Publications, Inc.
- Walther, J. B., & Parks, M. (2002). *Cues filtered out, cues filtered in*. Thoasand Oaks, CA, US: Sage Publications, Inc.

- Walther, J. B., Burgoon, B., & Judee, K. (1992). Relational Communication in Computer-Mediated Interaction. *Human Communication Research* , 19 (1), 50-88.
- Walther, J., & Parks, M. (2002). Cues filtered out, cues filtered in. In M. L. Knapp, & J. A. Daly, *Handbook of interpersonal communication* (p. 845). Thousand Oaks, CA. US: Sage Publications.
- Watts, R. J. (2003). *Politeness*. Cambridge: Cambridge University Press.

Bijlagen

Bijlage I	-	Neologismen: De verklarende woordenlijst voor Twitter
Bijlage II	-	Tabel uit “woorden van en voor emotie” Van Sterkenburg, 2007
Bijlage III	-	Veelgebruikte bestandstypen voor Twitteronderzoek
Bijlage IV	-	Shell scripts
Bijlage V	-	Scheldwoordenlijst
Bijlage VI	-	Meest frequent gebruikte scheldwoorden
Bijlage VII	-	Grafiek
Bijlage VIII	-	Grafiek
Bijlage IX	-	Matrixen van alle twitterdata
Bijlage X	-	Zoekopdracht Tjong Kim Sang
Bijlage XI	-	SPSS Output

Bijlage I – Neologismen en de verklarende woordenlijst voor Twitter

Dreigtweets: een dreigtweet is een bericht dat via Twitter wordt verspreid waarin een persoon een dreigement uit.

Hashtag: Ook wel het hekje (#). De hashtag geeft aan dat er een bepaald onderwerp in de tweet wordt besproken. Op deze manier hoopt degene die het bericht plaatst vaker wordt gevonden door mensen die daarnaar opzoek zijn (bijvoorbeeld bedrijven die een product onder de aandacht proberen te krijgen) of men probeert een bepaald onderwerp *trending* te maken. Wanneer een onderwerp trending is, is het een van de populairste getwitterde hashtags op Twitter.

Re-tweet: Een bericht dat interessant wordt gevonden, doorsturen en zodoende verspreiden onder jouw volgers.

Twadidos: Een echt neologisme. Twitter en het Spaanse „adidos’ zijn samengevoegd tot Twadidos, oftewel het afsluiten van een Twitterdag of Twittergesprek met iemand.

Twarewell: Twarewell is gelijk aan bovenstaande uitleg, echt gaat het hier om het Engelse „farewell’.

Tweep(s): De gebruikers van Twitter

Tweet: Een Twitterbericht van maximaal 140 karakters lang genaamd naar het geluid dat een vogel (het logo van Twitter) maakt.

Tweezer: Een oud persoon die Twitter gebruikt. „Oud’ is hierin echter niet gedefinieerd door de bedenker van het woord, Dinah Hunt.

Twello: Een welkomstwoord, *hello*, vervoegd met het woord Twitter.

Twestival: Een Twitter festival is een wereldwijde beweging die één dag duurt die compleet de kracht van het sociale medium gebruikt om een offline evenement neer te zetten Het doel is een lokaal (daar waar het festival wordt gehouden) probleem aanpakken.

Twexit: Hiermee wordt door een tweep aangegeven dat Twitter voor onbepaalde tijd wordt afgesloten om andere zaken te doen.

Twibbatical: Een poosje geen Twitter meer gebruiken heet ook wel een Twibbatical.

Twitter: Twitter is een social medium en microblog dienst.

Twitterology: De kunst en wetenschap van het gebruik van Twitter op (de meest) effectieve wijze.

Twitterverse: Het Twitter universum.

Twodellen: Roddelen op Twitter.

Twokkies: Typisch voor dit onderzoek: Tokkies op Twitter. Tokkie is een geuzennaam voor een groep mensen die zich asociaal opstelt in de samenleving, genaamd naar de Amsterdamse familie Ruijngaart-Tokkie uit de documentaire Familietrots. Een persoon die zich dus asociaal uitlaat op Twitter heet een Twokkie.

Tworellen: Borrelen.

Twunchen: Lunchen.

Meer neologisme en veelgebruikte woorden zijn te vinden op het zeer uitgebreide „woordenboek van Twitter”: <http://www.twittonary.com/>

Bijlage II – Tabel uit “woorden van en voor emotie” Van Sterkenburg
„07

Datum	Google pagina's in het Nederlands	Resultaten
10-9	shit	5.140.000
10-9	kut	6.280.000
10-9	fuck	3.150.000
10-9	godverdomme	418.000
10-9	gvd	588.000
10-9	verdomme	916.000
10-9	verdorie	465.000
10-9	godver	156.000
10-9	godallemachtig	15.400
10-9	jeetje	1.030.000
10-9	tering	876.000
10-9	potverdomme	19.000
10-9	potverdorie	73.400
10-9	jezus mina	25.900
10-9	jezusmina	1.350

Bijlage III - Veelgebruikte bestandstypen voor Twitteronderzoek

.XML bestanden

Een eXtensible Markup Language bestand, kortweg .XML⁴⁵, is een intelligent product van W3C⁴⁶. Zij hebben een bestandstype gecreëerd dat zowel door mensen als door machines begrepen kan worden en dient om gestructureerde data gemakkelijk te laten opzoeken door zijn gebruikers of in tekstbestanden op te slaan. Het zijn dus tekstbestanden, al zijn ze niet bedoeld om door mensen te worden gelezen. Iedere regel die beschreven wordt in een .XML bestand wordt vergezeld van een *tag* waarin de omschrijving van de regel wordt weergegeven in een metataal. Een metataal is een taal die iets zegt over de taal waarmee men inzicht krijgt in de inhoud, structuur en stijl van onze taalboodschap. Metatalen zijn dus (vrijwel) autologisch, heeft betrekking op zichzelf. Dat klinkt wellicht ingewikkeld, maar aan de hand van een voorbeeld is het snel duidelijk. Een metataal zoals .XML of HTML⁴⁷ is niet de boodschap, maar het zegt iets over de vorm van de boodschap. Metataal wordt in het Engels ook wel *markup* genoemd. Tot slot, .XML is afkomstig van de SGML⁴⁸, waarvan ook HTML is afgeleid.

.CSV bestanden

De Twitterdata waarmee het onderzoek in deze scriptie is gedaan, bestaat enkel uit .CSV bestanden. Kommagescheiden bestanden, kortweg de extensie .CSV wat afgekort is voor *Comma Separated Values*, zijn specificaties voor tabelbestanden. Het bestand wordt veelal gebruikt en is het oudste en eenvoudigste databaseformaat dat er bestaat. In .CSV bestanden staan geen moeilijke afkortingen of verwijzingen in andere kleuren, maar enkel tekstgegevens. De tekstgegevens worden gescheiden door een komma om zodoende een onderscheid te maken tussen categorieën. Door .CSV bestanden weer in te lezen in daarvoor geschikte software kan de informatie in het bestand weer vertaald worden naar klare, behapbare taal. Een bekend programma dat overweg kan met .CSV bestanden is het spreadsheet programma van Microsoft, Excel.

Normaliter krijg je bij het invoeren van een query in de Twitter API als output een .XML file. De data waar dit onderzoek gebruik van heeft gemaakt, zou via een *converter* (i.e.

⁴⁵ XML in 10 stappen <http://www.w3.org/XML/1999/XML-in-10-points.nl.html>.

⁴⁶ World Wide Web Consortium <http://www.w3.org/>.

⁴⁷ Hyper Text Markup Language.

⁴⁸ Standard Generalized Markup Language.

een applicatie op de computer die in dit geval .CSV bestanden om kan zetten naar .XML bestanden) omgezet kunnen worden, zodat van deze data ook weer .XML bestanden kunnen worden gemaakt. Echter, wanneer het bovenstaande niet noodzakelijk wordt bevonden dan kan er zonder problemen gebruik worden gemaakt van .CSV bestanden om analyses te maken.

.JSON bestanden

.JSON is een afkorting voor JavaScript Object Notation en is net als .XML een belangrijk bestandstype in de Twitteromgeving. Deze computertaal is voor mensen gemakkelijk om te lezen en te leren schrijven, terwijl machines het gemakkelijk kunnen verwerken vanwege de heldere structuur van het bestandstype. Vanwege deze feiten, en nog vele technische voordelen (het gebruikt bijvoorbeeld conventies die gelijk zijn aan populaire programmeringstalen C, C++, C#, Java, Javascript en Perl), is .JSON een ideale data uitwisselingstaal.

Bijlage IV – Scripts

Onderstaand script is gebruikt om het aantal tweets te tellen waarin een woord uit de woordenlijst voorkomt.

```
#!/bin/sh

function help()
{
cat <<stop_help
Usage: $0 [options]
  -h          : this help
  -w          : only complete words
  -c          : case sensitive
  -i          : show alleen hits
  -u          : add user
  -a          : show alleen hits en orgineel
  -f <file>   : file with words (default 'woordenlijst')

reads tweets with format

20120101 11 51.55436402 5.09562832  Snuffelmuis  Tijd voor nieuwjaarsflesje...

Counts the number of tweets containing a word from wordlist in
every file.
wordlist: file with one word or expression per line
output: a file with
yyyy mm dd hh ww #_tweets #tweets_with_words
if -i then tweets without a match are omitted from output
if -a then the tweet will be printed on a new line

Nota Bene: even when the -w option is not in effect, matching words should still be
preceded by a space

stop_help
  exit;
}

eval set -- `getopt -o hudaicwf: -- "$@"`
# set -- `getopt B:bmI:NMTLSs:p:hHd $*`

wl="woordenlijst"

for i
do
case "$i"
in

    -h) help; shift;;
    -w) hele_woorden=1; shift;;
    -c) case_sensitive=1; shift;;
    -u) user=1; shift;;

    -d) debug=1; shift;;
    -i) hits=1; shift;;
    -a) all=1; shift;;
    -f) wl=$2; shift; shift;;

    --) shift; break;;

esac
```

```

done
rm tempcol &> /dev/null

# rm file.?
x=$1

    awk 'BEGIN{
        s="Jan Feb Mar Apr May Jun Jul Aug Sep Oct Nov Dec"
        split(s,mm)
        for (x=1;x<=12;x++) maand[mm[x]]=x
        }

# -----

        function print_data()
        {
            if (term=="") term="_"
            printf("%d %02d %02d %02d %d %d %s", y,m,d,h,w,NR,belediging,term);
            if (user) printf(" %s",u)
            print ("\n");
        }

# -----

        {
            if (NR==1)
            {
                while (getline woordenlijst[++aantal] < naam);
            }
            ss=split($0,r,/[\ \.\,\/\!\;\:\/]);
            id=r[1]
            w=r[2]
#
            d=substr($1,7,2)
            m=substr($1,5,2)
            h=$2
            y=substr($1,1,4);
            u=$3
            while ("date -d "y"m"d" +%u"| getline w);
            close("date -d "y"m"d" +%u")

            term=""
            if (!case_sensitive) {s=tolower($0);ss=split(s,r,/[\ \.\,\/\!\;\:\/]);} else
s=$0
            belediging=0

            if (hele_woorden)
            {
                for (x=1;x<=aantal;x++)
                    for (y=1;y<=ss;y++) if (r[y]==woordenlijst[x])
{belediging++;term=term-"woordenlijst[x];}
                }
            else # ook substrings, maar wel met spatie voor term
                for (x=1;x<=aantal;x++)
                    if (index(s," "woordenlijst[x]))
                    {
                        belediging++;
                        term=term_"woordenlijst[x];
                    }

            if (!hits) print_data();
            if (hits)
            {
                if (term) print_data();
                if ((all) && (term)) print s
            }

        }' hele_woorden=$hele_woorden all=$all user=$user hits=$hits
case_sensitive=$case_sensitive naam=$w1 < "$x"

```


Hieronder staat het script dat gebruikt is om een overzicht te weergeven van het aantal tweets dat op een bepaalde dag, op een bepaalde tijdstip is verstuurd. Ook de twitteraar wordt hierin aangeduid.

```
awk '{
DAY[1] = "Sunday"; DAY[2] = "Monday"
DAY[3] = "Tuesday"; DAY[4] = "Wednesday"
DAY[5] = "Thursday"; DAY[6] = "Friday"; DAY[7] = "Saturday"

while ("cal "m" "y| getline)
{
if (NF==7) Sun = $1
}

dow = (Sun + 14 - d)%7
if (dow == 0) dow = 7
dow = 8 - dow
while ("date -d "y""m""d" +%u| getline a)
print "bla "a
printf ("%d/%d/%d was a %s\n", m, d, y, DAY[dow])

} ' y=$1 m=$2 d=$3 -
```

Om vervolgens een overzicht te krijgen van onbeleefde tweets is het volgende script gebruikt. De output was een overzichtelijke (en gemakkelijk te importeren) matrix.

```
# input format: 2011 05 20 09 00 2 0 kut jantje
# ofwel yyyy mm dd hh wk NR aant_scheldw scheldw user

awk '{
w=$5+0
h=$4+0
aant_bel=$7+0
matrix[w,h]++
if (aant_bel) matrix_grofheid[w,h]++
}
END{
weken=file"_weken"
weken_grof=file"_weken_grof"
for (x=0;x<=23;x++)
{
for (y=1;y<=7;y++)
{
if (!matrix[y,x]) matrix[y,x]=0
som+=matrix[y,x]
if (!matrix_grofheid[y,x]) matrix_grofheid[y,x]=0

printf ("%s ",matrix[y,x]) > weken
printf ("%s ",matrix_grofheid[y,x]) > weken_grof
}

printf("\n")> weken;
printf("\n")> weken_grof;
}
}
END{
print "totaal: "som
}' file=$1 < $1
```

Doordat er zo nu en dan bestanden tussen zitten die niet voldoen aan de vereiste layout, mogelijk doordat er gewerkt wordt met verschillende bestandstypen (zie bijlage III), is het middels dit script mogelijk corrupte tweets te repareren en in de juiste layout te zetten.

```
#!/bin/sh

function help()
{
cat <<stop_help
Usage: $0 input output

repairs stupid tweets that are cut somewhere in the middle

stop_help
    exit;
}

eval set -- `getopt -o hdcwf: -- "$@"`
# set -- `getopt hHd $*`

wl="woordenlijst"

for i
do
case "$i"
in

        -h) help; shift;;

        --) shift; break;;

        esac
done

awk '{
if (substr($0,1,1)~/[0-9]/) {print regel;regel=r=$0}
else regel=regel" "r
}' $1 > `basename $1 .csv`.repaired
```

Om snel een overzicht te krijgen van de onbeleefde uitingen in tweets wordt onderstaand script gebruikt om de woorden uit de woordenlijst te tellen. Dit is gedaan zodat er niet alleen wordt gekeken naar hoeveel tweets er een onbeleefde uiting bevatte, maar om ook de hoeveelheid onbeleefde uitingen te kunnen waarnemen.

```
awk '
{
if ($8!="_")
{
x=split($8,r,"_");
for (y=1;y<=x;y++) woord[r[y]]++
}
}
END{
for (n in woord)  print n,woord[n]
}
' < $1
```

Bijlage V – Scheldwoordenlijst

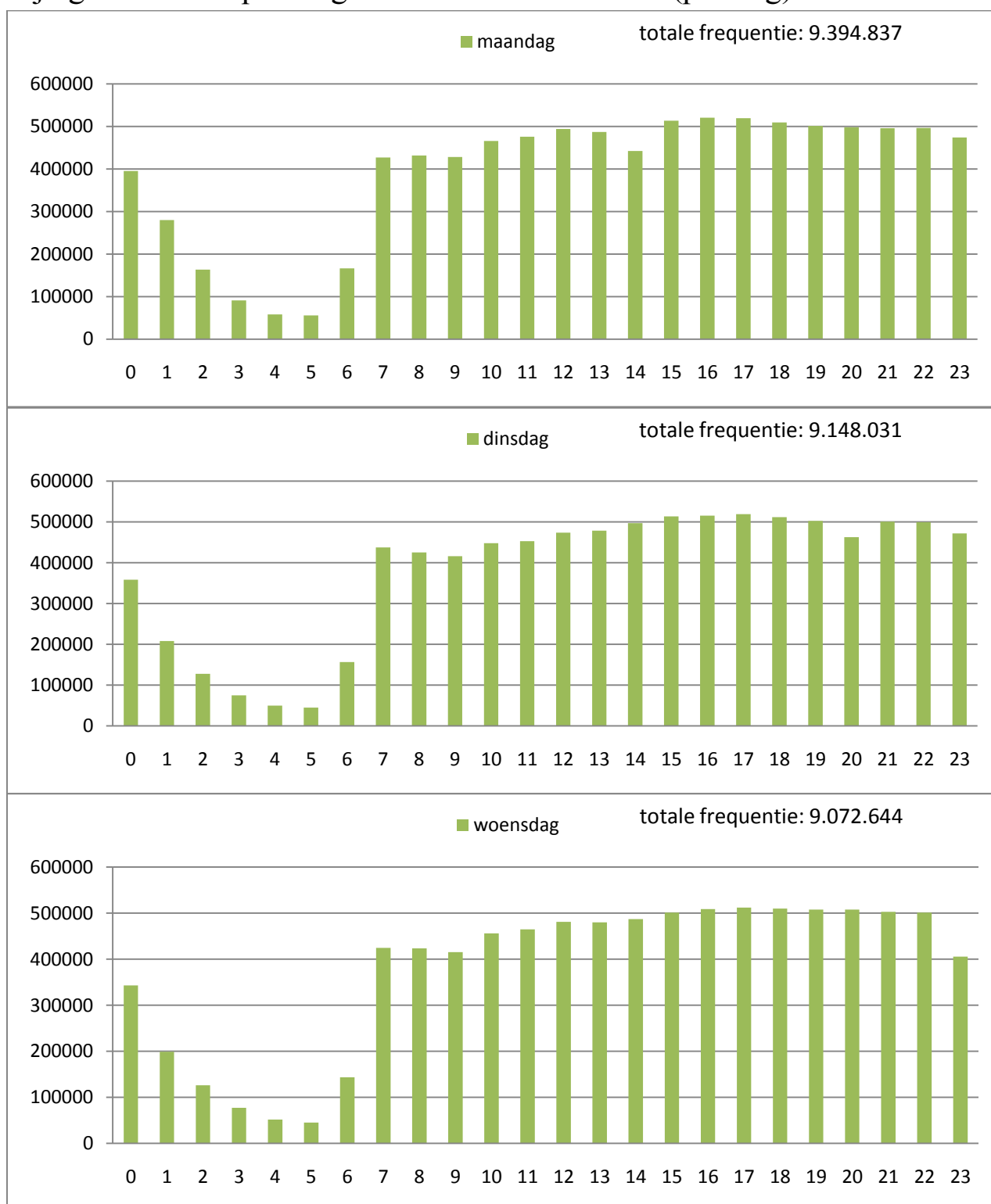
b!tch	greppelheks	rimpelzwijn
befgaaius	greppelslet	RTFM
befkoning	greppelhoer	rugridder
biaatch	gvd	scheisse
biatch	halvegare	scheiße
bitch	hoer	sh!t
bitchass	hoeren	sh*t
boeler	hoerenjong	shit
boerenlul	hoeruh	slet
boerenpummel	homofiel	sletje
boerentrien	homotje	sletjeeh
bokkelul	hondenlul	sletjeh
bosaap	k ut	sletjuh
breezerslet	k*t	sloeri
breezersletje	k_t	SMFH
broodmongool	kaaanker kaanker	snol
capsoneslijer	kankerhoer	sukkel
chocoladeprins	kankerjong	sukkeltje
chocoprins	kankerlijer	sukkol
cyberhoer	kankerteef	sukkoltje
cyberslet	kankerwif	takkewif
darmtoerist	kenker	teering
droplul	kenkerhoer	teef
eikel	kenkerrr	teering
eikels	kk	tering
eikol	kkhoer	tyfus
eikols	klerelijer	verdomme
f u c k	kloie	wrattenkop
f%ck	klooie	WTF
f*ck	klootzak	
f*cking	klote	
f_ck	klotuh	
faggot	kut	
flikker	kutlul	
focked op	kutwif	
focked up	lelijkerd	
fokked op	lesbo	
fokked up	LMFAO	
fokking	mafetel	
fuck	mongool	
fucking	neukfout	
gaytje	nikker	
geitenneuker	paardenpijper	
godver	pleuris	
godverdomme	pleurus	
godvur	ragkut	

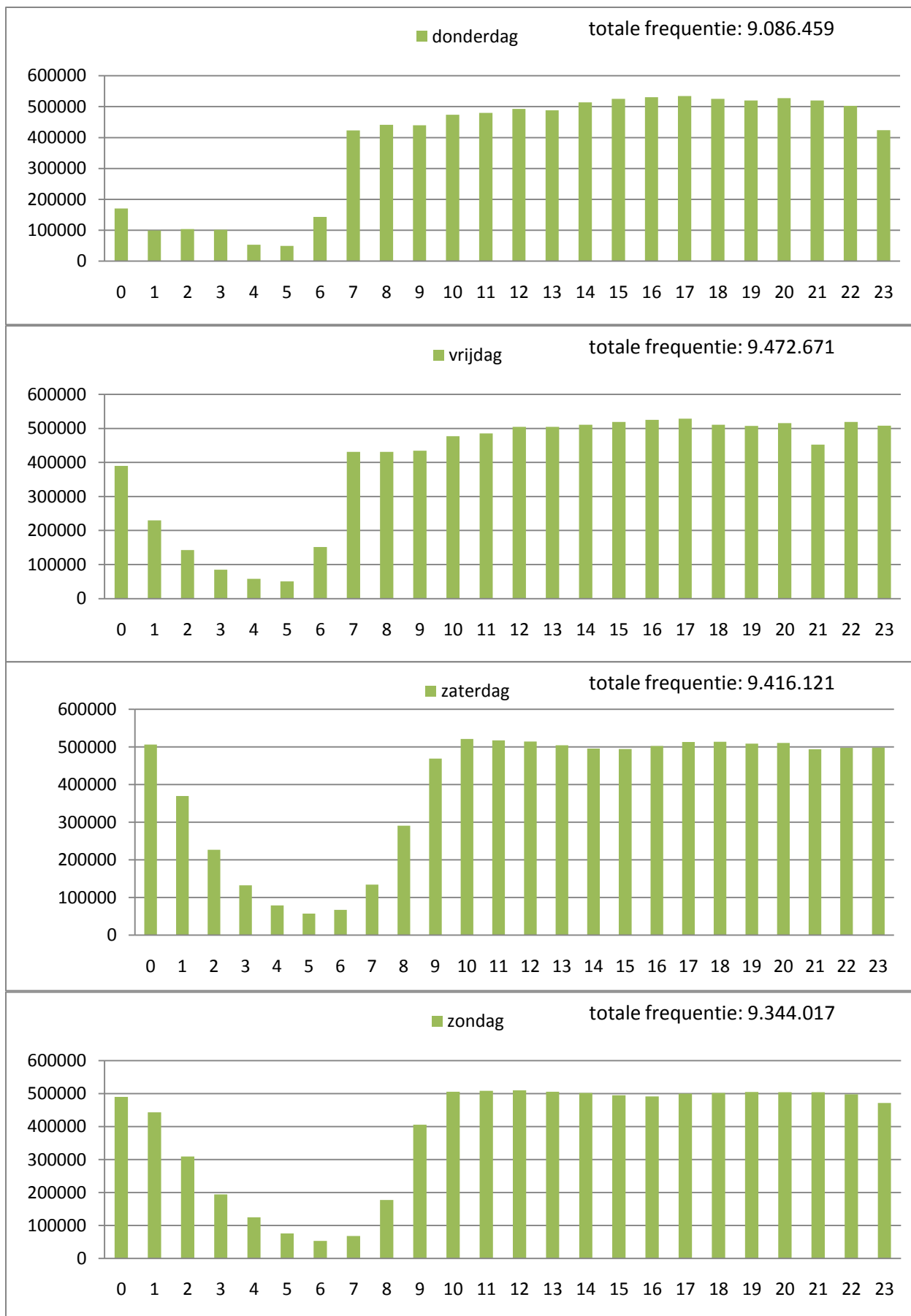
Bijlage VI – Meest frequent gebruikte scheldwoorden (absolute aantallen)

Onderstaande tabel geeft een overzicht van de vijftien meest voorkomende scheldwoorden in alle vier de groepen. Onderaan de tabel staat een percentage dat aangeeft hoeveel procent van het totaal aantal scheldwoorden in deze top 15 zit.

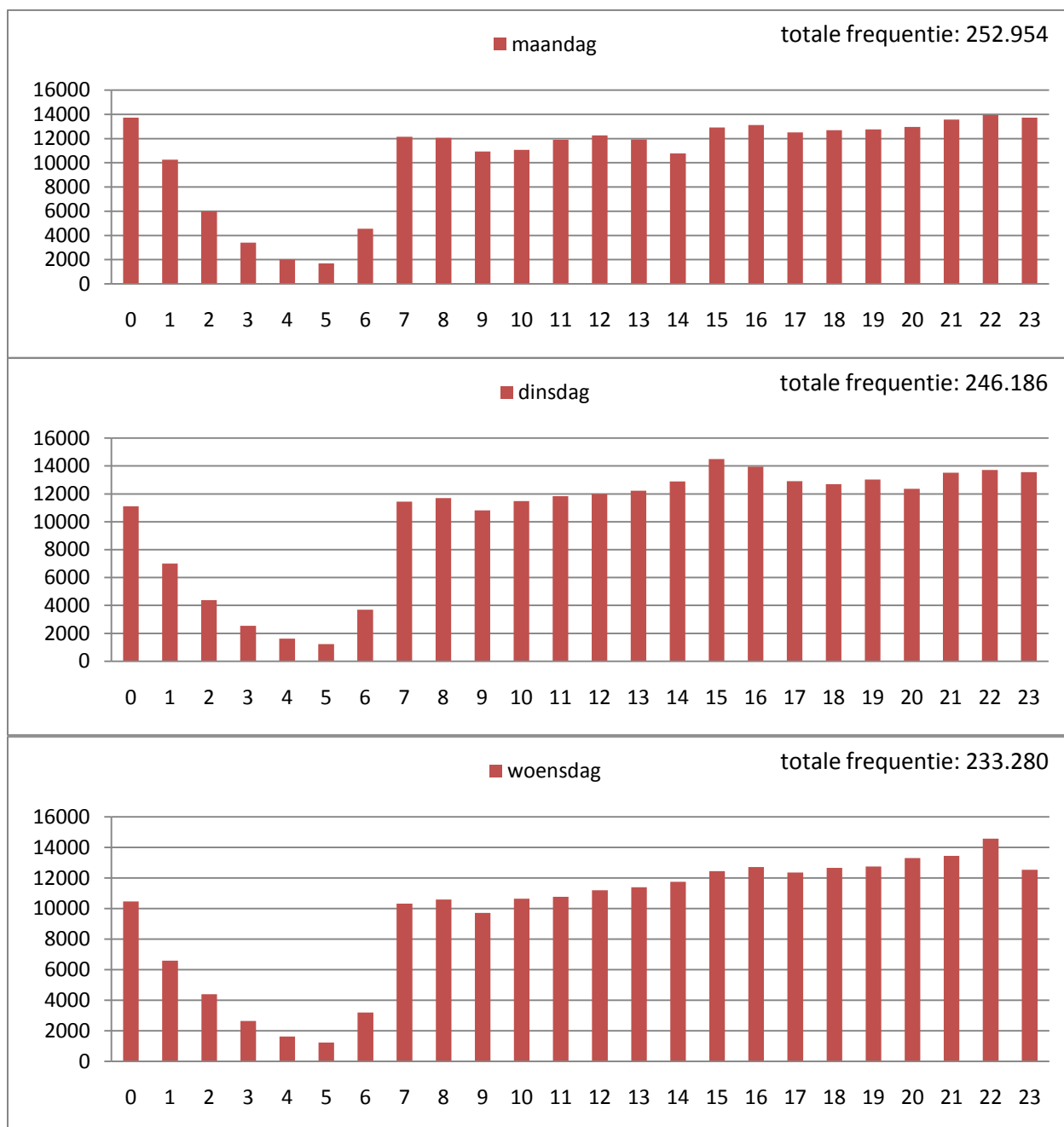
#	Scheldwoord	Nederland	Tilburg	Geen toestemming	Toestemming	Totaal
1	Kut	570.081	77	103	1.273	571.534
2	Kk	265.614	14	7	182	265.817
3	Fuck	215.463	24	77	1.076	216.640
4	Shit	125.654	8	105	943	126.710
5	Fucking	101.098	15	15	384	101.512
6	Bitch	83.165	6	22	300	83.493
7	Klote	69.145	12	30	346	69.533
8	Hoer	62.906	7	98	438	63.449
0	Tering	58.086	7	14	202	58.309
10	Gvd	47.717	13	14	95	47.839
11	Sukkel	41.001	3	15	139	41.158
12	Flikker	30.887	2	9	86	30.984
13	Verdomme	10.040	1	11	67	10.119
14	Klootzak	9.942	3	10	49	10.004
15	Eikel	5.482	0	9	84	5.575
	Totaal	1.696.281	192	539	5.664	1.702.676
	Totaal geheel	1.890.567	218	605	6.578	1.897.968
	Percentage	89,72%	88,07%	89,09%	86,11%	88,25 %

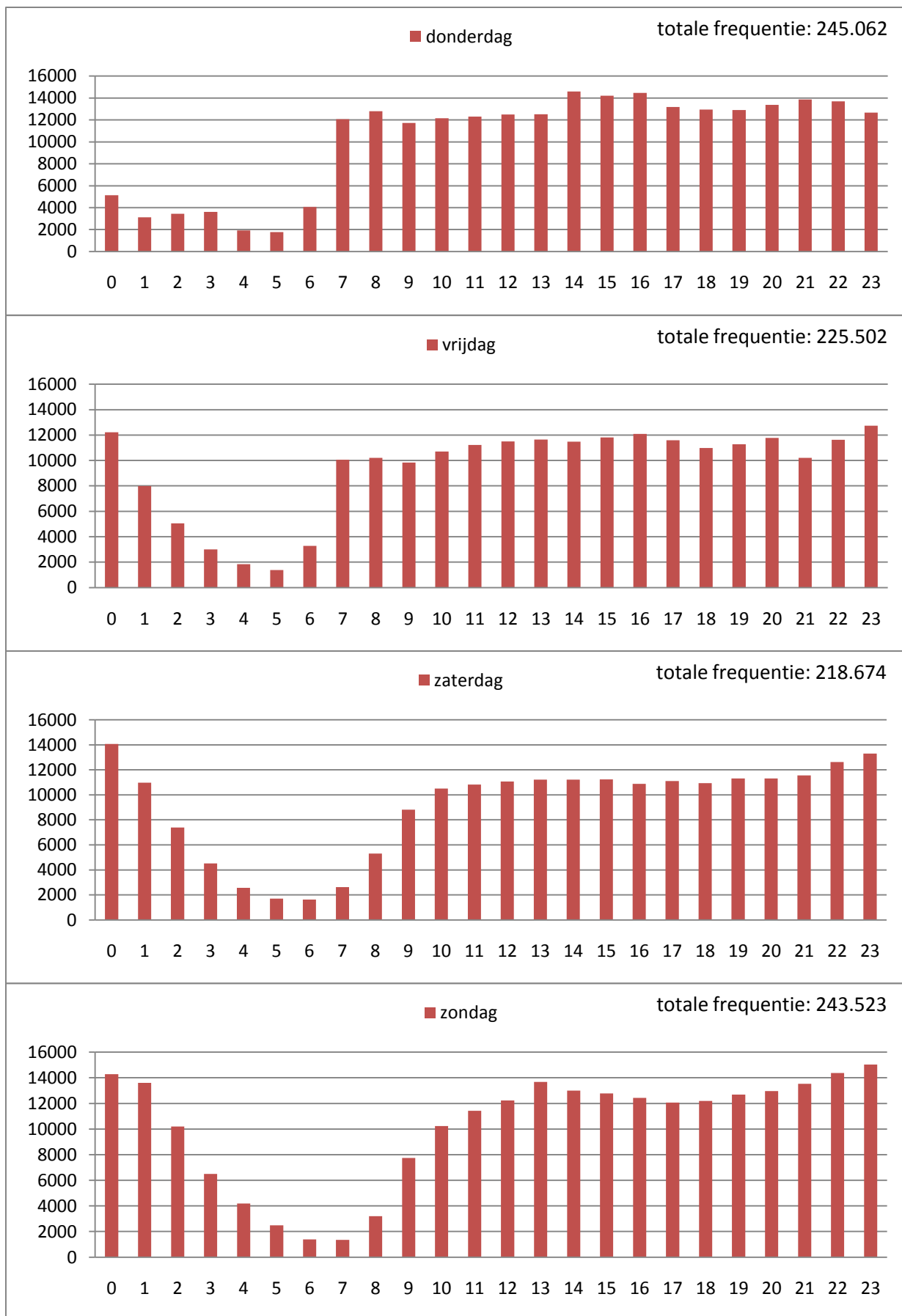
Bijlage VII – Frequentiegrafieken van alle tweets (per dag)





Bijlage VIII – Frequentiegrafieken van alle onbeleefde tweets (per dag)





Bijlage IX - Matrixen van alle twitterdata

Totale twitterdata van heel Nederland behorend tot de resultaten van hoofdstuk 5.1.1.

	maandag	dinsdag	woensdag	donderdag	vrijdag	zaterdag	zondag	totaal
0	395.297	358.126	343.078	170.727	390.190	506.050	490.252	2.653.720
1	280.100	207.914	198.700	99.713	229.768	369.521	443.435	1.829.151
2	163.608	127.756	126.041	104.264	142.260	226.855	309.082	1.199.866
3	91.251	75.001	77.035	101.844	84.914	132.110	193.930	756.085
4	58.194	50.100	51.387	53.407	57.772	79.102	124.530	474.492
5	55.996	45.022	44.931	49.712	50.249	56.978	75.676	378.564
6	166.743	156.523	143.164	143.009	151.189	66.849	52.853	880.330
7	427.370	437.556	424.624	423.317	431.368	134.120	67.598	2.345.953
8	432.132	425.545	423.564	441.182	431.203	290.412	177.159	2.621.197
9	428.378	416.328	415.477	440.167	434.645	468.821	405.693	3.009.509
10	465.983	447.777	455.889	473.855	476.861	521.336	505.650	3.347.351
11	476.027	452.732	464.956	479.956	485.313	517.232	508.573	3.384.789
12	494.468	473.993	480.919	493.035	504.616	514.417	510.087	3.471.535
13	487.338	478.924	480.102	488.557	504.336	504.589	505.513	3.449.359
14	442.461	497.217	486.956	513.974	511.147	495.745	502.272	3.449.772
15	513.520	513.482	501.034	525.159	518.983	494.709	494.826	3.561.713
16	520.952	515.358	508.799	530.547	525.447	502.252	491.688	3.595.043
17	519.519	519.287	511.843	534.263	528.784	512.880	499.998	3.626.574
18	509.298	511.924	509.925	525.342	510.554	513.467	502.107	3.582.617
19	501.394	502.747	507.605	520.357	507.212	508.416	504.912	3.552.643
20	498.183	462.634	507.491	527.354	515.980	510.485	504.449	3.526.576
21	495.838	500.540	502.693	520.245	452.558	493.974	504.029	3.469.877
22	496.500	499.450	501.215	502.507	519.095	497.788	498.302	3.514.857
23	474.287	472.095	405.216	423.966	508.227	498.013	471.403	3.253.207
	9.394.837	9.148.031	9.072.644	9.086.459	9.472.671	9.416.121	9.344.017	64.934.780

Onbeleeftde twitterdata van heel Nederland behorend tot de resultaten van hoofdstuk 5.1.1.

	maandag	dinsdag	woensdag	donderdag	vrijdag	zaterdag	zondag	totaal
0	13.722	11.109	10.459	5.145	12.227	14.062	14.274	80.998
1	10.268	6.997	6.581	3.119	7.982	10.966	13.612	59.525
2	6.005	4.371	4.394	3.440	5.047	7.394	10.194	40.845
3	3.413	2.541	2.632	3.614	3.004	4.516	6.502	26.222
4	2.014	1.610	1.628	1.915	1.847	2.558	4.195	15.767
5	1.694	1.224	1.234	1.783	1.379	1.706	2.480	11.500
6	4.565	3.703	3.188	4.054	3.281	1.633	1.377	21.801
7	12.145	11.435	10.313	12.078	10.058	2.620	1.342	59.991
8	12.068	11.703	10.589	12.804	10.207	5.304	3.200	65.875
9	10.925	10.812	9.714	11.727	9.843	8.814	7.733	69.568
10	11.078	11.492	10.650	12.152	10.706	10.502	10.232	76.812
11	11.896	11.829	10.764	12.296	11.215	10.827	11.426	80.253
12	12.269	12.000	11.189	12.506	11.497	11.059	12.233	82.753
13	11.919	12.229	11.399	12.525	11.638	11.215	13.683	84.608
14	10.770	12.887	11.744	14.595	11.480	11.218	12.995	85.689
15	12.925	14.497	12.453	14.215	11.803	11.243	12.775	89.911
16	13.110	13.955	12.720	14.464	12.095	10.883	12.437	89.664
17	12.502	12.905	12.353	13.187	11.594	11.099	12.072	85.712
18	12.691	12.693	12.668	12.953	10.980	10.945	12.193	85.123
19	12.755	13.025	12.744	12.894	11.275	11.305	12.691	86.689
20	12.963	12.363	13.309	13.380	11.781	11.308	12.962	88.066
21	13.581	13.522	13.453	13.873	10.205	11.558	13.522	89.714
22	13.957	13.719	14.567	13.687	11.621	12.631	14.369	94.551
23	13.719	13.565	12.535	12.656	12.737	13.308	15.024	93.544
	252.954	246.186	233.280	245.062	225.502	218.674	243.523	1.665.181

Totale twitterdata van Tilburg behorend tot de resultaten van hoofdstuk 5.1.2.

	maandag	dinsdag	woensdag	donderdag	vrijdag	zaterdag	zondag	totaal
0	48	33	54	27	114	121	122	519
1	34	33	17	12	18	82	158	354
2	10	14	11	12	15	49	58	169
3	3	8	10	5	14	45	46	131
4	2	14	2	6	13	13	38	88
5	6	7	1	8	3	4	17	46
6	26	24	14	16	16	5	16	117
7	86	66	77	88	94	11	17	439
8	100	102	106	79	61	38	29	515
9	75	90	101	75	95	76	73	585
10	73	97	67	92	87	105	72	593
11	69	82	65	78	62	115	68	539
12	83	76	65	85	73	89	71	542
13	80	67	72	96	103	80	102	600
14	64	80	72	72	68	75	96	527
15	91	80	75	51	71	87	98	553
16	68	63	63	54	87	80	92	507
17	67	57	74	73	68	81	104	524
18	81	96	58	58	90	106	64	553
19	80	54	73	47	76	82	48	460
20	73	53	55	59	72	74	57	443
21	59	71	69	58	102	93	70	522
22	64	77	65	58	96	91	108	559
23	57	48	52	86	84	72	111	510
	1.399	1.392	1.318	1.295	1.582	1.674	1.735	10.395

Onbeleeftde twitterdata van Tilburg behorend tot de resultaten van hoofdstuk 5.1.2.

	maandag	dinsdag	woensdag	donderdag	vrijdag	zaterdag	zondag	totaal
0	3	1	1	0	2	6	4	17
1	0	0	0	1	1	2	4	8
2	0	0	0	0	1	2	4	7
3	0	0	0	0	0	0	0	0
4	0	0	0	0	0	0	1	1
5	1	1	1	0	0	0	0	3
6	2	0	0	1	0	0	1	4
7	2	2	1	1	1	2	1	10
8	2	1	1	1	3	1	0	9
9	3	1	2	3	2	0	0	11
10	1	1	2	1	1	4	0	10
11	0	1	1	3	0	1	2	8
12	0	1	0	3	1	0	0	5
13	0	1	1	2	0	1	2	7
14	0	2	0	1	2	0	2	7
15	1	1	1	0	0	4	1	8
16	0	1	1	2	1	0	2	7
17	0	2	2	2	1	1	1	9
18	2	0	2	5	2	1	3	15
19	1	2	1	1	0	2	0	7
20	0	1	1	0	1	2	0	5
21	2	1	0	2	2	1	1	9
22	1	1	2	3	1	0	1	9
23	1	0	1	3	2	3	0	10
	22	21	21	35	24	33	30	186

Totale en onbeleefde twitterdata van de zondagen in de grote dataset behorend tot de resultaten van hoofdstuk 5.1.3. 1 januari 2012 is niet opgenomen in het onderzoek, terwijl 8, 15 en 22 januari 2012 wel in de onderzoeksdata zitten.

	onbeleefd		totaal	
	1-jan	8 15 22 jan	1-jan	8 15 22 jan
0	2.982	14.275	122.552	490.252
1	3.667	13.612	130.388	443.435
2	3.921	10.194	130.100	309.082
3	3.397	6.502	114.694	193.930
4	2.272	4.195	68.283	124.530
5	1.422	2.480	42.054	75.676
6	906	1.377	28.505	52.853
7	497	1.342	16.716	67.598
8	344	3.200	13.809	177.159
9	487	7.733	25.118	405.693
10	1.223	10.232	58.374	505.650
11	2.262	11.426	109.543	508.573
12	2.840	12.233	129.255	510.087
13	2.885	13.683	127.256	505.513
14	2.800	12.995	124.554	502.272
15	2.988	12.775	125.398	494.826
16	3.111	12.437	123.548	491.688
17	3.145	12.072	125.517	499.998
18	3.053	12.193	124.331	502.107
19	3.036	12.691	122.493	504.912
20	3.221	12.962	121.686	504.449
21	3.175	13.522	119.314	504.029
22	3.324	14.369	121.616	498.302
23	3.364	25.024	120.782	471.403
	60.322	253.524	2.345.886	9.344.017

Totale twitterdata van de „geen toestemming’ groep behorend tot de resultaten van hoofdstuk 5.2.1.

	maandag	dinsdag	woensdag	donderdag	vrijdag	zaterdag	zondag	totaal
0	128	276	423	507	612	492	386	2.824
1	56	45	97	217	112	216	105	848
2	54	64	53	78	51	129	68	497
3	54	59	62	58	64	59	31	387
4	126	81	126	108	66	92	41	640
5	351	354	421	400	311	145	83	2.065
6	958	1.084	1.138	1.046	889	476	193	5.784
7	1.770	2.062	2.018	1.888	1.645	887	622	10.892
8	2.063	2.313	2.568	2.305	2.163	1.638	1.034	14.084
9	2.090	2.218	2.436	2.497	2.424	1.882	1.483	15.030
10	1.946	2.170	2.445	2.597	2.474	1.911	1.765	15.308
11	1.744	1.892	2.381	2.453	2.359	2.170	2.042	15.041
12	1.957	2.079	2.533	2.489	2.430	2.097	2.072	15.657
13	2.075	2.211	2.544	2.346	2.284	1.922	1.637	15.019
14	2.215	2.429	2.259	2.422	2.325	1.941	1.582	15.173
15	2.291	2.397	2.426	2.462	2.564	2.171	1.530	15.841
16	2.250	2.351	2.285	2.216	2.432	1.709	1.337	14.580
17	2.337	2.235	2.382	2.141	1.872	1.545	1.471	13.983
18	2.249	2.582	2.520	2.412	1.973	1.736	1.512	14.984
19	2.280	3.010	3.065	2.650	1.967	1.788	1.807	16.567
20	2.588	3.273	3.166	2.759	1.976	1.725	1.991	17.478
21	2.908	3.502	3.070	3.039	2.362	1.585	1.852	18.318
22	2.249	2.615	2.739	2.436	1.912	1.292	1.254	14.497
23	943	1.255	1.633	1.314	920	815	590	7.470
	37.682	42.557	44.790	42.840	38.187	30.423	26.488	262.967

Onbeleefde twitterdata van de „geen toestemming’ groep behorend tot de resultaten van hoofdstuk 5.2.1.

	maandag	dinsdag	woensdag	donderdag	vrijdag	zaterdag	zondag	totaal
0	0	1	0	2	0	1	1	5
1	0	0	0	0	0	0	0	0
2	0	0	0	0	1	0	0	1
3	0	0	0	0	0	0	0	0
4	1	0	1	0	0	0	0	2
5	1	2	0	1	0	0	0	4
6	1	0	2	2	1	0	1	7
7	3	6	4	5	3	1	3	25
8	4	5	5	8	2	5	1	30
9	4	8	3	2	4	1	3	25
10	7	6	6	2	4	2	4	31
11	2	2	4	4	10	2	0	24
12	6	4	6	2	7	1	6	32
13	6	5	1	7	6	4	8	37
14	4	3	3	3	6	2	3	24
15	5	4	6	6	7	4	5	37
16	8	3	2	3	6	6	5	33
17	6	5	4	3	4	2	3	27
18	0	5	5	7	6	5	4	32
19	5	3	5	7	9	3	4	36
20	7	8	6	10	4	3	9	47
21	5	7	8	6	9	4	7	46
22	5	10	8	2	5	4	2	36
23	4	3	8	4	1	7	6	33
	84	90	87	86	95	57	75	574

Totale twitterdata van de „met toestemming’ groep behorend tot de resultaten van hoofdstuk 5.2.2.

	maandag	dinsdag	woensdag	donderdag	vrijdag	zaterdag	zondag	totaal
0	651	677	597	837	737	1.199	1.409	6.107
1	283	269	257	297	347	634	772	2.859
2	202	120	121	170	187	325	414	1.539
3	179	137	162	163	154	208	303	1.306
4	497	497	438	511	462	319	202	2.926
5	1.721	1.789	1.529	1.791	1.801	745	492	9.868
6	3.317	3.669	3.520	3.476	3.700	1.678	1.197	20.557
7	4.776	5.426	5.398	5.265	5.524	2.772	2.198	31.359
8	5.280	5.693	6.278	6.111	6.744	3.920	3.238	37.264
9	5.667	5.781	6.122	6.115	6.794	4.483	3.994	38.956
10	5.556	5.587	5.908	6.104	6.676	4.761	4.694	39.286
11	5.470	5.535	5.947	5.889	6.098	4.631	4.779	38.349
12	5.438	5.566	6.118	6.135	6.487	4.787	4.965	39.496
13	5.640	5.530	6.161	5.860	6.562	4.325	4.904	38.982
14	5.601	5.893	6.270	6.233	6.534	4.272	5.071	39.874
15	5.702	5.810	6.297	6.247	6.403	4.539	4.635	39.633
16	5.525	5.645	5.765	5.942	5.845	4.554	4.300	37.576
17	5.574	5.641	5.768	5.658	5.163	4.480	4.347	36.631
18	6.130	6.841	6.008	6.018	5.290	4.468	5.034	39.789
19	6.890	7.505	6.923	7.364	5.639	5.011	6.057	45.389
20	7.385	8.000	7.723	7.957	5.822	5.368	6.331	48.586
21	6.552	7.260	6.747	6.741	5.671	4.520	5.498	42.989
22	3.939	4.516	4.183	4.376	4.071	3.721	3.528	28.334
23	1.677	1.707	1.890	1.910	2.321	2.478	1.627	13.610
	99.652	105.094	106.130	107.170	105.032	78.198	79.989	681.265

Onbeleefde twitterdata van de „met toestemming’ groep behorend tot de resultaten van hoofdstuk 5.2.2.

	maandag	dinsdag	woensdag	donderdag	vrijdag	zaterdag	zondag	totaal
0	9	4	10	10	8	20	23	84
1	11	3	5	2	1	12	10	44
2	7	2	0	4	4	6	8	31
3	0	2	0	1	1	3	3	10
4	1	8	2	4	2	4	6	27
5	12	12	12	14	12	5	5	72
6	27	30	26	30	28	11	8	160
7	41	37	32	22	37	17	10	196
8	43	37	41	36	42	20	18	237
9	35	31	51	45	52	33	23	270
10	55	44	39	40	43	38	40	299
11	38	50	41	51	41	37	37	295
12	40	37	49	41	38	35	42	282
13	44	53	36	57	61	23	43	317
14	48	39	32	39	49	29	45	281
15	50	53	52	43	57	39	36	330
16	59	55	37	50	57	42	41	341
17	50	51	66	44	39	36	38	324
18	67	66	57	47	51	31	52	371
19	57	72	44	80	49	63	45	410
20	67	76	65	87	68	56	60	479
21	69	76	79	66	50	48	76	464
22	52	43	43	44	52	40	42	316
23	20	21	15	23	35	40	25	179
	902	902	834	880	877	688	736	5.819

Bijlage X – Zoekopdracht E.F. Tjong Kim Sang

Het gebruikte Linuxcommando voor het verzamelen van berichten:

```
curl -d '@sleutelwoorden.txt'  
http://stream.twitter.com/1/statuses/filter.json -u  
"ACCOUNTNAAM:WACHTWOORD"
```

Inhoud van het bestand sleutelwoorden.txt:

Track = een,het,ik,niet,maar,voor,ook,als,heb,naar,nog,echt,moet,weer,mijn,zijn,bij,jij,toch,
lekker,geen,gewoon,gaat,meer,slapen,weet,mensen,alleen,kijken,leren,heeft,vandaag,eens,
hoor,uur,jou,veel,denk,maken,leuk,heel,zou,daar,komt,eten,iets,vind,hebben,altijd,vanavond,
jullie,thuis,iemand,helemaal,waar,waarom,wakker,komen,beetje,nieuwe,worden,steeds,
gezellig,straks,kunnen,zeggen,iedereen,ofzo,omdat,werken,allemaal,moeten,andere,jaar,terug,
staat,kut,ging,erg,zien,vroeg,bijna,zelf,zegt,vrij,zeker,werk,#gtst,#tienerthings,#bzv,
#wiedoethet,#durftevragen,#nieuws,#tienerfeiten,#dutchteenagers,#dwdd,#penw,#widm,
#slajezelf,#voetbalfans,#pownews,#geenzin,#slapen,#lekker,#rtl7

Bijlage XI - SPSS Output

Pagina 35

ANOVA

		Sum of Squares	df	Mean Square	F	Sig.
Hoeveel onbeleefde tweets	Between Groups	2282152441,423	23	99224019,192	29,511	,000
	Within Groups	484175112,286	144	3362327,169		
	Total	2766327553,708	167			
Totaal aantal tweets	Between Groups	4204372820910,190	23	182798818300,443	56,751	,000
	Within Groups	463836936219,429	144	3221089834,857		
	Total	4668209757129,619	167			

Pagina 40

Paired Samples Statistics

	Mean	N	Std. Deviation	Std. Error Mean
Pair 1 weekend of doordeweek	,50	2	,707	,500
gem per uur 2	9812,5500	2	491,22708	347,35000

Paired Samples Correlations

	N	Correlation	Sig.
Pair 1 weekend of doordeweek & gem per uur 2	2	1,000	,000

Paired Samples Test

		Paired Differences					t	df	Sig. (2-tailed)
		Mean	Std. Deviation	Std. Error Mean	95% Confidence Interval of the Difference				
					Lower	Upper			
Pair 1	weekend of doordeweek - gem per uur 2	-9812,0500	490,51997	346,85000	-14219,19711	-5404,90289	-28,289	1	,022

Paired Samples Statistics

		Mean	N	Std. Deviation	Std. Error Mean
Pair 1	weekend of doordeweek	,50	2	,707	,500
	Gemiddelde per uur	1,1400	2	,15556	,11000

Paired Samples Correlations

		N	Correlation	Sig.
Pair 1	weekend of doordeweek & Gemiddelde per uur	2	-1,000	,000

Paired Samples Test

		Paired Differences				t	df	Sig. (2-tailed)
		Mean	Std. Deviation	Std. Error Mean	95% Confidence Interval of the Difference			
					Lower Upper			
Pair 1	weekend of doordeweek - Gemiddelde per uur	-,64000	,86267	,61000	-8,39078 7,11078	-1,049	1	,485

Paired Samples Test

		Paired Differences				t	df	Sig. (2-tailed)
		Mean	Std. Deviation	Std. Error Mean	95% Confidence Interval of the Difference			
					Lower Upper			
Pair 1	Onbeleefd_gem_res t - Onbeleefd_1jan	1007,708	1272,111	259,669	470,543 1544,874	3,881	23	,001

Dankwoorden

Graag sta ik nog even stil bij de totstandkoming van het eindproduct en het doorlopen van mijn drie en een half jaar op Tilburg University. Deze was nooit in deze hoedanigheid verschenen zonder de hulp, toewijding, ondersteuning, motivatie en frisse kijk op zaken van mijn vriendin Lindsay, mijn ouders Peter en Joke, mijn broer Dave die zijn academische knobbel kon botvieren op mijn scriptie, mijn zussen Barbara en Pascalle, mijn uitzonderlijk doortastende, goede en behulpzame vriend Björn, een motiverende en buitengewone scriptiebegeleider Hans Paijmans, tweede lezer Menno van Zaanen, Martin Reynaert, Anne Vermeer, Mehdi Aminian, mijn vrienden van SHOT heren 1, mijn medebestuurders uit het 26^e SHOT bestuur, „dispuut’ Groen 24 en in het speciaal Justin van Heesch en Michiel Jung, medestudenten Wesley, Yoeri en Michelle, mijn vrienden van Club Conjo; Bas, Jasper, Chris, Rudy en Boy, en tot slot dr. ir. E.F. Tjong Kim Sang van de Universiteit Groningen die mij aan een grote hoeveelheid Twitter data heeft geholpen.

BEDANKT!