

Een onderzoek naar de rol van sociale identiteit (in-group versus out-group) wat betreft het gebruik van haat zaaiende teksten op Twitter

Research into the role of 'social identity' (in-group versus out-group) in terms of the use of hate spreading texts on Twitter



Radboud Universiteit Nijmegen

Meliha Sander

1011110

Bachelor scriptie

Communicatie- en informatiewetenschappen

Faculteit Letteren

Radboud Universiteit, Nijmegen

Scriptiebegeleider: Lidwien van de Wijngaert

Tweede beoordelaar: Enny Das

8 juni 2020

Abstract

Dit onderzoek heeft betrekking op de rol die de sociale identiteit van een individu speelt in het gebruiken van online *hate speech*. Hate speech wordt in dit onderzoek gedefinieerd als “een vijandige, kwaadsprekende, offensieve, vooringenomen gemotiveerde, discriminerende uiting die beroep doet of gebruik maakt van stereotyperende kenmerken die de *out-group* in diskrediet brengt”. Door in kaart te brengen hoe gebruikers van verschillende sociale groepen hate speech op sociale media gebruiken wordt er inzicht verkregen in de mechanismen rondom hate speech en hoe mensen van de *in-group* zich uiten ten opzichte van de eigen groep en ten opzichte van de *out-group*. Door middel van een corpusanalyse is het gebruik van hate speech op sociale media (Twitter) onderzocht van de *in-group* versus de *out-group*. De vraag die centraal staat in dit onderzoek is ‘In hoeverre wordt er online hate speech ten opzichte van de andere groep gebruikt bij de *in-group* versus de *out-group*?’. Dit is onderzocht aan de hand van een specifieke casus. In het geval van deze casus staan twee groepen tegenover elkaar: individuen die werden ingedeeld bij de groep Baudetaanhangers en ‘Anderen’. De twitterprofielen zijn gecodeerd aan de hand van profielkenmerken en profielinformatie. De tweets zijn gecodeerd aan de hand van relevantie, onderwerp, toon, hate speech, het gebruik van scheldwoorden en ziektes, en abstract taalgebruik (*linguistic intergroup bias*). Uit de resultaten is gebleken dat er niet vaker sprake was van hate speech als de tweet betrekking had op de *in-group* dan op de *out-group*. Uit de resultaten is echter wel gebleken dat auteurs met een duidelijk profiel vaker hate speech gebruikten dan verwacht. Een verklaring hiervoor is dat zij vanwege een gedeelde gedachtegoed de mening van Baudet willen verdedigen en daarom de *out-group* in diskrediet brengen. Daarnaast is uit de resultaten gebleken dat tweets die betrekking hadden op de *out-group* van Baudetaanhangers (de Anderen) vaker abstracte taal bevatten dan tweets die betrekking hadden op de eigen *in-group* van Baudetaanhangers. Tot slot is gebleken dat er geen verband was tussen de *in-* en *out-*group van Anderen en abstract taalgebruik. Deze resultaten kunnen gebruikt worden om inzicht te krijgen in online hate speech en welke invloed sociale identiteit heeft op het gebruik van hate speech. Deze inzichten kunnen gebruikt worden om online hate speech, voor zover mogelijk, tegen te gaan.

Inleiding

In Nederland is vrijheid van meningsuiting vastgelegd in artikel 7 van de Grondwet.

“Niemand heeft voorafgaand verlof nodig om door de drukpers gedachten of gevoelens te openbaren, behoudens ieders verantwoordelijkheid volgens de wet. Er is geen voorafgaand toezicht op de inhoud van een radio- of televisie-uitzending. Voor het openbaren van gedachten of gevoelens door andere dan in de voorgaande leden genoemde middelen heeft niemand voorafgaand verlof nodig wegens de inhoud daarvan, behoudens ieders verantwoordelijkheid volgens de wet.” (De Nederlandse Grondwet, z.d.). Er zijn echter ook artikelen die ervoor zorgen dat niet álles kan, waaronder artikel 137c en artikel 137d. Deze artikelen hebben betrekking op het verbod tot belediging van een groep wegens hun ras, godsdienst of levelsovertuiging, seksuele gerichtheid of handicap en een verbod tot het openbaar aanzetten tot discriminatie, haat of geweld (Amnesty, 2019).

Ondanks deze wetten is er sprake van hate speech op sociale media. Hate speech wordt op verschillende manieren gedefinieerd. Cohen-Almagor (2015, p.148) definieert hate speech als vijandige en kwaadaardige taal die gericht is op een individu of een groep vanwege hun aangeboren kenmerken. Daarnaast zegt Cohen-Almagor (2015) dat hate speech discriminerend, intimiderend, afkeurend en bevooroordeeld is. Warner en Hirschberg (2012) noemen hate speech een vorm van offensieve taal, die gebruik maakt van stereotypes. Nockleby (2000) ziet hate speech als het in diskrediet brengen van een groep of individu op basis van ras, kleur, seksuele geaardheid, etc. Omdat al deze definities van hate speech verschillende componenten bevatten, zijn de aspecten van deze definities samengenomen tot een eigen definitie om hate speech zo goed mogelijk te kunnen onderzoeken. Uit de definities van Cohen-Almagor (2015), Warner en Hirschberg (2012), Nockleby (2000) en artikel 137d van de Nederlandse Grondwet (Amnesty, 2019) werd de volgende definitie van hate speech gevormd: “een vijandige, kwaadsprekende, offensieve, vooringenomen gemotiveerde, discriminerende uiting die beroep doet of gebruik maakt van stereotyperende kenmerken die de out-group in diskrediet brengt”.

Op het internet is er sprake van een toename van hate speech (Delgado & Stefancic, 2014). Hate speech is echter niet nieuw; het bestond ook al vóór het internet. Maar, ten opzichte van mondelinge opmerkingen, verdwijnen opmerkingen online niet zo snel en blijven ze meestal permanent staan (Delgado & Stefancic, 2014). Online opmerkingen kunnen bewaard worden, in tegenstelling tot mondelinge opmerkingen die verdwijnen zodra ze uitgesproken zijn. In enkele gevallen kunnen opmerkingen online zelfs viraal gaan, wat

duizenden of zelfs miljoenen kijkers trekt, waardoor de opmerkingen misschien wel nooit meer uit de *cyberspace* verdwijnen. De aard en de structuur van het internet maken het makkelijk om hate speech te verspreiden. Mensen kunnen namelijk snel en zonder filter of kritiek zichzelf uiten (Cohen-Almagor, 2017). Overigens constateerden Delgado en Stefancic (2014) dat hate speech ontstaat doordat we twee verschillende soorten normen en waarden hebben; een voor officiële gelegenheden en een voor privé gelegenheden. Eenzelfde persoon kan tijdens een publieke gelegenheid, bijvoorbeeld een werk borrel, sympathiek en aardig zijn tegen een homoseksuele man, maar wanneer deze persoon met zijn vrienden thuis op de bank zit, negatieve gevoelens hebben over homoseksualiteit. We wisselen dus af tussen deze twee verschillende waarden, afhankelijk van de situatie, context en omgeving. Dit zou kunnen betekenen dat iemand niet/weinig gebruik maakt van hate speech face-to-face, terwijl hij of zij dit wel doet op sociale media. Een bericht plaatsen op sociale media is vaak iets wat gebeurt tijdens privé gelegenheden. Vooral berichten die hate speech bevatten. Het verspreiden van hate speech gaat regelmatig in combinatie met het gebruik van scheldwoorden. Uit onderzoek van Warner en Hirschberg (2012) is gebleken dat het gebruik van scheldwoorden in combinatie met ziektes erger gevonden wordt dan het gebruik van scheldwoorden zonder ziektes.

Een voorbeeld van een medium waar hate speech wordt verspreid is Twitter. Twitter is een platform waarbij men korte berichtjes of reacties, genaamd tweets, online kan plaatsen. Hoewel veel tweets beschrijvingen van het dagelijks leven bevatten, zijn er ook tweets die nare opmerkingen bevatten. Volgens Delgado en Stefancic (2014) leiden de beperkte karakters (140) die je kunt gebruiken om een tweet op te stellen en de alomtegenwoordigheid van Twitter tot een vijandige en onbeleefde omgeving. Ook toonde eerder onderzoek aan dat men op Twitter eerder emotionele inhoud zal uiten vanwege het gebrek aan zelfbewustzijn en de anonimiteit van dit medium (Burnap & Williams, 2015). Platforms zoals Twitter, Instagram en Facebook worden ook gebruikt bij het verspreiden van hate speech in geval van sociale kwesties. Bij sociale kwesties zoals de ‘zwarte pietten’ discussie bijvoorbeeld, zien we dat er twee grote groepen tegenover elkaar staan: de Nederlanders die ‘zwarte piet’ willen behouden en de Nederlanders die dit niet willen. Dit is enkel één voorbeeld van twee groepen waarbij er sprake is van (online) hate speech ten opzichte van elkaar. De hate speech tussen deze groepen vindt plaats in het echte leven (face-to-face) en online, maar online is het beter zichtbaar (Papacharissi, 2011).

Een ander voorbeeld waarbij verschillende groepen online hate speech gebruiken is de ophef die ontstond nadat Thierry Baudet, fractievoorzitter van de Forum voor Democratie, de volgende tweet plaatste: “Vanavond zijn twee dierbare vriendinnen ernstig lastig gevallen door vier marokkanen in een trein. Aangifte doen natuurlijk volstrekt zinloos. Oh lieve, kinderlijk naïeve Nederlanders! Stem nou toch eindelijk voor verandering. Breek los van politiek correct gelul! Red dit land!” (zie bijlage 1). Na deze tweet kwamen er al snel allerlei berichten naar boven dat het verhaal van Baudet niet zou kloppen. De NS bevestigde deze berichten; het ging om conducteurs in burger die kaartjes aan het controleren waren (rtnieuws, 2020). Baudets tweet en de reactie van de NS leidde tot een uitbarsting aan reacties waarbij twee groepen, vanwege hun gedachtegoed, tegenover elkaar kwamen te staan.

De tweet van Baudet en de reacties daarop brengen een aantal vragen met zich mee. Wanneer wordt een tweet als hate speech gezien? Welke groep gebruikt meer hate speech? Welke rol speelt sociale identiteit bij online hate speech? In dit onderzoek richten wij ons specifiek op welke groep meer hate speech gebruikt en waarom dit zo zou kunnen zijn. Momenteel is er namelijk nog weinig onderzoek gedaan naar het gebruik van online hate speech in verschillende groepen, terwijl het belangrijk is om in kaart te brengen hoe gebruikers van sociale media hate speech gebruiken wanneer zij bij een specifieke groep horen. Het doel van dit onderzoek is om in kaart te brengen wat de verschillen zijn in het gebruik van hate speech tussen verschillende groepen en de rol die sociale identiteit hierin speelt. Door dit te onderzoeken kunnen deze verschillen in kaart worden gebracht en kan er inzicht verkregen worden in welke invloed sociale identiteit heeft op het gebruik van hate speech. De resultaten kunnen gebruikt worden om het gebruik van online hate speech te begrijpen en deze hate speech aan te vechten.

Voordat we ingaan op de manier waarop het onderzoek is opgezet, bespreken we in het theoretisch kader eerst de *social identity theory* en de *linguistic intergroup bias theory*. Ook zal er een aantal hypothesen worden geformuleerd.

Theoretisch kader

Social identity theory

Ten grondslag aan de *social identity theory* ligt dat een individu het gevoel heeft dat hij of zij lid is van een specifieke sociale groep; dit is hun sociale identiteit. De *social identity theory*

bevat drie belangrijke componenten: sociale categorisatie, sociale identificatie en sociale vergelijking. Deze theorie stelt dat mensen zichzelf in een bepaalde categorie plaatsen waarmee zij zich verbonden voelen. Daarnaast identificeert een individu zich met een bepaalde groep en geeft deze persoon de voorkeur aan deze groep ten opzichte van alle andere groepen, omdat hij affiniteit voelt met die bepaalde groep. Uit eerder onderzoek is gebleken dat een persoon zichzelf vaak vergelijkt met andere individuen. Diegene kan door middel van het vaststellen van vergelijkingen en verschillen met anderen een bepaalde groep kiezen waarmee hij zich identificeert (Mangum & Block, 2018). Dit wordt ook wel de *social comparison theory* (Tankard & Paluck, 2016) genoemd. De sociale groep waarin men zichzelf plaatst op basis van deze vergelijkingen wordt de *in-group* genoemd en de anderen waarmee we onszelf niet identificeren is de *out-group*. We zullen onszelf eerder vergelijken met andere in-group leden dan dat we onszelf zouden vergelijken met out-group leden. Een reden hiervoor is dat het gemakkelijker is om onszelf te vergelijken met in-group leden, omdat zij simpelweg dichterbij ons staan, bijvoorbeeld vanwege leefomstandigheden, vriendschappen en werk (Major, Sciacchitano & Crocker, 1993).

Individen die lid zijn van een specifieke groep hebben vaak dezelfde sociale normen (Pariser, 2011). De subjectieve norm, ook wel sociale norm, is de perceptie van een individu over wat anderen uit de directe omgeving vinden van bepaald gedrag. Dit gaat gepaard met de mate waarin een individu geneigd is om te conformeren aan de mening van de sociale omgeving (Ajzen, 1991). Individen hebben gewoonlijk de neiging om bij een groep te willen horen. Daarom zullen ze eerder geneigd zijn om te conformeren aan de mening van de groep(en) waarvan zij lid zijn (Tajfel & Turner, 1979). Aangezien mensen in groepen waarvan zij lid zijn vaak dezelfde overtuigingen, normen en waarden hebben, kan volgens Pariser (2011) een zogenaamde *echo chamber* ontstaan. In deze echo chamber worden bestaande overtuigingen versterkt door de herhaling van individuen die dezelfde soort meningen en overtuigingen hebben. Overtuigingen die niet in overeenkomst zijn met de opinies van de in-group worden als het ware ‘buitengesloten’. Leden van de out-group worden hierdoor nog meer als buitenstaanders gezien, wat kan leiden tot een verdere kloof tussen de in- en de out-group (Pariser, 2011). Bovendien blijkt uit verschillende onderzoeken dat leden van een in-group geneigd zijn om negatief te praten over leden van een out-group (Tajfel, 1982; Turner & Reynolds, 1987). Leden van de in-group doen dit om hun eigen sociale identiteit te bevestigen en om een positief zelfbeeld te behouden.

De eerste hypothese die moet worden getoetst betreft het verschil in het gebruiken van online hate speech tussen de in-groups en de out-groups. De verwachting is dat tweets over de out-group vaker hate speech bevatten dan tweets over de in-group. Deze verwachting is gebaseerd op eerdere onderzoeken naar de social identity theory (Tajfel, 1982; Turner & Reynolds, 1987) waaruit is gebleken dat leden van een in-group geneigd zijn om negatief te praten over leden van een out-group. Hieruit volgt onderstaande hypothese:

H1: Tweets die betrekking hebben op de out-group bevatten vaker hate speech dan tweets die betrekking hebben op de in-group.

De volgende hypothese heeft betrekking op de auteurs van de tweets. De verwachting is dat tweets over de out-group die afkomstig zijn van auteurs met een duidelijk profiel vaker hate speech bevatten dan tweets over de out-group van auteurs zonder duidelijk profiel. De auteurs die een duidelijk profiel hebben zijn lid van een duidelijke groep waarin zij over het algemeen dezelfde overtuigingen, normen en waarden hebben (Pariser, 2011). Tevens blijkt uit verschillende onderzoeken dat leden van een in-group negatief praten over leden van een out-group om hun eigen sociale identiteit te bevestigen en om een positief zelfbeeld te behouden (Tajfel, 1982; Turner & Reynolds, 1987). Uit bovenstaande argumentatie volgt onderstaande hypothese:

H2: Tweets over de out-group die afkomstig zijn van auteurs met een duidelijk profiel bevatten vaker hate speech dan tweets over de out-group die afkomstig zijn van auteurs zonder duidelijk profiel.

Naast het onderzoeken van de aanwezigheid van hate speech, wordt er ook gekeken naar de manier waarop hate speech vormgegeven wordt op talig niveau.

Linguistic intergroup bias

De *linguistic intergroup bias* voorspelt volgens Maass, Milesi, Zabbini en Stahlberg (1995) dat individuen geneigd zijn om positief gedrag van de in-group en negatief gedrag van de out-group met abstracte taal te beschrijven ('de donkere man is gewelddadig'). Daarnaast voorspelt de theorie dat individuen negatief gedrag van de eigen in-group en positief gedrag van de out-group met concrete taal beschrijven ('de blanke man heeft de slachtoffers waarschijnlijk pijn gedaan'). De uitspraak 'heeft waarschijnlijk pijn gedaan' is minder intens dan 'is gewelddadig'. Verder bleek uit het onderzoek van Maass et al. (1995) dat uitingen die

ongewenst gedrag van een vijand beschreven en uitingen die gewenst gedrag van een vriend beschreven een hogere mate van abstractie hadden. Abstract taalgebruik werd in het onderzoek van Maas en Salvi (1989) en Maass et al. (1995) gedefinieerd als het gebruiken van adjectieven en statische werkwoorden. Concreet taalgebruik werd gedefinieerd als het gebruiken van interpretatieve actie werkwoorden en descriptieve actie werkwoorden.

De verwachting is, op basis van het onderzoek van Maass et al. (1995), dat tweets met betrekking tot de out-group vaker abstracte taal bevatten dan tweets die betrekking hebben op de in-group. Hieruit volgen onderstaande hypothesen:

H3: Tweets die betrekking hebben op de out-group (van auteurs met een duidelijk profiel) bevatten vaker abstracte taal dan tweets die betrekking hebben op de in-group (van auteurs met een duidelijk profiel).

H4: Tweets die betrekking hebben op de out-group (van auteurs zonder duidelijk profiel) bevatten vaker abstracte taal dan tweets die betrekking hebben op de in-group (van auteurs zonder duidelijk profiel).

Methode

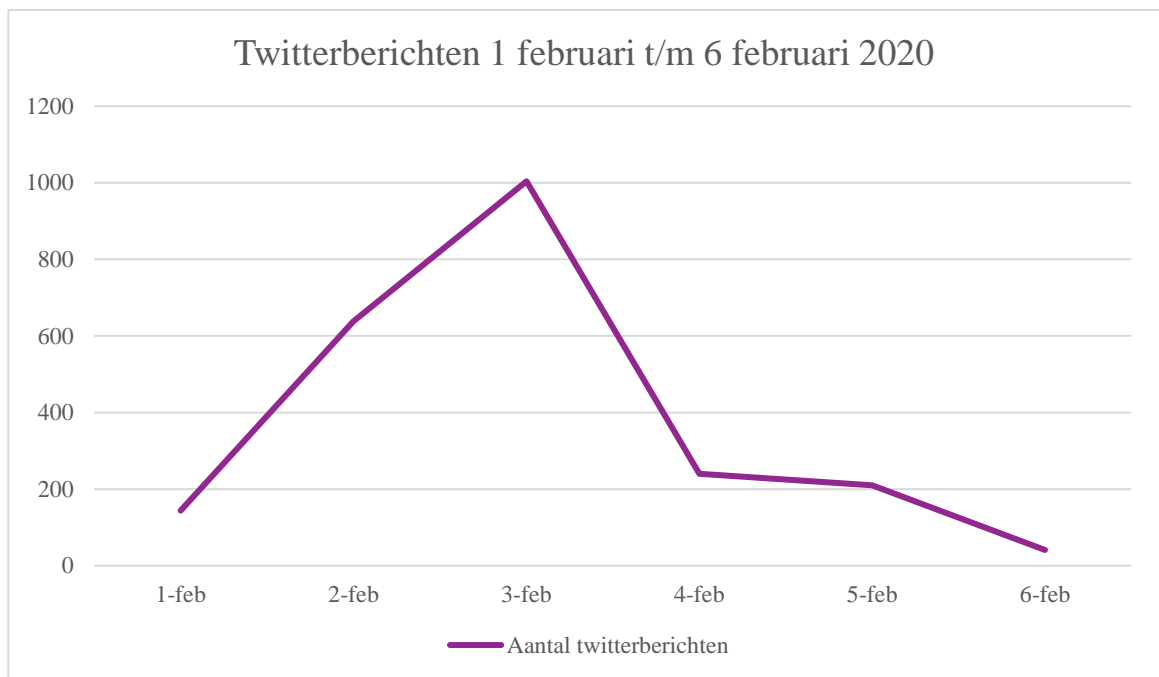
Om antwoord te kunnen geven op de onderzoeksvraag ‘in hoeverre wordt er online hate speech ten opzichte van de andere groep gebruikt bij de in-group versus de out-group?’, is een casus geanalyseerd. De voornaamste reden voor het kiezen van een corpusanalyse was om het daadwerkelijke gedrag van mensen te kunnen analyseren. In het geval van kwantitatief onderzoek kunnen participanten zich namelijk sociaal wenselijk gedragen, waardoor de resultaten beïnvloed zouden kunnen worden.

Materiaal

Het platform Twitter werd gebruikt omdat een beperkt deel van de Nederlandse bevolking zijn/haar mening uit via dit platform. De meningen die men deelt via Twitter zijn ook erg uitgesproken. Dit komt mede doordat Twitterprofielen niet per se persoonlijk zijn, waardoor men zich vrijer voelt om zijn/haar mening te uiten. Hierdoor springt in gevallen zoals deze casus de hate speech meer naar voren en kan de mate van het gebruik van online hate speech beter in kaart worden gebracht.

De casus die werd gebruikt is de reacties die ontstonden naar aanleiding van de tweet van Thierry Baudet (zie bijlage 1) over Marokkanen die zijn vriendinnen hebben lastiggevallen in de trein. Deze casus werd gebruikt omdat er in dit geval twee groepen tegenover elkaar stonden (sympathisanten van Baudet en ‘anderen’) die verschillende opinies en wereldbeelden hebben, mede vanwege verschillende politieke achtergronden. In totaal waren er vier groepen: de in-group van Baudet, de out-group van Baudet, de in-group van ‘anderen’ en de out-group van ‘anderen’ (zie tabel 2 voor verduidelijking van de groepen).

In maart 2020 werden er 2401 twitterberichten opgehaald via het programma Coosto van de periode 1 februari 2020 t/m 6 februari 2020. Ten eerste werd deze periode gebruikt omdat Thierry Baudet zijn tweet verstuurd heeft op 31 januari 2020 om 22:22 uur. De gebruikte periode begon dus anderhalf uur na het versturen van de tweet. Daarnaast was er via Coosto te zien dat op deze momenten de interesse bij mensen op Twitter rondom dit specifieke evenement hoog lag (zie figuur 1). Downs (1972) constateerde dan ook dat de interesse direct na een evenement of bepaalde situatie hoog is, maar daarna ook weer snel afneemt.



Figuur 1. Aantal twitterberichten tussen 1 februari 00:00 en 6 februari 23:59.

Alle tweets betroffen (de tweet van) Thierry Baudet die Marokkanen ervan beschuldigde vriendinnen van hem te hebben lastiggevallen in een NS-trein. De query die hiervoor gebruikt

is, om alleen specifieke tweets over deze casus op te halen, was dan ook: (Thierry OR Baudet OR FvD OR politici OR forum OR tweet OR Thierry Baudet) (Marokkanen OR trein OR incident OR vriendin OR NS OR conducteur OR treinincident OR vriendinnen OR lastiggevallen). Deze tweets zijn voortgekomen uit een combinatie van twee verschillende query's. De eerste helft van de query bestond uit woorden die tweets over Thierry Baudet en zijn tweet/gedachtegoed op moesten halen. De andere helft van de query bestond uit woorden die tweets over het treinincident op moesten halen. Gecombineerd leidden deze twee query's tot tweets over Baudet en (zijn tweet over) het treinincident. Het doel van deze query's was om tot zoveel mogelijk relevante tweets te leiden.

Procedure

In dit onderzoek werden 2401 tweets gecodeerd. Elke codeur heeft 1200 tweets gecodeerd. Iedere tweet werd dus dubbel gecodeerd. De tweets zijn random toegewezen aan de codeur; elke combinatie van codeurs kwam 400 keer voor. Op deze manier ontstond er geen bias in de codering. Als eerste werden de tweets gecodeerd op basis van relevantie. Hierna werden de andere variabelen en de profielen, die verderop uitgebreid worden besproken, gecodeerd. Profielen die niet meer bestonden of tijdelijk waren opgeschort, werden niet gecodeerd. De bijbehorende tweets werden ook niet meegenomen in de corpusanalyse.

Tweets

De tweets werden gecodeerd aan de hand van de eerdergenoemde definitie van hate speech. Hieraan werd ook nog de eerder besproken linguïstic intergroup bias (Maass et al., 1995) toegevoegd. In totaal werden er zeven items gebruikt: de relevantie van de tweet, het onderwerp van de tweet, de toon van de tweet, het gebruik van hate speech, het gebruik van scheldwoorden, het gebruik van ziektes, en het gebruik van abstracte taal (linguïstic intergroup bias). Zie tabel 1 voor een duidelijk overzicht van de variabelen.

De relevantie van de tweet werd gecodeerd aan de hand van de vraag 'is de tweet relevant (past het bij de casus) genoeg om op te nemen in het corpus?'. Een tweet werd als relevant gezien wanneer het over Thierry Baudet in combinatie met het treinincident ging. Wanneer de tweet niet bij de casus paste en dus niet relevant was, werd deze niet meegenomen in het corpus. De relevantie van de tweet was belangrijk omdat er op deze manier geen onnodige informatie mee werd genomen in het corpus en de toetsing.

Het onderwerp van de tweet werd gecodeerd aan de hand van de vraag ‘aan wie is de tweet gericht?’. Als de tweet gericht was naar beide Baudet en Anderen, dan werd de tweet gecodeerd op basis van de groep die het meest prominent aanwezig was in de tweet.

De toon van de tweet werd gecodeerd aan de hand van de vraag ‘welke toon heeft de tweet?’. De toon van de tweet werd als positief gecodeerd als het ging om een positief geladen tweet (#ThierryBaudet reageerde enkel op hetgeen hem verteld is... Als dan achteraf blijkt dat de twee dames niet het hele verhaal vertellen is dat hem niet te verwijten). Negatief betekende in deze casus dat het ging om een negatief geladen tweet (Stem racisme. Stem verdeeldheid. Stem op een lul. Stem op een relschopper. Stem op een feiten ontkenner. Stem Baudet. Stem FvD). Neutraal betekende in deze casus dat het ging om een neutrale tweet.

Het gebruik van online hate speech werd gecodeerd op basis van de eerder gevormde definitie van hate speech, aan de hand van de vraag ‘bevat de tweet offensief, stereotyperend en/of discriminerend taalgebruik en/of brengt het leden van de out-group in diskrediet?’.

Offensief werd geoperationaliseerd op basis van de definitie van Van Dale (z.d.):

‘aanvallend’. Stereotyperend werd geoperationaliseerd als het versimpelen, overdrijven, of generaliseren van de werkelijkheid (Fiske, 1998). Discriminerend werd geoperationaliseerd als het onterecht anders behandelen van mensen op basis van leeftijd, seksualiteit, godsdienst en levensovertuiging, geslacht, ras, nationaliteit, handicap of chronische ziekte, en politieke overtuiging (Ministerie van Algemene Zaken, 2019). Het in diskrediet brengen van de leden van de out-group werd geoperationaliseerd op basis van de definitie van Encyclo (z.d.): ervoor zorgen dat iedereen het vertrouwen in iemand of in een bepaalde groep verliest en dat deze persoon of groep een slecht imago eraan overhoudt. Als de tweet één of meer van deze kenmerken bevatte, dan werd de uiting als hate speech gezien.

Het gebruik van scheldwoorden werd gecodeerd aan de hand van de vraag ‘is/zijn er scheldwoorden richting bepaalde groepen of mensen aanwezig in de tweet?’. Een voorwaarde voor het gebruik van scheldwoorden was dat het scheldwoord aan iemand gericht moest zijn. Bij een tweet als [jullie noemen Baudet allemaal een vieze fascist, maar hij is een nationale held] werden er geen scheldwoorden gecodeerd, ondanks dat ‘vieze fascist’ wel een scheldwoord is. Daarnaast werd er gekeken naar of de scheldwoorden ziektes bevatten. Dit werd gecodeerd aan de hand van de vraag ‘zo ja, bevatten de scheldwoorden ziektes?’. Het gebruik van scheldwoorden en ziektes was van belang omdat er gekeken kon worden naar hoe

ernstig de vorm van hate speech was; ‘vieze vuile kanker Marokkaan’ is iets anders dan ‘Baudet is een klootzak’.

Het gebruik van abstract taalgebruik (linguistic intergroup bias) werd gecodeerd aan de hand van de vraag ‘bevat de tweet abstract taalgebruik?’. Er was sprake van abstract taalgebruik als er gebruik gemaakt was van adjectieven en/of statische werkwoorden.

Tabel 1. De codering van de tweets

Item	Tweet
Relevantie tweet	Relevant (0) / niet relevant (1)
Onderwerp tweet	Baudet (0) / Anderen (1) / niet duidelijk/neutral (2)
Toon tweet	Positief (0) / negatief (1) / neutraal (2)
Hate speech	Wel (0) / niet (1)
Gebruik van scheldwoorden	Wel (0) / niet (1)
Gebruik van ziektes	Wel (0) / niet (1)
Abstract taalgebruik	Wel (0) / niet (1)

Profielen

De twitterprofielen werden gecodeerd aan de hand van twee items: profielkenmerken en profielinformatie.

De profielkenmerken werden gecodeerd aan de hand van de vraag ‘heeft het twitterprofiel een Nederlandse vlag, Forum voor Democratie logo of PVV-logo als profielfoto?’. Indien een profiel een Nederlandse vlag, Forum voor Democratie logo of PVV-logo als profielfoto had, dan werd er geconcludeerd dat deze persoon tot de in-group van Baudet (en dus de out-group van de ‘anderen’) behoorde. Deze groep werden tot de “Baudetaanhangers” benoemd. Indien een profiel dit niet had en dus ‘anders’ was, dan werd er geconcludeerd dat deze persoon tot de out-group (van Baudet) behoorde; de ‘anderen’. Dit was geen genuanceerde keuze, maar in kader van dit onderzoek en de manier waarop het codeboek is opgesteld, was dit de manier waarop dit onderzocht zal worden (zie tabel 2 voor de vier verschillende groepen). Een individu werd namelijk toegewezen aan een bepaalde groep op basis van het profiel en de richting van zijn/haar tweet. Echter, er kon niet één duidelijke groep gevonden worden die tegenover de individuen die ingedeeld waren bij de

groep Baudetaanhangers stond. Linkse individuen bijvoorbeeld toonden hun politieke voorkeur vaak niet op hun profiel, maar enkel in hun tweets die geen betrekking hadden op deze casus.

Tabel 2. De groepen op basis van het profiel en de tweet van de auteur

	Tweet gaat over Baudet		Tweet gaat over Anderen
Profiel van de auteur	Baudet	In-group Baudet	Out-group Baudet
	Anders	Out-group Anderen	In-group Anderen

De profielinformatie werd gecodeerd aan de hand van de vraag ‘heeft het profiel informatie in de biografie staan waaruit een extreemrechtse politieke voorkeur blijkt?’ met 0 = ja, en 2 = nee. Extreemrechts werd geoperationaliseerd op basis van de gebruikte woorden in de biografie van het profiel zoals: FVD-aanhanger, PVV-aanhanger, (extreem) rechts, anti-links, haat aan linkse partijen, etc. Indien een profiel deze informatie in zijn/haar profiel had staan, dan werd er geconcludeerd dat deze persoon tot de in-group (van Baudet) behoorde. Indien een profiel geen informatie in de biografie had staan waaruit bleek dat hij/zij een voorkeur heeft voor extreemrechts, dan werd er geconcludeerd dat deze persoon tot de out-group (van Baudet) behoorde; de ‘anderen’.

Tabel 3. De codering van de twitterprofielen

Item	Profiel
Profielkenmerken	Nederlandse vlag/FvD-logo/PVV-logo (0) of ‘anders’ (1)
Profielinformatie	Extreemrechts (0) of ‘anders’ (1)

Cohens kappa

Om de interbeoordelaarsbetrouwbaarheid (IBB) van de codeurs te meten werd de Cohens kappa gebruikt. De IBB werd voor elke variabele gemeten: relevantie van de tweet, onderwerp van de tweet, toon van de tweet, hate speech, het gebruik van scheldwoorden en ziektes, het gebruik van abstracte taal, de profielkenmerken en de profielinformatie.

Eerst werden 150 tweets en de bijbehorende profielen gecodeerd. Nadat $\kappa > .7$ was bij alle variabelen, werden de rest van de tweets en profielen gecodeerd. De

interbeoordelaarsbetrouwbaarheid lag uiteindelijk tussen .72 en 1.00 (zie tabel 4). Dit betekende dat de coderingen betrouwbaar waren en we deze konden gaan analyseren.

Tabel 4. Interbeoordelaarsbetrouwbaarheid gecodeerde variabelen

Variabele	Cohens kappa
Profielfoto	.91
Profielinformatie	.91
Relevantie tweet	1.00
Onderwerp tweet	.81
Toon tweet	.72
Hate speech	.77
Gebruik van scheldwoorden	.75
Gebruik van ziektes	.80
Abstract taalgebruik	.72

Statistische toetsing

De hypothesen werden getoetst aan de hand van χ^2 -toetsen om de samenhang tussen twee nominale variabelen te meten. H1 werd getoetst door middel van een Chi-kwadraat toets tussen in-group/out-group en hate speech. Vooraf de toetsing is de variabele “In- en Outgroup” gecreëerd door de twee in-groepen samen te voegen en de twee uit-groepen samen te voegen. H2 werd getoetst aan de hand van een Chi-kwadraat toets tussen groep en hate speech. Vooraf de toetsing is de variabele “Groep” gecreëerd op basis van de groep waartoe iemand behoorde en de richting van de tweet. H3 werd getoetst door middel van een Chi-kwadraat toets tussen de in- en out-group van Baudet en abstract taalgebruik. Vooraf de toetsing is de variabele “In- en Outgroup van Baudet” gecreëerd op basis van de groep waartoe iemand behoorde (die van Baudet) en de richting van de tweet. H4 werd getoetst aan de hand van een Chi-kwadraat toets tussen in- en outgroup van Anderen en abstract taalgebruik. Vooraf de toetsing is de variabele “In- en Outgroup van Anderen” gecreëerd op basis van de groep waartoe iemand behoorde (die van de Anderen) en de richting van de tweet.

Resultaten

Ten eerste zullen we een aantal algemene resultaten bespreken wat betreft het aantal tweets, profielen en de indeling van de groepen. In totaal waren 2344 van de 2401 tweets (98%) relevant. Verder waren 57 van de 2401 profielen (2%) opgeschort of verwijderd. Uit een frequentie-analyse van de groepen bleek dat 194 (8%) van de profielen tot de groep Baudetaanhangers behoorde en 2344 (92%) tot de groep Anderen. Er was dus sprake van een scheve verdeling (zie tabel 5 hieronder).

Tabel 5. De verdeling van de groepen (Baudetaanhangers en Anderen) in percentages.

Groep	Percentage
Baudetaanhangers	8%
Anderen	92%

Ten tweede kijken we naar wat algemene resultaten wat betreft hate speech, abstract taalgebruik, het gebruik van scheldwoorden en het gebruik van ziektes. Uit een aantal frequentie-analyses bleek dat hate speech in 450 tweets voorkwam (19%), abstract taalgebruik in 491 tweets (21%), scheldwoorden in 199 tweets (9%), en ziektes in 2 tweets (0.1%). Zie tabel 6 hieronder voor een compleet overzicht.

Tabel 6. Het voorkomen van hate speech, abstract taalgebruik, scheldwoorden en ziektes in tweets (in percentages).

	Hate speech	Abstract taalgebruik	Scheldwoorden	Ziektes
Ja	19%	21%	9%	0.1%
Nee	81%	79%	91%	99.9%

Om de onderzoeksvraag ‘In hoeverre wordt er online hate speech ten opzichte van de andere groep gebruikt bij de in-group versus de out-group?’ te beantwoorden en de bijbehorende hypothesen te toetsen, zijn er een aantal Chi-kwadraat toetsen uitgevoerd.

Uit de χ^2 -toets tussen In-group/out-group en Hate speech bleek geen significant verband te bestaan ($\chi^2(1) = .86, p = .354$). Bij deze verwerpen de resultaten dat tweets die

betrekking hebben op de out-group vaker hate speech bevatten dan tweets die betrekking hebben op de in-group.

Uit de χ^2 -toets tussen de Groep waarin iemand werd ingedeeld en Hate speech bleek een significant verband te bestaan ($\chi^2(1) = 27.94, p < .001$). Individuen die tot de groep Baudetaanhangers behoorden gebruikten relatief vaker dan verwacht (op basis van toeval) wél hate speech richting Anderen (51%). Daarnaast gebruikten de individuen die tot de groep Baudetaanhangers behoorden relatief minder vaak dan verwacht (op basis van toeval) géén hate speech richting Anderen (49%). Het gebruik van hate speech van Anderen richting individuen die ingedeeld waren bij de groep Baudetaanhangers droeg niet bij aan het significante verband tussen de groep waartoe iemand behoort en hate speech. Bij deze bevestigingen de resultaten dat tweets over de out-group die afkomstig zijn van auteurs met een duidelijk profiel (Baudet) vaker hate speech bevatten dan tweets over de out-group afkomstig van auteurs zonder duidelijk profiel (Anderen).

Tabel 7. De percentages van het gebruik van hate speech door Baudetaanhangers richting Anderen.

Gebruik van hate speech	Baudet richting Anderen
Ja	51%
Nee	49%

Uit de χ^2 -toets tussen de In- en Outgroup van Baudet en Abstract taalgebruik bleek een significant verband te bestaan ($\chi^2(1) = 7.46, p = .006$). De in-group van Baudet gebruikte relatief vaker abstracte taal als het over de out-group ging (48%) dan over de eigen in-group (25.8%). Daarnaast gebruikte de in-group van Baudet relatief vaker géén abstracte taal als het over de eigen in-group ging (74%) dan over de out-group (52%). Bij deze bevestigingen de resultaten dat tweets die betrekking hebben op de out-group (van auteurs met een duidelijk profiel) vaker abstracte taal bevatten dan tweets die betrekking hebben op de in-group (van auteurs met een duidelijk profiel).

Tabel 8. De percentages van het gebruik van abstracte taal van de in-group van Baudetaanhangers.

Gebruik van abstracte taal	In-group Baudet over in-group Baudet	In-group Baudet over out-group Baudet
-----------------------------------	---	--

Ja	26%	48%
Nee	74%	52%

Uit de χ^2 -toets tussen de In- en Outgroup van Anderen en Abstract taalgebruik bleek geen significant verband te bestaan ($\chi^2(1) = .01, p = .915$). De resultaten verwerpen bij deze dat tweets die betrekking hebben op de out-group (van auteurs zonder duidelijk profiel) vaker abstracte taal bevatten dan tweets die betrekking hebben op de in-group (van auteurs zonder duidelijk profiel).

Conclusie

Nu alle resultaten besproken zijn, kan er een antwoord worden gegeven op de onderzoeksvraag ‘In hoeverre wordt er online hate speech ten opzichte van de andere groep gebruikt bij de in-group versus de out-group?’. Ook konden de bijbehorende hypothesen verworpen dan wel bevestigd worden op basis van de resultaten.

Ten eerste werd er geen significant verband gevonden tussen de groep waartoe iemand behoort (in-group of out-group) en het gebruik van online hate speech. H1: ‘Tweets die betrekking hebben op de out-group bevatten vaker hate speech dan tweets die betrekking hebben op de in-group’ werd op basis van deze resultaten dus verworpen.

Ten tweede werd er wel een significant verband gevonden tussen de groep waartoe iemand behoort (en de richting van de tweet) en het gebruik van online hate speech. Op basis van deze resultaten kon H2: ‘Tweets over de out-group die afkomstig zijn van auteurs met een duidelijk profiel bevatten vaker hate speech dan tweets over de out-group afkomstig van auteurs zonder duidelijk profiel’ worden bevestigd. Het gebruik van hate speech van Anderen richting individuen die tot de groep Baudetaanhangers behoorden droeg niet bij aan het significante verband. Hieruit kunnen we concluderen dat auteurs met een duidelijk profiel vaker hate speech online gebruiken dan verwacht. Een voorbeeld van een tweet die hate speech bevatte van een individu die ingedeeld was bij de groep Baudetaanhangers richting zijn out-group:

“[Baudet] zwijgt na ophef over [lastiggevallen] [vriendinnen]. Uiteraard blijf ik zeker op [FvD] stemmen want laat me door die linkse idioten niets aanpraten. Wel hoop ik dat [Thierry] zijn excuses maakt want zit er behoorlijk naast en ik zou dat zeker waarderen. Forza [FvD]”.

Ten derde werd een significant verband gevonden tussen het gebruik van abstracte taal en de in- en out-group van Baudet. H3: ‘Tweets die betrekking hebben op de out-group (van auteurs met een duidelijk profiel) bevatten vaker abstracte taal dan tweets die betrekking hebben op de in-group (van auteurs met een duidelijk profiel)’ kon op basis van deze resultaten dus worden bevestigd. Een voorbeeld van een tweet m.b.t. de out-group die abstract taalgebruik bevatte:

“Leugens van de @telegraaf Je mag dus niks zeggen als je vrienden gemolesteerd worden door criminele [Marokkanen]. :>> Framing: [Baudet] zwijgt na ophef over [lastiggevallen] [vriendinnen] <https://t.co/RJkmLZu2Ys>”

Tot slot werd er geen significant verband gevonden tussen het gebruik van abstracte taal en de in- en out-group van de Anderen. H3b: ‘Tweets die betrekking hebben op de out-group (van auteurs zonder duidelijk profiel) bevatten vaker abstracte taal dan tweets die betrekking hebben op de in-group (van auteurs zonder duidelijk profiel)’ kon op basis van deze resultaten dus worden verworpen.

In conclusie, er is niet één antwoord op de vraag ‘In hoeverre wordt er online hate speech ten opzichte van de andere groep gebruikt bij de in-group versus de out-group?’. Het belangrijkste resultaat is dat auteurs met duidelijke profielen vaker online hate speech gebruiken dan verwacht. In dit geval was dat de in-group van Baudet. Daarnaast werd er door de in-group van Baudet vaker abstracte taal gebruikt in tweets die betrekking hebben op hun out-group (Anderen).

Discussie

Verwacht werd dat tweets die betrekking hadden op de out-group vaker hate speech bevatten dan tweets die betrekking hadden op de in-group. Echter, wanneer we kijken naar de resultaten zien we dat deze hypothese niet klopt. De resultaten komen niet overeen met de resultaten van eerdere onderzoeken naar de social identity theory (Tajfel, 1982; Turner & Reynolds, 1987) waaruit bleek dat dat leden van een in-group geneigd zijn om negatief te praten over leden van een out-group. Niettemin, de resultaten zijn niet geheel onverwacht. Ten eerste werden de twee in-groepen (die van Baudet en die van Anderen) samengevoegd in ons onderzoek voor deze hypothese. Ook werden de twee uit-groepen samengevoegd. De groep ‘Anderen’ is echter heel divers, wat invloed gehad kon hebben op de resultaten. Van

twee gelijk verdeelde groepen was namelijk ook geen sprake; de verdeling was juist erg scheef. Dit zou ervoor gezorgd kunnen hebben dat er geen significant verband aanwezig was. Om deze reden werd er een tweede hypothese opgesteld, die de twee in-groepen en twee uit-groepen wel van elkaar scheidt.

Verwacht werd dat tweets over de out-group afkomstig van auteurs met een duidelijk profiel vaker hate speech bevatten dan tweets over de out-group afkomstig van auteurs zonder duidelijk profiel. Deze verwachting was gebaseerd op onderzoek van Pariser (2011) waaruit bleek dat mensen met over het algemeen dezelfde overtuigingen, normen en waarden lid zijn van een duidelijke groep. Daarnaast was deze verwachting gebaseerd op eerdere onderzoeken van Tajfel (1982) en Turner en Reynolds (1987) waaruit bleek dat leden van een in-group geneigd zijn om negatief te praten over leden van een out-group. Wanneer we naar de resultaten kijken zien we dat dit klopt. Een significant verband werd gevonden tussen de groep waartoe iemand behoort (en de richting van de tweet) en het gebruik van online hate speech. Individuen die tot de groep Baudetaanhangers behoorden bleken vaker dan verwacht hate speech te gebruiken richting hun out-group (de Anderen). Een verklaring hiervoor zou kunnen zijn dat zij een gedeeld gedachtegoed hebben met Baudet en hem om deze reden verdedigen tegenover de andere groep.

Verder werd verwacht dat tweets die betrekking hadden op de out-group (van auteurs met een duidelijk profiel) vaker abstracte taal bevatten dan tweets die betrekking hadden op de in-group (van auteurs met een duidelijk profiel). Op basis van de resultaten kan er gezegd worden dat deze hypothese klopt. Een significant verband werd gevonden tussen het gebruik van abstracte taal en de in- en out-group van Baudet. De in-group van Baudet gebruikte vaker abstracte taal als het over de out-group ging dan over de eigen in-group. Bovendien gebruikte de in-group van Baudet vaker géén abstracte taal (maar concrete taal) als het over de eigen in-group ging. In combinatie met het resultaat van hypothese 2, kan er gezegd worden dat de resultaten overeenkomen met onderzoek van Maass et al. (1995) waaruit is gebleken dat individuen geneigd zijn om positief gedrag van de in-group en negatief gedrag van de out-group met abstracte taal te beschrijven.

Tot slot werd er verwacht dat tweets die betrekking hadden op de out-group (van auteurs zonder duidelijk profiel) vaker abstracte taal bevatten dan tweets die betrekking hadden op de in-group (van auteurs zonder duidelijk profiel). Wanneer we kijken naar de resultaten zien we dat deze aanname niet klopt. Er werd geen significant verband gevonden tussen het gebruik van abstracte taal en de in- en out-group van de Anderen. Dit komt niet

overeen met het bovengenoemde onderzoek van Maass et al. (1995). Wederom zou de oorzaak kunnen zijn dat niet alle leden uit de groep Anderen in het dagelijks leven een ingroep vormen. De groep Anderen is namelijk erg divers en bestaat uit individuen met verschillende politieke achtergronden en meningen. Dit zou effect gehad kunnen hebben op de resultaten van dit onderzoek.

Beperkingen

Data

De data die is gebruikt voor dit onderzoek zijn tweets, gedownload via het programma Coosto. Ondanks dat Twitter een platform is waarbij hate speech goed te analyseren valt, zit er maar een deel van de Nederlandse bevolking op Twitter, waardoor de resultaten van dit onderzoek niet per se zouden gelden voor de hele populatie van Nederland. Deze resultaten zijn hierdoor niet te generaliseren naar een grotere groep; we weten namelijk niets over de leeftijd, leefstijl, woonplaats, etc. van de auteurs van de tweets. Het enige wat we weten is hun politieke voorkeur gebaseerd op wat zij weergeven op hun twitterprofiel. Daarnaast zijn de resultaten ook niet te generaliseren naar een andere groep, bijvoorbeeld mensen op Facebook. Twitter is namelijk, zoals eerder besproken, een medium waarbij er sprake is van een gebrek aan zelfbewustzijn. Daarnaast kan je op Twitter anoniem zijn, waardoor je eerder emotionele inhoud zal uiten (Burnap & Williams, 2015). Twitter is in dat opzicht dus anders dan andere sociale media platforms. Daarnaast is er in dit onderzoek voor een specifieke casus gekozen, namelijk de ophef die ontstond naar aanleiding van Baudet en zijn “Marokkanen-tweet” (zie bijlage 1). Er is geen garantie dat we dezelfde resultaten zouden krijgen als de casus bijvoorbeeld het bouwen van nieuwe moskeeën in Nederland betrof. Het gebruiken van Twitter en deze specifieke casus zouden ervoor kunnen hebben gezorgd dat de externe validiteit van dit onderzoek niet volledig in orde is.

Ook valt er iets te zeggen over de ecologische validiteit van dit onderzoek. De data die is geanalyseerd is namelijk van de periode 1 februari 2020 t/m 6 februari 2020. Deze periode is gebruikt omdat dit de periode betrof meteen na het versturen van de tweet van Baudet. Dit kan er echter wel voor hebben gezorgd dat online hate speech vaker voorkwam dan normaal het geval is, aangezien er werd gereageerd op een specifieke situatie die niet altijd voorkomt. Echter, er kan wel gezegd worden dat de tweet van Baudet overeenkomt met zijn normale,

politieke ideologie. De reacties die ontstonden naar aanleiding van zijn tweet zullen dus hoogstwaarschijnlijk meningen zijn geweest die de meesten op voorhand al hadden.

Indeling van de groepen

In dit onderzoek werden vier groepen van elkaar onderscheiden: de in-group van Baudet, de out-group van Baudet (de Anderen), de in-group van Anderen en de out-group van Anderen (Baudetaanhangers). De verdeling van deze groepen was echter erg scheef; 8% behoorde tot de Baudetaanhangers, terwijl 92% tot de Anderen behoorde. Dit zou invloed kunnen hebben gehad op de resultaten. Daarnaast was de groep Anderen veel meer divers dan de individuen die tot de groep Baudetaanhangers behoorden. Iemand behoorde namelijk tot de Anderen als hun profiel géén extreemrechtse kenmerken bevatte. Binnen de groep Anderen kwamen veel verschillende groepen voor; extreemlinkse mensen, rechtse mensen, mensen die niets om politiek geven, Marokkanen, Nederlanders, enzovoorts. De groep Anderen kon zelfs Baudetaanhangers bevatten, omdat zij niet voldeden aan de criteria om tot de groep Baudetaanhangers te behoren.

De codering

Tijdens het coderen ontstonden er wat onduidelijkheden wat betreft de coderingen. Ten eerste kwamen sarcastische uitingen redelijk vaak voor in de dataset, maar dit was geen variabele die werd meegenomen in het onderzoek. Het was echter moeilijk te bepalen of sarcasme onder hate speech viel, en of de toon van de tweet hierdoor negatief was. Om deze reden werd sarcasme in de meeste gevallen gecodeerd als geen hate speech en neutraal.

Een andere beperking van dit onderzoek is dat, in één tweet, een persoon iets positiefs kan zeggen over de eigen in-group en iets negatiefs over de out-group. Er is voor gekozen om de toon van de tweet (positief, negatief, neutraal) te coderen op basis van het hoofdonderwerp van de tweet. Daarnaast is er in dit onderzoek voor gekozen om bij de richting van de tweet de meest prominent aanwezige groep te coderen, maar dit was niet altijd duidelijk. Een voorbeeld hiervan was de volgende tweet: “[NS] weerspreekt '[Marokkanen]-[tweet]’ [Baudet]: 'Het was een kaartcontrole' <https://t.co/1OqAglJPoQ> via @NOS”. Bovendien waren de meeste tweets gericht aan Baudet, wat logisch is aangezien de reacties ontstonden

naar aanleiding van zijn tweet. Dit zorgde echter wel voor een scheve verdeling van de richting van de tweet.

In de dataset kwamen vaak retweets van neutrale nieuwsberichten voor, zonder dat de persoon hier iets aan had toegevoegd. In totaal waren 1089 (47%) tweets neutraal. Deze werden wel meegenomen in het corpus, omdat het onderwerp van de tweet de casus betrof en deze dus relevant was. Hier hadden we achteraf gezien echter niet veel aan, aangezien we hate speech onderzochten en niet neutraliteit.

De codering van de variabelen is een subjectieve taak van de onderzoeker(s). De onderzoeker kiest ervoor welke uitingen hij op welke manier codeert en welke uitingen hij wel en niet meeneemt in het onderzoek. Per tweet en profiel hebben de codeurs beslist welke uitingen wel en niet werden meegenomen en hoe deze werden gecodeerd. Echter, de manier waarop de codering heeft plaatsgevonden is duidelijk gerapporteerd in het codeboek, waardoor deze subjectiviteit geen groot probleem vormt. Het onderzoek is namelijk replicerbaar en er kan door andere onderzoekers worden onderzocht of zij dezelfde resultaten verkrijgen als dit onderzoek.

Vervolgonderzoek

Ten eerste zou er in het vervolg onderzoek gedaan kunnen worden waarbij er sprake is van een betere verdeling van in- en out-group. Zo kan er namelijk beter in kaart worden gebracht wat de verschillen zijn in het gebruik van hate speech tussen deze twee groepen. Daarnaast zouden twitterprofielen in het vervolg weggelaten kunnen worden om de groep waartoe iemand behoort te bepalen. In plaats daarvan, zou er alleen naar de tweet gekeken kunnen worden om vast te stellen bij welke groep een individu hoort.

In het vervolg zou ook er gekeken kunnen worden naar andere variabelen bij het coderen van het gebruik van online hate speech, zoals bijvoorbeeld sarcasme. Dit kwam in ons databestand vaak voor, maar is niet gecodeerd of geanalyseerd. Het is interessant om te kijken of/wanneer het gebruik van sarcasme iets toevoegt aan de mate van hate speech. Daarnaast zou het interessant zijn om in te zoomen op het gebruik van abstracte taal. In dit onderzoek is er gecodeerd of er sprake was van abstract taalgebruik als er statische werkwoorden of adjectieven aanwezig waren in een tweet. Echter, er is niet gekeken wanneer er abstracte (of concrete) taal werd gebruikt en door wie. Dit zou eventueel interessant kunnen zijn voor vervolgonderzoek. Op deze manier zou er ook gekeken kunnen worden of de

resultaten overeenkomen met de resultaten van het onderzoek van Maass et al. (1995) en Maass en Salvi (1989).

In het vervolg zou er ook een andere casus gebruikt kunnen worden, met dezelfde onderzoeksvraag en variabelen. Het is mogelijk dat de context veel verschil maakt in hoeverre er online hate speech aanwezig is. Door het gebruiken van een andere casus zou er vastgesteld kunnen worden of dezelfde resultaten verkregen zouden worden in een andere situatie.

Tot slot zou het interessant zijn om, in plaats van een corpusanalyse, een andere methode te gebruiken. Door een experiment of survey te gebruiken, kan de sociale norm in de in- en out-group bevraagd en gemeten worden. In het geval van onze dataset met tweets was het lastig te bepalen of er wel of niet sprake was van een sociale norm.

Implicaties onderzoek

De resultaten van dit onderzoek en de ideeën voor vervolgonderzoek kunnen gebruikt worden door de overheid in de strijd tegen online hate speech. Op basis van de resultaten kan er geconcludeerd worden dat er relatief vaker hate speech wordt gebruikt tegenover de out-group dan tegenover de in-group. De overheid zou op sociale media kanalen kunnen opletten of er sprake is van bepaalde groepen die tegenover elkaar staan wat betreft hun sociale normen, waarden en hun gedachtegoed. Dit zou namelijk erop kunnen duiden dat er sprake is van online hate speech tussen deze groepen. Door te kijken naar de aanwezigheid van in- en out-groups kan online hate speech gemonitord worden. Daarnaast kan er op basis van de resultaten geconcludeerd worden dat er relatief vaker abstract taalgebruik is tegenover de out-group dan tegenover de in-group. Dit betekent dat wanneer er online sprake is van abstract taalgebruik, de overheid hier op zou moeten letten. Dit kan namelijk duiden op de aanwezigheid van een out-group. Door te letten op abstract taalgebruik op online platformen kan de eventuele aanwezigheid van hate speech en discriminatie in de gaten gehouden worden. In conclusie, de strijd tegen online hate speech kan met behulp van dit onderzoek worden voortgezet.

Literatuur

Ajzen, I. (1991). The theory of planned behavior. *Organizational behavior and human decision processes*, 50(2), 179-211.

Artikel 7: Vrijheid van meningsuiting; censuurverbod. (z.d.). Geraadpleegd op 11 maart 2020, van https://www.denederlandsegrondwet.nl/id/vgrnbj1z0qzw/artikel_7_vrijheid_van_meningsuiting

Begrip – diskrediet. (z.d.). Geraadpleegd op 12 maart 2020, van <https://www.encyclo.nl/begrip/diskrediet>

Bevolking; geslacht, leeftijd, generatie en migratieachtergrond, 1 januari. (2019, 27 september). Geraadpleegd op 10 maart 2020, van <https://opendata.cbs.nl/statline/#/CBS/nl/dataset/37325/table?ts=1583872701991>

Burnap, P., and Williams, M. L. (2015). Cyber hate speech on twitter: An application of machine classification and statistical modelling for policy and decision making. *Policy & Internet*, 7(2), 223–242.

Cohen-Almagor, R. (2015). *Introduction to: Confronting the Internet's Dark Side: Moral and Social Responsibility on the Free Highway*. NY and Washington DC : Woodrow Wilson Center Press and Cambridge University Press. 148.

Cohen-Almagor, R. (2017). Balancing Freedom of Expression and Social Responsibility on the Internet. *Philosophia*, 45, 973-985.

Delgado, R., & Stefancic, J. (2014). Hate speech in cyberspace. *Wake Forest L. Rev.*, 49, 319.

Downs, A. (1972). Up and down with ecology-The 'issue-attention cycle'. *Public Interest*, 28, 28-50.

Fiske, S. (1998). Stereotyping, prejudice and discrimination. In: D. Gilbert, S. Fiske & G. Lindzey (eds.), *The handbook of social psychology*. New York: Mc Graw-Hill.

Kamerleden halen uit naar Baudet vanwege "Marokkanen-tweet": "Walgelijk". (2020, 3 februari). Geraadpleegd op 26 mei 2020, van <https://www.rtlnieuws.nl/nieuws/politiek/artikel/5008171/kamerleden-over-marokkonen-tweet-baudet-walgelijk-en-laf>

- Maass, A. & Salvi, D. (1989). Language use in intergroup context: the linguistic intergroup bias. *Journal of Personality and Social Psychology*, 57(6), 981-993.
- Maass, A., Milesi, A., Zabbini, S. & Stahlberg, D. (1995). Linguistic Intergroup Bias. *Journal of Personality and Social Psychology*, 68(1), 116–126. doi: 10.1037/0022-3514.68.1.116.
- Major, B., Sciacchitano, A. M., & Crocker, J. (1993). In-Group Versus Out-Group Comparisons and Self-Esteem. *Personality and Social Psychology Bulletin*, 19(6), 711–721.
<https://doi.org/10.1177/0146167293196006>
- Mangum, M., & Block, R. (2018). Social Identity Theory and Public Opinion towards Immigration. *Social Sciences*, 7(3), 41. <https://doi.org/10.3390/socsci7030041>
- Ministerie van Algemene Zaken. (2019, 21 februari). *Wat is discriminatie?* Geraadpleegd op 12 maart 2020, van <https://www.rijksoverheid.nl/onderwerpen/discriminatie/vraag-en-antwoord/wat-is-discriminatie>
- Nockleby, J. T. (2000). *Hate speech* (2de editie). New York, Verenigde Staten: Macmillan Publishers. *Offensief*. (z.d.). Geraadpleegd op 12 maart 2020, van <https://www.vandale.nl/gratis-woordenboek/nederlands/betekenis/offensief#.XmoDcahKhPY>
- Papacharissi, Z. (2011). *A Networked Self: Identity, Community and Culture on Social Network Sites*. Abingdon, Verenigd Koninkrijk: Taylor & Francis.
- Pariser, E. (2011). *The Filter Bubble: What the Internet Is Hiding From You*. London: Penguin UK.
- Tankard, M. E., & Paluck, E. L. (2016). Norm perception as a vehicle for social change. *Social Issues and Policy Review*, 10(1), 181-211.
- Tajfel, H. (1982). *Social identity and intergroup relations*. Cambridge: Cambridge University Press. doi: 10.1007/978-3-658-13213-2_75
- Turner, J. C., & Reynolds, K. J. (1987). Self-Categorization Theory. In P. A. M. Van Lange, A.W. Kruglanski, & T. Higgins (Red.), *Handbook of Theories of Social Psychology* (1e ed., pp. 399–417. London: SAGE Publications Ltd. doi: 10.4135/9781446249222.n46

Vrijheid van meningsuiting - mensenrechten. (2019, 5 juli). Geraadpleegd op 11 maart 2020, van <https://www.amnesty.nl/encyclopedie/vrijheid-van-meningsuiting>

Warner, W., & Hirschberg, J. (2012). Detecting Hate Speech on the World Wide Web. *LSM '12*, 2, 19-26.

Bijlagen

Bijlage 1 – Casus Baudet (tweet 31-01-2020)



Bijlage 2 – Verklaring geen fraude en plagiaat

Bachelorstudent Communicatie- en Informatiewetenschappen aan de Letterenfaculteit van de Radboud Universiteit Nijmegen, verklaart met ondertekening van dit formulier het volgende:

a. Ik verklaar hiermee dat ik kennis heb genomen van de facultaire handleiding (www.ru.nl/stip/regels-richtlijnen/fraude-plagiaat), en van artikel 16 “Fraude en plagiaat” in de Onderwijs- en Examenregeling voor de BA-opleiding Communicatie- en Informatiewetenschappen.

b. Ik verklaar tevens dat ik alleen teksten heb ingeleverd die ik in eigen woorden geschreven heb en dat ik daarin de regels heb toegepast van het citeren, parafraseren en verwijzen volgens het Vademecum Rapporteren.

c. Ik verklaar hiermee ook dat ik geen teksten heb ingeleverd die ik reeds ingeleverd heb in het kader van de tentaminering van een ander examenonderdeel van deze of een andere opleiding zonder uitdrukkelijke toestemming van mijn scriptiebegeleider.

d. Ik verklaar dat ik de onderzoeksdata, of mijn onderdeel daarvan, die zijn beschreven in de BA-scriptie daadwerkelijk empirisch heb verkregen en op een wetenschappelijk verantwoordelijke manier heb verwerkt.