

***Of* versus *als* ter inleiding van afhankelijke *ja/nee*-vragen in het Surinaams-Nederlands: een kwantitatieve verkenning**

Anne-Sophie Ghyselen¹

Abstract

Even though Dutch has been the official language of government and education in Suriname since 1876 and about 50% of the Surinamese households reports speaking Dutch at home, defining or delineating Surinamese Dutch remains highly problematic. ‘Typical’ Surinamese Dutch lexical, phonetic, morphological and syntactic features have been identified, but it is unclear to what degree these features are used (cf. de Kleine 2013, 855), by whom, how they correlate with social, stylistic and other parameters and whether they are accepted as standard Dutch by the Surinamese. The present study addresses this lacuna by quantitatively analyzing a feature often considered to be typical of Surinamese Dutch: the use of the subordinating conjunction *als* in dependent *yes/no*-clauses (e.g. Ik weet niet *als* hij komt, ‘I don’t know whether he’s coming’). While in Netherlandic and Belgian Standard Dutch the conjunction *of* is expected, the conjunction *als* is said to be typical of Surinamese Dutch. An analysis of the distribution of *als* in two corpora – a new corpus of spoken Surinamese Dutch and the Surinamese component of the Corpus Hedendaags Nederlands – however indicates that the subordinator *of* is actually dominant in Suriname: in the written corpus, with mainly texts of linguistic professionals, *als* is attested in only 7% of 7722 relevant clauses, whereas in the spoken corpus – which contains the speech of 22 highly educated Creole women – its frequency is higher, viz. 44% (n=161). Machine learning based multivariate analyses show that the variation between *als* and *of* is mostly individually determined. In the spoken dataset, the individual speaker emerges as overpowering predictor, and interestingly, the interspeaker variation is difficult to explain in terms of the speakers’ language backgrounds or age. Linguistic variables such as the matrix verb, the polarity of the sentence and the frequency of the matrix verb have little predictive power. In the written dataset, the interauthor variation could not be estimated due to methodological issues, but here too, the studied

© Anne-Sophie Ghyselen

This is an open access article distributed under the terms of the CC BY-NC-ND 4.0 license
<https://creativecommons.org/licenses/by-nc-nd/4.0/>

linguistic variables have hardly any explanatory power. In the discussion, it is argued that the considerable individual variation might be a sign of endoglossic standardization, whereby the conjunction *als* is establishing itself slowly as part of the Surinamese Dutch standard norm.

Keywords: Surinamese Dutch, endoglossic standardization, subordinating conjunctions, dependent *yes/no*-questions, interspeaker variation, random forest modelling, conditional inference tree

1. Inleiding

Hoewel het Nederlands sinds 1876 de overheids- en onderwijstaal is in Suriname en hoewel het de belangrijkste thuistaal zou zijn van ongeveer 50% van de Surinaamse huishoudens (Algemeen Bureau voor de Statistiek Suriname 2014), zijn er nog veel vragen over de vorm en functie van het Nederlands in het Zuid-Amerikaanse land. Het Surinaams-Nederlands wordt doorgaans omschreven als een “diffuse” taalvariëteit (Muysken 2013, 745), die gekenmerkt wordt door veel inter- en intraspreker variatie en invloeden van de andere (minstens 19) talen gesproken in Suriname. Vooral het Sranan, de op het Engels gebaseerde creooltaal die als informele lingua franca fungeert in Suriname, zou als donortaal dienen. ‘Typisch’ Surinaams-Nederlandse lexicale, fonetische, morfologische en syntactische kenmerken werden in het verleden al gesignaleerd (zie bijvoorbeeld Gobardhan-Rambocus 1993; Van Donselaer 2005; de Kleine 2007; de Bies 2009; Van der Sijs 2014; Muysken 2017) en zouden ervoor zorgen dat het Surinaams-Nederlands zich duidelijk onderscheidt van het ‘Europees-Nederlands’ (EN) – een overkoepelende term voor Nederlands- en Belgisch-Nederlands – maar het is niet zo duidelijk in welke mate die kenmerken worden gebruikt (cf. de Kleine 2013, 855), door wie, wanneer, en of ze door de Surinamers zelf ook als Standaardnederlands worden geaccepteerd. In standaardwerken over het Standaardnederlands – zoals de nieuwste editie van de Algemeen Nederlandse Spraakkunst of het online platform Taalportaal – blijft het Surinaams Standaardnederlands daarom vaak buiten beschouwing, ook wanneer expliciet een pluricentrische aanpak geambieerd wordt. De e-ANS3 schrijft in dat verband het volgende:

[D]e beschrijving van het Surinaamse Nederlands [is] in ANS3 door een gebrek aan voorstudies veel minder systematisch dan die van het Nederlandse en het Belgische Nederlands. In de pluricentrische benadering van de

Nederlandse grammatica die de ANS₃ voorstaat, is er alvast de conceptuele ruimte voor een beschrijving van de grammaticale verschijnselen die typisch zijn voor de standaardtaal in Suriname. Alleen beschikken we voor Suriname vooralsnog niet over de geschikte data om betrouwbare uitspraken te doen over de status van eventuele nationale varianten die daar tot de standaardtaal zouden behoren. De redactie hoopt dat het op termijn, naarmate er meer onderzoek beschikbaar komt naar de verscheidenheid van het Nederlands in Suriname, mogelijk zal zijn om de zich nog volop ontwikkelende Surinaamse variëteit van het Standaardnederlands een uitgebreidere plaats te geven in de beschrijving. (<https://e-ans.ivdnt.org/over>)

In dit artikel wil ik graag een bijdrage bieden aan de beschrijving van het Surinaams-Nederlands door in te gaan op één grammaticaal fenomeen: het voegwoord dat gebruikt wordt om afhankelijke *ja/nee*-vragen in te leiden. In het Nederlands en Belgisch Standaardnederlands worden afhankelijke vraagzinnen ingeleid door het voegwoord *of* (cf. voorbeeldzin 1, e-ANS₃ 10-3-2-1). Enkel bij nevenschikking met *of* van twee vragende bijzinnen verschijnt in de Europese standaardvariëteiten soms *dat* in de tweede bijzin (cf. voorbeeldzin 2, e-ANS₃ 10-3-2-1). In het Surinaamse-Nederlands echter – zo rapporteren meerdere bronnen (Van Wel & Vervoorn n.d. – ca. 1974, 54; Gobardhan-Rambocus 1993, 147; de Kleine 2007, 129-130) – zou veeleer het voegwoord *als* gangbaar zijn (cf. voorbeeldzin 3, uit de Kleine 2013, 129-130).

- (1) *Ik wist niet of hij thuis was.*
- (2) *Hij wist niet meer of hij dat werkelijk had meegemaakt of dat hij het gedroomd had.*
- (3) *Dus ik zou dingen, Helen vragen als we dat mogen hebben.*
- (4) *Of neem ik het, ik scheur het!* (EN: ‘Als ik het neem, scheur ik het’)

Communicatie met Surinaamse docenten Nederlands leert dat de normatieve status van het dubitatieve *als* onduidelijk is. In het secundaire onderwijs bijvoorbeeld zou de afspraak onder vele docenten Nederlands zijn om de constructie in geschreven teksten weliswaar aan te duiden, maar scoregewijs niet te sanctioneren. Andere docenten zouden dan weer geen opmerkingen maken over de constructie.

Discussie is mogelijk over de oorsprong van het dubitatieve *als* in Suriname. Muysken (2017, 304-305), Sluisdom (1992) en Ventura (2013) suggereren invloed van het Sranan *efi*, dat in het Sranan zowel voorwaardelijke bijzinnen inleidt als afhankelijke *ja/nee*-vragen. De veronderstelling is dan

vaak dat Surinamers onder invloed van *efi* in Nederlandse voorwaardelijke zinnen dikwijls de vergelijkbare vorm *of* gebruiken (cf. voorbeeldzin 4, uit de Kleine 2007, 129). De vorm *als* in afhankelijke *ja/nee*-vragen – waar je in het Europees-Nederlands *of* verwacht – zou dan een hypercorrecte vorm zijn, waarbij Surinamers de regel dat het Nederlands *als* gebruikt voor het Sranan *efi* overgeneraliseren naar niet-voorwaardelijke bijzinnen (Sluisdom 1992). Van Wel & Vervoorn (n.d. – ca. 1974, 54) suggereren dan weer invloed van het Engelse *if*, dat net zoals het Sranan *efi* zowel afhankelijke *ja/nee*-vragen als voorwaardelijke zinnen kan inleiden. De Kleine (2007, 130) merkt echter op dat dubitatief *als* niet zomaar als het resultaat van hypercorrectie gezien kan worden. De zogenaamd hypercorrecte *als*-vormen kwamen in haar data immers veel frequenter voor dan de voorwaardelijke *of*-vormen (respectievelijk 11 keer bij 4 van haar 21 Creoolse informanten versus 1 keer), wat volgens haar een analyse van *als* als hypercorrecte vorm ongelooftwaardig maakt. Hoe het ook zij, het is wel aannemelijk dat Surinamers onder invloed van het Sranan, of het ermee verwante Engels – twee talen die voegwoordelijk geen onderscheid maken tussen voorwaardelijke bijzinnen en afhankelijke vraagzinnen – dat onderscheid vaak ook laten varen in het Nederlands. Dat dan het voegwoord *als* veralgemeend wordt, en niet het meer op *efi* en *if* lijkende voegwoord *of*, is wellicht een frequentie-effect: aangezien *als* in het Nederlands frequenter voorkomt dan *of* – zowel in het algemeen als binnen de categorie van de onderschikkende voegwoorden² – is het vanuit cognitief perspectief logisch dat *als* veralgemeend wordt.

We moeten echter ook de optie overwegen dat het dubitatieve *als* al aanwezig was in het taalgebruik van de kolonisten die het Nederlands in Suriname introduceerden. Uniek Surinaams is het gebruik van *als* in plaats van *of* in afhankelijke vraagzinnen immers niet; ook in verschillende Europees-Nederlandse dialecten is het geattesteerd. Figuur 1 – dialectkaart IX uit De Rooy (1965) – toont hoe het voegwoord *als* (witte cirkeltjes en vierkantjes) en de verkorte vorm *a* (zwarte cirkeltjes en vierkantjes) vooral in de dialecten in West- en Oost-Vlaanderen en in Friesland voorkomen. De data in de *Dynamische Syntactische Atlas van de Nederlandse Dialecten* (DynaSAND, Barbiers & al. 2006, zin 203) bevestigen dat kaartbeeld.³ De veronderstelling van De Rooy (1965, 85) dat het dubitatieve *als* vroeger over een veel groter gedeelte van het Nederlandse taalgebied verspreid moet zijn geweest, voedt het idee dat *als* niet puur aan Sranan-invloed (of invloed van het Engels) toegeschreven moet worden, maar al vanouds in het Nederlands aanwezig was. Mijn vermoeden is dat een samenspel van factoren – de historische aanwezigheid van dubitatief *als* in Europees-Nederlandse dialecten en de invloed van het Sranan – het dubitatieve *als* in Suriname in de hand heeft gewerkt.

Hoe sterk het dubitatieve *als* vandaag precies staat in Suriname en of we het als een kenmerk van het Surinaamse Standaardnederlands kunnen zien, is niet duidelijk. De Kleine (2007, 130) stelde in het gesproken taalgebruik van haar 21 Creoolse participanten vast dat de “grote meerderheid” van de voorwaardelijke bijzinnen en afhankelijke vraagzinnen *niet* afwijkt van het Europees-Nederlands, al kwamen afhankelijke *als*-vraagzinnen volgens haar “veel frequenter” voor dan voorwaardelijke *of*-zinnen (cf. supra). Ook Muysken (2017, 304) merkt op dat het dubitatieve *als* in zijn dataset – 15 taalsamples van interviews, een verhaaltaak en een videobeschrijvingstaak, met soms ook sprekers die Nederlands als tweede taal hebben verworven – “niet zo frequent” voorkomt. Hoe infrequent is echter niet duidelijk (relatieve frequenties worden niet gegeven). Daarnaast is ook niet duidelijk in welke mate het *als*-gebruik bepaald wordt door taalinterne factoren (zoals het werkwoord waar de vraag van afhangt), door de talige achtergrond van de taalgebruikers (met bvb. Nederlands als eerste versus als tweede of derde taal), hun sociale profiel, het soort situatie waarin ze zich bevinden en het medium. Dubitatief

als is weliswaar geattesteerd in zowel spreektaal (Muysken 2017, 304; de Kleine 2007) als schrijftaal (Van Wel & Vervoorn n.d. – ca. 1974, 54), maar het is niet duidelijk of het in beide registers even sterk staat.

Om meer inzicht te bieden in het gebruik van dubitatief *als* in Suriname, focust dit artikel op de volgende vragen:

- (1) Hoe vaak realiseren Surinamers de voegwoorden *als* en *of* in afhankelijke vraagwoordzinnen?
- (2) Wordt het *als*-gebruik geconditioneerd door (a) taalinterne factoren, (b) sociale factoren, (c) de taalachtergrond van de taalgebruiker, (d) het medium (spreek- versus schrijftaal) en/of (e) de formaliteit van de situatie?

Een antwoord op die vragen zal hopelijk toelaten in te schatten of dubitatief *als* als een kenmerk van het Surinaamse Standaardnederlands beschouwd kan worden. In wat volgt, wordt in wat volgt eerst ingegaan op de Surinaamse taalcontext, om vervolgens de opzet en de resultaten van het onderzoek toe te lichten.

2. Het Nederlands in Suriname

Suriname is een republiek van ongeveer 598.000 inwoners aan de noordkust van Zuid-Amerika. De bevolking is etnisch divers, zeker in vergelijking met landen als Nederland en België. Geen enkele etnische groep heeft in Suriname immers een absolute meerderheid. Volgens censusdata uit 2012 (cf. Menke & Sno 2016) zijn de vier grootste groepen (1) de Hindostanen (27% van de bevolking), (2) de Marrons (22%), (3) de Creolen (16%) en (4) de Javanen (14%), hoewel een steeds grotere groep Surinamers zichzelf als ‘Gemengd’ identificeert (13% in 2012). De verschillende etnische groepen hebben elk hun eigen ta(a)l(en), al is er geen een-op-eenverband tussen (moeder)taal en etniciteit. Individuele meertaligheid is er immers de regel, en de voorbije decennia werd ook domeinverlies gerapporteerd voor zowel de inheemse talen (Arawak, Trio en Carib) als het Sarnámi (ook wel Hindostaans) en het Javaans (Algemene Bureau voor de Statistiek Censuskantoor 2006). Dat verlies speelt doorgaans in het voordeel van het Nederlands en het Sranan, de belangrijkste lingua franca in Suriname.

De geschiedenis van het Surinaams-Nederlands begint bij de Nederlandse kolonisatie van het Surinaamse grondgebied in 1667. In de 17e eeuw was het gebruik van het Nederlands echter nog beperkt, aangezien het Nederlands

toen niet gepropageerd werd; het werd veeleer beschouwd als voorrecht van een witte, koloniale elite (Gobardhan-Rambocus 1997, 246). In de loop van de 18e eeuw won het Nederlands enigszins aan populariteit, toen een nieuwe middenklasse van gemengd Creools-Kaukasische origine zich in Paramaribo vormde. Deze groep – doorgaans *Stadscreolen* genoemd – maakte zich het Nederlands eigen, vermoedelijk om zich te onderscheiden van andere (Sranan-sprekende) niet-witte Surinamers (de Kleine 2013, 843). Een echte explosie van het aantal Nederlandstaligen kwam er in de 19e eeuw, met de invoering van de leerplicht in 1876 als belangrijkste katalysator. Van dan af aan voerde de Nederlandse overheid een uitgesproken pro-Nederlands taalbeleid in Suriname (Gobardhan-Rambocus 1997, 246). Het werd toen bijvoorbeeld de enige onderwijstaal. Op die manier werd het Nederlands voor steeds meer Surinamers – uit alle etnische groepen – de tweede of eerste taal. Dat de tweetalig Nederlands-Sranansprekende Stadscreolen in de 20e eeuw een erg prominente socio-politieke rol speelden (De Bruijne & Schalkwijk 2005) droeg bij aan die evolutie; het Nederlands werd zo immers steeds meer de taal die je moest beheersen om de socio-economische ladder te beklimmen.

Vandaag, meer dan 40 jaar na de onafhankelijkheid, is het Nederlands nog steeds de belangrijkste taal van bestuur, wetgeving en onderwijs in Suriname, ondanks aanhoudende discussies over het taalbeleid dat de jonge natie zou moeten nastreven (cf. Diepeveen & Hüning 2016). Ook in de privésfeer lijkt het Nederlands een van de meest gesproken talen te zijn. In censusdata van 2012 rapporteert bijvoorbeeld 49,7% van de huishoudens het Nederlands als belangrijkste omgangstaal te hebben; 24,1% noemde het de tweede taal van het gezin (cf. Tabel 1). In een recentere enquête van de Nederlandse Taalunie rapporteerde 88% van de bevraagde 1187 Surinamers Nederlands te spreken met familie, vrienden en kennissen, al dan niet in combinatie met een andere taal (Rys et al. 2021, 20). Met dergelijk cijfermateriaal moet uiteraard voorzichtig omgesprongen worden – gerapporteerd taalgebruik vertelt vaak meer over taalattitudes dan over het werkelijke taalgedrag (cf. Légise & Migge 2015) – maar het toont wel dat het Nederlands nog steeds een belangrijke positie inneemt in Suriname. Een belangrijke consoliderende factor daarbij zijn de aanhoudende migratiebewegingen tussen Nederland en Suriname (Yakpo, van den Berg, & Borges 2015, 172). Vele Surinamers hebben familie of kennissen in Nederland, gaan er op bezoek, volgen een deel van hun opleiding in Nederland of brengen er (een deel van) hun professionele carrière door, en dat werkt het behoud van het Nederlands natuurlijk in de hand.

**Tabel 1: Meest gesproken taal en tweede taal in Surinaamse huishouden
(Algemeen Bureau voor de Statistiek Suriname, 2014)**

	Meest gesproken taal		Tweede taal		Meest gesproken of tweede taal
	<i>n huishouden</i>	%	<i>N</i>	%	<i>Totaal %</i>
Nederlands	69.715	49,7	33.883	24,1	73,8
Sranan(tongo)	11.802	8,4	50.143	35,7	44,1
Sarnámi	19.830	14,1	10.758	7,7	21,8
Javaans	6.531	3,9	7.436	5,3	9,2
Marrontalen ⁴	21.699	15,5	3.508	2,5	18
Andere talen	8.597	6,1	5.693	4,1	10,2
Geen tweede taal	-	-	25.163	17,9	17,9
Onbekend	3.193	2,3	3.783	2,7	5
Totaal	140.367	100	140.367	100	200

Van het Surinaams-Nederlands wordt gesteld dat het een endoglossisch standaardiseringsproces ondergaat, en dat er zich dus ‘van binnenuit’ een eigen, endogene standaardvariëteit van het Nederlands ontwikkelt, vooral door taalcontact met het Sranan (cf. Muysken 2017). Het bestaan van die variëteit wordt onder andere erkend door de Nederlandse Taalunie, waarvan Suriname sinds 2003 geassocieerd lid is. Binnen Suriname zelf getuigen codificeringsprojecten (zoals de woordenboeken van Van Donselaar 1989 en de Bies 2009) van endoglossische standaardisering, net als attitudedata die aantonen dat Surinamers – en jonge Surinamers in het bijzonder – het Surinaams-Nederlands verkiezen boven het Nederlands of Belgisch Standaardnederlands, op zowel prestige- als solidariteitsschalen (Ghyselen et al. 2022).

Wat niet duidelijk is, is wat de formele kenmerken van het Surinaams Standaardnederlands dan precies zijn en of we echt van een formeel standaardiseringsproces kunnen spreken. In de studie van Ghyselen, Grondelaers, Doerga Misier-Patadien & Balesar (2022) bevatten de beoordeelde Surinaams-Nederlandse audiofragmenten enkel accentkenmerken, geen lexicale, morfologische of syntactische Surinamismen. In welke mate lexicale of morfo-syntactische Surinamismen algemeen aanvaard worden, is dan ook voer voor discussie. Een complicerende factor in die discussie is dat evenmin duidelijk is welke lexicale of morfo-syntactische verschijnselen het label ‘Surinamisme’ verdienen. Zoals in de inleiding al werd gesignaleerd, zijn in beschrijvingen van het Surinaams-Nederlands wel lijstjes te vinden van typerende kenmerken (cf. de Kleine 2007; Muysken 2017; de Bies 2009;

Van der Sijs 2014) – waarin het dubitatieve *als* vaak opduikt – maar in welke mate die algemeen voorkomen in Suriname (en bijvoorbeeld ook in formeel taalgebruik), is op basis van het huidige onderzoek moeilijk te bepalen. Tot slot is onduidelijk of de endoglossische standaardisering in Suriname ook een formele homogenisering van het taalgebruik inhoudt. Neemt het zogenoemd ‘diffuse’ karakter van het Surinaams-Nederlands (Muysken 2017, 286) af? Om die vraag te kunnen beantwoorden, zijn *real of apparent-time* data (cf. Bailey et al. 1992) nodig, waarbij de intersprekervariatie doorheen de tijd bestudeerd wordt.

Kortom: willen we het pluricentrische karakter van het Nederlands en de manier waarop dat evolueert goed begrijpen, dan is er nood aan kwantitatief onderzoek naar het Surinaams-Nederlands (cf. oproep van de Kleine 2013, 855). Daar wil deze studie aan bijdragen met een multivariate analyse van het dubitatieve *als*.

3. Opzet

Dat het kwantitatieve onderzoek naar de formele kenmerken van het Surinaams-Nederlands nog grotendeels in de kinderschoenen staat, is aan diverse factoren toe te schrijven, waaronder het gebrek aan een grootschalig corpus van gesproken en geschreven Surinaams-Nederlands. Sinds 2013 bevat het Corpus Hedendaags Nederlands (CHN)⁵ echter een component Surinaams-Nederlands, en dat biedt mogelijkheden voor een kwantitatieve analyse van het dubitatieve *als-of*-gebruik in Surinaams-Nederlandse schrijftaal. Aangezien voor het gesproken Surinaams-Nederlands geen systematisch corpus voor handen is, maakte ik in 2020 zelf opnames in Suriname, met een kleinschalig ‘Corpus Gesproken Surinaams-Nederlands’ (CGSN) als resultaat. In wat volgt, worden beide corpora eerst toegelicht, om vervolgens in te gaan op de data-annotatie en -analyse.

3.1. Het Corpus Hedendaags Nederlands (CHN) en het Corpus Gesproken Surinaams-Nederlands (CGSN)

Het Corpus Hedendaags Nederlands (CHN) is een steeds groeiende tekstverzameling van inmiddels meer dan 2.700.000 teksten, goed voor zo’n miljard woorden uit kranten, tijdschriften, journaaluitzendingen, blogs en boeken. Het CHN is automatisch taalkundig verrijkt met woordsoort- en lemma-informatie en bevat ook de nodige metadata (titel, auteur (soms), taal, publicatiejaar, medium, land van herkomst), wat gerichte zoekacties toelaat. Sinds de tweede uitgave bevat het corpus ook Surinaams-Nederlandse

teksten (grotendeels krantentaal), op dit moment goed voor 6,3% van het corpus (66.084.678 woorden). Het corpus wordt beschikbaar gesteld via het Instituut voor de Nederlandse Taal en is toegankelijk voor iedereen met een CLARIN-account.⁶ Het CHN biedt zo veel mogelijkheden, maar tegelijkertijd heeft het ook een aantal beperkingen die in het achterhoofd gehouden moeten worden. Zo laat de dataset op dit moment moeilijk toe om de sociale conditionering van specifieke taalkenmerken te onderzoeken (zoals de impact van de leeftijd, het geslacht, en de etniciteit van de taalgebruiker) en biedt het voorlopig enkel inzicht in het geschreven taalgebruik van Surinaamse taalspecialisten (journalisten en auteurs) – niet in dat van ‘doorsnee’ Surinamers en niet in spreektaal.

Om het hedendaagse gesproken Surinaams-Nederlands te kunnen onderzoeken, maakte ik in 2020 opnames bij 22 Creoolse vrouwen⁷ die opgegroeid zijn in het district Paramaribo en het Nederlands als ‘moedertaal’ hebben, wat wil zeggen dat het de belangrijkste voertaal was in de communicatie met minimaal één ouder in het gezin. Ongeveer de helft van de sprekers was op het moment van de opnames tussen 25 en 37 jaar oud (n=10, geboren tussen 1995 en 1983); de andere helft was tussen 50 en 60 jaar oud (n=11, geboren tussen 1970 en 1960). Eén respondent, die 42 jaar oud was op het moment van de opnames (geboren in 1978), valt tussen die twee leeftijdscategorieën in. Alle sprekers hadden een vorm van hoger onderwijs gevolgd (hoger beroepsonderwijs of universiteit), maar geen taalgerelateerde opleiding (dus geen lerarenopleiding, journalistiek, taal- en/of letterkunde, vertalen of tolken, ...). Wie meer dan een jaar in Nederland of België had gewoond, kwam niet in aanmerking voor het onderzoek.

Bij iedere respondent werden vijf opnames gemaakt. Eerst werd de sprekers gevraagd (i) een gesprek van 60 minuten op te nemen met een creoolse vriendin van ongeveer hun eigen leeftijd en vervolgens ook (ii) een gesprek van 60 minuten met een vriendin met een andere, niet-creoolse achtergrond (ook hier weer van dezelfde generatie). Bij die opnames was de onderzoeker niet aanwezig, om de opnames zo spontaan en informeel mogelijk te laten verlopen. De respondenten kregen ook de instructie te “spreken zoals ze normaal met de vriendin in kwestie zouden spreken”. Er werd dus geen taal(variëteit) opgelegd; respondenten konden switchen naar Sranan of een andere taal als dat natuurlijk aanvoelde. Eens die twee opnames afgerond waren, vond er (iii) een sociolinguïstisch interview plaats (30 à 60 minuten), afgenomen door de auteur van dit artikel. Tijdens dat interview werd onder andere gepraat over het taalgebruik in de opnames en de visie van de respondent op taalvariatie en meertaligheid in Suriname. Ook hier werd geen taal(variëteit) opgelegd. Na het interview legden de respondenten nog (iv)



Figuur 2: Stimuli in de grammaticatest om afhankelijke *ja/nee*-vraagzinnen te eliciteren.

een grammaticatest af (8 à 10 minuten), waarin gevraagd werd 20 vragen te beantwoorden op basis van een afbeelding. Daarbij werden ze verzocht te antwoorden met één (volledige) zin “in het Surinaams Standaardnederlands”. In deze opdracht werd dus wel heel expliciet om Standaardnederlands gevraagd. Soms werd een deel van de zin al gesuggereerd, en moesten de respondenten aanvullen. Het doel hier was om een aantal grammaticale taalkenmerken te eliciteren die in de literatuur vermeld zijn als ‘typisch Surinaams’. Figuur 2 toont de twee stimuli die in de grammaticatest opgenomen werden om afhankelijke *ja/nee*-vraagzinnen te eliciteren. Tot slot kregen de respondenten nog (v) een voorleestaak waarin ze 92 zinnen (grotendeels nonsenszinnen, gebaseerd op Van Hout et al. 1999) moesten voorlezen “in het Surinaams Standaardnederlands”. De bedoeling van deze opdracht was om data te verzamelen waarmee de uitspraak van het Nederlands in Suriname akoestisch bestudeerd kan worden. Op deze datacomponent wordt in wat volgt niet verder ingegaan. Alle sprekers in het onderzoek – niet enkel de 22 ‘hoofdsprekers’, maar ook de 40 gesprekspartners – gaven een *informed consent* en vulden een vragenlijst in waarin informatie werd opgevraagd over hun sociolinguïstische achtergrond.

Alle gesprekken werden orthografisch getranscribeerd met het programma ELAN volgens een licht aangepaste versie van het protocol dat bij het Corpus Gesproken Nederlands werd gehanteerd (Goedertier & Goddijn 2000). Van de vriendengesprekken werden telkens 40 minuten getranscribeerd; de sociolinguïstische interviews en de grammatica- en uitspraaktests werden helemaal getranscribeerd. Het resultaat is een corpus – dat ik bij dezen het ‘Corpus Gesproken Surinaams-Nederlands’ wil dopen – van 48 uur getranscribeerde spraak, goed voor 301.630 woorden (cf. Tabel 2).

Tabel 2: Statistieken Corpus Gesproken Surinaams-Nederlands

	Aantal minuten opgenomen	Aantal minuten getranscribeerd	Aantal woorden in transcripties
Grammaticatest	120	120	8.675
Voorleestaak	148	148	12.093
Sociolinguïstisch interview	976	976	116.220
Gesprek met Creoolse vriendin	1.336	814	89.558
Gesprek met niet- Creoolse vriendin	1.278	819	75.084
Totaal	3.858 min (64,3 u)	2.877 min (48u)	301.630

Het grote voordeel van het corpus is dat het (i) toelaat stilistische intrasprekervariatie te onderzoeken, (ii) veel metadata bevat over de sociolinguïstische achtergrond van de sprekers en (iii) ook leeftijdsgebonden patronen aan het licht kan brengen, die op hun beurt op taalverandering kunnen wijzen. De dataset heeft echter ook beperkingen. Het corpus bevat op dit moment bijvoorbeeld data van ‘maar’ 62 sprekers (22 hoofdrespondenten en 40 gesprekspartners) en de hoofdrespondenten hebben bovendien allen hetzelfde geslacht, dezelfde opleidingsgraad, moedertaal en etniciteit. Het taalgebruik in de dataset is dan ook niet representatief voor de variabiliteit in de Surinaamse taalgemeenschap, al biedt het wel een mooi vertrekpunt om het Surinaams-Nederlands te beschrijven. Voor Creoolse sprekers uit Paramaribo werd geselecteerd omdat die etnische groep al het langst Nederlands zou hebben als moedertaal (naast het Sranan), en zo de sterkste invloed zou uitoefenen op de vorming van het Surinaams-Nederlands (de Kleine 2013, 849). De keuze voor hoogopgeleide vrouwen werd gestuurd door het idee dat die vaak de voortrekkers van taalverandering zijn in standaardiseringsprocessen (cf. Labov 1990). Het is de hoop dat het corpus in de toekomst uitgebreid kan worden met data van sprekers uit andere sociale en etnische groepen.

3.2. Data-annotatie

Vanuit een interesse voor de syntactische kenmerken van bijzinnen in het gesproken Surinaams-Nederlands (o.a. woordvolgorde, voegwoorden, ...) werden alle bijzinnen gerealiseerd door de 22 hoofdrespondenten in het Corpus Gesproken Surinaams-Nederlands manueel uit het corpus gefilterd. Daarbij werden de afhankelijke *ja/nee*-vraagzinnen – 146 werden

er geattesteerd – als afzonderlijk type gemarkeerd. Die werden vervolgens op 11 factoren geannoteerd:

- (i) het gerealiseerde voegwoord (*als* versus *of*, of uitzonderlijk, in gevallen van nevenschikking met samentrekking, ook wel eens géén voegwoord),
- (ii) sprekergebonden kenmerken:
 - a. de spreker,
 - b. de leeftijd van de spreker,
 - c. de moederta(a)l(en) die opgegeven werd(en) door de spreker (Nederlands versus Nederlands en Sranan),
 - d. de zelfingeschatte kennis van het Sranan (op een schaal van 0 tot 7),
 - e. de zelfingeschatte kennis van het Engels (op een schaal van 0 tot 7),
- (iii) het type gesprekssituatie (intra-etnisch vriendengesprek, inter-etnisch vriendengesprek, interview versus grammaticatest),
- (iv) de eventuele combinatie met een nevenschikkend voegwoord (niet na nevenschikkend voegwoord versus wel na nevenschikkend voegwoord),
- (v) het werkwoord van de dominerende zin (in wat volgt: het matrixwerkwoord),
- (vi) de polariteit van de matrixzin (negatief versus positief),
- (vii) de frequentie van het matrixwerkwoord in de Surinaamse component van het Corpus Hedendaags Nederlands⁸,
- (viii) de grammaticale functie van de afhankelijke zin in de dominerende zin (subject versus object versus nabepaling in een voorzetselconstituent).

Factoren (ii) en (iii) laten ons toe om inter- en intrasprekervariatie in het *als/of*-gebruik in kaart te brengen en die eventueel aan de leeftijd of taalachtergrond van de spreker te linken of aan het type situatie waarin de spreker zich bevindt. Aangezien het voegwoord *als* soms aan invloed van het Sranan en het Engels wordt toegeschreven (cf. sectie 1), lijkt het logisch kennis van beide talen mee te nemen in de analyses. Wat factor (iv) betreft, weten we dat er bij nevenschikking met *of* in de Europese standaardvariëteiten soms *dat* verschijnt in plaats van *of* in de tweede (of derde of daarop volgende) deelzin (cf. inleiding en ANS 10-3-2-1). Het is dan ook relevant deelzinnen die na een nevenschikkend voegwoord verschijnen als een aparte categorie te behandelen. Factoren (v)-(viii) worden meegenomen omdat die in kwantitatief onderzoek

naar andere types voegwoordvariatie – en dan vooral bij studies naar de *dat-zero*-alternantie bij afhankelijke mededelende zinnen (zie bvb. Shank, Van Bogaert, & Plevaets 2016; Kruger 2019) – als relevant naar voren kwamen.

Voor het Corpus Hedendaags Nederlands was een manuele filtering gezien de grotere omvang van de dataset niet haalbaar. Hier werd via een *Corpus Query Language*-commando gezocht op de onderschikkende voegwoorden *als* en *of*, met als voorwaarde dat er aan het voegwoord een werkwoord vooraf moest gaan dat in het Corpus Gesproken Surinaams-Nederlands met een afhankelijke *ja/nee*-vraagzin werd geattesteerd (*aangeven, afhangen, afvragen, begrijpen, bepalen, checken, controleren, kijken, nadenken, nagaan, opletten, spreken, twijfelen, uitmaken, vragen, weten, zien of gaan*)⁹ – met maximaal vier tokens tussen werkwoord en voegwoord. Tussen werkwoord en onderschikkend voegwoord mocht er ook geen ander voegwoord staan en geen leesteken.¹⁰ Voor elk zoekresultaat werd achteraf manueel geverifieerd of het ook effectief om een dubitatieve bijzin ging.

Met de zoekopdracht werd getracht een evenwicht te vinden tussen volledigheid en haalbaarheid. Met het werkwoords criterium en het afstandscriterium van vier tokens glipten sowieso relevante constructies door de mazen van het net. Zonder die criteria zou het manuele verificatiewerk achteraf echter te arbeidsintensief zijn geweest. Voor de scheidbaar samengestelde werkwoorden leverde de zoekformule ook alleen maar de ongescheiden vormen op; de delen van gescheiden vormen krijgen in het CHN immers elk een lemma. Zoeken op combinaties van de werkwoordelijke en niet-werkwoordelijke delen van scheidbare werkwoorden zou evenwel opnieuw te veel manueel verificatiewerk vergen. De huidige zoekactie resulteerde in 15.777 hits, waarvan er 7.772 (49%) dubitatieve afhankelijke bijzinnen bleken te zijn.

De afhankelijke vraagzinnen werden vervolgens geannoteerd op dezelfde factoren als de zinnen in de gesproken dataset. De sprekers-/schrijversgebonden kenmerken konden echter door een gebrek aan metadata niet geannoteerd worden. De functie van de afhankelijke zin en de eventuele combinatie met een nevenschikkend voegwoord werden evenmin geannoteerd aangezien de zoekopdracht enkel objectszinnen opleverde en bij nevenschikking enkel de eerste deelzin weergaf. Tot slot werd het ‘type gespreks situatie’ voor de geschreven dataset vervangen door het ‘type tekst’ (krantenartikel – boek – website – ‘geschreven om gesproken te worden’),

3.3. Data-analyse

Aangezien er voor het *als/of*-gebruik vrij veel potentieel relevante predictoren zijn, terwijl sommige delen van de dataset (vooral voor het gesproken Surinaams-Nederlands) weinig observaties tellen, werden de data geanalyseerd

aan de hand van *conditional random forest modelling* en *conditional inference tree modelling* (Levshina 2020; Tagliamonte & Baayen 2012). Deze *machine-learning* technieken trachten met recursieve partitie de waarde van een responsvariabele te voorspellen op basis van de verschillende predictoren.

Als eerste stap wordt bij recursieve partitie gezocht naar de predictor die het meest helpt om de respons te voorspellen. Daarbij wordt met permutatie gewerkt, zodat de data niet – in tegenstelling tot bijvoorbeeld bij lineaire regressie – aan een specifieke statistische verdeling hoeven te beantwoorden. Permutatie houdt in dat je de observaties van de afhankelijke variabele in de dataset willekeurig van volgorde verandert, terwijl de waarden van de predictoren niet meegepermuteerd worden. De informatie over de predictoren wordt dus losgekoppeld van de afhankelijke variabele. Vervolgens gaat het algoritme na of de associatie tussen de afhankelijke variabele en een specifieke predictor in de oorspronkelijke dataset groter is dan die in de gepermuteerde dataset. Hoe groter het verschil tussen de associaties in de oorspronkelijke en de gepermuteerde dataset, hoe groter de voorspellingskracht van de predictor.

Eens bepaald is welke predictor de grootste voorspellingskracht heeft, wordt de dataset in een tweede stap op basis van die predictor in tweeën gesplitst. Dat gebeurt vrij willekeurig; het algoritme beslist dan bijvoorbeeld – laat ons even bij wijze van denkoefening veronderstellen dat de gesprekssituatie bij onze data het grootste effect zou hebben – om de data op te splitsen in interviewdata enerzijds en de data van de andere types gesprekken anderzijds. De vraag luidt dan: helpt deze opsplitsing om de gevallen van *als* en die van *of* duidelijker van elkaar te scheiden of is een andere opsplitsing (in bijvoorbeeld grammaticatests enerzijds en andere gesprekssituaties anderzijds) meer zinvol? Het algoritme bepaalt de meest zinvolle splitsing (op basis van verschillende criteria) en gaat vervolgens verder met het opsplitsen van een of beide subdatasets, om na te gaan of het bijvoorbeeld relevant is nog fijnmazigere opsplitsingen te maken in de dataset volgens de opgegeven predictor. Als de opsplitsing het voorspellingssucces niet verbetert, gaat het algoritme terug naar stap 1 om op zoek te gaan naar een andere predictor die de overgebleven variantie kan helpen te verklaren. Het algoritme herhaalt stappen 1 en 2 tot verder splitsen niet meer zinvol is. Het resultaat is een beslissingsboom (*conditional inference tree*) waarin het belang van de verschillende predictoren en de manier waarop ze interageren schematisch is gevisualiseerd.

Een *random forest* analyse gaat een stapje verder. Deze *ensemble*-methode genereert verschillende aparte bomen, telkens op basis van een willekeurige steekproef van de oorspronkelijke dataset en van de oorspronkelijke

predictoren. Het basisidee is dat patronen die in veel bomen naar voren komen robuust zijn. Per predictor wordt een *conditional permutation-based variable importance score* berekend, die het voorspellende belang van de predictor weerspiegelt, rekening houdend met alle andere predictoren en de manier waarop ze interageren. Hierbij wordt een gemiddelde genomen van de verschillende bomen in de analyse. Doordat per boom telkens andere combinaties van predictoren in rekening worden gebracht, zouden de effecten van die predictoren met een *random forest* beter in kaart gebracht worden dan met een individuele *conditional inference tree* (Levshina 2020, 616-617; Strobl, Malley, & Tutz 2009). Voor mathematische details verwijst ik graag naar Levshina (2020) en Strobl et al. (2009).

4. Resultaten

In wat volgt, bespreek ik eerst de resultaten voor het Corpus Gesproken Surinaams-Nederlands, om vervolgens in te gaan op die voor het geschreven Surinaams-Nederlands.

4.1. Als vs. of in het Corpus Gesproken Surinaams-Nederlands

In totaal werden er in het gesproken corpus 161 afhankelijke vraagzinnen geattesteerd, waarvan er 43,5% met *als* werd gerealiseerd en 54,0% met *of*; in vier nevengeschikte deelzinnen (2,5%) werd het voegwoord samengetrokken. Data-inspectie voorafgaand aan de statistische modellering leidde tot de conclusie dat noch de combinatie met een nevenschikkend voegwoord noch de functie van de deelzin (subject versus object versus nabepaling in een voorzetselconstituent) meegenomen konden worden als predictor. Het corpus telde immers maar 14 nevengeschikte zinnen, wat te weinig is voor een zinvolle statistische analyse. Ook op het vlak van de grammaticale functie van de deelzinnen was er te weinig variatie in de dataset: 98% van de geattesteerde vraagzinnen waren voorwerpszinnen (cf. voorbeeldzin 5); de dataset telde maar 2 subjectzinnen (cf. voorbeeldzin 6) en 1 nabepaling bij het woord *ongeacht* (cf. voorbeeldzin 7). Om die reden werden enkel de objectzinnen meegenomen in de statistische analyse.

(5) *Doe hem open en zie of er een rat in is.*

(6) *Maar als het elke dag lukt is de vraag.*

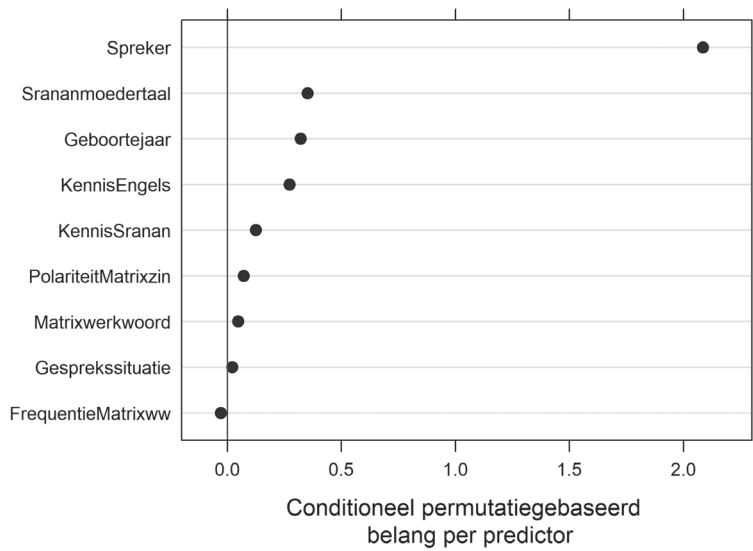
(7) *Ongeacht als mijn ouders Surinaams spraken was ik verplicht Nederlands te praten.*

Tabel 3: Combinaties van nevenschikkende voegwoorden en dubitatieve voegwoorden in het Corpus Gesproken Surinaams-Nederlands

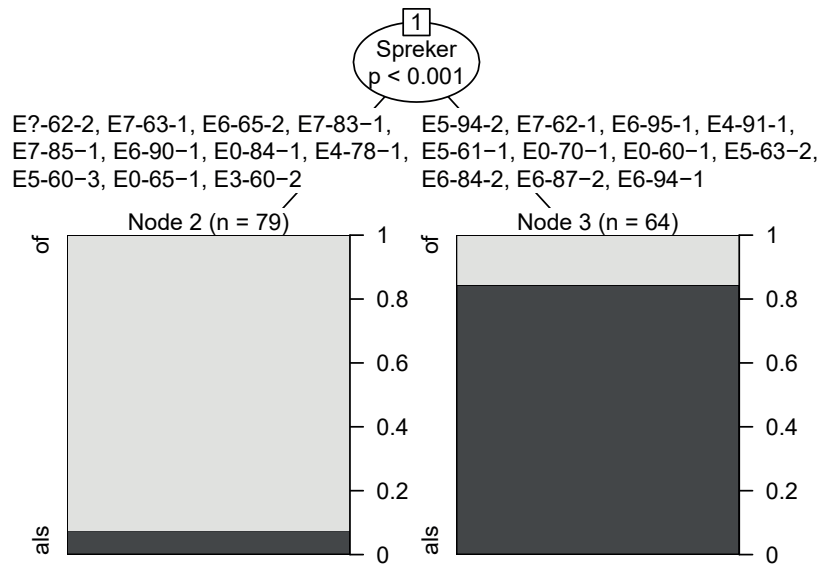
Nevenschikkend voegwoord	Onderschikkend voegwoord	Aantal	Voorbeeldzin uit corpus
Nevenschikking met voegwoord en	<i>Als</i> in zowel bijzin 1 als 2	1	<i>Ik weet niet als het al is gebeurd en als ze het hebben geapprecieerd.</i>
	<i>Of</i> in zowel bijzin 1 als 2	1	<i>Ze vroeg de leerkracht of de oefening juist is en of de oefening fout is.</i>
Nevenschikking met voegwoord of	<i>Als</i> in zowel bijzin 1 als 2	1	<i>Hangt ervan af als jij hebt gereden of als je moeder heeft gereden.</i>
	<i>Als</i> in bijzin 2, maar niet in 1	4	<i>Is gewoon even een moment om te weten of het nog in mij zit of als het nog kan.</i>
	<i>Of</i> in zowel bijzin 1 als 2	1	<i>Ik weet niet of de manier van het uit... van bepaalde woorden anders is of... of het accent misschien anders is.</i>
	Samentrekking van voegwoord in bijzin 2	4	<i>Dan kan je aan de hand hiervan kijken van... of het hetzelfde was of wat anders is.</i>
Nevenschikking zonder voegwoord	<i>Als</i> in zowel bijzin 1 als 2	2	<i>Je kan eerst komen kijken als je 't leuk vindt als het iets voor jou is.</i>

Een kwalitatieve inspectie van de nevenschikte zinnen levert wel een aantal interessante inzichten op. Tabel 3 toont hoe bij nevenschikking het dubitatieve voegwoord in de regel hetzelfde is in alle deelzinnen, met 64% van de zinnen dat in de eerste deelzin *of* krijgt en 36% *als*. Het voegwoord wordt dan ook soms samengetrokken in de tweede of derde deelzin. Enkel bij de nevenschikking met *of* verschijnt in de tweede of derde deelzin soms *als* – terwijl in de eerste deelzin *of* werd gebruikt – om een botsing van nevenschikkend *of* en dubitatief *of* te vermijden (n=4, cf. Tabel 3). Interessant is dat de vermijdingsstrategie hier anders is dan in het Europees-Nederlands: terwijl volgens de ANS in het Nederlands- en Belgisch-Nederlands de combinatie *of dat* verschijnt (cf. voorbeeldzin 2 in de inleiding), en soms ook het archaischere *dan wel of*, attesteren we in de Surinaamse data enkel *of of* (met een pauze tussen de twee *of*s) en vooral *of als* (cf. Tabel 3). Aangezien de keuze voor *of* versus *als* in de tweede of derde deelzin bij nevenschikte *ja/nee*-vraagzinnen veel minder vrij is dan in de eerste deelzin, nemen we die tweede en derde deelzinnen niet mee in de verdere statistische analyses.

In de c-tree- en random forest-analyse gaan we na in welke mate het type gesprekssituatie, de spreker, de leeftijd en taalachtergrond van de spreker, het matrixwerkwoord, de frequentie van het matrixwerkwoord en de polariteit van de matrixzin de keuze voor *als* dan wel *of* beïnvloedt. Figuur 3 toont



Figuur 3: Conditioneel permutatiegebaseerd belang per predictor voor de variatie tussen *als* en *of* in afhankelijke *ja/nee*-vragen in het Corpus Gesproken Surinaams-Nederlands (enkel voorwerpszinnen)

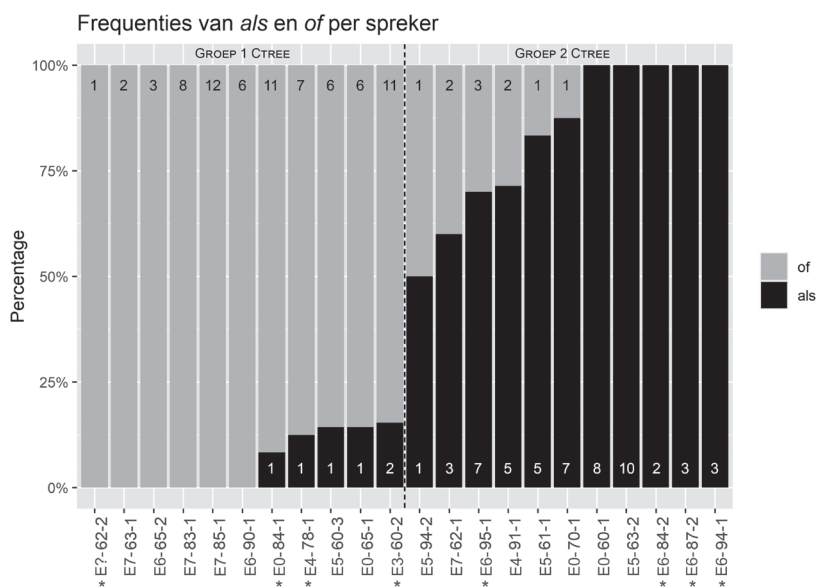


Figuur 4: Conditional inference tree voor de variatie tussen *als* en *of* in afhankelijke *ja/nee*-vragen in het Corpus Gesproken Surinaams-Nederlands

de *conditional permutation-based variable importance scores* die op basis van een random forest analyse¹¹ werden berekend voor de objectszinnen in onze dataset. Met een C-waarde van 0.97 heeft het model een duidelijk voorspellende kracht.¹² Opvallend is dat de predictor 'spreker' met een conditional importance score van 2,127 het sterkst doorweegt in het model. Daarna volgen – op ruime afstand – de predictoren 'Srananmoedertaal', 'Geboortelaar' en 'KennisEngels' met scores van respectievelijk 0,351, 0,307 en 0,319. De andere predictoren – de ingeschatte kennis van het Sranan, het matrixwerkwoord, de polariteit van de matrixzin, de gesprekssituatie en de frequentie van het matrixwerkwoord – leunen dicht aan bij de nullijn die 'geen belang' aangeeft.

Kijken we naar het resultaat van de *conditional inference tree* analyse¹³ (cf. Figuur 4, C=0,89), dan komt enkel spreker als significante predictor naar voren. Mijn aanvankelijke vermoeden was dat het effect van spreker – een heel fijnmazige variabele – hier het effect van de andere factoren (zoals leeftijd of moedertaal) verdoezelt. Herhalen we de random tree analyse echter zonder de predictor 'spreker', dan nog komen geen andere predictoren in beeld. Het resultaat is dan een boom zonder takken, met een C-waarde van 0,5. Ook wanneer het significantieniveau naar beneden wordt gehaald – van 0.05 naar 0.01 – blijven de resultaten ongewijzigd. Conclusie: de variatie tussen *als* en *of* is in de gesproken dataset voornamelijk individueel bepaald.

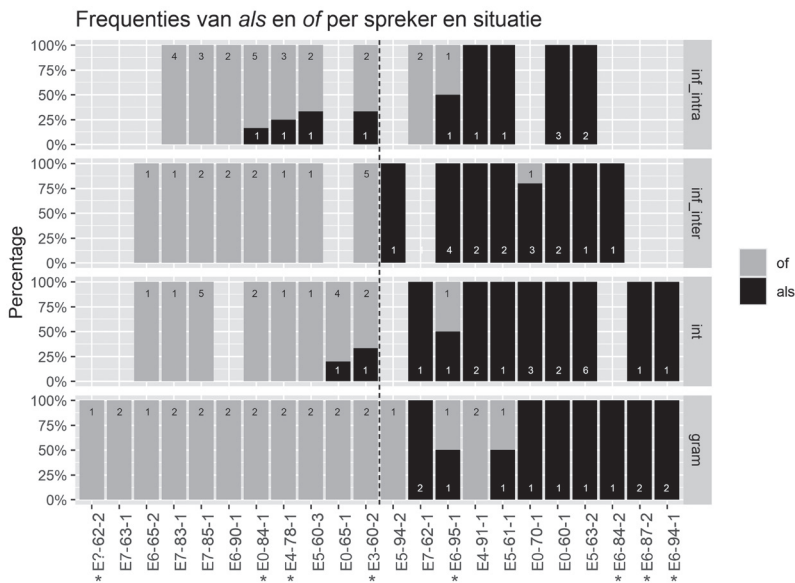
De inference tree toont twee groepen sprekers: (1) sprekers E?-62-2, E7-63-1, E6-65-2, E7-83-1, E7-85-1, E6-90-1, E0-84-1, E4-78-1, E5-60-3, E0-65-1 en E3-60-2, die heel sterk naar het voegwoord *of* neigen, en (2) sprekers E5-94-2, E7-62-1, E6-95-1, E4-91-1, E5-61-1, E0-70-1, E0-60-1, E5-63-2, E6-84-2, E6-87-2 en E6-94-1, die veel frequenter het voegwoord *als* realiseren. Figuur 5 toont hoe slechts de helft van de sprekers in de opnames varieerde tussen *als* en *of*; bij zes sprekers uit groep 1 werd enkel het voegwoord *of* geattesteerd, terwijl we bij vijf sprekers in groep 2 enkel het voegwoord *als* aantreffen. De vraag rijst of er factoren zijn die de variatie tussen de sprekers kunnen verklaren. De random forest analyse suggereert een bescheiden rol voor het geboortelaar van een spreker, het hebben van Sranan als moedertaal naast het Nederlands en de zelfingeschatte kennis van het Engels, maar die factoren komen in de conditional inference tree niet aan bod. Het is ook moeilijk in het ruwe cijfermateriaal in Figuur 5 – waarin het cijfer dat volgt op 'E' in de sprekerscode het zelfingeschatte niveau van Engels weergeeft (0=geen kennis, 7='vloeiend'), het middelste cijfer het geboortelaar representeert en een asterisk de sprekers markeert die aangaven naast Nederlands



Figuur 5: Absolute en relatieve frequenties van *als* en *of* in voorwerpszinnen per spreker in het Corpus Gesproken Surinaams-Nederlands. De stippellijn markeert de tweedeling die gemaakt werd in de conditional inference tree analyse; de asterisken markeren sprekers die naast Nederlands ook Sranan opgaven als moedertaal.

Sranan als moedertaal te hebben – leeftijds- of taalgerelateerde tendensen te ontwaren. In de ordening van sprekers van laag naar hoog *als*-gebruik wisselen sprekers van alle leeftijden, niveaus van Engels en moedertaalachtergronden elkaar immers af. Een kwalitatieve analyse van het interviewmateriaal – waarbij ik onder andere naging in welke mate de sprekers aangaven op Europa georiënteerd te zijn of veel familie in Europa te hebben – brengt evenmin factoren aan het licht die de sprekergebonden variatie kunnen verklaren.

Hoewel gesprekssituatie in de random forest en conditional inference tree analyses niet naar voren kwam als significante predictor, blijkt bij stratificatie van het ruwe cijfermateriaal per spreker en gesprekssituatie (cf. Figuur 6) wel dat er een implicatiele structuur in het *als*-gebruik te ontdekken valt: wie in de grammaticatest – waarin expliciet om ‘Surinaams Standaardnederlands’ werd gevraagd – het voegwoord *als* gebruikte (9 van de 22 sprekers), realiseerde het voegwoord sowieso ook in één of meer van de spontanere opnames (interview of vriendengesprekken). Omgekeerd geldt evenwel niet dat wie in de spontane gesprekken *als* hanteert, dat ook in de grammaticatest doet.

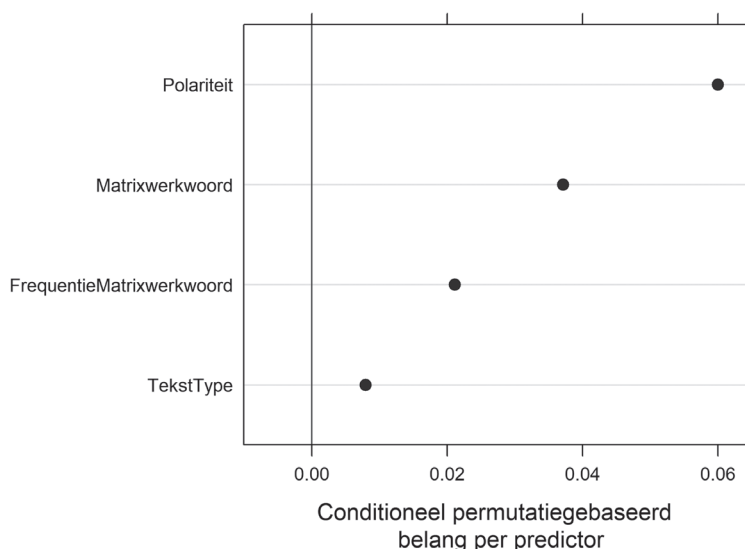


Figuur 6: Absolute en relatieve frequenties van *als* en *of* in voorwerpszinnen per spreker en gesprekssituatie in het Corpus Gesproken Surinaams-Nederlands. De stippellijn markeert de tweedeling die gemaakt werd in de conditional inference tree analyse; de asterisken markeren sprekers die naast Nederlands ook Sranan opgaven als moedertaal.

4.2. *Als* vs. *of* in het Corpus Hedendaags Nederlands

De zoekopdracht in de Surinaamse component van het Corpus Hedendaags Nederlands leverde 7722 afhankelijke *ja/nee*-vragen op, waarvan er 547 (7.1%) door het voegwoord *als* worden ingeleid en 7175 (92.9%) door het voegwoord *of*. Het voegwoord *als* staat met andere woorden in de geschreven dataset een heel stuk minder sterk dan in het Corpus Gesproken Surinaams-Nederlands.

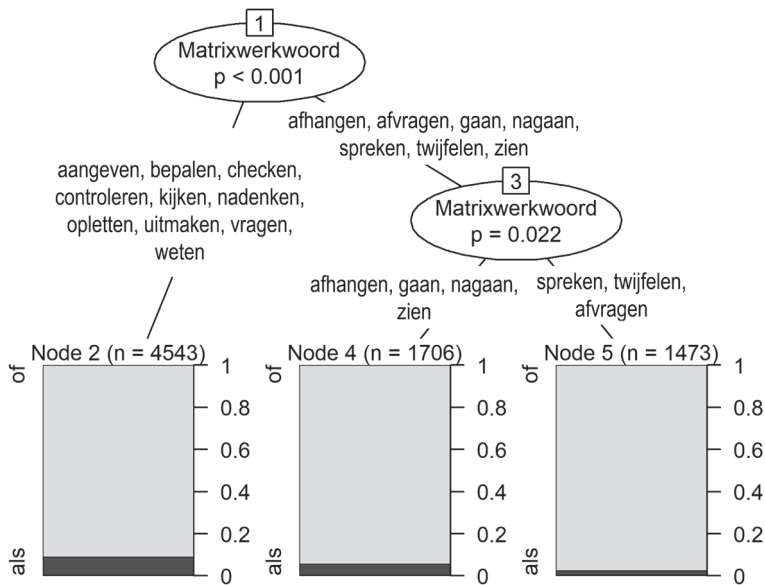
Een random forest analyse met het matrixwerkwoord, de frequentie van het matrixwerkwoord, de polariteit van de matrixzin en het teksttype als predictoren toont aan dat de polariteit van de zin de belangrijkste predictor is (conditioneel permutatiegebaseerd belang = 0,060, cf. Figuur 7), gevolgd door het matrixwerkwoord (0,037), de frequentie van het matrixwerkwoord (0,021) en het teksttype (0,008).¹⁴ De C-waarde van het model is echter vrij laag (0.64), wat aangeeft dat de opgenomen predictoren de voegwoordvariatie maar voor een klein deel kunnen verklaren. Om de richting van de effecten te kunnen interpreteren, kijken we in eerste instantie naar de conditional inference tree (cf. Figuur 8). Daar komt echter enkel het matrixwerkwoord naar voren als significante predictor; de polariteit van de zin en de frequentie van het



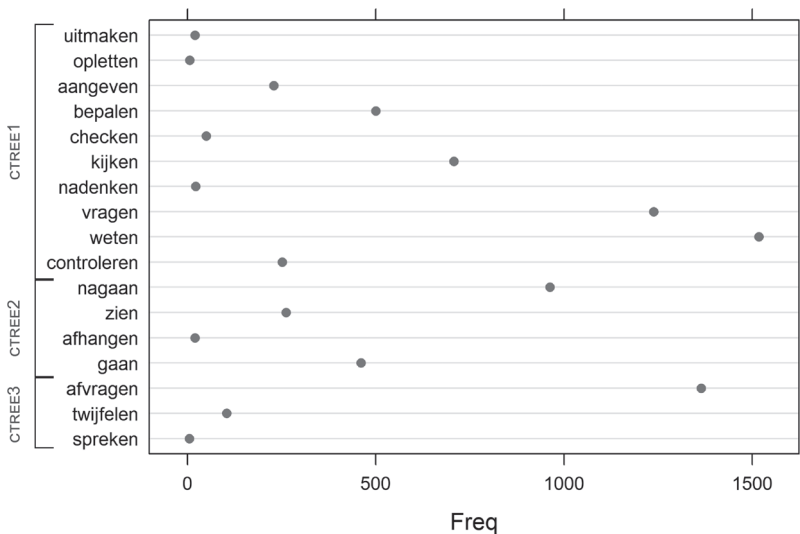
Figuur 7: Conditioneel permutatiegebaseerd belang per predictor voor de variatie tussen *als* en *of* in afhankelijke *ja/nee*-vragen in de Surinaamse component van het Corpus Hedendaags Nederlands (enkel voorwerpszinnen)

matrixwerkwoord blijken hier geen relevante predictoren. De grootste kans op *als* vinden we in de boom bij de werkwoorden *aangeven*, *bepalen*, *checken*, *controleren*, *kijken*, *nadenken*, *opletten*, *uitmaken*, *vragen* en *weten*, een wat lagere kans bij *afhangen*, *gaan*, *nagaan* en *zien* en de laagste kans is er ten slotte bij *afvragen*, *spreken* en *twijfelen*. In die werkwoordsindeling valt niet meteen een semantisch patroon te herkennen. De categorieën lijken evenmin samen te hangen met de algemene frequentie van de werkwoorden in de Surinaamse component van het Corpus Hedendaags Nederlands: Figuur 9 toont hoe in elk van de groepen zowel hoog- als laagfrequente werkwoorden te vinden zijn. Ook bij de ctree is de C-waarde echter erg laag (0.596), wat er opnieuw wijst dat het ctree-model net als de random forest weinig verklarende kracht heeft. Bekijken we de ‘ruwe data’ uitgesplitst per werkwoord, polariteit van de matrixzin en per teksttype (cf. Figuur 10), dan is geen duidelijk patroon te zien wat bijvoorbeeld polariteit en teksttype betreft; het effect van matrixwerkwoord – zoals dat in de ctree naar voren kwam – is wel te herkennen.

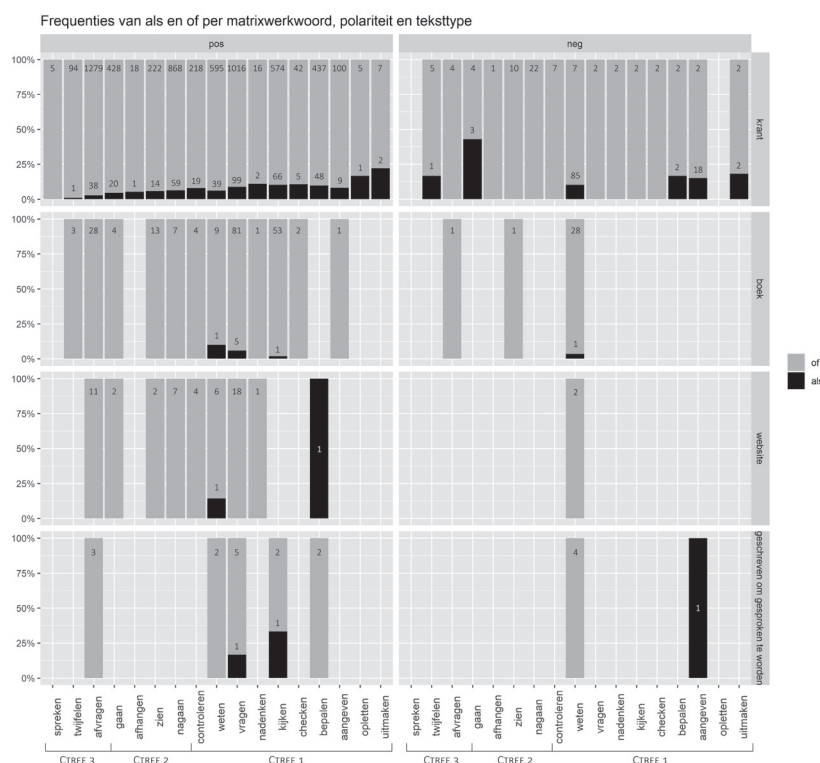
De lage C-waarden van de random forest en random tree voor de geschreven data geven aan dat de bestudeerde predictoren niet volstaan om de variatie



Figuur 8: Conditional inference tree voor de variatie tussen *als* en *of* in afhankelijke *ja/nee*-vragen in de Surinaamse component van het Corpus Hedendaags Nederlands



Figuur 9: Frequentie van de verschillende matrixwerkwoorden in de Surinaamse Component van het Corpus Gesproken Nederlands. De werkwoorden zijn op de y-as gerangschikt volgens hoe vaak ze bij afhankelijke *ja/nee*-vragen het voegwoord *als* krijgen (*uitmaken*: in 20% van de mogelijke gevallen met *als*; *spreken*: nooit met *als*). De haakstructuren markeren de clusters uit de conditional inference tree.



Figuur 10: Absolute en relatieve frequenties van *als* en *of* in voorwerpszinnen per matrixwerkwoord, polariteit van de matrixzin en teksttype in de Surinaamse component van het Gesproken Hedendaags Nederlands. De werkwoorden zijn op de x-as gerangschikt volgens hoe vaak ze globaal bij afhankelijke *ja/nee*-vragen het voegwoord *als* krijgen (*spreek*: nooit met *als*; *uitleggen*: in 20% van de mogelijke gevallen met *als*). De haken onderaan markeren de driedeling die gemaakt werd in de conditional inference tree analyse.

in het *als*-gebruik te verklaren. Mogelijk zijn er andere, hier niet meegenomen taalinterne predictoren die een sterkere verklarende kracht hebben, maar gezien de resultaten voor de gesproken data – en vooral de dominantie van ‘spreker’ daar als predictor – is het erg waarschijnlijk dat de keuze voor *als* ook in schrijftaal meer door individuele, taalgebruikergerelateerde factoren wordt bepaald dan door taalinterne factoren. In de ideale wereld maken we die hypothese hard aan de hand van een mixed-effects logistische regressie, een techniek die als voordeel heeft dat het systematisch de effecten van

‘willekeurige’ variabelen – zoals de individuele taalgebruiker en het werkwoord waarbij de *ja/nee*-zin verschijnt – in rekening kan brengen (Gries 2015).¹⁵ Een vergelijking van een model mét en zonder die willekeurige variabelen – en dan vooral de voorspellingskracht van die modellen – zou ons veel kunnen vertellen. De dataset is echter niet geschikt voor een dergelijke analyse, aangezien hij niet aan de voorwaarden voor mixed-effects modellering voldoet (cf. Tagliamonte & Baayen 2012). Eerst en vooral is er niet voor alle CHN-teksten auteursinformatie voorhanden (slechts voor 5,3% van de titels, en voor 6,9% van de observaties). Een analyse op enkel de data waarvoor auteursinformatie voorhanden is, is weinig zinvol wegens erg lage tokenaantallen. Als oplossing voor dat probleem werd overwogen om de teksttitel – die wel voor alle observaties beschikbaar is – als proxy te hanteren voor auteur. Daarbij rijst echter het probleem dat slechts een fractie van de teksten (15%) meer dan één *ja/nee*-zin bevat, en dat bemoeilijkt de modellering; de observaties zijn immers veeleer onafhankelijk dan ‘genest’ binnen de willekeurige variabele ‘titel’ (cf. Gries 2015). Ook hier biedt *subsetting* geen soelaas: selecteer je enkel de titels met meer dan twee observaties – pas dan zijn de resultaten van collineariteitsanalyses niet meer alarmerend – dan blijven er nog maar 55 zinnen met het voegwoord *als* over. De ‘limiting sampling size’ (Harrell 2015, 73), i.e. de frequentie van de minst voorkomende variant, is dan met andere woorden te beperkt voor een betrouwbare regressie-analyse.

5. Discussie & besluit

Aan het begin van dit artikel werden twee vragen vooropgesteld over het dubitatieve *als* in het Surinaams-Nederlands, namelijk (1) hoe vaak Surinamers de voegwoorden *als* en *of* realiseren in afhankelijke vraagwoordzinnen en (2) in welke mate het *als*-gebruik geconditioneerd wordt door taalinterne factoren, sociale factoren, de taalachtergrond van de taalgebruiker, het medium (spreek- versus schrijftaal) en/of de formaliteit van de situatie.

Wat vraag 1 betreft, kan geconcludeerd worden dat het dubitatieve *als* globaal genomen minder vaak voorkomt dan het dubitatieve *of*, maar dat de frequentie sterk medium- en individugebonden is. In de geschreven data – met vooral teksten van taalprofessionals – werd het dubitatieve *als* in 7% van de relevante zinnen geattesteerd. In het gesproken taalgebruik van 22 hoogopgeleide, maar niet talig gespecialiseerde Creoolse vrouwen lag het aandeel van *als* een stuk hoger, namelijk op 44%.

Wat de conditionering van het *als*-gebruik betreft (vraag 2), is de belangrijkste conclusie dat de variatie sterk persoonsgebonden is. Bij de

gesproken data kwam ‘spreker’ als allesoverheersende predictor naar voren. Opvallend daarbij was dat sommige sprekers (n=5) in alle relevante zinnen voor *als* kozen, anderen het nooit deden (n=6) en de helft (n=11) afwisselde tussen de twee vormen. Die intersprekervariatie bleek moeilijk te verklaren aan de hand van de gerapporteerde talenkennis van de sprekers, hun leeftijd of de mate waarin ze naar eigen zeggen in contact komen met het Europees-Nederlands. Het type gesprekssituatie (gesprek met vriendin van eigen etniciteit vs. met vriendin van andere etniciteit vs. met onbekende interviewer vs. grammatica-test) bleek evenmin een significante invloed uit te oefenen, al stelden we kwalitatief wel vast dat wie in de grammaticatest – waarin expliciet om ‘Surinaams Standaardnederlands’ werd gevraagd – het voegwoord *als* gebruikte, dat sowieso ook een of meerdere keren deed in de spontanere opnames. Bij de geschreven data kon de interauteurvariatie om methodologische redenen moeilijk becijferd worden, maar bleken modellen met vooral taalinterne variabelen (het matrixwerkwoord, de frequentie en de polariteit van het matrixwerkwoord) en het teksttype (krant vs. roman vs. website vs. geschreven om gesproken te worden) als predictoren weinig verklarende kracht te hebben. Het matrixwerkwoord had weliswaar een invloed – het werkwoord *uitmaken* komt bijvoorbeeld vaker voor met dubitatief *als* dan het werkwoord *twijfelen* – maar het effect was relatief klein.

Uiteraard moeten we bij deze conclusies de beperkingen van de datasets goed indachtig zijn. De gesproken dataset is bijvoorbeeld beperkt in omvang; met meer sprekers, meer sociale variatie in de sprekerselectie en meer afhankelijke vraagzinnen hadden mogelijk duidelijkere sociale, stilistische en intern-linguïstische patronen vastgesteld kunnen worden. De geschreven dataset telde veel meer observaties, maar door een gebrek aan metadata over de auteurs bleef daar de vraag onbeantwoord hoeveel variatie individueel bepaald was. De geobserveerde proporties van het dubitatieve *als* zijn bovendien niet noodzakelijk representatief voor heel Suriname. In de gesproken dataset hadden alle sprekers hetzelfde geslacht, dezelfde opleidingsgraad, moedertaal en etniciteit; in de geschreven dataset waren alle auteurs taalprofessionals. Het is goed denkbaar dat bij andere types sprekers en schrijvers het percentage dubitatief *als* een stuk hoger of lager zou liggen. In de toekomst zullen grootschaligere corpora van het gesproken en geschreven Surinaams-Nederlands, met voldoende sociale variatie en spreker-/schrijvergerelateerde metadata, hopelijk soelaas kunnen bieden. Bovendien moet ook meer aandacht besteed worden aan de invloed van culturele identiteit, taalbewustzijn en de taalattitudes op de keuze voor *als* dan wel *of*. Mogelijk identificeren sprekers die consequent voor *of* kiezen zich bijvoorbeeld meer met Europa, terwijl de *als*-gebruikers zich bewust

of onbewust afzetten tegen Nederlandse culturele normen. De culturele oriëntatie van de sprekers in het Corpus Gesproken Surinaams-Nederlands werd echter maar oppervlakkig bevraagd, en aan de mate waarin sprekers zich bewust zijn van het onderscheid tussen *als* en *of* werd evenmin aandacht besteed. Diepgaander kwalitatief vervolgonderzoek zal hier hopelijk meer inzicht in brengen, en zo ook helpen de geobserveerde intersprekervariatie te verklaren.

De insteek van dit artikel was dat kwantitatief onderzoek nodig is om “betrouwbare uitspraken [te] doen over de status van eventuele nationale varianten die daar [in Suriname] tot de standaardtaal zouden behoren” (inleiding e-ANS₃). De hier gepresenteerde analyses wijzen er alvast op dat de status van het dubitatieve *als* in Suriname voorlopig complex is, hoe we het begrip ‘standaardtaal’ ook invullen. Benaderen we het begrip vanuit productieperspectief, en definiëren we standaardtaal als het taalgebruik dat gangbaar is in formele situaties (cf. Auer 2011, 490), dan kunnen we stellen dat dubitatief *als* deel van de standaard is – het kwam voor in formelere interviews én in krantentaal – maar dat het slechts een perifere positie inneemt en meer typerend is voor gesproken dan geschreven taal. Volgens de maatstaf van de taaladvieswebsite *Taaladvies.net* verdient het dubitatieve *als* in Suriname op basis van de hier geanalyseerde data het eindoordeel ‘status onduidelijk’. De website, die ontwikkeld is door de Taalunie en op vraag de standaardtaligheid van taalvarianten beoordeelt, hanteert immers een relatieve frequentie van 50% in krantencorpora als criterium voor standaardtaligheid; varianten met relatieve frequenties tussen 5% en 50% krijgen er het label ‘status onduidelijk’ (Caluwé & Verreycken 2012, 174-179). Dergelijke grenswaarden zijn uiteraard onvermijdelijk arbitrair (cf. Dhondt et al. 2020; Ghyselen 2022), maar hoe dan ook is duidelijk dat dubitatief *als* in formele registers in Suriname minder sterk staat dan het dubitatieve *of*.

Standaardtaligheid kan ook vanuit perceptie- of attitudeperspectief benaderd worden (cf. Ghyselen 2022). Standaardtaal is dan wat taalgebruikers als standaardtaal of prestigieus percipiëren. In deze studie werden taalpercepties of taalattitudes niet systematisch bevraagd, maar er zijn wel enkele indicaties dat de status van het dubitatieve *als* ook op dat vlak niet eenduidig is. Zo lieten vele respondenten die in de spontane gesprekken constructies met dubitatief *als* realiseerden dat *als* achterwege bij de grammaticatest (cf. Figuur 6), wat impliceert dat ze die vorm niet als standaardtalig percipiëren. Eén respondente begon in de grammaticatest haar deelzin met een dubitatief *als*, maar wisselde die vorm vervolgens snel in voor een dubitatief *of*, wat in dezelfde richting wijst. Zes andere sprekers realiseerden dan weer vrij consequent – ook in de grammaticatest – *als*-zinnen. Zij zijn zich dus veeleer

onbewust van de vorm of ervaren die mogelijk wel als ‘standaardtaal in Suriname’.

De grote individuele variatie in (de perceptie van) het *als*-gebruik – en het feit dat ook Surinaamse onderwijzers niet goed weten wat ermee aan te vangen (cf. inleiding) – kan een teken zijn van endoglossische standaardisering, waarbij dubitatief *als* gaandeweg steeds meer geaccepteerd wordt als kenmerk van het Surinaams Standaardnederlands. In de gesproken dataset vonden we weliswaar geen leeftijdseffecten, maar mogelijk levert grootschaliger productie- en perceptie-onderzoek, met aandacht voor *real* en *apparent time* effecten, op dat vlak wel relevante patronen op. Ik hoop met deze studie alvast te hebben aangetoond dat het dubitatieve *als* in onderzoek naar de standaardisering van het Surinaams-Nederlands een interessante graadmeter kan zijn en hoop ook dat het hier gepresenteerde cijfermateriaal in de toekomst als ijkpunt kan dienen bij de interpretatie van nieuwe data.

Dankwoord

Dit onderzoek kwam tot stand met de steun van het Vlaamse Fonds voor Wetenschappelijk Onderzoek [beurs voor postdoctoraal onderzoek FWO 1232520N]. Ik dank ook (i) alle sprekers die hun stem leenden aan het Corpus Gesproken Surinaams-Nederlands, (ii) Usha Balesar, Carol Bobson, Helen Chang, Widya Matawlie en Sen Nandoe voor hun reflecties over de *als*/of-alternantie in Suriname, (iii) Katrien Depuydt en Jesse de Does van het Instituut voor de Nederlandse Taal voor hun hulp bij het exporteren van de zoekresultaten uit het Corpus Hedendaags Nederlands, (iv) Gert De Sutter en Ludovic De Cuypere voor hun statistisch advies, (v) Gunther De Vogelaer en Eva-Marie Thebes om in hun eigen materiaal even te gaan vissen naar voorbeelden van dubitatief *als* en (v) Timothy Coleman, Christa De Kleine en een anonieme reviewer voor hun nuttige commentaren bij een eerdere versie van dit artikel.

Noten

1. Dit onderzoek kwam tot stand met de steun van het Vlaamse Fonds voor Wetenschappelijk Onderzoek [beurs voor postdoctoraal onderzoek FWO 1232520N]. Ik dank ook (i) alle sprekers die hun stem leenden aan het Corpus Gesproken Surinaams-Nederlands, (ii) Usha Balesar, Carol Bobson,

Helen Chang, Widya Matawlie en Sen Nandoe voor hun reflecties over de *als/of*-alternantie in Suriname, (iii) Katrien Depuydt en Jesse de Does van het Instituut voor de Nederlandse Taal voor hun hulp bij het exporteren van de zoekresultaten uit het Corpus Hedendaags Nederlands, (iv) Gert De Sutter en Ludovic De Cuypere voor hun statistisch advies, (v) Gunther De Vogelaer en Eva-Marie Thebes om in hun eigen materiaal even te gaan visen naar voorbeelden van dubitatief *als* en (v) Timothy Coleman, Christa De Kleine en een anonieme reviewer voor hun nuttige commentaren bij een eerdere versie van dit artikel.

2. Ter illustratie: zoek je in het Corpus Hedendaags Nederlands naar zinnen met het onderschikkende voegwoord *als*, dan krijg je 3.545.374 hits; bij het onderschikkende voegwoord *of* zijn dat er maar 379.377. Kijk je naar de frequenties van *als* en *of* in het algemeen (los van de woordsoort), dan is *als* met 5.553.719 attestaties opnieuw frequenter dan *of* (3.209.182 hits).
3. Ook in het Gesproken Corpus van de Zuidelijk-Nederlandse Dialecten – dat momenteel nog volop in ontwikkeling is (cf. Breitbarth et al. 2020; Ghyselen et al. 2020) – zijn vele voorbeelden van dubitatief *als* te vinden. Een simpele zoekopdracht met de reguliere expressie ‘(weet|weten|wist|wisten) niet als’ in 233 GCND-opnames levert bijvoorbeeld 31 zinnen op, zoals ‘ik weet niet als ik ik twee keren boter gegeten heb van al dat ik ik klein was’ (Knokke, België) en ‘over een jaar ik weet niet als het niet veertigduizend en was in de Belgiquen’ (Clairmaraix, Frankrijk). De hits komen allemaal uit West-Vlaanderen, Frans-Vlaanderen en het westen van Oost-Vlaanderen.
4. Saramaccaans, Aucaans en Paramaccaans zijn hier samengenomen; de talen van kleinere Marrongroepen werden ondergebracht in de categorie ‘andere talen’.
5. ivdnt.org/corpora-lexica/corpus-hedendaags-nederlands
6. CLARIN (‘Common Language Resources and Technology Infrastructure’) bouwt aan stabiele, permanente, toegankelijke en uitbreidbare onderzoeksinfrastructuur voor taalbronnen en taaltechnologie. Het doel van het programma is om de bestaande versnippering op te heffen en het gebruik van computationele technieken in de geesteswetenschappen (Digital Humanities) te stimuleren. Aan CLARIN nemen ondertussen 22 EU-landen deel. Meer info: ivdnt.org/onderzoek-projecten/clarin.
7. Als criterium gebruikten we zelfidentificatie: wie zichzelf als ‘Creools’ beschouwt, kwam in aanmerking voor het onderzoek.
8. Aangezien het Corpus Gesproken Surinaams-Nederlands niet gelemmatiseerd is, is het moeilijker om werkwoordfrequenties op basis van de gesproken data te berekenen.
9. Bij *spreken* en *gaan* verwacht je op basis van het valentieschema misschien niet meteen dubitatieve afhankelijke bijzinnen, maar ze werden wel degelijk in het CGSN geattesteerd: “ik heb nog niet met DEREK gesproken of h... of hij het wil doen” en “het ging over uh of ik mijn kinderen een bok gaf uh wanneer ze Surina...”. Om die reden werden ze ook in de zoekopdracht meegenomen.

10. Zoekopdracht zoals die in Corpus Query Language werd geformuleerd:
[lemma = "aangeven|afhangen|afvragen|begrijpen|bepalen|checken|controleren|kijken|nadenken|nagaan|oplekken|spreken|twijfelen|uitmaken|vragen|weten|zien|gaan" & pos="vrb"] [punct != '.*[.,;:]*' & pos!= "CONJ "] {0,4}{ (word = "als"|word="of") & pos = "conj" & pos_type = "sub"&punct != '.*[.,;:]*'] within <s/>
11. De analyse werd uitgevoerd in R met de *cforest*-functie van het *Partykit*-pakket. Volgend commando werd gegeven: `partykit::cforest(voegwoord ~ Spreker + Geboortejaar + Matrixwerkwoord + Srananmoedertaal + KennisSranan + KennisEngels + PolariteitMatrixzin + FrequentieMatrixww + Gesprekssituatie, data = data, mtry=3)`.
12. C is een maat voor de overeenstemming tussen voorspelde waarden en geobserveerde waarden. Wanneer C een waarde van 0,5 aanneemt, dan zijn de voorspellingen willekeurig; een C-waarde van 1 wijst daarentegen op perfecte voorspellingen. Aangezien volgens Baayen (2008, 2014) modellen met een C-waarde boven 0,8 echte voorspellende kracht hebben, zit het met het random forest-model dat we opstelden zeker goed wat voorspellende kracht betreft.
13. Deze analyse werd uitgevoerd in R met de *ctree*-functie van het *Partykit*-pakket. Volgend commando werd gegeven: `partykit::ctree(data$voegwoord ~ Spreker + Geboortejaar + Srananmoedertaal + KennisSranan + KennisEngels + Matrixwerkwoord + PolariteitMatrixzin + Gesprekssituatie, data = data, control = partykit::ctree_control(mincriterion = 0.95))`.
14. R-commando: `partykit::cforest(voegwoord ~ Matrixwerkwoord + FrequentieMatrixwerkwoord + Polariteit + TekstType, data = data2, mtry=2)`.
15. Dit soort willekeurige effecten kan doorgaans moeilijker meegenomen worden in random forests en inference trees, aangezien de algoritmes vastlopen wanneer predictoren te veel categorieën (*levels*) hebben. Voor de gesproken data konden we de variabele 'spreker' wel meenemen, omdat die maar 22 categorieën telde.

Over de auteur

Anne-Sophie Ghyselen, Universiteit Gent. Adres: Vakgroep Taalkunde, Blandijnberg 2, 9000 Gent, België. E-mail: Annesophie.ghyselen@ugent.be.

Referenties

Algemeen Bureau voor de Statistiek Suriname. 2014. *Suriname census 2012. Volume III. Huishoudens, Woonverblijven en Gezinnen, Milieu, Criminaliteit*. Algemeen Bureau voor de Statistiek (Paramaribo).

- Algemene Bureau voor de Statistiek Censuskantoor. 2006. *Zevende Algemene Volks- en Woningtelling in Suriname. Landelijke resultaten volume IV. Huishoudens, Gezinnen en Woonverblijven*. Algemeen Bureau voor de Statistiek / Censuskantoor (Paramaribo).
- Auer, Peter. 2011. "Dialect vs. standard: A typology of scenarios in Europe." In *The languages and linguistics of Europe. A comprehensive guide*, geredigeerd door Bernd Kortmann & Johan Van der Auwera, 485-500. Berlijn: De Gruyter.
- Baayen, R.H. 2008. *Analyzing Linguistic Data. A Practical Introduction to Statistics using R*. Cambridge: Cambridge University Press.
- Bailey, Guy, Tom Wikle, Jan Tillery, & Lori Sand. 1992. "The apparent time construct." *Language Variation and Change* 3: 241-264.
- Barbiers, Sjef, & et al. 2006. "Dynamische Syntactische Atlas van de Nederlandse Dialecten (DynaSAND)." Meertens Instituut. <http://www.meertens.knaw.nl/sand/>.
- Breitbarth, Anne, Melissa Farasyn, Anne-Sophie Ghyselen, & Jacques Van Keymeulen. 2020. "Het gesproken corpus van de zuidelijk-Nederlandse dialecten." *Handelingen van de Koninklijke Zuidnederlandse Maatschappij voor Taal- en Letterkunde en Geschiedenis* 72: 23-38.
- Caluwé, Dirk, & R. Verreycken. 2012. "Standaardtaal en tussentaal in de taaladvisering." In *De manke usurpator: over Verkavelingsvlaams*, geredigeerd door Keven Absillis, Jürgen Jaspers & Sarah Van Hoof, 167-185. Gent: Academia Press.
- de Bies, Renata. 2009. *Prisma Woordenboek Surinaams-Nederlands*. Houten: Prisma.
- De Bruijne, Ad, & Aart Schalkwijk. 2005. "The position and residential patterns of ethnic groups in Paramaribo's development in the twentieth century." *New West Indian Guide/Nieuwe West-Indische Gids* 79, no. 3x4: 239-271.
- de Kleine, Christa. 2007. *A Morphosyntactic Analysis of Surinamese Dutch*. München: Lincom Europa.
- . 2013. "Dutch in Suriname." In *Language and Space. An International Handbook of Linguistic Variation. Dutch*, geredigeerd door Frans Hinskens & Johan Taeldeman, 841-857. Berlijn/Boston: Walter de Gruyter.
- De Rooy, J. 1965. *Als, of, dat : enkele conjuncties in ABN, dialect en Fries*. Assen : Van Gorcum.
- Dhondt, Arne, Timothy Coleman, Johan de Caluwe, & Gauthier Delaby. 2020. "Naar een pluricentrische Algemene Nederlandse Spraakkunst: het gebruik van productiedata voor de beschrijving van nationale variatie." *Handelingen van de Koninklijke Zuidnederlandse Maatschappij voor Taal- en Letterkunde en Geschiedenis* LXXIII: 85-134.
- Diepeveen, Janneke, & Matthias Hüning. 2016. "The status of Dutch in post-colonial Suriname." In *Aspects of (post)colonial linguistics. Current perspectives and new approaches*, geredigeerd door Daniel Schmidt-Brücken, Susanne Schuster & Marina Wienberg, 131-155. Berlin/Boston: De Gruyter.

- Ghyselen, Anne-Sophie. 2022. "Attitudinal and perceptual research as part of the methodological toolbox to define standard languages: Advances, issues and perspectives in research on Belgian Dutch." In *Standard Languages in Germanic-Speaking Europe: Attitudes and Perception*, geredigeerd door Alexandra Lenz, Barbara Soukup & Wolfgang Koppensteiner, 151-187. Oslo: Novus.
- Ghyselen, Anne-Sophie, Anne Breitbarth, Melissa Farasyn, Jacques Van Keymeulen, & Arjan Van Hessen. 2020. "Clearing the transcription hurdle in dialect corpus building: the corpus of Southern Dutch dialects as case-study." *Frontiers in Artificial Intelligence* 3: 1-17.
- Ghyselen, Anne-Sophie, Stef Grondelaers, Sita Doerga Misier-Patadien, & Usha Balesar. 2022. "Standard language dynamics in postcolonial Suriname. Measuring language attitudes and ideologies in Paramaribo." *Lingua* 273, no. 103340a: 1-27.
- Gobardhan-Rambocus, Lila. 1993. "Het Surinaams Nederlands." In *Immigratie en ontwikkeling. Emancipatie van contractanten*, geredigeerd door Lila Gobardhan-Rambocus & Maurits S. Hassankhan, 140-158. Paramaribo: Anton de Kom Universiteit.
- . 1997. "Suriname en het Nederlands." In *Koloniale Taalpolitiek in Oost en West. Nederlands-Indië, Suriname, Nederlandse Antillen en Aruba*, geredigeerd door Kees Groeneboer, 225-249. Amsterdam: Amsterdam University Press.
- Goedertier, Wim, & Simo Goddijn. 2000. *Corpus Gesproken Nederlands: Protocol voor Orthografische Transcriptie*. Bezocht op 14-11-2011.
- Gries, Stefan. 2015. "The most under-used statistical method in corpus linguistics: multi-level (and mixed-effects) models." *Corpora* 10 (1): 95-125.
- Harrell, Frank E. 2015. *Regression Modeling Strategies. With Applications to Linear Models, Logistic and Ordinal Regression, and Survival Analysis. Second Edition*. New York/Dordrecht/Londen: Springer.
- Instituut voor de Nederlandse Taal. 2021. *Algemene Nederlandse Spraakkunst*. Versie 3.1.
- Kruger, Haidee. 2019. "That Again: A Multivariate Analysis of the Factors Conditioning Syntactic Explicitness in Translated English." *Across Languages and Cultures* 20 (1): 1-33.
- Labov, William. 1990. "The Intersection of Sex and Social Class in the Course of Linguistic Change." *Language Variation and Change* 2: 205-254.
- Léglise, Isabelle, & Bettina Migge. 2015. "Language Practices and Linguistic Ideologies in Suriname: Results from a School Survey." In *In and Out of Suriname. Language, Mobility and Identity*, geredigeerd door Eithne B. Carlin, Isabelle Léglise, Bettina Migge & Paul B. Tjon Sie Fat, 13-57. Leiden/Boston: Brill.
- Levshina, Natalia. 2020. "Conditional Inference Trees and Random Forests." In *A Practical Handbook of Corpus Linguistics*, geredigeerd door Magali Paquot & Stefan Gries, 611-643. Cham: Springer International Publishing.

- Menke, Jack, & Iwan Sno. 2016. "Ras en etniciteit in volkstellingen van Suriname." In *Mozaïek van het Surinaamse volk. Volkstellingen in demografisch, economisch en sociaal perspectief*, geredigeerd door Jack Menke. Paramaribo: Algemeen Bureau voor de Statistiek/Institute for Graduate Studies and Research.
- Muysken, Pieter. 2013. "Ethnolects of Dutch." In *Language and Space. An International Handbook of Linguistic Variation. Dutch*, geredigeerd door Frans Hinskens & Johan Taeldeman, 739-761. Berlijn/Boston: De Gruyter Mouton.
- . 2017. "The transformation of a colonial language: Surinamese Dutch." In *Boundaries and Bridges: Language Contact in Multilingual Ecologies*, geredigeerd door Kofi Yakpo & Pieter Muysken, 283-310. Boston/Berlijn: Walter de Gruyter.
- Rys, Kathy, Wilbert Heeringa, Janneke Sif Rutten, Frans Hinskens, Johan De Caluwe, Usha Balesar, & Sita Doerga Misier. 2021. *Staat van het Nederlands. Over de verschuivende taalkeuzes van Surinamers, Vlamingen en Nederlanders in het dagelijks leven. Onderzoeksrapport 2021*. De Nederlandse Taalunie (Amsterdam).
- Shank, Christopher, Julie Van Bogaert, & Koen Plevoets. 2016. "The diachronic development of zero complementation: A multifactorial analysis of the that/zero alternation with think, suppose, and believe." *Corpus Linguistics and Linguistic Theory* 12 (1): 31-72.
- Sluisdom, Renate. 1992. "Norm en standaard in Suriname." Masterscriptie, Vrije Universiteit Amsterdam.
- Strobl, Carolin, James D. Malley, & Gerhard Tutz. 2009. "An introduction to recursive partitioning: Rationale, application, and characteristics of classification and regression trees, bagging, and random forests." *Psychological Methods* 14 (4): 323-348.
- Tagliamonte, Sali A., & Harald R. Baayen. 2012. "Models, forests, and trees of York English: Was/were variation as a case study for statistical practice." *Language Variation and Change* 24: 135-178.
- Van der Sijs, Nicoline. 2014. "'Laat-me-er-dit-van-zeggen'. Grammaticale bijzonderheden van het Surinaams-Nederlands." *Onze Taal* 11: 314-316.
- Van Donselaar, Jan. 1989. *Woordenboek van het Surinaams-Nederlands*. Muiderberg: Coutinho.
- Van Donselaar, Jaap. 2005. "Het Surinaams-Nederlands in Suriname." In *Wereld-nederlands. Oude en jongere variëteiten van het Nederlands*, geredigeerd door Nicoline Van der Sijs, 111-130. The Hague: sdu.
- Van Hout, Roeland, Georges De Schutter, E. De Crom, W. Huinck, Hanne Kloots, & Hans Van de Velde. 1999. "De uitspraak van het Standaard-Nederlands: variatie en varianten in Vlaanderen en Nederland." In *Artikelen van de Derde Sociolinguïstische Conferentie*, geredigeerd door E. Huls & B. Weltens, 183-196. Delft: Eburon.
- Van Wel, F. J., & A. J. Vervoorn. n.d. – ca. 1974. *Het Nederlands in Suriname*. Den Haag: Kabinet voor Surinaamse en Nederlands-Antilliaanse zaken.

- Ventura, Welmoed. 2013. “‘Wat gebeurt?’ – Interferenties en de syntaxis van het Surinaams-Nederlands.” *Conferentie Neerlandistiek in het Caribisch Gebied* 2013: 46-60.
- Yakpo, Kofi, Margot van den Berg, & Robert Borges. 2015. “On the Linguistic Consequences of Language Contact in Suriname: The Case of Convergence.” In *In and Out of Suriname*, geredigeerd door Eithne B. Carlin, Isabelle Légise, Bettina Migge & Paul B. Tjon Sie Fat, In *Language, Mobility and Identity*, 164-195. Leiden/Boston: Brill.