# Granulariteit in taal

Een bestaande granulariteitstheorie toegepast op de locatieve indexical ‘hier’

Lotje Bosman

Begeleider: Rick Nouwen
Tweede beoordelaar: Johannes Korbmacher

Bachelor Kunstmatige Intelligentie
7,5 ECTS

28 juni 2019

Universiteit Utrecht

# Inhoudsopgave

## Abstract

Deze scriptie gaat over granulariteit. Granulariteit gaat over verschillen in detailniveau. Die verschillen zijn terug te vinden in zowel kennisrepresentatie als taal. Hobbs (1985) heeft een theorie over granulariteit geschreven die vooral over kennisrepresentatie gaat. De vraag is of deze kan worden toegepast op meer talige fenomenen. Er is specifiek gekeken naar het woord 'hier'. 'Hier' bevat namelijk veel verschillende detailniveaus en dus veel granulariteit. De onderzoeksvraag is dan ook: '(Hoe) kunnen we Hobbs' theorie van granulariteit toepassen op de locatieve indexical 'hier'?'. Om deze te beantwoorden, zijn de relevante delen van Hobbs' theorie uitgelegd en er is een beeld geschetst van de semantiek van 'hier'. Uit het onderzoek is gebleken dat de theorie kan worden toegepast op 'hier' na een kleine aanpassing: er worden granulariteiten toegekend aan 'hier'.

**Trefwoorden**: granulariteit, imprecisie, vaagheid, simplificatie, indexicals, hier, locatives, kunstmatige intelligentie, tekstbegrip

# 1 Inleiding

## 1.1 Imprecisie en granulariteit in taal

In taal is een hoop vaagheid te vinden. Als je zegt dat iemand lang is, is hij dan 2 meter of 1,80 meter? En hoort dat laatste wel tot de categorie 'lang'? Betekent 'even' in de zin 'hij komt even langs' dat hij één minuut langskomt of een uur? En waar is eigenlijk 'hier' als je zegt "ik ben hier geboren"? Ondanks al deze vaagheid kunnen mensen elkaar toch meestal goed begrijpen. Het is vaak niet nodig om iets heel precies, tot op zeer gedetailleerd niveau, over te brengen. Je hoeft kennelijk niet te weten dat iemand 196,43 centimeter is; het is genoeg om te weten dat hij 'lang' is, grofweg ergens boven de 180 centimeter. Sterker nog, het kost onnodig veel tijd en moeite om die cijfers achter de komma te achterhalen en uit te spreken. Het is dus vaak beter om wat impreciezer te zijn. Er is een link tussen imprecisie en granulariteit. Het verschil in detailniveau, zoals tussen 199,3743 en 199 en 200, noemen we een verschil in granulariteit: hoe impreciezer, hoe 'grofkorreliger'.

## 1.2 Probleem in de kunstmatige intelligentie

Voor een computer is het achterhalen en opnoemen van een gedetailleerd getal veel simpeler dan voor mensen. Zonder moeite kan elke computer 20! uitrekenen ($20 * 19 * 18 * ... * 2 * 1$), maar omdat de meeste beeldschermen niet zo groot zijn en mensen dat soort getallen niet aankunnen, noemt de computer het voor ons grofweg $2.432902e + 18$. Een computer heeft natuurlijk een limiet, maar die ligt veel hoger dan bij mensen. In programmeertalen kun je vaak aangeven op welk granulariteitsniveau je wilt dat de computer rekent. Zo geven integers alleen hele getallen, floats zeven getallen achter de komma en doubles vijftien getallen achter de komma. Maar veel verder dan een paar ingeprogrammeerde granulariteitsniveaus komt de computer niet. Mensen zijn continu (al dan niet onbewust) bezig met granulariteitsniveaus kiezen. Een computer heeft dus ook wel verschillende granulariteiten, maar manipuleert die niet, zoals een mens dat wel doet. Dat is op dit moment ook niet nodig. Maar stel dat we een computer en een mens willen laten communiceren met elkaar, stel dat een computer de menselijke taal met al z'n imprecisie moet kunnen begrijpen en op het juiste granulariteitsniveau taal moet kunnen genereren? Kunnen we granulariteit logisch modelleren?

In 1985 publiceerde Jerry Hobbs het paper 'Granularity' dat een kader biedt om simpele theorieën te verkrijgen uit complexe theorieën. Met andere woorden: een kader voor granulariteit. Hij geeft hierin een algemene methode om complexe en precieze theorieën te kunnen versimpelen naar grofkorrelige theorieën. Hij geeft dus een manier om van een exacte 199,3743 naar een 200 te gaan. Ook legt hij uit dat dit versimpelen iets is wat mensen kunnen zonder er ook maar over na te denken en dat het cruciaal is om dit versimpelen te kunnen modelleren als we een intelligente machine willen maken. Bij het begrijpen van een simpele zin als ‘de banaan ligt in de fruitschaal’, wordt het grootste deel van de (latente) kennis direct weggestopt onder ‘niet belangrijk’. Grofweg alleen de kennis van de woorden in de zin en hoe taal werkt, blijft achter. Misschien komen ook gelijk de begrippen ’lekker’ of ’honger’ bij de spreker op, maar bijvoorbeeld niet wat de naam is van de jongste broer van zijn vader; dat is nu niet belangrijk.

Eén van de dingen die Hobbs wel noemt maar waar hij niet verder op ingaat, is granulariteit in taal, terwijl mij dat juist zo interessant en relevant lijkt in het kader van kunstmatige intelligentie. Hoe kunnen een computer en een mens goed communiceren als de computer geen granulariteitsgrenzen heeft? Een computer zal bij ‘vijf minuten’ ervan uitgaan dat je precies 300,00 seconden bedoelt, het niet begrijpen als je ‘even naar de buren’ gaat en je vertellen dat het precies twee maanden, drie weken, één dag, tweeëntwintig uur, tweeënveertig minuten en elf seconden geleden is dat je voor het laatst bij de kapper was. Dit is niet wenselijk. De computer gaat uit van de fijnst mogelijke granulariteit, terwijl mensen hun uiting vaak impreciezer bedoelen of naar een grofkorreliger antwoord op zoek zijn (zoals ‘ongeveer drie maanden’).

## 1.3 Hier

Om granulariteit in taal verder te onderzoeken, willen we een talig fenomeen dat zowel imprecisie als granulariteit bevat. Zonder granulariteit kun je immers geen granulariteit in taal onderzoeken en zonder imprecisie is het minder relevant, omdat dan meestal wel duidelijk is wat er bedoeld wordt. Een van die fenomenen is het woordje ‘hier’. ‘Hier’ duidt een bepaalde locatie aan: de locatie van de spreker. ‘Hier’ kan bijvoorbeeld betekenen: de kamer waar iemand zich bevindt (“De kat was net nog hier”), de stad waar iemand is (“Ik ben hier geboren”) of zelfs de planeet (“We zijn hier met 7,5 miljard mensen”). Dit zijn verschillende granulariteitsniveaus en wat er precies bedoeld wordt, is niet altijd even duidelijk, dus er is ook imprecisie. We beperken ons tot het *locatieve* ‘hier’, dus waarbij ‘hier’ een locatie aanduidt (Quer, 2011). ‘Hier’ kan ook verwijzen naar een situatie zoals in “Ik heb hier last van”. In dat geval vindt er wel imprecisie plaats (waar heb je precies last van?). Er is ook granulariteit te vinden. Iemand kan bijvoorbeeld last hebben van gefluit (fijn), van geluid dat gemaakt werd (groffer) of van dat er überhaupt iemand de kamer in kwam (grofkorrelig). Echter, dit werkt heel anders dan het locatieve ‘hier’, dus ik ga alleen op het locatieve ‘hier’ focussen.

De vraag is of we Hobbs’algemene theorie kunnen gebruiken voor dit talige, semantische fenomeen en zo ja, hoe? Zo kom ik op de onderzoeksvraag: (hoe) kunnen we Hobbs’ theorie van granulariteit toepassen op de locatieve indexical ‘hier’? Om antwoord te geven, zal ik eerst kijken naar granulariteit en relevante delen halen uit de theorieën die er op dit moment over zijn. Vervolgens ga ik de betekenis en semantiek van ‘hier’ uiteenzetten en kijken naar hoe imprecisie en granulariteit bij dit woord werken. Tot slot ga ik proberen de koppeling tussen deze twee delen te maken door de theorie toe te passen op de semantiek.

# 2 Theorie

## 2.1 Wat is granulariteit?

Om antwoord te kunnen geven op de onderzoeksvraag, geef ik eerst een definitie van granulariteit.

Bij granulariteit gaat het om het opsplitsen van iets in kleinere delen of juist het samenvoegen tot een groter geheel. Dit kan gaan over verschillende dingen zoals acties, groottes, situaties, locaties, et cetera. Je kunt bijvoorbeeld 'naar je werk gaan' uitsplitsen in jas aandoen, tas pakken, deur openen, door de opening heengaan, deur sluiten, naar je fiets lopen, et cetera. Andersom kun je de drie acties met de deur samenvoegen tot 'naar buiten gaan'. Granulariteit komt overal voor. Kerkklokken geven meestal geen seconden aan, maar wel minuten (dus niet alleen uren), terwijl stopwatches vaak tot op milliseconden precies zijn. Er is dus een granulariteitsverschil tussen kerkklokken en stopwatches: stopwatches zijn fijn en kerkklokken zijn grof. En je bedenkt dat je boodschappen moet doen en denkt niet aan alle losse handelingen zoals een tas pakken, je jas aandoen, naar buiten gaan, et cetera. Aardappelen koken is een proces bestaande uit kleinere processen (schillen, eventueel snijden, water in een pan doen, fornuis aandoen, et cetera) en is onderdeel van het grotere proces eten maken. Kortom, granulariteit is een dagelijks fenomeen.

Granulariteit speelt ook een grote rol in natuurlijke taal. Afhankelijk van de situatie, vertel je iets op een ander granulariteitsniveau of wordt iets op een ander niveau geïnterpreteerd. Bijvoorbeeld het woordje 'schoon' in 'je moet je handen schoonmaken', betekent iets heel anders wanneer het tegen een chirurg wordt gezegd voor een operatie, dan wanneer het tegen een kind wordt gezegd voor het eten. En als ik zeg: "Tim heeft een labrador en is er blij mee, dus ik wil dezelfde", zal men niet denken dat ik Tims hond wil stelen, maar dat ik ook een labrador wil. Dat is het verschil tussen een type en een token. Een token is het exacte object, in dit geval Tims labrador. Een type is een soort object, in dit geval een labrador in het algemeen. Een token is dus een fijnere granulariteit dan een type.

## 2.2 Simplificatie en Hobbs

We willen een logische theorie hebben die een representatie van onze wereld kan versimpelen, deze grofkorreliger kan maken. De wereld is erg complex. Als je een kamer zou willen beschrijven, zou je elk item erin moeten opnoemen, de afmetingen en afstanden tot elkaar, de kleur, schaduwen, noem maar op. Veel hiervan is vaak niet relevant. Als iemand aan mij vraagt hoe mijn woonkamer eruitziet, ga ik niet de afstand van de tafel tot de boekenkast aangeven en vertel ik niet hoeveel boeken daar precies in staan. Hobbs (1985) is met een simplificatietheorie gekomen waarmee je een complexe wereld $T_0$ kan versimpelen tot een grofkorrelige wereld $T_1$.

Stel, je hebt vier boeken: Harry Potter (HP), Animal Farm (AF), Turks Fruit (TF) en Lotte Weeda (LW). Nu gaan we kijken naar de schrijvers van deze boeken (respectievelijk Rowling (R), Orwell (O), Wolkers (W) en 't Hart (H)). De eerste twee zijn Brits en de laatste twee zijn Nederlands. Nu kunnen we de wereld omschrijven als vier entiteiten $S_0 = \{HP, AF, TF, LW\}$ en zes predicaten $P_0 = \{R, O, W, H, Brits, Nederlands\}$. Hierbij gelden de volgende proposities: R(HP), Brits(HP), O(AF), Brits(AF), W(TF), Nederlands(TF), H(LW) en Nederlands(LW). De omschrijving van deze wereld heet $T_0$. Stel dat iemand nu alleen geïnteresseerd is in de vraag of een boek van een Nederlandse of Britse schrijver is en verder maakt het hem niet uit wie die schrijver is. De relevante predicaten zijn dus $\mathbf{R} = \{Brits, Nederlands\}$. Dan kun je simplificeren en de versimpelde $T_1$ krijgen, waarbij $S_1 = \{[HP, AF],[TF, LW]\}$ en $P_1 = \{Brits, Nederlands\}$[1]. Wie de schrijver is, is nu niet meer relevant, dus deze predicaten vallen weg. Harry Potter en Animal Farm zijn beide van Britse schrijvers, die worden toegewezen aan dezelfde equivalentieklasse die we labelen met 'Brits'. Het verschil is niet relevant, dus we noemen het geen verschil meer, maar gewoon twee boeken van een Britse schrijver, twee entiteiten die voldoen aan hetzelfde predicaat.

$S_0$ is het domein van interpretatie. Hierin bevinden zich alle entiteiten die een rol spelen. In ons voorbeeld zijn dat dus de vier boeken. Verder heb je een verzameling predicaten, $P_0$, die iets kunnen zeggen over de entiteiten. In bovenstaand voorbeeld hebben we alleen schrijver en land van oorsprong genomen, maar we hadden hier bijvoorbeeld ook genre of aantal pagina's bij kunnen zetten. Om te versimpelen kijken we welke van deze predicaten relevant zijn. Dat wordt dus een deelverzameling van $P_0$ en we noemen deze verzameling $\mathbf{R}$. Bij dit voorbeeld hadden we besloten dat alleen land van oorsprong relevant is. Dit wordt vervolgens $P_1$: de predicaten in je nieuwe wereld $T_1$. Deze predicaten zaten dus ook al in $P_0$, maar nu valt er een hele klasse met entiteiten onder. $S_1$ is nu de verzameling equivalentieklassen met entiteiten die door de keuze van de relevante predicaten hetzelfde zijn. We noemen de entiteiten binnen een klasse ononderscheidbaar van elkaar, omdat er geen relevant verschil tussen is.

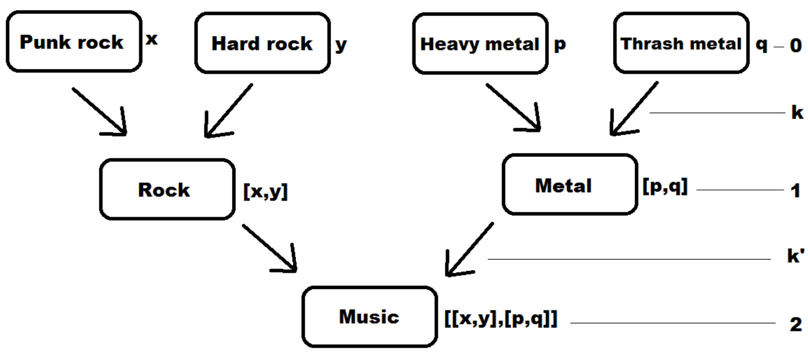
Nu kan er een ononderscheidbaarheidsrelatie $\sim$ worden gedefinieerd die aangeeft dat de entiteiten $x$ en $y$ ononderscheidbaar zijn:

**(1)** $(\forall x, y) x \sim y \equiv (\forall p \in \mathbf{R})(p(x) \equiv p(y))$

Dit betekent dus dat x $\sim$ y (dat $x$ en $y$ ononderscheidbaar zijn) als voor elk predicaat p dat een relevant predicaat is (en dus in $\mathbf{R}$ zit), geldt dat $p$ waar is voor $x$ dan en slechts dan als $p$ waar is voor $y$. Stel, je wilt luisteren naar bands die rockmuziek maken. Er wordt je onder andere $x$: Green Day (punkrock) en $y$: Led Zeppelin (hardrock) aangeraden. Deze bands (entiteiten) hebben een

[1]Dit klopt niet helemaal, maar dat wordt bij definitie (**2**) en (**3**) verder toegelicht.

hele set eigenschappen (predicaten). Deze bands maken allebei muziek, allebei rock, de een is Engels, de ander Amerikaans, de een is hardrock, de ander punkrock. Maar het enige wat relevant is, is of het rock is of niet, dus dat is het relevante predicaat. Green Day en Led Zeppelin maken allebei rock en dus zijn ze ononderscheidbaar; rock(Led Zeppelin) $\equiv$ rock(Green Day) en dus Green Day $\sim$ Led Zeppelin (sorry muziekliefhebbers). De bands komen hiermee in dezelfde equivalentieklasse $[x, y]$ terecht, die voldoet aan het predicaat rock. Als je als ref




Figuur 1: Simplificatie

Het naar een grofkorreligere granulariteit toe gaan (zoals van hard- en punkrock naar rock), heet simplificatie (figuur 1). Het is simpeler om alle subgenres van rock gewoon rock te noemen. Het proces zelf kunnen we een ‘toewijzing’ $k$ noemen. Dit kunnen we als volgt definiëren:

(**2**) $(\forall x, y) x \sim y \equiv k(x) = k(y)$

Dit zegt dat als $x$ en $y$ ononderscheidbaar zijn, dan en slechts dan is toewijzing $k$ op $x$ hetzelfde als toewijzing $k$ op $y$. Dus als Green Day en Led Zeppelin ononderscheidbaar zijn, dan k(Green Day) = k(Led Zeppelin). Entiteiten met dezelfde toewijzing komen in dezelfde equivalentieklasse. Predicaten ondergaan zo’n zelfde toewijzing:

(**3**) $(\forall p \in \mathbf{R})\ k(p)(k(x)) \equiv p(x)$

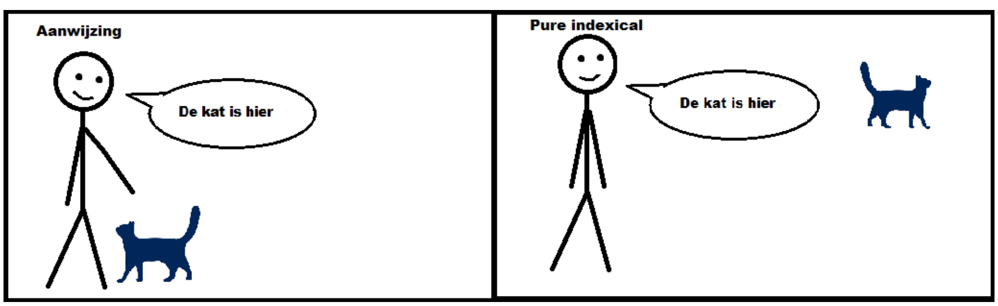
$k(p)(k(x))$ is waar dan en slechts dan als $p(x)$ waar is. Dus de toewijzing k op een predicaat $p$ (dat is dus $k(p)$) is waar voor een toegewezen entiteit $(k(x))$ dan en slechts dan als $p(x)$ waar is in $\mathrm{T}_0$. Deze toewijzing is noodzakelijk omdat het predicaat nu niet de eigenschap is van een entiteit, maar van een hele equivalentieklasse. In het voorbeeld van de boeken hierboven gold dus eigenlijk: $\mathrm{P}_1 = \{$k(Brits), k(Nederlands)$\}$. Als we het voorbeeld over muziekgenres volgen, krijgen we: $k$(genre)($k$(band)) is waar dan en slechts dan als genre(band) waar is. Dit betekent dat de toewijzing alleen waar is als de band onder dat genre valt. Dus $k$(rock)($k$(Green Day)) is waar want rock(Green Day), maar $k$(rock)($k$(Iron Maiden)) is niet waar en rock(Iron Maiden) niet (Iron Maiden is metal). Doordat de toewijzing niet gelijk is, kunnen Green Day en Iron Maiden niet in dezelfde equivalentieklasse komen en zijn ze onderscheidbaar van elkaar. Door alleen te kijken naar de equivalentieklassen in plaats van alle entiteiten, heb je de wereld versimpeld.

# 3 Het fenomeen ‘hier’

## 3.1 Betekenis

In deze scriptie pas ik de theorie uit het vorige hoofdstuk toe op het talige fenomeen ‘hier’. In dit hoofdstuk zet ik de betekenis en semantiek van ‘hier’ uiteen en wordt duidelijk hoe imprecisie en granulariteit bij dit woord werken.

Het woord ‘hier’ is een indexical. Daarmee bedoel ik dat de betekenis van het woord afhangt van de situatie waarin het geuit wordt. De betekenis van de zin ‘ik woon hier’ is niet te achterhalen zonder te weten wie ‘ik’ (ook een indexical) is en waar die persoon zich bevindt. ‘Hier’ kan zowel een aanwijzing (Engels: demonstrative) als een pure indexical zijn (Kaplan, 1977). Bij een aanwijzing moet de spreker een tweede actie verrichten, namelijk een plek aanwijzen. Iemand kan zeggen “De kat is hier” terwijl hij naar de kat wijst (zie figuur 2). Hij kan het ook zeggen zonder naar de kat te wijzen; dan is het een pure indexical. Een pure indexical houdt in dat duidelijk is wat bedoeld wordt met ‘hier’ in een bepaalde context, doordat we weten dat het woord ‘hier’ betekent: ‘op de plek van de spreker’. Bij de pure indexical ‘ik’ is dit heel duidelijk te zien: we weten dat ‘ik’ altijd ‘de spreker’ betekent, dus in de zin ‘ik ben jarig’ weten we dat het de spreker is die jarig is. Bij ‘hier’ is dat net zo. Je weet dat ‘hier’ betekent ‘de locatie van de spreker’, dus als iemand zegt: “Jan is hier getrouwd”, is Jan getrouwd op de locatie van de spreker.




Figuur 2: Aanwijzing en pure indexical.

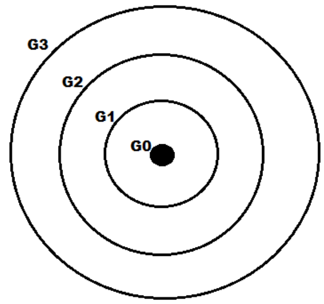
## 3.2 Imprecisie en granulariteit

Maar stel dat iemand inderdaad zonder aan te wijzen zegt: “Jan is hier getrouwd”, waar is Jan dan precies getrouwd? Misschien sta je op dat moment wel op een grote open plek in het bos. Is Jan op die open plek getrouwd? Of is hij echt op de vierkante meter getrouwd waar de spreker nu staat? Misschien is hij wel gewoon ergens in dat bos getrouwd, of in de stad in de buurt van dat bos. De zin ‘Jan is hier getrouwd’ is dus imprecies en dat komt doordat je niet weet waar ‘hier’ is. Echter, als de spreker op de grond aan zou wijzen op welke plek

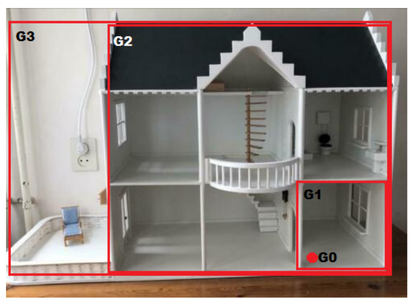
Jan getrouwd is, zou het preciezer zijn. Er is nog enige imprecisie waar precies naar gewezen wordt, maar dat dan niet 'Nederland' bedoeld wordt, is duidelijk. Doordat deze imprecisie veel minder optreedt als 'hier' een aanwijzing is en een aanwijzing veel minder vatbaar is voor granulariteit, focus ik mij alleen op 'hier' als pure indexical.

Bij deze imprecisie komt ook weer granulariteit kijken. De exacte plek waar iemand is, zou je kunnen zien als $G_0$ (granulariteit 0). Elk stapje groter (grover) is $G_{(n+1)}$. Dus bijvoorbeeld bij de zin 'het is hier mooi' kan iemand $G_0$, de precieze plek waar hij staat, bedoelen (maar dat is wel vrij narcistisch), of $G_1$, de kamer waar iemand net is binnengelopen, of $G_2$, het huis waar hij is, of misschien wel $G_3$, de hele buurt. Wat $G_x$ precies betekent, is een keuze. Je zou ook kunnen zeggen dat $G_1$ de stad is en $G_2$ het land, het hangt af van waar je naar op zoek bent. 'Hier' blijft in dit geval nog steeds erg ambigu. Het kan nog steeds elke plaats (hoe groot dan ook) zijn, zolang het maar onder andere de locatie van de spreker bevat.

Het probleem hierbij ten opzichte van bijvoorbeeld getallen is dat de vergroting niet lineair is. Een granulariteitsverschuiving zoals in figuur 3a zou gemakkelijker te modelleren zijn, maar de realiteit zit dichter bij 3b. In figuur 3a zien we namelijk dat de stap naar een volgende granulariteit steeds even groot is, terwijl in 3b de grootte van die stappen heel erg kan verschillen. Het is hierdoor niet alleen moeilijk om te ontdekken wat relevante predicaten zijn, maar ook wat de predicaten überhaupt zijn. Iemand zou bijvoorbeeld met 'hier' wel de kamer kunnen bedoelen, maar niet de kamer en de tien centimeter daaromheen; dat gebied hoort duidelijk niet bij de kamer. Bij getallen is dat wat gemakkelijker. Als je gaat afronden op tientallen, dan rond je alles tussen de 45 en de 54,999 af op 50, alles tussen de 55 en de 64,999 af op 60, et cetera. Dit is lineair: de stappen zijn even groot. Dat is bij 'hier' niet het geval en dat maakt het modelleren wat ingewikkelder, omdat je nooit weet wat de volgende stap gaat zijn en hoeveel groter die is.




(a)




(b)

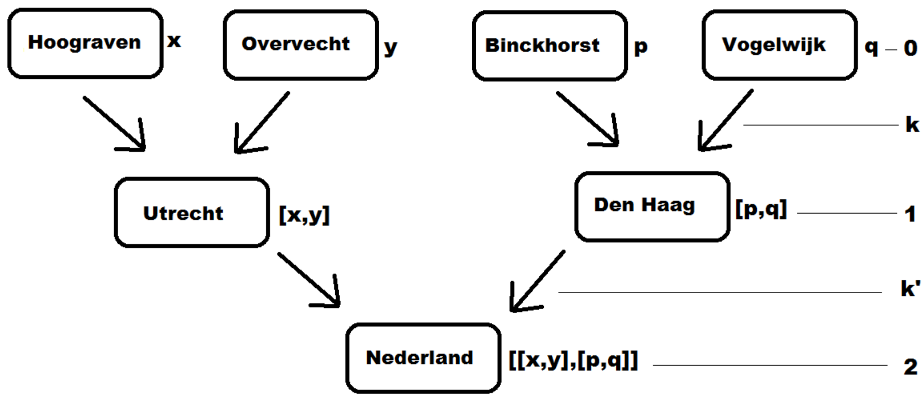
Figuur 3: Granulariteit 'hier'

## 3.3 Semantiek

Voor de semantiek van 'hier' kijken we eerst naar de semantiek van locatieven in het algemeen. Een locatie definieer ik als een verzameling objecten die zich op die locatie bevinden. Zo is Utrecht bijvoorbeeld {Drift 21, Domtoren, Lotje, et cetera}. Zoals Flieger (2009) uitlegt, heb je in een locatieve zin een Figuur (Figure) en een Grond (Ground). Het Figuur is het onderwerp, waarover je iets over de locatie gaat zeggen, en de Grond is een constante waardoor je weet over welke plek het gaat. In de zin 'Rosalie is in Antwerpen' is Rosalie het Figuur en Antwerpen de Grond. Je weet nu waar Rosalie is, doordat je weet waar Antwerpen is. Daarnaast is aan de verzameling objecten van die locatie iets toegevoegd: Rosalie. Als iemand hierna zegt dat Albert bij Rosalie is, weet je ook dat Albert in Antwerpen is. Een ander voorbeeld is 'de bal vliegt door de lucht' waar de bal het Figuur is en de lucht de Grond. Het gegeven werkwoord maakt in principe niet uit: ook 'Rosalie vindt het leuk in Antwerpen' of 'de bal verandert van richting in de lucht' zijn goede zinnen waarbij de locatie van het Figuur gegeven wordt door de Grond. De relatie tussen het Figuur en de Grond is over het algemeen zo dat het 'meest verplaatsbare' het Figuur is ('Rosalie is in Antwerpen' en niet 'Antwerpen is om Rosalie'). Bij 'hier' zijn er ook een Figuur en een Grond aan te wijzen. Het Figuur is het onderwerp en 'hier' is de Grond. De Grond lijkt nu geen constante, maar verwijst altijd naar de locatie van de spreker.

Zinnen met locatieven zijn dus over het algemeen opgebouwd uit een Figuur, een actie (zijn, vliegen, het leuk vinden) en een Grond. Het Figuur (F) is hierbij degene die de actie (act) verricht en de Grond (G) is het 'resultaat' daarvan, dus act(F) = G. Dit is ook te zien in onze voorbeelden: zijn(Rosalie) = Antwerpen, vliegen(bal) = lucht. Aangezien de actie zelf niet erg relevant is voor ons model, kunnen we deze weglaten en van G ons predicaat maken. Hierdoor krijg je G(F) met een bepaalde waarheidswaarde, bijvoorbeeld Antwerpen(Rosalie) = waar en tafel(bal) = onwaar.

We hebben nu eenzelfde soort predicaat-entiteitsstructuur voor locatieve zinnen als we eerder zagen bij de muziekgenres (zie figuur 4). Stel, ik ben een entiteit en ik zeg "Ik woon in Utrecht". Dan weet je weliswaar dat ik niet in Den Haag woon, maar je weet niet of ik in Hoograven of Overvecht (of een andere wijk van Utrecht) woon, maar kennelijk is dat niet relevant. Op dezelfde manier kunnen we het woord 'hier' zien. 'Hier' is een locatie en dus een predicaat. De propositie hier(kat) kan waar of onwaar zijn. Of het waar is of niet, hangt af van de granulariteit die je kiest. Als de kat van de buren je huis is binnengewandeld, kun je hen opbellen en zeggen: "De kat is hier". Dit is dan waar voor de granulariteit van je hele huis (in figuur 3b is dat $G_2$), maar misschien niet waar voor de kamer waarin je bent ($G_1$). De waarheid van de zin is dus afhankelijk van het granulariteitsniveau.




Figuur 4: Granulariteit locaties

# 4 'Hier' en simplificatie

In dit hoofdstuk ga ik de granulariteitstheorie toepassen op 'hier'. Onze wereld $T_0$ bestaat uit plekken waar iemand kan zijn en een persoon die ergens is. Voor het gemak houden we $S_0$ op {Jan} en $P_0$ op {Nederland, België, Utrecht, Den Haag, Gent, Hoograven, Overvecht, Vogelwijk}. Stel dat het waar is dat Jan in Overvecht is, dan krijg je deze waarheidstabel in $T_0$:

| | | | |
|---|---|---|---|
| Nederland(Jan) | 1 | België(Jan) | 0 |
| Utrecht(Jan) | 1 | Den Haag(Jan) | 0 |
| Gent(Jan) | 0 | Hoograven(Jan) | 0 |
| Overvecht(Jan) | 1 | Vogelwijk(Jan) | 0 |

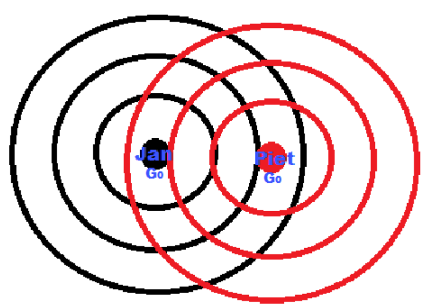
Stel dat het ons niet uitmaakt in welke wijk Jan zich bevindt, maar alleen in welke stad. Onze relevante predicaten worden dan: {Utrecht, Den Haag, Gent}. We hebben nu een toewijzing k nodig, zodat alle entiteiten (in ons geval maar één) in dezelfde equivalentieklasse komen dan en slechts dan als ze in dezelfde stad zijn (en dus in $T_1$ ononderscheidbaar zullen zijn). In $T_0$ geldt Utrecht(Jan) dus $k$(Utrecht)($k$(Jan)) is waar. Nu hebben we één equivalentieklasse met predicaat $k$(Utrecht): [Jan]. $T_0$ is succesvol versimpeld, dus met locatieve zinnen kunnen we de granulariteitstheorie toepassen. Is dit ook mogelijk met het woord 'hier'?

'Hier' is een locatie, dus in ons model een predicaat. Het doel is om een wereld $T_0$ en een versimpelde wereld $T_1$ te kunnen geven zodat 'hier' in $T_0$ heel precies is, bijvoorbeeld alleen de vierkante meter waar de spreker staat, en $T_1$ wat groffer, bijvoorbeeld de kamer waarin iemand is. Als we 'hier' direct in ons granulariteitsmodel zouden willen passen, lopen we tegen een probleem aan. Als we 'hier' als predicaat kiezen, is er slechts één predicaat. Dat betekent dat er in $T_1$ ($T_2$, $T_3$, et cetera) ook alleen maar het predicaat 'hier' kan bestaan, dat niet iets anders betekent dan in $T_0$.

Om de theorie toch toe te kunnen passen, hebben we een manier nodig om 'hier' op verschillende granulariteiten aan te kunnen geven. We gaan ervan uit dat zich in $T_0$ verschillende predicaten bevinden met een bepaalde betekenis van 'hier' en die duiden we aan met hier$^n$. Hier$^0$ betekent de exacte plek waar de spreker zich bevindt en meer niet en kunnen we dus ook zien als de verzameling {spreker}. Hier$^1$ is dan bijvoorbeeld de verzameling objecten die zich in de kamer bevinden, bijvoorbeeld {spreker, schommelstoel, kat, et cetera}. Hier$^n$ is de grootste granulariteit die we ons kunnen voorstellen en is de verzameling van alles wat we kunnen bedenken. De verzameling van $n - 1$ is altijd een strikte deelverzameling van de verzameling $n$; de verzameling (en dus de locatie) wordt steeds groter bij het toenemen van de granulariteit. Verder geldt dat de verzameling V = {hier$^0$, hier$^1$, ..., hier$^n$} een partiële ordening is. Als hier$^{14}$ de stad is en iemand staat in het zuidoosten van deze stad, kan hier$^{13a}$ bijvoorbeeld het oosten van de stad zijn en hier$^{13b}$ het zuiden van de stad. Deze delen, hier$^{13a}$ en hier$^{13b}$, overlappen elkaar gedeeltelijk, maar zijn beide een deelverzameling van hier$^{14}$ (de stad) en beide een superset van hier$^{12}$ (de wijk). Het predicaat hier$^{13}$ kan het best beschreven worden als 'een deel van de stad'.

De waarde van $n$ wordt (meestal onbewust) vastgesteld door de spreker en moet geïnterpreteerd worden door de luisteraar. Hierbij gaan we ervan uit dat de spreker zo informatief mogelijk is en dus de kleinst mogelijke granulariteit gebruikt. Als een spreker in Utrecht zegt dat hij 'hier' geboren is terwijl hij in Groningen geboren is, is dat alleen relevant als hij wil aangeven dat hij in Nederland (of in een groffere granulariteit) geboren is. Nederland is namelijk de kleinste granulariteit waar zowel de locatie van de spreker als de plek waar hij daadwerkelijk geboren is, toe behoren. Dit valt onder Grice's maxim of quantity (Grice, 1989): "One tries to be as informative as one possibly can, and gives as much information as is needed, and no more." Kortom, geef zo veel informatie als nodig en niet meer dan dat. Als iemand zegt: "De kat is hier", terwijl hijzelf elders in Nederland op vakantie is, dan is de zin waar bij de granulariteit hier$^{50}$ (Nederland), maar toch zullen de meeste mensen vinden dat de zin niet waar is, omdat ze uitgaan van een andere granulariteit.

Verder gaan we er in ons model van uit dat een zin met 'hier' erin geuit wordt door een spreker en dat die spreker hetzelfde blijft in het model. In $T_0$ zal deze spreker altijd in de verzameling van hier$^0$ zitten. Per keer dat het woord 'hier' voorkomt, kan het zijn dat er een andere spreker is. Jan kan bijvoorbeeld vragen: "Ben je hier?", waarop Piet roept: "Nee, ik ben hier!"(zie figuur 5).




Figuur 5: Twee verschillende uitingen van hier.

Ter verduidelijking geef ik een voorbeeld met drie granulariteiten: hier$^0$, hier$^1$ (de kamer waarin de spreker zich bevindt) en hier$^2$ (het gebouw waarin de spreker zich bevindt). We kunnen nu in $T_0$ onze entiteiten kiezen: $S_0$ = {Piet, Jan, Sam}. We stellen $P_0$ = {hier$^0$, hier$^1$, hier$^2$} zoals net gedefiniëerd. Stel, Piet en Jan zijn in dezelfde kamer en Sam is elders in het gebouw. Piet doet de uitspraak: "Jan is hier" en daarna: "Sam is hier". Beide uitspraken zijn waar, echter zij zijn waar op een ander granulariteitsniveau. Pragmatisch gezien is het raar dat Piet ineens van granulariteitsniveau wisselt, maar dat laten we voor nu buiten beschouwing. We krijgen nu de volgende waarheidstabel in $T_0$:

| | | | | | |
|---|---|---|---|---|---|
| $\text{hier}^0$(Piet) | 1 | $\text{hier}^0$(Jan) | 0 | $\text{hier}^0$(Sam) | 0 |
| $\text{hier}^1$(Piet) | 1 | $\text{hier}^1$(Jan) | 1 | $\text{hier}^1$(Sam) | 0 |
| $\text{hier}^2$(Piet) | 1 | $\text{hier}^2$(Jan) | 1 | $\text{hier}^2$(Sam) | 1 |

We kunnen nu een toewijzing k maken, waarbij alleen $\text{hier}^1$, de kamer, een relevant predicaat is. We krijgen dan $S_1 = \{$ [Piet, Jan], [Sam] $\}$ (dus 2 equivalentieklassen, één met alleen Sam en één met Jan en Piet, die nu ononderscheidbaar zijn) en $P_1 = \{k(\text{hier}^1)\}$. De equivalentieklasse waar Jan en Piet toe behoren, kunnen we nu weer simpelweg 'hier' noemen. Immers, 'hier' is de uiting die is gedaan om duidelijk te maken wie er in de kamer zijn. We weten nu dat Jan ook in de verzameling objecten van $\text{hier}^1$ zit en Sam niet. Als we hadden gesteld dat $S_0$ het domein van alle mensen hadden is, dan behoorde Sam samen met alle andere mensen tot in één equivalentieklasse zonder naam: 'de rest', 'niet hier'. Nu hebben we onze wereld ($T_0$) versimpeld: we weten wie er in de kamer is bij de spreker en wie niet en dat was ons enige doel.

We hadden ook kunnen versimpelen (laten we deze wereld $T_2$ noemen) met als relevant predicaat $\text{hier}^2$: 'hier' is het gebouw waar de spreker is. In dat geval krijgen we equivalentieklasse $\{$[Piet, Jan, Sam]$\}$. Voor hen is allemaal $k(\text{hier}^2)$ waar. Op dezelfde manier kunnen we nu zeggen dat "Sam is hier" en "Jan is hier" allebei waar zijn als ons relevante predicaat $\text{hier}^2$ is.

## 5 Conclusie en discussie

We hebben gekeken naar de granulariteit van 'hier' en succesvol Hobbs' simplificatietheorie toegepast op het woord 'hier'. Om dit te doen, moesten we een extra variabele aan het predicaat toevoegen die de granulariteit van 'hier' aangeeft. Een vraag is echter, of we de granulariteitstheorie nog wel nodig hebben, als deze versimpeling alleen maar lukt als hier$^n$ gebruikt wordt. Om die variabele toe te voegen en te weten wat de relevante predicaten worden, moeten we immers eerst weten op welk granulariteitsniveau er gecommuniceerd wordt of moet worden. Ik denk daarom dat we niet veel hebben aan Hobbs' theorie bij dit fenomeen. Deze theorie lijkt niet veel nuttigs toe te voegen. Wel lijkt het me belangrijk om dit soort onderwerpen vanuit de logica te blijven bekijken; als taalproblemen enkel worden aangepakt met machine learning, is interpreteerbaarheid een probleem. We weten het probleem en de oplossing die de computer geeft, maar niet <u>waarom</u> hij die oplossing biedt en niet een andere. Misschien interpreteert de computer een bepaalde 'hier' als hier$^7$, maar waarom niet als hier$^6$ of hier$^8$? Daarnaast kan het zo zijn dat de computer het goed doet voor granulariteiten hier$^0$ tot en met hier$^{50}$, maar voor hier$^{51}$ niet, omdat dat een granulariteit is die heel weinig gebruikt wordt. Dit maakt het gebruiken hiervan moeilijker te verantwoorden en het is is ook lastiger kleine aanpassingen te maken.

Zoals ik in de inleiding al zei, werkt de computer meestal op $T_0$-niveau, met een heel fijne granulariteit: te precies en te uitgebreid. Door 'hier' (en andere talige fenomenen) te modelleren, kan een computer beter met die granulariteit spelen, op een manier zoals mensen dat ook doen. Dit is bevorderlijk voor de communicatie tussen mens en computer.

Ik heb in deze scriptie niet gekeken naar het achterhalen van wat relevante predicaten zijn. Het blijft een erg lastig probleem om er achter te komen welke predicaten uit $T_0$ er in $\mathbf{R}$ komen. Iets wat zou kunnen helpen zijn zogenaamde approximatoren (Engels: approximators)(Sauerland & Stateva, 2007). Dat zijn woorden als 'ongeveer' en 'precies' die voor het woordje 'hier' kunnen worden gezet. Approximatoren zouden kunnen aangeven hoe groot je granulariteit ongeveer gaat zijn. Vergelijk de volgende twee zinnen:

**(i)** "Ik ben precies hier"

**(ii)** "Ik ben ongeveer hier"

Bij zin (i) moeten we kijken naar een precieze granulariteit, bijvoorbeeld hier$^0$ of hier$^1$. Dat weten we door het woordje 'precies'. Net zo goed kunnen we uit 'ongeveer' in zin (ii) afleiden dat we te maken hebben met een groffere granulariteit bijvoorbeeld rond hier$^{10}$. Er zou in vervolgstudie dus gekeken kunnen worden naar approximatoren en of die hulp bieden bij het bepalen van de granulariteit.

Er is de laatste tijd binnen het vakgebied van kunstmatige intelligentie veel onderzoek gaande naar tekstbegrip door computers. Verschillende technieken worden gebruikt om tekstbegrip sterker te maken en relaties tussen woorden te vinden. Zoals ik al eerder zei, lijkt het mij belangrijk dit niet alleen met machine learning aan te pakken. Het logisch modelleren van woorden als 'hier' kan helpen bij het verkrijgen van meer tekstbegrip van computers. De theorie die in deze scriptie is gegeven is een goede basis hiervoor. Wel moet er nog verder gekeken worden naar hoe we de theorie daadwerkelijk aan een computer kunnen leren en of dat werkt. Een probleem dat hierbij nog kan optreden, is de keuze welke predicaten er precies in $T_0$ zitten en wat die betekenen. $T_0$ bevat de predicaten $\text{hier}^0$ tot en met $\text{hier}^n$, maar hoeveel is n? En wat betekent $\text{hier}^5$ in de praktijk?

Een volgend verkennend onderzoek kan worden uitgevoerd naar andere taalfenomenen die granulariteit bevatten en het modelleren hiervan. De manier waarop ik heb aangetoond dat het bij 'hier' kan, kan misschien ook worden gebruikt voor andere fenomenen. Het is waarschijnlijk op dezelfde manier toepasbaar bij andere indexicale woorden als 'daar' en 'wij'. Buisman (2019) heeft succesvol op ongeveer dezelfde manier de theorie toegepast op bepaalde hoofdtelwoorden wat weer wat verder afligt van het fenomeen dat ik bekeken heb. Dat het met verschillende fenomenen mogelijk is, biedt hoop voor taal in het algemeen.

# 6 Bronnen


1. Buisman, V.C. (2019). Een theorie van granulariteit toegepast op telwoorden. Geraadpleegd in scriptiearchief Universiteit Utrecht, http://studenttheses.library.uu.nl.

2. Flieger, J.C. (2009). Gradable Adjectives and the Semantics of Locatives, Institute for Communicating and Collaborative Systems, University of Edinburgh.

3. Grice, P. (1989). Logic and Conversation, Studies in the Way of Words, Harvard University Press.

4. Hobbs, J.R. (1985). Granularity. In Readings in qualitative reasoning about physical system, 542-545. Morgan Kaufmann.

5. Kaplan, D. (1977), Demonstratives, An Essay on the Semantics, Logic, Metaphysics, and Epistemology of Demonstratives and Other Indexicals, 489-492.

6. Lasersohn, P. (1999). Pragmatic Halos. Language, 75(3), 522-551. doi:10.2307/417059.

7. Quer, J. (2011), Reporting and Quoting in Signed Discourse, ICREA & Universitat Pompeu Fabra.

8. Sauerland, U. & Stateva, P. (2007). Scalar vs. epistemic vagueness: Evidence from approximators. In Semantics and Linguistic Theory, Vol. 17, pp. 228-245.

9. Wetzel, L. (2006). Type and tokens, Stanford Encyclopedia of Philosophy.