



**Universiteit Utrecht**

Verbetering van Natural Language Generation: Een vergelijking van refererende expressies door mensen en computers

Britt Leker  
5912520

8 november 2019

Bachelor Kunstmatige Intelligentie

Bachelor Eindwerkstuk (7,5 ECTS)

Begeleider: dr. Rick Nouwen

Tweede beoordelaar: dr. Stella Donker

# Inhoud

Abstract .....	2
1. Inleiding.....	3
2. Natural Language Generation.....	4
2.1 Achtergrond.....	4
2.2 Refererende expressies.....	5
2.3 Soorten algoritmen.....	7
2.3.1 Klassieke regelgebaseerde benadering.....	7
2.3.2 Probabilistische NLG-systemen .....	9
2.3.3 Verschillen .....	10
2.3.4 Wat is “beter”? .....	10
3. Hoe doen mensen NLG en waarom.....	10
3.1 Overspecificatie.....	11
3.2 Mogelijke verklaringen voor overspecificatie .....	12
3.3 Wat betekent dit voor NLG algoritmen? .....	14
4. Discussie .....	14
Bronnen.....	17

## **Abstract**

Een onderdeel van Natural Language Generation (NLG) is het genereren van refererende expressies. Er bestaan verscheidene systemen met verschillende aanpakken om zulk soort expressies te genereren. In dit verslag worden deze verschillende manieren bekeken en vergeleken met de manier waarop mensen refererende expressies genereren. Mensen doen vaak aan overspecificatie en geven dus in hun refererende expressies meer informatie dan strikt noodzakelijk is om het object waar naar gerefereerd wordt te identificeren. Dit verslag stelt de vraag in hoeverre het wenselijk is dat NLG-systemen hun refererende expressies op dezelfde manier genereren als mensen dat doen. Om een antwoord op deze vraag te vinden wordt gekeken naar de manier waarop sprekers refererende expressies genereren, waarom ze dit op deze manier doen en of het beter zou zijn voor NLG-systemen om dit ook te doen of juist niet.

# 1. Inleiding

Kunstmatige intelligentie is in de laatste jaren hard aan het groeien en in al die jaren is er ook een discussie gaande over wat kunstmatige intelligentie precies moet zijn. Er wordt gesproken over zogenaamde *weak AI* en *strong AI*. Strong AI wil zeggen computers die beschikken over meer menselijke intelligentie die kan worden toegepast op ieder soort probleem, of zelfs machines die bewustzijn ervaren. Weak AI is enkel een algemene intelligentie dat bedoeld is om specifieke problemen op te lossen en de mens te assisteren. Om te bepalen wat kunstmatige intelligentie zou moeten zijn moet er nagedacht worden over het precieze doel van kunstmatige intelligentie. Is het een hulpmiddel bedoeld om mensen te assisteren met bepaalde taken, of moet het meer een echt mens voorstellen met een completere intelligentie en een eigen bewustzijn?

Dezelfde vraag kan gesteld worden voor een bepaald onderdeel van kunstmatige intelligentie, zoals taal. Dit verslag zal zich dan ook focussen op Natural Language Generation (NLG), het genereren van natuurlijk klinkende taal. Moet een computer taal produceren op de meest efficiënte manier mogelijk en dus enkel gebruik maken van de nodige kennis om dit probleem op te lossen, of moet de computer meer zoals de mens 'denken' en expressies genereren die op die van mensen lijken en dus niet altijd de meest informatieve keuze zijn? In dit verslag zal gefocust worden op het genereren van refererende expressies.

Wanneer er in een conversatie naar een bepaald object verwezen wordt met behulp van taal gebeurt dit door middel van refererende expressies. Voorbeelden van refererende expressies kunnen zijn "de blauwe beker", "de lange man" of "de labrador". Een goede refererende expressie stelt de luisteraar in staat het object waar de spreker naar refereert te identificeren in de context. Het object waar naar verwezen wordt door de spreker wordt de referent, of het *target* genoemd. De overige objecten in de context waar niet naar verwezen wordt zijn de *distractors*, of de contrastverzameling. Om naar een object te verwijzen moet er een beschrijving worden gemaakt van het object. De eigenschappen die deel uitmaken van deze beschrijving worden *attributen* genoemd. Tijdens het genereren van deze refererende expressies moet de spreker een overweging maken welke attributen hij aan de beschrijving van het target toe wil voegen om de communicatie te bevorderen. Wat een afweging kan zijn in het kiezen van attributen voor de beschrijving is de onderscheidende kracht van de attributen. Wanneer het target een attribuut bevat dat zich onderscheidt van dat attribuut in de contrastverzameling heeft het een grote onderscheidende kracht. In bijvoorbeeld een context met twee blauwe bekers van verschillende grootten heeft het attribuut *grootte* een grotere onderscheidende kracht dan het attribuut *kleur*. Het kan efficiënt zijn om in de beschrijving de attributen te gebruiken met de meeste onderscheidende kracht zodat er zo min mogelijk woorden nodig zijn om toch een goede refererende expressie te genereren. Dit doen mensen echter niet altijd. Wanneer er meer attributen worden gebruikt in de refererende expressie dan noodzakelijk is om het target te identificeren is er sprake van overspecificatie. De beschrijving bevat dus meer specificaties dan noodzakelijk om te achterhalen naar welk object gerefereerd wordt. Dit lijkt niet altijd even efficiënt en daarom is voor sommige NLG-systemen gekozen om deze manier van refererende expressies genereren niet aan te houden.

Op dit moment bestaan er verschillende NLG-systemen die allemaal een andere aanpak hebben. De verschillen in de systemen resulteren ook tot verschillen in de beschrijvingen die ze genereren voor bepaalde gevallen. Sommige van deze gegenereerde beschrijvingen lijken erg op de menselijke refererende expressies. Er zijn ook systemen die minder kijken naar hoe mensen NLG uitvoeren en zich focussen op het genereren van de meest effectieve refererende expressie. Dit verslag zal zich focussen op de volgende vraag: In hoeverre is het wenselijk dat computers NLG uitvoeren op de manier waarop mensen het doen?

In dit verslag zal door middel van literatuuronderzoek een poging gedaan worden dichter bij een antwoord op deze vraag te komen. Twee verschillende benaderingen voor NLG-systemen zullen nader bekeken worden: de klassieke regelgebaseerde benadering, en de Bayesiaanse benadering. De verschillen tussen deze systemen zullen bekeken worden en er zal een vergelijking worden gemaakt met de manier waarop mensen refererende expressies genereren. Dit bevat een uitleg over wat overspecificatie precies is en mogelijke verklaringen waarom mensen aan overspecificatie doen. Ook wordt er gekeken naar wat dit betekent voor automatische systemen en of het gunstig kan zijn voor NLG om de menselijke NLG na te bootsen.

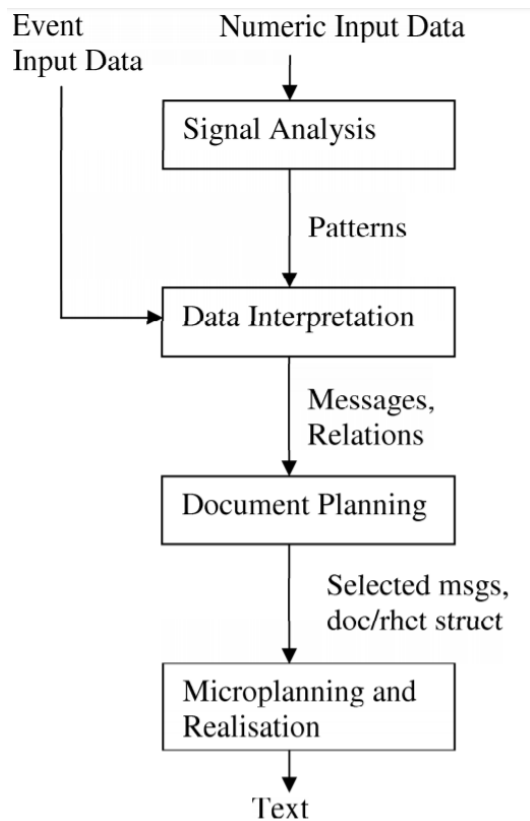
Eerst zal er wat achtergrondinformatie worden gegeven over Natural Language Generation en refererende expressies in het algemeen. Vervolgens zullen er een aantal verschillende soorten NLG-systemen om refererende expressies te genereren besproken worden en met elkaar vergeleken worden. Daarna zal dieper worden ingegaan op de manier waarop mensen refererende expressies genereren en wat mogelijke verklaringen zijn voor dit taalgedrag. Ook zal vergeleken worden hoe de refererende expressies van mensen verschillen van die van computers en wat deze verschillen betekenen voor de NLG-algoritmen. Tot slot zal de discussie terugkomen op de vragen die in dit verslag gesteld zijn en zal de relevantie van dit onderwerp met betrekking tot kunstmatige intelligentie verder besproken worden.

## 2. Natural Language Generation

### 2.1 Achtergrond

Over een algemene definitie van Natural Language Generation bestaat enige discussie. Dit is voornamelijk te wijten aan het feit dat de input in NLG vaak verschilt per systeem (Evans et al. 2002). Wel kan gezegd worden dat NLG zich bezig houdt met het construeren van computersystemen die teksten in natuurlijke taal kunnen produceren. Dit wordt bijvoorbeeld gebruikt in *data-to-text* systemen. Het genereren van deze teksten komt meestal voort uit een onderliggende niet-linguïstische representatie van informatie. Zulke systemen maken gebruik van data analyse en taalverwerking om allerlei soorten teksten te genereren. Data-to-text generatie wordt bijvoorbeeld gebruikt om samenvattingen te maken van langere teksten, teksten te vertalen naar een andere taal, spelling- en grammaticacontrole of om nieuwsberichten te genereren met behulp van sensordata (Gatt & Krahmer, 2018). Reiter and Dale (2000) ontwierpen bijvoorbeeld een model dat tekstuele weersvoorspellingen geeft vanuit numerieke weersvoorspellingdata. In dit model bestaat de input uit vier delen: een kennisbron, een communicatief doel, een gebruikersmodel en een conversatiegeschiedenis. Al deze inputdata is dus weergegeven in een niet-linguïstische representatie.

*Data-to-text* systemen zijn één van de toepassingen van NLG. Dit zijn systemen die tekst genereren met behulp van niet-linguïstische input door gebruik te maken van data-analyse en taalverwerking. In deze algemene architectuur is NLG onderverdeeld in drie verschillende stadia: *Document Planning*, *Microplanning* en *Realisation*. Later heeft Reiter hier nog twee stadia aan toegevoegd: *Signal Analysis* en *Data Interpretation*. Dit resulteerde in het ontwerp voor een algemene NLG dat te zien is in Figuur 1 (Reiter, 2007). Deze toevoeging maakte het mogelijk voor de input data van het systeem om meer dan alleen kennis te zijn.



**Figuur 1: De algemene NLG architectuur van Reiter (2007)**

Het is de taak van de computer om informatie aan mensen over te brengen op een begrijpelijke manier. In sommige gevallen kan de data die het systeem als output geeft alleen worden geïnterpreteerd door experts. Echter, vaak is het noodzakelijk dat systemen deze normaalgesproken ingewikkelde data zodanig kunnen weergeven dat ook gebruikers zonder deze expertise het kunnen begrijpen (Reiter & Dale, 1997). Daarom is NLG erg belangrijk in de kunstmatige intelligentie. Door het genereren van natuurlijke taal kunnen computers gemakkelijker communiceren met hun gebruikers. Aangezien 'natuurlijke taal' een erg breed veld is zal er in de rest van dit verslag gefocust worden op een van de meest voorkomende taken binnen NLG: het genereren van refererende expressies.

## 2.2 Refererende expressies

Een refererende expressie is een zin of zinsdeel dat bedoeld is om een bepaald object te identificeren. Om te begrijpen hoe deze expressies gegenereerd kunnen worden is het belangrijk dat eerst het een en ander wordt uitgelegd over het taalgedrag van mensen. Volgens Grice et al. (1975) hoeven mensen in een conversatie niet alles letterlijk tegen elkaar te zeggen om toch begrepen te worden. Wanneer persoon A bijvoorbeeld tegen persoon B zegt: "Ik heb geen benzine meer" en persoon B reageert met "Er is een tankstation hier om de hoek" dan impliceert B dat hij denkt dat het tankstation open is en dat A er zal kunnen tanken, zonder dat alles letterlijk hoeft worden uitgesproken. Dit soort niet-letterlijke inferenties worden conversationale implicaturen genoemd. En hiervoor is het coöperatieve principe van Grice ontworpen, bestaande uit vier maxims waaraan zowel de spreker als de luisteraar zich over het algemeen aan houden in een sociale situatie.

1. Maxime van kwantiteit: wees zo informatief mogelijk, maar niet meer dan nodig is, gezien het doel van het gesprek
2. Maxime van kwaliteit: zeg niets waarvan je gelooft dat het onwaar is of waar je onvoldoende bewijs voor hebt.

3. Maxime van relevantie: wees relevant.
4. Maxime van stijl: vermijd onduidelijkheden en ambiguïteit en wees kort en ordelijk.

Deze maximes zijn ontworpen om een maximaal effectieve informatie-uitwisseling te creëren. We gaan er over het algemeen vanuit dat de mensen met wie we spreken zich aan deze maximes houden (Dale & Reiter, 1995). Dit is dus belangrijk wanneer er een NLG-systeem wordt gemaakt dat taal op een begrijpelijke manier moet overbrengen aan luisteraars. De uitspraak moet genoeg informatie bevatten om de boodschap over te brengen, maar hoeft dus blijkbaar niet alles letterlijk uit te spreken om toch begrepen te worden. Daarom baseren veel systemen hun algoritme voor het genereren van refererende expressies op dit coöperatieve principe. Maar dan nog zijn er verschillende manieren om expressies te genereren waarmee naar een object wordt verwezen.

Een manier om naar objecten te refereren is met behulp van relationele beschrijvingen. Dit zijn beschrijvingen die een relatie met een ander object bevatten. Een voorbeeld van een dergelijk relationele expressie is “de hond in het hondenhek”, waarbij er naar het target (de hond) wordt gerefereerd door gebruik zijn relatie met een ander object (het hondenhek) te omschrijven. (Krahmer et al., 2003) Het *Relation Algorithm* (Dale & Haddock, 1991) genereert refererende expressies die zulke relationele beschrijvingen bevatten. Omdat de relationele eigenschappen vaak de meeste onderscheidende kracht hebben lijkt het logisch om deze te gebruiken in de beschrijving. Wanneer er verwezen zou moeten worden naar lade nummer 2 in Figuur 3 kan het algoritme een expressie genereren zoals “de lade boven de lade boven de lade boven de roze lade”. Dit zou alleen voor mensen geen typische beschrijving zijn om te produceren als het ook anders kan. Zij zouden eerder een expressie genereren zoals “de oranje lade boven de blauwe lade”, of “de oranje lade in de bovenste rij”. Sprekers maken dan ook minder gebruik van relationele beschrijvingen in hun refererende expressies dan het *Relation Algorithm* en gebruiken eerder de geprefereerde attributen zoals kleur (Viethen & Dale, 2006).

1 (blue)	2 (orange)	3 (pink)	4 (yellow)
8 (blue)	7 (blue)	6 (yellow)	5 (pink)
9 (orange)	10 (blue)	11 (yellow)	12 (orange)
16 (yellow)	15 (pink)	14 (orange)	13 (pink)

**Figuur 2: De ladekast (Viethen & Dale, 2006).**

Hier zal dit verslag echter verder niet op focussen. Er zal enkel gekeken worden naar simpele objectverwijzingen bestaande uit een zelfstandig naamwoord (zoals “lade”) en eventueel één of meerdere bijvoeglijk naamwoorden (zoals “oranje”) en geen relationele of lokaliserende beschrijvingen. Er bestaan verschillende soorten NLG-systemen om zulk soort refererende expressies te genereren. In de volgende sectie zullen enkele van deze systemen besproken worden en zal een onderscheid worden gemaakt tussen twee categorieën binnen deze systemen.

## 2.3 Soorten algoritmen

Het eerste echte algoritme dat zich bezig hield met het genereren van refererende expressies was het *Full Brevity* algoritme (Dale, 1989). Dit algoritme is bedoeld om de kortst mogelijke refererende expressies te genereren waarmee de referent kan worden geïdentificeerd en houdt daarom erg streng vast aan Grice's maxims van kwantiteit, relevantie en stijl. Hierdoor bevatten de expressies genoeg informatie, maar niet meer dan nodig en is deze informatie relevant voor het identificeren van de referent en duidelijk en ordelijk genoeg voor een snelle en correcte identificatie.

Hieruit zijn velen andere algoritmen ontstaan. Dale en Haddock (1991) wilden af van de beperkingen van de beschrijvingen bestaande uit één eigenschap en introduceerden het *Relation Algorithm* dat expressies genereerde die relaties bevatten. De *Greedy Heuristic Interpretation* van Dale & Reiter (1992) leidde tot een algoritme dat ook meerdere attributen gebruikte in zijn refererende expressies. Maar soms leidde dit tot een langere beschrijving dan nodig was om het target te identificeren. Vervolgens ontwierpen zij het *Incremental Algorithm* (Dale & Reiter, 1995): een veel simpeler algoritme dat meer lijkt op hoe mensen refererende expressies genereren. Dit is in essentie het algoritme dat hieronder in sectie 2.3.1 *Klassieke regelgebaseerde benadering* wordt besproken.

Binnen NLG bestaan er twee brede categorieën waar de meeste systemen onder vallen: de kennisgebaseerde systemen (ook wel regelgebaseerde systemen genoemd) en de statistische systemen. Kennisgebaseerde systemen hebben expertise van een bepaald domein nodig om zo regels te maken voor iedere stap van de NLG. Statistische systemen maken gebruik van corpus data om zo de regels te leren voor de verschillende componenten van NLG-systemen.

Beide soorten systemen hebben zowel voordelen als nadelen. De kennisgebaseerde systemen produceren over het algemeen teksten van een hogere kwaliteit dan probabilistische systemen en bevatten minder fouten. Een nadeel is dat deze systemen gelimiteerd zijn in hun taalkundige omvang, data uit het domein niet kunnen hergebruiken, en erg duur kunnen zijn om te maken. Hoewel de probabilistische systemen goedkoper zijn, data uit het domein kunnen hergebruiken en een grotere taalkundige omvang kunnen hebben, bevat de output vaak fouten en kan het gebrek aan regels de tekst onsamenhangend maken (Kondadadi et al. 2013).

In deze scriptie zal de focus liggen op twee verschillende benaderingen van NLG: binnen de kennisgebaseerde systemen zal gekeken worden naar de klassieke regelgebaseerde benadering die refererende expressies genereert door het stapsgewijs volgen van vooropgestelde regels; en binnen de statistische systemen zal enkel gefocust worden op de moderne Bayesiaanse benadering die gebruik maakt van probabilistische berekeningen om refererende expressies te genereren.

### 2.3.1 Klassieke regelgebaseerde benadering

Alle NLG-systemen streven naar het produceren van natuurlijke refererende expressies. Dit betekent in veel gevallen dat er geprobeerd wordt om het systeem dezelfde soorten principes te laten volgen als mensen lijken te gebruiken wanneer zij spreken. Een voorbeeld van een dergelijk principe is het aanhouden van Grice's maxims (Grice, 1975) tijdens het spreken. Dale en Reiter (1995) beweren dat er eigenlijk nog een vijfde maxime aan de lijst kan worden toegevoegd. Wat namelijk ook deel uitmaakt van conversationeel gedrag is het gebruik van woorden op het taxonomische basisniveau.

Sprekers kunnen met verschillende termen naar eenzelfde object verwijzen. Bijvoorbeeld in een context waar slechts één hond aanwezig is zal hier waarschijnlijk naar worden gerefereerd met het woord *hond*. *Hond* is namelijk een basisniveau in de taxonomische hiërarchie. Wanneer er op een lager taxonomisch niveau naar de hond wordt verwezen, bijvoorbeeld met het woord *dalmatiër*, zou dit kunnen impliceren dat de spreker het noodzakelijk vindt om de soort hond te specificeren naar de luisteraar. Beide termen (*hond* en *dalmatiër*) vervullen in deze context het doel van de spreker om naar de hond te refereren, maar toch is het beter voor de spreker om het basisniveau te gebruiken



wanneer er geen reden is voor dergelijke specificatie. Het vijfde maxime is dan ook het maxime van lexicale voorkeur. Sprekers geven in veel gevallen ook de voorkeur aan een basisniveau (*hond*) wanneer ze niet worden beperkt door de context (Graf et al., 2016).

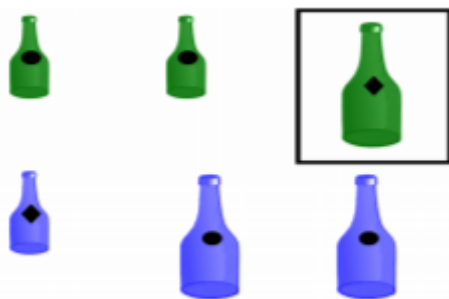
5. Maxime van lexicale voorkeur: maak gebruik het taxonomische basisniveau en andere geprefereerde klassen wanneer mogelijk.

Grice's maximes kunnen ook worden toegepast op refererende expressies.

1. Kwantiteit: een refererende expressie moet genoeg informatie bevatten zodat de luisteraar de referent kan identificeren, maar geen onnodige informatie.
2. Kwaliteit: een refererende expressie moet een accurate beschrijving zijn van de bedoelde referent.
3. Relevantie: een refererende expressie moet geen attributen bevatten die geen onderscheidingskracht hebben en dus niet helpen in het onderscheiden van de referent in de contrastverzameling.
4. Stijl: een refererende expressie moet zo kort mogelijk zijn.
5. Lexicale voorkeur: een refererende expressie moet gebruik maken van het taxonomische basisniveau en andere geprefereerde klassen wanneer mogelijk.

Het *Incremental Algorithm* van Dale en Reiter (1995) heeft als input een object waarnaar gerefereerd wordt en een lijst van objecten die corresponderen met de contrastverzameling. Als output geeft het een lijst met paren van de attributen en hun waarde waar de refererende expressie uit zal bestaan. Het algoritme bevat een verzameling met attributen waar menselijke sprekers de voorkeur aan geven, waaronder kleur, grootte, type en vorm. Het algoritme itereert door de geprefereerde attributen uit de input en controleert voor elk attribuut of zijn waarde andere leden van de contrastverzameling (die nog niet zijn uitgesloten) uitsluit. Als dit zo is wordt het attribuut toegevoegd aan de refererende expressie. Dit proces stopt zodra de gevormde refererende expressie alle leden uit de contrastverzameling uitsluit.

Neem bijvoorbeeld de context in Figuur 3, bestaande uit zes verschillende vazen. Deze vazen verschillen in kleur, grootte en in het symbool dat er op staat. De vaas in het kader is het target en de andere vijf maken de contrastverzameling. Wanneer het algoritme begint met het attribuut kleur zal de waarde *groen* drie vazen uitsluiten. Vervolgens kan het algoritme controleren op grootte, dit zal dan de overige twee distractors uitsluiten en het target is gevonden. De refererende expressie zal dan dus zijn: "groene grote vaas".



**Figuur 3: Zes verschillende vazen (Gatt et al., 2013).**

De refererende expressie bevat altijd een zelfstandig naamwoord, ook als dit geen onderscheidende kracht heeft. Dit systeem doet niet aan *backtracking*, wat betekent dat het mogelijk is voor een

refererende expressie om attributen te bevatten die overbodig worden gemaakt door opvolgende attributen. Zodra er dus gekozen is voor de eigenschap *groen* kan deze later niet meer uit de beschrijving worden geschrapt. De gegenereerde refererende expressie kan dus overspecificatie bevatten. Dit algoritme is snel, het houdt rekening met menselijke voorkeuren en capaciteiten, en het genereert korte expressies als output.

Dale en Reiter concluderen dat deze simpele en snelle manier waarschijnlijk ook de beste is, omdat dit het meest lijkt op wat mensen doen. Veel systemen zijn al gebaseerd op deze aanname dat het beter is om menselijk taalgedrag na te bootsen, met overspecificatie en al, zelfs als dit op het eerste gezicht niet de meest effectieve manier lijkt. Er zouden vraagtekens gezet kunnen worden bij de claim dat wat mensen doen automatisch beter is. Er zou ook gekeken moeten worden naar de situatie en wat precies het doel van de refererende expressies is.

### 2.3.2 Probabilistische NLG-systemen

De probabilistische systemen maken gebruik van corpus data om zo de regels te leren voor de verschillende componenten van NLG-systemen. Het Bayesiaanse model van Frank en Goodman (2012) berekent de waarschijnlijkheid dat een bepaalde referent wordt omschreven met een bepaald woord in een gegeven context. Om dit te berekenen is de *prior probability* nodig, en de *likelihood*. De *prior probability* is in dit model de kans dat er überhaupt naar een bepaalde referent *R* verwezen wordt. Er wordt aangenomen dat dit een gegeven is in de context. De *likelihood* is de kans dat een bepaald woord *W* gebruikt wordt om naar de referent *R* te verwijzen in context *C*. Dit wordt bepaald met behulp van het aantal referenten waar *W* naar kan refereren. Neem in de context van Figuur 3 het woord "fles", dit kan naar meer objecten refereren dan het woord "groen". Dat betekent dat de *likelihood* van "fles" kleiner is dan van "groen", omdat de kans kleiner is dat een spreker het woord "fles" zal gebruiken om de referent te omschrijven. Wanneer de *likelihood* met de *prior probability* wordt vermenigvuldigd is dit proportioneel aan de waarschijnlijkheid dat de referent *R* wordt bedoeld wanneer het woord *W* wordt gebruikt in context *C*. Hierbij wordt er aangenomen dat de spreker zich houdt aan Grice's coöperatieve principe en dus zo informatief mogelijk is, maar niet meer dan nodig, en dat de enkel relevante informatie wordt verschaft. Hoe informatief een woord is wordt hier bepaald door vast te stellen hoeveel onzekerheid over de referent het wegneemt. Om de expressies zo kort en informatief mogelijk te maken wordt er in dit model geen gebruik gemaakt van overspecificatie. Enkel de strikt noodzakelijke informatie die nodig is om de referent te identificeren wordt in de beschrijving opgenomen. In de context van Figuur 3 zou er dus gekozen worden voor het attribuut van het symbool, aangezien "vaas met de ruit" de meeste van de distractors uitsluit (namelijk vier) en dus de meeste onzekerheid over de referent wegneemt. Ook neemt dit model aan dat sprekers alleen refererende expressies genereren bestaande uit één eigenschap. Dit betekent dat er alleen expressies als "de groene vaas" of "de grote vaas", maar geen expressies als "de groene grote vaas" kunnen worden gegenereerd. Dit kan dus in sommige contexten problemen veroorzaken omdat één eigenschap niet genoeg is om alle distractors uit te sluiten. Daarbij is er geen voorkeurindeling van de eigenschappen, waardoor er mogelijk minder 'nuttige' eigenschappen worden gebruikt in de refererende expressie (Gatt, Kraemer, et al., 2013).

Een model dat wel rekening houdt met de voorkeur van sprekers is het *Probabilistic Referential Overspecification* (PRO) model (Gatt, Van Gompel et al., 2013). Dit model is ontworpen op basis van experimenten die referentiële keuzes, voorkeursgebaseerde heuristiek en non-determinisme nader onderzochten bij sprekers van Nederlands en Engels. Uit deze experimenten bleek dat sprekers liever gebruik maken van kleur dan aan grootte in hun refererende expressies. Het PRO model gaat ervan uit dat sprekers als eerste de eigenschap zullen kiezen die de referent volledig onderscheidt van de andere objecten. Wanneer er meerdere van deze eigenschappen zijn zal er een probabilistisch gekozen worden op basis van voorkeur. Vervolgens kunnen sprekers een tweede eigenschap kiezen waarvan de waarschijnlijkheid opnieuw afhangt van voorkeur. Dit model zou dus voor de situatie in

Figuur 3 als eerste kiezen voor de eigenschap *groen*, aangezien mensen de meeste voorkeur geven aan het attribuut *kleur*. Vervolgens zal bepaald moeten worden wat het tweede favoriete attribuut is. Stel dat *grootte* verkozen wordt boven *symbool*, dan zal dit resulteren in de beschrijving “groene grote vaas”. PRO maakt dus gebruik van zowel onderscheidende kracht als voorkeur en kan overgespecificeerde expressies genereren. Het model van Frank en Goodman onderschat de waarschijnlijkheid dat gekozen zal worden voor de geprefereerde eigenschap, namelijk kleur, en gaat er daarom vanuit dat er meer gekozen wordt voor de minder geprefereerde eigenschap grootte. Dit model gaat ervan uit dat sprekers enkel rationale keuzes maken op basis van nut, wat niet zo is. Sprekers maken in hun keuzes van refererende expressies meer gebruik van voorkeur dan nut, wat suggereert dat deze keuzes gemaakt worden op basis van simpelere heuristiek.

Degen et al. (2019) hebben een model gemaakt dat ontworpen is in het kader van de *Rational Speech Act (RSA)* en daarom aanneemt dat sprekers rationale wezens zijn die de kosten van een uitspraak afwegen tegen hoe informatief de uitspraak is. Normaalgesproken vormt het simpele RSA-model geen overinformatieve expressies, het model van Degen is gebaseerd op de Bayesiaanse benadering en kan daarom wel refererende expressies genereren die overinformatief zijn. Degen definieert overinformatieve expressies als ‘expressies die specifiek zijn dan noodzakelijk is voor het vaststellen van een unieke referentie’.

### 2.3.3 Verschillen

De regelgebaseerde systemen volgen een stappenplan om zo tot een refererende expressie te komen dat het mogelijk maakt voor de luisteraar om het target te identificeren. De probabilistische systemen daarentegen maken gebruik van Bayesiaanse waarschijnlijkheden om zo te berekenen wat de kans is dat de referent omschreven zal worden met een bepaald woord om zo de beschrijving te vinden die de beste kans heeft op correcte identificatie. In beide categorieën bestaan er systemen die hun algoritme baseren op de manier waarop mensen refererende expressies genereren en systemen die simpelweg op zoek zijn naar de meest informatieve beschrijvingen. De expressies van deze twee soorten systemen verschillen dan ook van elkaar. Wanneer het systeem enkel op zoek zijn naar de kortste en meest informatieve beschrijvingen van de referent is de expressie over het algemeen kort, bestaande uit weinig attributen en bevat het zeker geen overtollige eigenschappen van het object. De systemen die het menselijke taalgedrag proberen te simuleren geven expressies vaak juist wel de mogelijkheid om overinformatief te zijn.

### 2.3.4 Wat is “beter”?

Zowel in het geval van de regelgebaseerde als de probabilistische systemen wordt meestal gesuggereerd dat NLG-systemen het taalgedrag van mensen na zouden moeten bootsen. Maar waarom zou wat mensen doen de beste manier zijn? Computers kunnen exact bepalen welke informatie noodzakelijk is om de boodschap over te brengen door de beschrijvingen met de meeste onderscheidende kracht uit te kiezen. Dit is niet de manier waarop mensen hun keuzes baseren in het genereren van refererende expressies, waardoor deze ook niet altijd enkel de noodzakelijke informatie bevatten. Mensen lijken soms meer informatie te geven dan nodig is. Dit zou een teken kunnen zijn dat de manier waarop mensen NLG uitvoeren niet per se de beste is voor NLG-systemen. In het volgende deel wordt gekeken naar hoe mensen aan NLG doen en wat de mogelijke verklaringen zijn waarom het op deze manier gebeurt.

## 3. Hoe doen mensen NLG en waarom

Mensen kunnen refererende expressies gebruiken in verschillende situaties. Bijvoorbeeld voor het identificeren van objecten, of het geven van een routebeschrijving. In deze verschillende situaties genereren sprekers ook verschillende refererende expressies. Er wordt weliswaar gebruik gemaakt

van dezelfde soorten attributen (kleur, grootte, vorm, etc.), maar er zijn verschillen in de formulering. Wanneer sprekers objecten identificeren genereren ze langere expressies met meer postnominale eigenschappen (Baltaretu et al., 2019). Postnominale eigenschappen staan in een zin achter het zelfstandig naamwoord (1). In tegenstelling tot prenominale eigenschappen (2), die in de zin juist vóór het zelfstandig naamwoord komen.

(1) het gebouw met grote ramen

(2) het witte gebouw

(3) het huis aan de linker kant

Sprekers zullen tijdens de identificatie van een object dus eerder (1) produceren, omdat de eigenschap (*grote ramen*) hier volgt op het zelfstandig naamwoord (*het gebouw*). Wanneer sprekers een routebeschrijving genereren, geven zij echter eerder de voorkeur aan het gebruik van prenominale eigenschappen en zullen zij dus eerder (2) produceren. Ook maken sprekers tijdens het identificeren van objecten meer gebruik van lokaliserende eigenschappen, waar dus beschrijvingen worden gegeven die te maken hebben met de plaats die het object inneemt in de situatie, zoals bijvoorbeeld in (3).

Ook heeft de afstand van de spreker ten opzichte van het target invloed op de refererende expressie. Hoe verder de spreker van het target af staat, hoe gedetailleerder de beschrijving wordt. Dit alles laat zien dat sprekers tijdens het genereren van refererende expressies rekening houden met het doel van deze expressie. Voor de luisteraar maakt het echter geen verschil. Mensen begrijpen het op beide manieren en beschouwen de verschillende soorten refererende expressies als even adequaat (Baltaretu et al., 2019). Iets wat wel effect lijkt te hebben op beide gesprekspartners is het gebruik van overspecificatie in het refereren naar objecten.

### 3.1 Overspecificatie

Wanneer mensen refererende expressies genereren moeten ze ook keuzes maken over de hoeveelheid en soort informatie die ze geven om succesvol te communiceren. Mensen doen dit echter niet altijd op de meest voor de hand liggende manier. In veel gevallen doen mensen namelijk aan overspecificatie, waarbij ze dus meer informatie geven dan noodzakelijk is voor de identificatie van het object. (Koolen et al., 2011) Wanneer iemand in een context met twee mensen zoals in Figuur 2 naar persoon A (het target) wil verwijzen kan dit door de attributen van persoon A te omschrijven die zich onderscheiden van de eigenschappen van persoon B (de distractor). Een goede omschrijving zou bijvoorbeeld kunnen zijn “De man met het witte shirt” of “De man met de bril”. Als de spreker naar het target zou refereren met enkel “De man” is dit niet informatief genoeg. Maar als de spreker bijvoorbeeld zou refereren naar het target met “De zwartharige man met de bril en zonder das” is er sprake van overspecificatie. In dit geval worden er meer eigenschappen van persoon A benoemd dan strikt noodzakelijk is om hem te onderscheiden van persoon B.



**Person A**



**Person B**

**Figuur 4: Twee mogelijke target referenten (Koolen et al., 2011).**

Hoe beslissen mensen welke informatie ze gebruiken in hun targetomschrijving? Sprekers identificeren een target object door er een combinatie van attributen aan toe te schrijven die niet toegeschreven kunnen worden aan de andere objecten in de context, om zo alle distractors uit te sluiten. Deze attributen worden gekozen op basis van zowel kennis over de huidige context als algemene kennis over de wereld. Wanneer mensen aan overspecificatie doen lijkt dit de Maxim of Kwantiteit van het coöperatieve principe van Grice te overschrijden.

Maxime van kwantiteit: maak je contributie zo informatief als nodig is, niet meer, niet minder.

Wanneer mensen hun refererende expressie overspecificeren geven ze dus meer informatie dan gevraagd is in de desbetreffende context en overschrijden ze hiermee Grice's Maximes voor conversationale implicatuur. Dit lijkt onwenselijk te zijn in een conversatie. Waarom doen mensen dan toch aan overspecificatie?

### **3.2 Mogelijke verklaringen voor overspecificatie**

Degen et al. (2019) concludeerden in hun onderzoek dat in sommige situaties het toevoegen van overtollige informatie aan een refererende expressie succesvolle communicatie juist verzekert. Bijvoorbeeld wanneer de overtollige eigenschappen opvallender of duidelijker te identificeren zijn voor de luisteraar dan alleen de noodzakelijke eigenschappen. Dit zou dus bijvoorbeeld kunnen resulteren in een refererende expressie als "groene grote vaas met de ruit". Niet alle eigenschappen van de objecten zullen de refererende expressie namelijk informatiever maken. Wanneer een context veel visuele variatie bevat en de overtollige eigenschap duidelijk en niet te ambigu is, zal deze eigenschap meer informatie over de referent toevoegen en dus eerder gekozen worden door de spreker. *Kleur* is een minder ambigu attribuut dan bijvoorbeeld *grootte*, want *kleur* is over het algemeen een absolute eigenschap, terwijl *grootte* relatief is aan de context. Om deze reden worden kleuren vaker gekozen als overtollige eigenschap van een object om aan de refererende expressie toe te voegen. Wanneer

de luisteraar een bepaalde overtollige expressies verwacht gezien de context wordt deze expressie ook sneller verwerkt dan een minimaal gespecificeerde expressie. Mensen zijn niet onnodig overinformatief, maar voegen overtollige attributen toe aan een expressie of refereren op een lager taxonomisch niveau dan strikt noodzakelijk is en dit leidt tot een weldegelijk meer informatieve refererende expressie. Mensen genereren rationeel overtollige refererende expressies (Degen et al., 2019). Dit zou dus Grice's maxime van kwantiteit, dat suggereert dat de sprekers' bijdrage aan een conversatie "niet meer informatief dan noodzakelijk" moet zijn, overtreden. Toch maken mensen hun refererende expressies vaak informatiever dan nodig. Er bestaan verschillende theorieën over de reden waarom mensen dit doen. In het volgende deel zullen er zes mogelijke verklaringen voor het gebruik van overspecificatie in refererende expressies door mensen besproken worden.

1. Er bestaat een mogelijkheid dat de luisteraar een *conceptual gestalt* creëert van het gerefereerde object waar ze naar moeten zoeken in de context. De opvallende kenmerken van het object, zoals type en kleur zijn belangrijk voor het *gestalt* en maken het gemakkelijker voor de luisteraar om een mentale representatie van het object samen te stellen. Daarom kan het voordelig zijn voor de spreker om in de refererende expressie deze eigenschappen op te nemen, zodat de luisteraar een betere *conceptual gestalt* kan maken en het target sneller kan identificeren. Echter, er is ook gebleken uit verscheidene studies dat overspecificatie in sommige gevallen juist nadelig kan zijn voor luisteraars en het identificeren van het target juist vertraagt (Koolen et al., 2011). Verschillende eye-trackingexperimenten (Engelhardt et al., 2010; Tanenhaus et al., 1995) hebben uitgewezen dat de luisteraar de binnenkomende informatie verwerkt tijdens het luisteren en zo de distractors een voor een uitsluit op zoek naar de mogelijke referent. Dit suggereert dat overtollige informatie niets zou toevoegen aan het proces van identificatie en de luisteraar alleen maar af zou leiden tijdens het uitsluiten van distractors en dus het identificeren vertraagt.

2. Wat ook een rol kan spelen in overspecificatie zijn sprekergeoriënteerde processen, zoals het contrasteren van het target met voorgaande objecten, zelfs als deze niet langer in beeld zijn. In zulke gevallen bevat de omschrijving van het huidige target het onderscheidende attribuut van het vorige target, ook als dit attribuut in de huidige context geen onderscheidende kracht heeft (Yoon, 2018).

3. Een andere mogelijke verklaring voor overspecificatie is dat sprekers gebruik maken van zogenaamde *referentiescripten* in het genereren van refererende expressies. In plaats van een expressie uit het niets te maken wordt een verzameling bestaande uit een vaste groep attributen gekozen, ook als sommige van deze attributen geen onderscheidende kracht in de desbetreffende context hebben (Dale & Reiter, 1995).

4. Ook is het zo dat sprekers voorkeuren hebben voor bepaalde attributen en deze daarom vaak gebruiken in hun refererende expressies, ongeacht de relevantie voor het onderscheiden van het target van de contrastverzameling. Mensen gebruiken bijvoorbeeld graag het attribuut kleur in hun omschrijving van objecten. Dit is over het algemeen een opvallend en absoluut kenmerk dat makkelijk te herkennen is, in tegenstelling tot bijvoorbeeld het attribuut *grootte*, dat relatief is aan de andere objecten in de context. Mensen maken meer gebruik van overtollige kleurbeschrijvingen wanneer de kleur atypisch is voor het object. Bijvoorbeeld in het geval van een roze banaan (Rubio-Fernández, 2016). In sommige categorieën is *kleur* belangrijker dan in anderen. Voor bijvoorbeeld kleren en auto's is het veel relevanter om de kleur te benoemen dan bijvoorbeeld gereedschappen of geometrische figuren. Wanneer voor sprekers *kleur* een centrale eigenschap van het object lijkt te zijn is het niet meer dan logisch dat de kleur van het object gespecificeerd wordt. Op dezelfde manier vindt men het over het algemeen logisch dat iemand naar zijn huisdier verwijst als "mijn hond" in plaats van het hogere taxonomisch niveau "mijn dier".

5. Het gebruik van overspecificatie in refererende expressies kan in sommige gevallen efficiënt zijn. Het kan namelijk makkelijker zijn voor sprekers om in plaats van ieder attribuut te beoordelen op

onderscheidende kracht bepaalde overtollige attributen te gebruiken in de refererende expressie. Ook verkort het de zoektocht naar potentiële concurrenten. Sprekers produceren woorden namelijk zo snel als ze kunnen en kiezen daarom graag voor het attribuut kleur omdat dit makkelijker te herkennen is dan bijvoorbeeld grootte (Pechmann, 1989). Dit is een reden voor overspecificatie wanneer de contrastverzameling groot is en het de spreker veel tijd kost om alle objecten te bestuderen. Om dezelfde reden neemt de kans op overspecificatie toe wanneer sprekers een beperkte tijd hebben om alle objecten te bestuderen dan wanneer deze tijd ongelimiteerd is. Bij een beperkte kijktijd kiezen sprekers meer voor inherent opvallende attributen en wordt er minder rekening gehouden met de onderscheidende kracht van deze attributen (Koolen et al., 2016).

6. De laatste mogelijke reden waarom sprekers hun refererende expressies overinformatief maken is dat de luisteraar baat kan hebben bij het gebruik van overspecificatie. Over het algemeen zullen de kenmerken die uitspringen voor de spreker ook uitspringen voor de luisteraar, aangezien de fysieke omgeving van een communicatie *common ground* is. Dit is een aanname die sprekers mogen doen (Grice, 1975). Daarom zal het gunstig zijn voor beide gespreksdeelnemers wanneer deze uitspringende kenmerken worden opgenomen in de refererende expressie.

### 3.3 Wat betekent dit voor NLG algoritmen?

Natural Language Generation is het genereren van natuurlijk klinkende taal. Aan de hand van deze definitie is het logisch om aan te nemen dat het taal zou moeten genereren wat zo veel mogelijk lijkt op hoe mensen natuurlijke taal genereren. Dale en Reiter (1995) opperden dan ook dat het misschien beter zou zijn om een algoritme te creëren dat gebaseerd op menselijke spraak dan het zo streng mogelijk interpreteren van Grice's maxims. Maar waarom zou wat mensen doen beter zijn? Om deze vraag te beantwoorden is het belangrijk om vast te stellen wat het uiteindelijke doel van NLG systemen zijn. Als het doel is om een zo effectief mogelijke refererende expressie te genereren met zo min mogelijk woorden is het beter om Grice's maxims aan te houden en gebruik te maken van de probabilistische NLG systemen. In dit geval zou overspecificatie een systeem minder goed maken en relationele beschrijvingen juist beter. Een ander doel van NLG systemen zou kunnen zijn om de boodschap zo duidelijk mogelijk over te brengen naar de luisteraar zodat er zo min mogelijk tijd nodig is om het target correct te identificeren. Omdat beschrijvingen met uitspringende kenmerken (ook als deze niet noodzakelijk zijn voor het onderscheiden van het target) makkelijker worden geïdentificeerd door de luisteraar kan overspecificatie in dit geval dus weldegelijk toegevoegde waarde hebben voor refererende expressies.

Wat het doel ook zal zijn, uiteindelijk zal Natural Language Generation altijd teksten genereren die bedoeld zijn voor mensen om te lezen en begrijpen. Dus zou het een refererende expressie als *effectief* gezien kunnen worden als mensen de boodschap correct kunnen interpreteren. Onderzoek van Van der Meulen et al., (2010) heeft uitgewezen dat mensen over het algemeen de voorkeur geven aan teksten gegenereerd door mensen in plaats van automatisch gegenereerde teksten. Deze menselijk gegenereerde teksten waren langer, hadden een meer verhalende structuur en werden beoordeeld als effectiever door de lezers. Dit suggereert dat er weldegelijk iets te zeggen valt voor NLG-systemen die menselijk taalgedrag nabootsen.

## 4. Discussie

Veel NLG-systemen zijn al gebaseerd op het taalgedrag van mensen om zo refererende expressies te genereren die lijken op die van menselijke sprekers. Dit zijn niet de meest informatieve of korte beschrijvingen, maar bevatten in veel gevallen juist overtollige informatie die niet noodzakelijk is voor het onderscheiden van het target. Toch lijken sprekers hier de voorkeur aan te geven. Ook wijzen verschillende onderzoeken uit dat luisteraars de voorkeur geven aan menselijke beschrijvingen. Dus wanneer het systeem bedoeld is om de luisteraar een refererende expressie te geven waarmee de

referent zo snel en makkelijk mogelijk geïdentificeerd kan worden is het gebruik van overspecificatie een goed idee. Maar niet ieder NLG-systeem zal hetzelfde doel hebben. Het hangt dus af van de situatie waarvoor de refererende expressies gegenereerd moeten worden of het beter is om de beschrijvingen zo kort en informatief mogelijk te maken, of juist meer op menselijk taalgedrag te laten lijken. Ook kan de vraag worden gesteld of het wel met zekerheid te zeggen is dat luisteraars de referent echt makkelijker identificeren wanneer er gebruik wordt gemaakt van menselijke elementen als overspecificatie? Is het niet simpelweg omdat luisteraars gewend zijn van sprekers dat zij niet aan de meest korte en effectieve manier van spreken doen? Veel onderzoek naar refererende expressies is gefocust op het produceren ervan en minder op de interpretatie. Toekomstig onderzoek zou zich kunnen focussen op hoe luisteraars refererende expressies met en zonder gebruik van elementen als overspecificatie interpreteren om erachter te komen wat de identificatie van referenten vergemakkelijkt. Als de luisteraar toch voordelen lijkt te hebben bij refererende expressies die niet enkel de kortste en meest informatieve beschrijvingen maar juist overtollige informatie bevatten suggereert dit dat het weldegelijk wenselijk zou zijn voor NLG-systemen om refererende expressies te genereren op de manier waarop mensen het doen. Dan zou ruimte maken in het algoritme voor menselijke eigenschappen zoals overspecificatie een goed idee zijn.

Maar waarom is dit belangrijk? De meest voor de hand liggende reden dat dit verslag relevant is voor de kunstmatige intelligentie is dat taal onlosmakend verbonden is met KI. Taal is de manier waarop de computers en gebruikers met elkaar communiceren en is dus essentieel. Wanneer taalkundige processen, waaronder het genereren van refererende expressies, worden geoptimaliseerd in computers zal dit tot betere communicatie leiden en dus betere intelligente systemen.

Maar dit onderwerp is nog op een andere manier gerelateerd aan kunstmatige intelligentie, namelijk met de vraag wat precies het doel is van de NLG-systemen. Is het de bedoeling dat het systeem refererende expressies genereert om de mens te assisteren in bepaalde taken? Moeten de beschrijvingen voornamelijk nuttig zijn en rationeel, en dus enkel de noodzakelijke informatie bevatten om de boodschap over te brengen? Of is het doel van NLG-systemen om meer op mensen te lijken en hun taalgedrag exact na te bootsen? Gaat het er niet om dat de refererende expressies zo kort en informatief mogelijk zijn, maar dat ze het meest lijken op de expressies die mensen zouden genereren, gebaseerd op voorkeuren en met overspecificaties? Dit is te vergelijken met de vraag "Wat is kunstmatige intelligentie?" Is kunstmatige intelligentie het gebruik van methoden gebaseerd op intelligent gedrag om complexe problemen op te lossen? In andere woorden moeten de computers enkel de illusie geven van menselijke intelligentie om de gebruiker te kunnen assisteren in bepaalde situaties, zonder de daadwerkelijke kennis te hebben waar mensen over beschikken? Of is een kunstmatige intelligentie een computer die bezit over werkelijk menselijke intelligentie en dus op dezelfde manier kan denken als een mens doet? Dan zou simpelweg het oplossen van problemen niet langer genoeg zijn om te kwalificeren voor een kunstmatige intelligentie maar zou het ook over allerlei andere menselijke eigenschappen moeten beschikken, zoals een bewustzijn en emoties. Degenen die deze definitie aanhangen horen bij de beweging die geloven in zogeheten sterke KI, wat beweert dat machines werkelijke mentale staten kunnen hebben net als mensen. Aanhangers van de zwakke KI daarentegen geloven dat computers nooit over echt bewustzijn zullen beschikken en dat kunstmatige intelligentie enkel het modelleren van intelligent probleemoplossend gedrag is. Bijvoorbeeld een systeem dat betekenis uit gesproken teksten van mensen kan afleiden. Dit wordt wel het begrijpen van menselijke spraak genoemd, maar de computer heeft geen echt begrip van de taal. Het past enkel bepaalde principes toe die gebaseerd zijn op menselijk taalbegrip om zo enige betekenis af te kunnen leiden en de illusie van taalbegrip te creëren. Om als sterke KI gezien te worden zou het dus wel hetzelfde taalvermogen als dat van mensen moeten hebben. Veel KI-onderzoekers en filosofen geloven niet in het idee van sterke KI dat computers een bewustzijn zouden kunnen hebben en zien dit alleen mogelijk in science fiction (Coppin, 2004).



De vraag wat het doel is van NLG-systemen verschilt dus niet zo veel van de zwakke versus sterke KI kwestie. Is het de bedoeling dat NLG-systemen refererende expressies genereren zoals mensen het doen, en dus beschrijvingen produceren die een mens ook zou kunnen maken? Of hoeven de systemen niet te denken zoals mensen doen en is hun doel enkel het assisteren van mensen door de meest effectieve refererende expressies te genereren? Voorlopig blijven deze vragen nog onbeantwoord, maar toekomstig onderzoek zou het mogelijk kunnen maken om erachter te komen wat het doel van Natural Language Generation (en kunstmatige intelligentie) precies is.

## Bronnen

Baltaretu, A., Krahmer, E., & Maes, A. (2019). Producing Referring Expressions in Identification Tasks and Route Directions: What's the Difference?. *Discourse Processes*, 56(2), 136-154.

Coppin, B. (2004). *Artificial intelligence illuminated*. Jones & Bartlett Learning.

Dale, R., & Reiter, E. (1995). Computational interpretations of the Gricean maxims in the generation of referring expressions. *Cognitive science*, 19(2), 233-263.

Degen, J., Hawkins, R. X., Graf, C., Kreiss, E., & Goodman, N. D. (2019). When redundancy is rational: A Bayesian approach to 'overinformative' referring expressions. *arXiv preprint arXiv:1903.08237*.

Engelhardt, P. E., Demiral, Ş. B., & Ferreira, F. (2011). Over-specified referring expressions impair comprehension: An ERP study. *Brain and cognition*, 77(2), 304-314.

Evans, R., Piwek, P., & Cahill, L. (2002, July). What is NLG?. In *Proceedings of the International Natural Language Generation Conference* (pp. 144-151).

Frank, M., & Goodman, N. (2012). Predicting pragmatic reasoning in language games. *Science*, 336, 998.

Gatt, A., & Krahmer, E. (2018). Survey of the state of the art in natural language generation: Core tasks, applications and evaluation. *Journal of Artificial Intelligence Research*, 61, 65-170.

Gatt, A., Krahmer, E., Van Gompel, R.P., & van Deemter, K. (2013). Production of referring expressions: Preference trumps discrimination. In *Proceedings of the Annual Meeting of the Cognitive Science Society* (Vol. 35, No. 35).

Gatt, A., van Gompel, R. P., van Deemter, K., & Krahmer, E. (2013). Are we Bayesian referring expression generators. In *Proceedings of CogSci* (Vol. 35).

Graf, C., Degen, J., Hawkins, R. X., & Goodman, N. D. (2016). Animal, dog, or dalmatian? Level of abstraction in nominal referring expressions. In *CogSci*.

Grice, H. P., Cole, P., & Morgan, J. (1975). Logic and conversation. 1975, 41-58.

Kondadadi, R., Howald, B., & Schilder, F. (2013, August). A statistical nlg framework for aggregated planning and realization. In *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)* (pp. 1406-1415).

Koolen, R., Gatt, A., Goudbeek, M., & Krahmer, E. (2011). Factors causing overspecification in definite descriptions. *Journal of Pragmatics*, 43(13), 3231-3250.

Koolen, R., Gatt, A., van Gompel, R. P., Krahmer, E., & Van Deemter, K. (2016). Viewing time affects overspecification: Evidence for two strategies of attribute selection during reference production. In *CogSci*.

Krahmer, E., Erk, S. V., & Verleg, A. (2003). Graph-based generation of referring expressions. *Computational Linguistics*, 29(1), 53-72.

- Pechmann T. (1989). Incremental speech production and referential overspecification. *Linguistics* 27 89-110.
- Reiter, E. (2007, June). An architecture for data-to-text systems. In *Proceedings of the Eleventh European Workshop on Natural Language Generation* (pp. 97-104). Association for Computational Linguistics.
- Reiter, E., & Dale, R. (1997). Building applied natural language generation systems. *Natural Language Engineering*, 3(1), 57-87.
- Reiter, E., & Dale, R. (2000). *Building natural language generation systems*. Cambridge university press.
- Rubio-Fernández, P. (2016). How redundant are redundant color adjectives? An efficiency-based analysis of color overspecification. *Frontiers in psychology*, 7, 153.
- Tanenhaus, M. K., Spivey-Knowlton, M. J., Eberhard, K. M., & Sedivy, J. C. (1995). Integration of visual and linguistic information in spoken language comprehension. *Science*, 268(5217), 1632-1634.
- Van Der Meulen, M., Logie, R. H., Freer, Y., Sykes, C., McIntosh, N., & Hunter, J. (2010). When a graph is poorer than 100 words: A comparison of computerised natural language generation, human generated descriptions and graphical displays in neonatal intensive care. *Applied Cognitive Psychology: The Official Journal of the Society for Applied Research in Memory and Cognition*, 24(1), 77-89.
- Viethen, J., & Dale, R. (2006, July). Algorithms for generating referring expressions: do they do what people do?. In *Proceedings of the fourth international natural language generation conference* (pp. 63-70). Association for Computational Linguistics.
- Yoon, S. (2018). The role of discourse context in reference production and comprehension: Insights from the lexical differentiation effect. *Psychology of Learning and Motivation*, 68, 361-392.