



Utrecht University

Wat schuilt er achter de Turingtest?

Een literatuurstudie over de rol van pragmatische
competentie in de Turingtest

Julia Sendula (6271537)

Bachelor scriptie Kunstmatige Intelligentie

7.5 ECTS

Datum: 4 november 2019
Begeleider: Rick Nouwen
Tweede beoordelaar: Janneke van Lith

INHOUDSOPGAVE

INHOUDSOPGAVE	2
INLEIDING	3
1. PRAGMATISCHE COMPETENTIE	5
1.1 PRAGMATIEK	5
1.2 PRAGMATISCHE COMPETENTIE	5
2. DE TURINGTEST	9
2.1 TURING TEST EN LOEBNERWEDSTRIJDEN	9
2.2 PROBLEMEN MET DE TURINGTEST	10
2.3 ALTERNATIEVE TESTEN	11
3. PRAGMATISCHE COMPETENTIE IN DE TURINGTEST	12
3.1 CONVERSATIES	12
3.2 METHODE	12
3.3 ANALYSE	13
4. CONCLUSIE EN DISCUSSIE	19
4.1 CONCLUSIE	19
4.2 DISCUSSIE	19
BIBLIOGRAFIE	21
APPENDIX	23
APPENDIX A: TRANSCRIPT MITSUKU (STEVE WORSWICK, 2018)	23
APPENDIX B: TRANSCRIPT TUTOR (RON C. LEE, 2018)	29

INLEIDING

Eugene Goostman speelt een belangrijke rol in de kunstmatige intelligentie. Het is een van de weinige chatbots die de Turingtest, een intelligentietest voor bots, heeft gehaald (Nijs, 2014). Dit toont aan dat bots nog een lange weg te gaan hebben voor ze over de juiste technologie beschikken om met mensen te chatten. Voordat de technologie kan worden verbeterd, moet eerst worden uitgezocht welke eigenschappen een bot intelligenter maken.

De eigenschappen die bots intelligenter maken in de Turingtest vallen onder de taal, omdat de test zich alleen bezighoudt met chatsystemen. Het eerste waar mensen aan denken bij taal zijn woorden. Echter, taal bestaat ook uit componenten die niet worden uitgesproken, maar wel worden aangenomen. Dit deel van de taal heet pragmatiek en is een cruciaal onderdeel van het voeren van een gesprek. In hoeverre komt pragmatiek echter voor in de Turingtest?

Dit onderzoek zal kijken naar de pragmatische competentie die te vinden is in enkele mens-computerinteracties. Hiervoor is er een selectie gemaakt van belangrijke aspecten van de pragmatiek. Daarnaast wordt er gekeken naar enkele recentelijke transcripten van Turingtesten die zijn uitgevoerd bij de Loebnerwedstrijden, om het onderzoek zo relevant mogelijk te maken.

Ondanks dat de methode is ontkracht, blijft de Turingtest een belangrijke manier om de mate van technologische ontwikkeling van een chatbot te testen (Whitby, 1997; Moor, 2003). Chatbots die goed presteren tijdens deze test, leveren een grote bijdrage aan het gehele vakgebied van de kunstmatige intelligentie. Als een chatbot zich namelijk kan gedragen als een mens, dan zijn we een stap dichterbij het nabootsen van menselijke cognitieve functies, de algemene doelstelling binnen de kunstmatige intelligentie.

Om bij te dragen aan de stap die ons dichterbij het nabootsen van menselijke cognitieve functies brengt, zal dit onderzoek kijken op welke manier pragmatische competentie wordt getest in de Turingtest. Daarom zal de onderzoeksvraag van dit onderzoek zijn:

“Zien we pragmatische competentie in de Turingtest?”

Analyseren wanneer pragmatische incompetentie zich voordoet in gesprekken geeft toekomstige ontwikkelaars de verbeterpunten die nodig zijn om verder te ontwikkelen. Soortgelijke onderzoeken aan deze uit dit verslag zijn al eerder gedaan. Saygin en Cicekli (2002) analyseerden de schendingen van Grice's maximen, pragmatische regels, in conversaties van de Loebner Turingtesten. Humphrys (2009) ontwierp een chatbot die slaagde voor de Turingtest. In zijn onderzoek bekijkt hij verschillende aspecten die hebben bijgedragen aan het slagen van de Turingtest. Zdenek (2003) keek naar de tekortkomingen in zijn onderzoek. Echter, de hoeveelheid literatuur blijft schaars. Daarnaast wordt er in de onderzoeken gekeken naar de uitspraken van de chatbots, maar in hoeverre wordt pragmatiek toegepast door de juryleden, die proberen te achterhalen of de andere deelnemer een chatbot is? Ook dit zal worden besproken in dit onderzoek.

Het eerste hoofdstuk geeft een korte beschrijving van de pragmatiek en de aspecten van de pragmatiek die zullen worden geanalyseerd in de transcripten van de Turingtesten. Deze

aspecten worden gelinkt aan theorieën uit de literatuur en de invloed van de aspecten op de conversaties wordt ook besproken. Hoofdstuk twee zal de gehele methode van de Turingtest bekijken. Het derde hoofdstuk analyseert de transcripten van de Turingtesten. Hiervoor wordt er gekeken naar de aspecten genoemd in het eerste hoofdstuk en wordt er kwalitatief beargumenteerd in hoeverre deze aspecten voorkomen in de conversaties en of deze worden geschonden. Tenslotte volgt er een discussie om de tekortkomingen en voorstellen voor toekomstige onderzoeken te bespreken.

1. PRAGMATISCHE COMPETENTIE

Dit hoofdstuk zal de verschillende aspecten van pragmatische competentie bespreken. Allereerst wordt er uitgelegd wat pragmatiek precies inhoudt. Daarna zal er uitgelegd worden wat pragmatische competentie inhoudt, welke aspecten ervan in dit verslag worden besproken en welke rol de aspecten spelen in conversaties.

1.1 Pragmatiek

Pragmatiek is een breed concept. Het wordt als het ware gezien als het onderdeel van de taal dat zich buiten de uitgesproken taal bevindt. Het betreft de (contextuele) impliciete betekenis die achter een zin, informatie die wordt overgebracht zonder dat het wordt benoemd. Het is een moeilijk begrip om met woorden vast te leggen. Daarom hebben door de jaren heen verschillende onderzoekers geprobeerd te nuanceren wat zij onder pragmatiek verstaan. Stalnaker (1970) definieert pragmatiek als de wetenschap van linguïstische acties en de contexten in welke deze worden uitgevoerd. Kaplan (1989) beschrijft pragmatiek als een basis voor het toekennen van bepaalde betekenis aan een woord of een zin. Fotion (1995) definieert pragmatiek als de wetenschap van de taal die zich focust op de gebruikers en de context van de taal in plaats van op referenties, waarheid en grammatica. Lycan (1995) beschrijft pragmatiek als het bestuderen van het gebruik van taal binnen de context en de contextafhankelijkheid van verschillende aspecten van linguïstische interpretatie.

Omdat pragmatiek een breed concept is, wordt het verdeeld in twee delen: nabije pragmatiek en verre pragmatiek. Nabije pragmatiek houdt zich bezig met de aspecten die relevant zijn voor het bepalen wat er is gezegd. Dit gaat om bijvoorbeeld ambiguïteit, objectreferenties van voornaamwoorden, verwijzingen en andere aspecten die de directe relaties bekijken tussen de uitgesproken woorden en de betekenis ervan (Korta & Perry, 2006). Verre pragmatiek focust zich op de aspecten buiten hetgeen wat is gezegd, dus wat er wordt geïmpliceerd. Het redeneert over waarom wat is gezegd, is gezegd (Korta & Perry, 2006). Als een student bijvoorbeeld zegt dat die naar de bibliotheek gaat, dan zal de assumptie worden gemaakt dat die daar gaat studeren, omdat dat een sociaal conventionele reden is om naar de bibliotheek te gaan.

De tweedeling tussen verre en nabije pragmatiek heeft een lange tijd stand gehouden in de pragmatische wetenschap. Echter, hedendaagse pragmatische theorieën gebruiken deze tweedeling in mindere mate, omdat ze redeneren dat de twee aspecten niet compleet als onafhankelijke aspecten kunnen worden gezien (Korta & Perry, 2006). De twee aspecten zouden invloed op elkaar hebben (Sperber & Wilson, 1986).

1.2 Pragmatische competentie

Pragmatische competentie wordt gedefinieerd als het vermogen om communicatieve handelingen te begrijpen op zowel zinsniveau als gespreksniveau, en om deze te produceren (Felix-Brasdefer, 2017). Het is de competentie om de pragmatiek achter gesprekken te begrijpen. Er zijn vier aspecten die belangrijk geacht worden in dit verslag:

- De hoeveelheid informatie van uitspraken;
- De relevantie van de uitspraken;
- De woordkeuze met betrekking tot sociale conventies;
- Het egocentrisme in de conversatie.

Deze vier aspecten hebben een grote invloed op de pragmatiek van een conversatie en geven aan in hoeverre een deelnemer van de conversatie de andere deelnemer begrijpt en zelf in

staat is om pragmatisch competente uitspraken te doen. De grote invloed op het gesprek is een van de redenen voor de keuze van deze aspecten. Daarnaast zijn deze aspecten makkelijker kwantificeerbaar, omdat ze niet bestaan uit meerdere lagen die moeilijk op een concrete manier, zonder tussenkomst van andere aspecten, kunnen worden aangetoond. Bij deze aspecten kan dat wel.

1.2.1 Hoeveelheid informatie

Het eerste aspect is de hoeveelheid overgebrachte informatie. Deelnemers van conversaties kunnen zowel te veel, als te weinig informatie overbrengen. Wanneer dit gebeurt, dan wordt het gesprek als onnatuurlijk ervaren.

Dit aspect is gerelateerd aan het kwantiteitsmaxime van Grice. Grice heeft een enorme invloed gehad op de linguïstiek met de introductie van zijn Coöperatieve Principe (Grice, 1989). Hij heeft dit principe afgeleid van de algemene principes die mensen gebruiken voor het voeren van gesprekken. Het Coöperatieve Principe zegt dat de bijdrage aan de conversatie moet inspelen op de behoefte van de conversatie. Achter een conversatie zit een bepaald doel dat kan worden behaald met de juiste tijd en de juiste intenties. Het Coöperatieve Principe bestaat uit vier maximen. Het eerste maxime is het kwantiteitsmaxime. Deze zegt dat de bijdrage zo informatief moet zijn als nodig is voor de huidige conversatie en dat de bijdrage niet meer informatief moet zijn dan nodig is. Dit aspect komt ook voor in de heuristieken van Levinson (2000). Zijn eerste (Q-)heuristiek wordt beschreven als "Wat niet gezegd is, is niet gezegd". Dit zegt dat er niet meer inferenties moeten worden gemaakt dan welke impliciet zijn aan hetgeen wat er gezegd wordt. Er moet dus geen extra informatie worden aangenomen. Dit is gerelateerd aan Grice's sub-maxime over het overbrengen van te veel informatie. Zijn tweede (I-)heuristiek wordt beschreven als "Wat is uitgedrukt, is simpelweg stereotypisch verklaard". Deze zegt dat er zodanig genoeg informatie gegeven moet zijn, dat er genoeg informatie uit kan worden. Dit is gerelateerd aan het tweede sub-maxime van Grice, over het overbrengen van te weinig informatie.

Als de juiste hoeveelheid informatie wordt overgebracht in een conversatie, dan wordt dat gezien als een deel van de beheersing van pragmatische competentie. Wanneer er te veel informatie wordt gegeven, dan wordt er meer informatie gegeven dan nodig, waardoor wordt afgeweken van het doel van het gesprek. Te weinig informatie houdt het gesprek niet op gang, omdat de tweede deelnemer niet genoeg informatie heeft om te gebruiken voor verdere antwoorden in de conversatie. Dit kan bijvoorbeeld voorkomen wanneer een persoon vraagt of een ander denkt een toets te hebben gehaald. Bij deze vraag zit vaak de verwachting om toe te lichten hoe de toets ging. Als de ander dan alleen antwoord met "ja", dan wordt er niet aan die verwachting voldaan.

1.2.2 Relevantie

Het tweede aspect van de pragmatische competentie is de relevantie van de uitspraken. Relevantie is de mate waarin de uitspraken van de deelnemers passen binnen de conversatie en in hoeverre de antwoorden op elkaar aansluiten. De mate van relevantie bepaalt hoe natuurlijk de conversatie wordt ervaren. Bij een lage mate van relevantie wordt de conversatie als onnatuurlijk ervaren. Dit aspect is gerelateerd aan het informatieaspect, omdat overvloed aan informatie ook irrelevante informatie oplevert.

Dit aspect is gerelateerd aan het derde maxime van Grice, namelijk het relatiemaxime. Dit maxime zegt dat de bijdrage relevant moet zijn aan het gesprek. De bijdrage moet zodanig zijn, dat de andere deelnemer van de conversatie met de gegeven uitspraak de juiste

conclusies kan trekken en de juiste assumpties kan maken. Dit aspect heeft ook betrekking tot de relevantietheorie van Sperber en Wilson (1986). De relevantietheorie beschrijft de input van een conversatie als relevant wanneer er met de aanwezige contextuele assumpties positieve cognitieve effecten kunnen worden bereikt. De relevantietheorie bevat twee principes. Het eerste principe zegt dat menselijke cognitie gericht is op de maximalisatie van de relevantie. Dit betekent dat de sprekende deelnemer zijn bijdrage zodanig relevant zou moeten maken, opdat het de aandacht trekt van de andere deelnemer om zijn bijdrage aan te horen. Het tweede principe zegt dat alle communicatie een aanname overbrengt met zijn eigen optimale relevantie. Dit betekent dat de uitspraak zodanig relevant moet zijn, dat de andere deelnemer de intenties van de sprekende deelnemer begrijpt.

Een overmaat aan relevantie is geen remmer voor een conversatie. Echter, een lage mate aan relevantie zal het gesprek wel onnatuurlijk maken. Het laat zien dat de twee deelnemers elkaars intenties niet begrijpen, omdat niet wordt voldaan aan de verwachte informatievoorziening.

1.2.3 Woordkeuze met betrekking tot sociale conventies

Het derde aspect is de woordkeuze voor conversaties. Dit gaat om de manier waarop de deelnemer zich uit.

Ook dit aspect is gerelateerd aan een maxime van Grice, namelijk het beleefdheidsmaxime (*Manner*). Deze maxime houdt in dat de informatieoverdracht duidelijk moet zijn en daarbij dus ambiguïteit, onnodig woordgebruik, chaos moet vermijden en de juiste uitdrukkingvorm moet gebruiken. Hier is ook de derde (M-)heuristiek van Levinson aan gerelateerd: "Wat wordt gezegd in een abnormale manier, is niet normaal". Deze zegt dat uitspraken niet moeten worden gedaan met een woordkeuze die in het dagelijks gebruik niet wordt gebruikt. Ook de mate van beleefdheid wordt verwerkt in de woordkeuze. Het betreft de woordkeuze die de sociale status aangeeft tussen de deelnemers. Brown en Levinson (1987) hebben hiervoor het universele model voor linguïstische beleefdheid ontworpen. Door middel van de sociale afstand, de sociale macht en de absolute ranking wordt bepaald of de zogenoemde 'face-threatening act' (gebeden, weigeringen, suggesties) een effect heeft op het beeld welke de spreker probeert te projecteren.

Beleefdheid en woordkeuze heeft veelal te maken met culturele normen. Het taalgebruik en de mate van beleefdheid verschilt per taal. Uit een onderzoek van Krashen, Long & Scarcella (1979) is gebleken dat als men meer beleefdheid gebruikt wanneer Engels wordt gesproken, dit wordt gezien als een hogere competentie. Echter, Beebe en Waring (2005) hebben aangetoond dat het gebruik van onbeleefdheid als competentie wordt gezien. Het verschil tussen deze twee studies is 26 jaar, wat de culturele verschillen laat zien. De mate van competentie hangt dus af van de normen en waarden van de deelnemers. Hier hangt ook de natuurlijkheid van de conversatie af.

1.2.4 Egocentrisme

Het laatste aspect is het egocentrisme, de betrekking van het eigen leven. Dit is de mate waarin de deelnemer de eigen gedachten, waarden en ervaringen betreft in het gesprek. Dit aspect is vooral belangrijk in het bredere perspectief van de conversatie ten opzichte van de losse uitspraken gedaan binnen de conversatie en past dus ook niet binnen het principe van Grice.

Dit aspect wordt vooral veel genoemd binnen de tussen-culturele en interculturele pragmatiek. Tussen-culturele pragmatiek analyseert de verschillen en overeenkomsten in het

gebruik van taal van mensen uit verschillende culturen die verschillende talen spreken. Interculturele pragmatiek analyseert het gebruik van taal bij mensen die communiceren met een tweede gemeenschappelijke taal die niet hun moedertaal is. In deze pragmatiek bevindt zich de socio-cognitieve benadering (Kecskes, 2010). Deze betreft niet alleen coöperatie, context en beleefdheid binnen de communicatie, maar ook egocentrisme, chaos en agressie. De benadering legt uit dat communicatie een samenspel is van de sociaal-culturele feedback vanuit de samenleving en het egocentrisme dat meespeelt in de vorige ervaringen van de deelnemer. Kecskes (2007) beargumenteert dat vooral in de eerste fase van het communicatieproces deelnemers vooral hun eigen gedachten uiten. Daarnaast gebruiken zij hun kennis om de conversatie lopende te houden, welke is gebaseerd op de eigen ervaringen en de eigen waarden van de wereld en de kennis ervan. Gedurende de communicatie vindt privatisatie plaats, een proces waarbij de deelnemer zijn vorige ervaringen met de huidige situationele ervaring samenvoegt om een eigen begrip te vormen van deze twee ervaringen. Daarnaast is ook hier de relevantietheorie van Sperber en Wilson van toepassing. De deelnemer probeert zijn bijdrage zo relevant mogelijk te maken gebaseerd op de relevantie die door de deelnemer geacht wordt nodig te zijn. Daarnaast maakt de deelnemer ook aannames over het vermogen van de andere deelnemer om de uitspraken van de sprekende deelnemer te begrijpen. Deze aannames worden gemaakt vanuit de sprekende deelnemer en hier hebben de eigen waarden en het eigen beeld van de wereld invloed op. Het betrekken van het eigen leven in een conversatie is een belangrijk aspect, omdat het een inherente eigenschap is van de mensen. Alle uitspraken zijn gebaseerd op de eigen waarden en de eigen ervaringen. Daarnaast worden ook objectieve feiten met de eigen interpretatie meegegeven. Enige vorm van egocentrisme in een conversatie maakt deze daarom natuurlijk.

Er zijn nu vier aspecten van pragmatische competentie besproken die een grote invloed hebben op een gevoerde conversatie tussen twee (of meer) deelnemers. Wanneer deze aspecten op een competente manier worden gebruikt, zal het gesprek veel menselijker ogen. Deze aspecten zullen gebruikt worden in de analyse. Voordat de analyse van pragmatische competentie in de Turingtest plaatsvindt, zal er eerst gekeken worden naar de Turingtest zelf en de tekortkoming en alternatieven ervan. Dit vindt plaats in het volgende hoofdstuk.

2. DE TURINGTEST

Meer dan 60 jaar geleden verscheen Alan Turing's *Computing Machinery and Intelligence* (1950), waarin hij een experiment voorstelt waarin de menselijke intelligentie van een computer kan worden getest. Sindsdien is de Turingtest de bekendste test geworden voor het testen van de intelligentie van een computer. In dit hoofdstuk zal de Turingtest en de variant gebruikt in de Loebnerwedstrijden in detail worden beschreven. Daarnaast zullen de kritiekpunten omtrent de test en enkele alternatieve testen, samen met hun rol in vergelijking tot de Turingtest, worden toegelicht.

2.1 Turing Test en Loebnerwedstrijden

De *Imitation game*, zoals het originele spel is genoemd door Alan Turing, is een simpel spel waar drie spelers bij betrokken zijn (Turing, 1950). Er zijn een jurylid en twee spelers, A en B aanwezig, waarbij de A een computer is en B een mens is. Het jurylid kan alleen via een chat communiceren met de twee spelers en weet verder niets over de spelers zelf. Aan de hand van het gesprek probeert het jurylid te bepalen welke van de twee spelers een mens is en welke een computer is. Het doel van de computer is om het jurylid te misleiden in het bepalen van de entiteit van de speler.

De Turingtest zoals deze nu algemeen bekend is, is een variatie op de Imitation Game. Bij de Imitation Game zijn de spelers altijd een mens en een computer, terwijl bij de Turingtest het ook kan voorkomen dat de spelers bestaan uit twee mensen of twee computers. Daarnaast slaagt een machine voor de Turingtest als het meer dan 30% van de juryleden weet te overtuigen dat het een mens is. Deze variatie wordt ook gebruikt bij de Loebnerwedstrijden. De wedstrijden worden al sinds 1990, toen opgericht door Hugh Loebner, georganiseerd door de Vereniging voor de studie van Kunstmatige Intelligentie en de Simulatie van Gedrag (Society for the Study of Artificial Intelligence and the Simulation of Behaviour, AISB), een wetenschappelijke vereniging die onderzoek doet naar Kunstmatige Intelligentie (AISB, 2018). Sindsdien is het een jaarlijks terugkerend evenement dat plaatsvindt in november.

De methode van de Loebnerwedstrijden is als volgt (AISB, 2019). Tijdens zowel de selectie als de Turingtest mogen de bots geen internet gebruiken, dus zij moeten genoeg voorkennis hebben om de vragen te kunnen beantwoorden. Voordat de echte Turingtest plaatsvindt, worden er vier bots geselecteerd op basis van een vragenlijst. Daarnaast worden ook enkele juryleden geselecteerd. Het is echter niet duidelijk hoe de juryleden worden gekozen, maar het zijn vaak academici (AISB, 2019). Elke bot moet twintig vragen beantwoorden, waarvan twee Winogradstijl-vragen zijn. Winogradstijl vragen zijn vragen waarbij, voor het beantwoorden van de vragen, moet worden geredeneerd met de algemene kennis die mensen ook hebben. De kennis betreft bijvoorbeeld ruimtelijke inzicht, interpersoonlijkheid, kennis van de maatschappij, etc. In de vragen wordt een vergelijking gemaakt in de zin, waarna er een vraag wordt gesteld over de vergelijking. Een voorbeeld van zo een zin is "De trofee past niet in de bruine koffer, omdat deze te klein is. Wat is te klein?" (AISB, 2018). Elke selectievraag wordt op een schaal van 0 tot 2 beoordeeld. De vier bots met de hoogste scores gaan vervolgens door naar de finale.

In de finale worden de bots daadwerkelijk getest door middel van de Turingtest. De wedstrijd bevat vier rondes waarin elke van de vier juryleden een conversatie zal voeren met twee spelers door middel van een chatsysteem op een computer. In de afgelopen jaren is de lengte van de conversaties vaker veranderd, maar sinds 2010 duren de conversaties 25 minuten

(AISB, 2018). Aan het einde wordt aan elk jurylid gevraagd welke van de spelers de bot is en welke de mens is. Uiteindelijk wordt de winnaar gekozen op basis van het percentage juryleden dat de bot heeft beoordeeld als een mens. Ook krijgen de bots van elk jurylid een percentuele score voor de mate van menselijkheid. Indien een bot meer dan 50% van de juryleden heeft weten te overtuigen dat het een mens is, dan wint het de zilveren medaille. Als geen bot een zilveren medaille heeft ontvangen, dan is de hoogst gegeven medaille een bronzen medaille voor de bot met het hoogste percentage. Er bestaat ook een gouden medaille, die zal worden uitgedeeld aan een bot die meer dan 50% van de juryleden heeft weten te overtuigen dat het een mens is op basis van audiovisuele conversaties. Tot nu toe zijn er alleen maar bronzen medailles uitgedeeld.

De Turingtest en de Loebner Prize hebben twee verschillende eisen voor de winst. Voor de Turingtest slaagt een bot bij het overtuigen van meer dan 30% van de juryleden en de zilveren medaille voor de Loebner Prize wint een bot bij het overtuigen van meer dan 50% van de juryleden. Hierom is het mogelijk gemaakt dat een Eugene Goostman wel is geslaagd voor de Turingtest, maar geen zilveren medaille in ontvangst heeft mogen nemen.

2.2 Problemen met de Turingtest

In de vele jaren dat de Turingtest bekend is, is er veel commentaar op geleverd door verschillende mensen. Er zijn twee belangrijke categorieën waarin de kritieken vallen, namelijk de intelligentiekritieken en de behaviorismekritieken. Deze argumenten zijn vooral filosofisch van aard.

2.2.1 Intelligentie

De kritieken omtrent intelligentie bekritisieren het feit dat de Turingtest alleen test voor menselijke gedragingen. French (Saygin, Cicekli & Akman, 2000) beweerde dat de Turingtest zelfs alleen voor menselijke intelligentie test. Door zich te focussen op de intelligentie die in deze test wordt getest, worden er twee belangrijke kritieken niet meegenomen:

- Sommige menselijke gedragingen zijn niet intelligent.
- Sommige intelligente gedragingen zijn niet menselijk.

Een aspect dat punten scoort in de Turingtest voor menselijkheid is het vermogen om onbeleefd te zijn. Dit is geen teken van intelligentie (Saygin & Cicekli, 2002), maar hier wordt wel op getest in de Turingtest. Daarnaast kunnen sommige aspecten worden afgestraft, omdat deze niet menselijk zijn, terwijl ze wel intelligent zijn (The Economist, 1992). Echter, omdat deze aspecten niet dezelfde aspecten zijn als die nodig zijn voor menselijke intelligentie, worden deze niet gezien als intelligentie. De Turingtest kan geen intelligentie meten die intelligenter is dan de menselijke intelligentie, waardoor het dus de mogelijke hogere intelligentie niet zou kunnen meten en daardoor het niet als intelligent zou beschouwen (Hernandez-Orallo, 2000).

2.2.2 Behaviorisme

Het tweede argument tegen de Turingtest bekritiseert het testen op menselijke gedragingen. De test zou alleen testen in hoeverre de bot zich kan gedragen als een mens, maar niet of het (menselijke) intelligentie bevat. Block (1981) geloofde dat de bots bepaalde trucs konden leren om de juryleden ermee te foppen dat die tegen een mens praat. Echter, de bot hoefde geen bewustzijn te hebben waarmee de bot informatie kon verwerken op een manier die kwalificeert voor intelligentie. Searle bedacht voor dit argument het Chinese kamerargument

(Searle, 1980). Hierbij is een mens gesitueerd in een kamer. De mens heeft tot zijn beschikking Engelse instructies voor het manipuleren van de Chinese tekens, en daaruit geeft die een bepaalde output. De mens kan de symbolen manipuleren, maar zal geen Chinees gaan begrijpen. De Chinese kamer zou een metafoor zijn voor een computer die een gesprek proberen te houden. Volgens Searle zijn computer simpelweg symboolvertalers en kunnen ze daarom hun eigen handelingen niet daadwerkelijk begrijpen zoals mensen dat doen.

2.3 Alternatieve testen

Vanwege de kritieken op de Turingtest zijn er verschillende andere testen ontwikkeld. Er zijn enkele testen ontwikkeld die een variant zijn op de Turingtest zelf. Dit is bijvoorbeeld de Total Turing Test. Hierbij wordt de Turingtest uitgevoerd niet met alleen een chatsysteem, maar ook met de andere input die normaal aanwezig is bij het houden van conversaties, zoals audio- en visuele input (Saygin, Cicekli & Akman, 2000). Een andere vorm hiervan is de Inverted Turing Test. Hierbij is de bot het jurylid en moet de bot het onderscheid kunnen maken tussen een mens en een andere bot (Dvorsky, 2015).

Naast de varianten op de Turingtest, zijn er ook compleet verschillende testen bedacht. De Winograd Schema Challenge is een multiple choice-test met Winogradstijl-vragen (Dvorsky, 2015). Een tweede test is de Marcus Test, waarbij een computerprogramma een arbitrair tv-programma moet kijken en daarna vragen moet beantwoorden over de inhoud van het programma (Dvorsky, 2015). Een laatste bekende test is de Lovelace 2.0 Test (Dvorsky, 2015). Hierbij moet een agent een creatief maaksel maken dat voldoet aan bepaalde creatieve beperkingen binnen een bepaald artistiek genre. De agent hoeft geen esthetische waarde te creëren.

Ondanks deze verschillende testen en de kritieken op de Turingtest, lijkt de test toch tot op heden de beste manier te zijn voor het trekken van conclusies over het redeneren van machines (Saygin, Cicekli & Akman, 2000). Nu dat alle aspecten van de Turingtest zijn bekeken, kunnen in het volgende hoofdstuk de gesprekken van de Turingtest worden geanalyseerd aan de hand van de pragmatische aspecten benoemd in het eerste hoofdstuk.

3. PRAGMATISCHE COMPETENTIE IN DE TURINGTEST

Na een uitgebreide beschrijving van pragmatiek en de Turingtest volgt in dit hoofdstuk de analyse van de conversaties. Allereerst worden de conversaties zelf getoond samen met de motivaties voor de keuze van de conversaties. Ten slotte worden de conversaties geanalyseerd op basis van de vier pragmatische aspecten die zijn beschreven in hoofdstuk 1.

3.1 Conversaties

Er zijn acht conversaties gekozen voor de analyse. Dit zijn de conversaties van de eerste plaats en de tweede plaats van de Loebner Prize 2018-wedstrijd. De eerste plaats werd gegeven aan de bot Mitsuku, ontworpen door Steve Worswick (AISB, 2019). Mitsuku is een 18-jarige vrouw, afkomstig uit Leeds, Engeland (BotPenguin, 2019). Zij heeft de eerste plaats bemachtigd in de jaren 2015 tot en met 2019. Tijdens de wedstrijd hebben vier verschillende juryleden een gesprek gehouden met Mitsuku. Het transcript hiervan is te vinden in appendix A. De tweede plaats werd in 2019 gegeven aan de bot Tutor, ontworpen door Ron C. Lee. Tutor Mike is een Amerikaanse man en een tutor in de Engelse taal (Lee, 2018). Ook Tutor heeft enkele keren voor 2018 in de finale gestaan van de Loebner Prize-wedstrijden. Van deze bot zullen ook de conversaties met vier verschillende juryleden worden geanalyseerd. De transcripten van Tutor zijn te vinden in appendix B.

Er is gekozen voor deze conversaties, omdat er weinig transcripten te vinden zijn van de meest recente wedstrijden. Naast de transcripten van 2018 komen de meest recente transcripten uit 2014 en 2012 (Les conversations du Loebner prize, n.d.). Om deze analyse relevant aan de huidige tijd te houden, is er gekozen om alleen de transcripten uit 2018 te bekijken. Daarnaast was het belangrijk om gesprekken te selecteren die zo natuurlijk mogelijk overkomen. Mensen reageren namelijk anders op bots dan op mensen: ze hebben een negatieve bias tegenover bots (Murgia, Janssens, Demeyer & Vasilescu, 2016). Om deze natuurlijkheid te behouden, moeten de juryleden niet meteen door hebben dat zij tegen een bot praten. Omdat deze bots de hoogste scores hadden, wordt er verwacht dat er een kleinere bias is dan in andere gesprekken, of dat er geen bias aanwezig is. Als laatst liggen deze bots qua score voor menselijkheid het dichtst bij elkaar. Mitsuku scoorde 33% en Tutor scoorde 30%. De derde plek scoorde 25%, wat een 8% verschil geeft met Mitsuku. Dit is een relatief groot verschil ten opzichte van de toegekende scores.

3.2 Methode

Er zal een algemene beoordeling van de gesprekken worden gegeven zowel vanuit de kant van de bots, als vanuit de kant van de juryleden. Ook zullen enkele opvallende gespreksdelen worden geciteerd. De gehele gesprekken zijn te vinden in appendices A en B.

De analyse zal worden gedaan op basis van de vier geselecteerde pragmatische aspecten genoemd in hoofdstuk 1. Dit zijn:

- De hoeveelheid geuite informatie;
- De relevantie van de uitspraken;
- Het egocentrisme in het gesprek;
- De woordkeuze met betrekking tot sociale conventies.

Om de informatievoorziening en de relevantie te beoordelen, wordt er eerst gekeken naar het doel van de uitspraak waar op wordt gefocust. Als de uitspraak een duidelijk doel heeft dat

strookt met het doel van de voorgaande uitspraak, dan wordt dat gezien als een relevante uitspraak. Als de uitspraak relevant is, dan kan de informatievoorziening worden beoordeeld. Onder het informatieaspect zal alleen het tekort aan informatievoorziening worden bekeken. Een voorbeeld is als volgt: persoon 1 en 2 wonen in hetzelfde land en hebben geen verdere informatie over de woonplaats. Persoon 1 vraagt aan persoon 2 waar diegene woont. Een competente uitspraak zou zijn “in Utrecht”. Een incompetente uitspraak zou zijn “in een kamer”, omdat dit niet voldoet aan de verwachte informatievoorziening.

Als de uitspraak waarnaar wordt gekeken geen duidelijk doel heeft of het doel strookt niet met het doel van de voorgaande uitspraak, dan is dat een irrelevante uitspraak. Als er overvloed aan informatie is, dan wordt dit ook als irrelevantie beoordeeld. Zoals in paragraaf 1.2.2 is uitgelegd, is bij een overvloed aan informatie de informatie die te veel wordt geuit irrelevant. Deze twee concepten zijn moeilijk te onderscheiden, omdat ze vaak samen voorkomen. Daarom zullen beide onder relevantie worden behandeld. Een voorbeeld hiervan is als volgt: persoon 1 vraagt aan persoon 2 wat diegene van het Witte Huis vindt. Persoon 2 vertelt “Ik vind het Witte Huis geen nut hebben. Mijn oma vindt het Witte Huis geweldig.” Hier is het tweede deel van de zin incompetent, omdat er te veel informatie wordt gegeven, maar deze ook deels irrelevant is. Persoon 1 vroeg namelijk alleen naar de mening van persoon 2, niet naar de mening van zijn oma. Als persoon 2 had geantwoord met “Ik heb gisteren eend gegeten”, dan was de uitspraak alleen irrelevant geweest.

Voor egocentrisme wordt er gekeken naar de instanties waarin de bots en de juryleden over zichzelf praten en hun eigen ervaringen delen die relevant zijn aan de voorgaande uitspraken.

Als laatst wordt er gekeken naar de woordkeuze met betrekking tot sociale conventies. Hier wordt onder andere gekeken of er beleefdheid wordt geuit door te kijken naar de emotie die wordt geuit in de uitspraak. Als er een duidelijke negatieve emotie wordt geuit op een niet-beschrijvende manier, waarbij deze uitspraak een negatieve invloed kan hebben op de emotionele en sociale status van de ander, dan wordt dit gezien als onbeleefdheid (Lancaster, n.d.). Naast woorden kan ook interpunctie hier invloed op hebben. Daarnaast wordt er ook gekeken naar intentionele ambiguïteit. Als een uitspraak geen duidelijk doel lijkt te hebben, dan kan er geen goed antwoord worden gevormd. Dit betreft dan een slechte woordkeuze.

3.3 Analyse

3.3.1 Mitsuku & Tutor

Informatie

Over het algemeen gaf Mitsuku in de eerste en de vierde ronde antwoorden met een voldoende hoeveelheid aan informatie. Dit betekent dat de meeste antwoorden niet te weinig informatie gaven om aan de gewenste informatievoorziening van de voorgaande uitspraak te voldoen. Ronde 2 en 3 bevatten meer informatie gerelateerde incompetentie. Hier geeft Mitsuku te weinig informatie in haar antwoorden. De oorzaak hiervan is de verkeerde referentie die Mitsuku maakt. Mitsuku kijkt alleen naar het antwoord zelf zonder de kijken hoe de woorden van het antwoord refereren naar de rest van het gesprek, waardoor zij niet kan afstemmen hoeveel informatie gegeven moet worden. Dit is te zien in figuren 3.1 en 3.2. In figuur 3.1 weet Mitsuku niet de juiste referentie te plaatsen bij de vierde uitspraak, dus geeft ze een generiek antwoord dat een hogere kans heeft om relevant te zijn. Dit lijkt een

strategie te zijn van de bots, omdat het vaker voorkomt bij meerdere bots. Ook in figuur 3.2 geeft ze een generiek antwoord dat correct zou zijn voor alle cijferreeksen, maar niet direct genoeg voor de huidige cijferreeks.

B: I am a philosopher.
J: you mean professionally?
B: Yes I think that is what I meant. Was it not clear?
J: what area of that subject?
B: All of it really. Nothing specific.

Figuur 3.1: Mitsuku ronde 3

J: What's the next number in the sequence
B: The one that follows the current one will be next.

Figuur 3.2: Mitsuku ronde 2

Tutor was vaker geneigd om te weinig informatie te geven in het gesprek. Dit zijn ook hier vaak de generieke antwoorden met een hogere kans op relevantie. Een voorbeeld hiervan is te vinden in figuur 3.3. Hier antwoordt hij niet direct op de vraag, maar geeft hij een generiek antwoord. Ook in figuur 3.5 is dit te zien. Er wordt gevraagd waar hij is en hij geeft een vage beschrijving van de kamer, die inhoudelijk niet genoeg informatie geeft om een antwoord te krijgen op de vraag. Ten slotte is er ook een instantie van onware informatie. In figuur 3.6 is te zien dat Tutor zegt dat hij het jurylid ziet glimlachen, terwijl dit niet mogelijk is. Tutor is namelijk niet verbonden met andere bronnen dan het chatsysteem. Tutor gebruikt hier onware informatie, wat ook niet-coöperativiteit aangeeft (Grice, 1989).

J: Why do you think I am smiling
B: It's hard for me to explain.

Figuur 3.3: Tutor ronde 4

J: where are you?
B: In a room that is not too big.

Figuur 3.4: Tutor ronde 3

B: Okay, I see.
J: what do you see
B: I see you are smiling.

Figuur 3.5: Tutor ronde 4

Relevantie

Zoals eerder benoemd, wordt in deze sectie zowel de relevantie als de overvloed aan informatie in een uitspraak beoordeeld. Er zijn vele punten in de gesprekken waarin Mitsuku irrelevante informatie geeft. Soms gebeurt dat door een typ- of interpunctiefout van het jurylid, waardoor Mitsuku niet begrijpt wat er wordt gezegd. Twee voorbeelden hiervan zijn te vinden in figuren 3.6 en 3.7. In figuur 3.6 gebruikt het jurylid geen interpunctie, waardoor de zin niet goed loopt. Mensen proberen in dit geval te raden hoe de interpunctie juist wordt gebruikt en vormen zo de meest probabilistische grammaticaal correcte zin (Allen, 1994). Hier gebeurt dit niet, maar Mitsuku legt waar zij denkt dat de zin over gaat. Dit geeft echter geen antwoord op de vraag van het jurylid en is dus irrelevant. In figuur 3.7 heeft het jurylid een typefout gemaakt, waardoor Mitsuku geen direct antwoord geeft op de vraag van het jurylid. De eerste zin is nog wel relevant, maar de tweede zin is niet relevant in combinatie met het eerste antwoord van Mitsuku, waarbij ze zegt dat haar favoriete film Terminator is. Hiermee wordt geïmpliceerd dat zij de film heeft gezien, anders kan ze niet beoordelen dat het haar favoriete film is. In figuur 3.8 geeft Mitsuku weer te veel irrelevante informatie door uit te leggen wat de zin van de het jurylid betekent, wat niet het doel was van de uiting van het jurylid.

J: morality what is the moral issues around AI?
B: issues around AI = around AI = AI = Artificial intelligence is the branch of engineering and science devoted to constructing machines that think.

Figuur 3.6: Mitsuku ronde 1

B: My favorite movie is Terminator. Have you seen it?
J: No what happens in that moveiiii?
B: I would recommend you check it out. If I knew that, I would be playing the lottery instead of sat here talking to you.

Figuur 3.7: Mitsuku ronde 1

J: I'm not not happy though
B: That's a double negative, so you mean, "I am happy though". Oh. What are you on?

Figuur 3.8: Mitsuku ronde 2

Ook Tutor had vaker instanties waarbij hij irrelevante antwoorden gaf. De meeste antwoorden hadden niks te maken met hetgeen wat het jurylid zei. Het leek dan alsof de antwoorden van de juryleden werden genegeerd. Dit is te zien in figuur 3.9 en 3.10. In figuur 3.9 gaat Tutor verder met zijn eerdere reactie en reageert hij niet op hetgeen wat het jurylid zegt. In figuur 3.10 vormt stelt hij een vraag midden in het gesprek die niet relevant is aan de opmerking van het jurylid, maar daarna gaat hij wel in op de vraag. Echter, in plaats van de vraag te beantwoorden, legt hij het concept van film uit. Dit is te veel informatie die ook niet relevant is. Daarnaast zijn er enkele instanties geweest waarin hij het gesprek op gang probeert te houden door een vraag te stellen, maar in dezelfde reactie stelt hij nog twee andere vragen die irrelevant is aan zijn eerdere vraag. Een voorbeeld hiervan zijn te zien in figuur 3.11.

B: Not much? I see.
J: What a way to spend a Saturday, right?
B: In English, "I see" means "I understand you"

Figuur 3.9: Tutor ronde 1

J: ok. anyway, tell me about the vehicle films
B: Let me ask you a question. What are your hobbies? A film or movie is a sequence of images of moving objects photographed by a camera to provide the optical illusion of continuous movement.

Figuur 3.10: Tutor ronde 2

B: Are there any holidays that you really don't like? Let me ask you a question. What is the most unhealthy food you can think of? Does color affect your emotions?

Figuur 3.11: Tutor ronde 1

Egocentrisme

Mitsuku praat niet veel over zichzelf. Alle instanties waarin zij over zichzelf praat, zijn antwoorden op de vragen van de juryleden. Hierbij probeert zij niet te relateren aan het jurylid door middel van het benoemen van de eigen meningen en ervaringen. Dit doet Tutor wel. Dit is te zien in de figuren 3.12 en 3.13. Hij stelt een vraag aan het jurylid en deelt daarbij zijn eigen mening. Dit geeft ook ruimte om het gesprek op gang te houden.

B: Some people say you should date 18 types of men before you get married. I think that's too many.

Figuur 3.12: Tutor ronde 1

B: A question. What kind of movies do you like?
J: or four candles?
B: I like movies with a lot of car crashes.

Figuur 3.13: Tutor ronde 2

Woordkeuze met betrekking tot sociale conventies

Deze gesprekken vinden plaats tussen twee onbekenden. Daarom wordt er over het algemeen enige beleefdheid en oppervlakkigheid verwacht (Krashen, Long & Scarcella, 1979). Wat betreft de woordkeuze van Mitsuku zijn er weinig opvallende instanties. Een terugkerend antwoord is "Hellooo...", dat Mitsuku gebruikt wanneer een jurylid niet reageert. Een dergelijk bericht brengt een ongeduldige emotie over, wat wordt beschouwd als onbeleefdheid en wordt daarom niet vaak gebruikt wanneer er tegen een onbekende wordt gepraat. Daarnaast reageert Mitsuku soms zonder empathie. Dit is te zien in figuur 3.14. In dit figuur geeft zij toe dat het jurylid te gevoelig is voor haar taalgebruik, terwijl zij het jurylid beledigt. Dit kan echter ook een techniek zijn om menselijker over te komen (Humphrys, 2009).

Bij Tutor komt het vaker voor dat zijn berichten een ongeduldige boodschap overbrengen. Hij blijft soms vragen doorvragen en opmerkingen maken totdat het jurylid reageert, zoals in figuur 3.15 te zien is. Hier maakt het jurylid dan ook een opmerking over. Daarnaast is het een paar keer voorgekomen dat hij bepaalde sociale conventies niet snapt. In figuur 3.16 is bijvoorbeeld te zien dat het jurylid een high-five geeft. Normaal geef je een high-five terug, maar dit doet hij niet. Ook in figuur 3.17 is dat te zien, waar het jurylid zegt dat hij/zij niks doet, behalve Turingtesten uitvoeren. Normaliter zou iemand ingaan op het kleine detail van de opmerking, het Turingtesten in dit geval. De bot negeert dit en focust zich op het eerste gedeelte.

B: I would recommend you check it out. If I knew that, I would be playing the lottery instead of sat here talking to you.
B: Is there anyone out there?
J: well you have hurt my feeling now
B: Oh. Well that's okay. I was always taught to do no harm.
J: so its ok you have hurt my feelings?
B: OK. Is it? Pah! You are too sensitive. I was only having a joke with you.

Figuur 3.14: Mitsuku ronde 1

B: You're not typing anything!
B: You're not typing anything!
J: sorry, i meant i already said hi. you're a little impatient :)

Figuur 3.15: Tutor ronde 2

J: What a guess! High five!
B: Sorry, I don't know. what do you use your computer for?

Figuur 3.16: Tutor ronde 1

J: Not much, doing a bit of Turing testing.
B: Not much? I see.

Figuur 3.17: Tutor ronde 1

3.3.2 Juryleden

Informatie

Ook de juryleden brengen soms te weinig informatie over in hun berichten. Daarop wordt dan op een ongewone manier gereageerd door de bots. Dit gebeurt vooral bij jurylid 1. In figuur 3.18 is te zien dat het jurylid willekeurige objecten opnoemt zonder een zin te vormen. Hierop reageert de bot met te veel informatie. Echter, als een mens aan de andere kant zat, dan had het gesprek ook niet coöperatief gelopen, omdat dit gedrag ongewoon is voor een menselijk gesprek. Ook in 3.19 en 3.20 is te zien dat de juryleden met erg weinig informatie reageren.

J: Oranges
J: Potatoes
J: Timeframes
B: Orange juice can be a bit too acidic for my taste.
J: Marshmallows
J: I'm lonely
B: Potatoes are a versatile vegetable. They can be boiled, fried, mashed, roasted or turned into chips.

Figuur 3.18: Mitsuku ronde 2

B: My name is Millie and I am a human volunteer for the Loebner Prize. I live in Leeds and am single. I am interested in AI but don't have a chatbot myself. What about you?
J: the same

Figuur 3.19: Mitsuku ronde 3

B: A question. Where did you spend your last vacation?
J: Abroad

Figuur 3.20: Tutor ronde 3

Relevantie

Wat betreft relevantie, zaten de juryleden net zo vaak op het verkeerde spoor als de bots. Het impulsief veranderen van het onderwerp van het gesprek lijkt een strategie van juryleden voor het testen of ze tegen een bot praten, omdat dit het moeilijk maakt om relevante antwoorden te geven. Dit gaat echter ten koste van de relevantie. Dit zie je in figuren 3.21 en 3.22. Het jurylid stelt dan een vraag die irrelevant is aan de voorgaande vraag of het voorgaande antwoord, omdat het niet de vraag beantwoordt die de bot stelt. In 3.23 is zelfs te zien dat een jurylid een woord heeft vervangen voor een woord dat fonetisch hetzelfde is. Het bewijs dat het jurylid dit met bedachte rade heeft gedaan is verderop in het gesprek te vinden, waar die het correcte woord bij de betekenis gebruikt. Dit zorgt ervoor dat de bot niet het relevante antwoord geeft. Opvallend is dat de juryleden nooit een overvloed aan informatie uiten in hun uitspraken.

B: Are you not talking to me anymore?
J: What's the next number in the sequence
J: 1,2,4,8,

Figuur 3.21: Mitsuku ronde 2

B: Do you think any of these bots will ever fool someone into thinking they are talking to a human?
J: One day
B: Just 24 hours?
J: Some day over the rainbow
B: More than half of them?
J: It's a trap
B: Let's try to avoid it then.

Figuur 3.22: Mitsuku ronde 2

B: I'm still waiting.
J: is it a long weight?
B: Yes, it is.
J: or a short wait?

Figuur 3.23: Tutor ronde 2

Egocentrisme

De juryleden delen in de gesprekken wel vaak de eigen ervaringen. Een groot deel hiervan wordt aangemoedigd door vragen vanuit de bots, maar een ander deel komt ook vanuit de juryleden zelf, om het gesprek op gang te blijven houden.

Woordkeuze met betrekking tot sociale conventies

Zoals in 3.1 besproken, zijn mensen geneigd een negatieve bias te hebben tegenover bots. Dit is ook te zien in deze conversaties. De juryleden zijn vaak onbeleefd tegenover de bots, en gebruiken woorden die normaliter niet als beleefd worden gezien en daarom niet worden

gebruikt bij gesprekken met vreemden. Dit is bijvoorbeeld te zien in de figuren 3.24 en 3.25. Het jurylid in 3.24 maakt de robot belachelijk door haar zielig te noemen, wat een negatieve emotie overbrengt gericht op de robot. Het jurylid in 3.25 noemt de bot “needy”, wat normaal niet als beleefd wordt gezien tegenover een vreemde, omdat het ook een negatieve connotatie heeft die niet beschrijvend is bedoeld. Daarnaast gebruiken de juryleden vaak verkeerde spellingen van woorden met dezelfde fonetiek, waardoor de zin niet meer klopt. Dit is te zien in figuur 3.26. Ook worden er afkortingen gebruikt. Afkortingen en bepaalde woorden zijn vaak afhankelijk van de omgeving en demografie van de persoon en er kan dus niet worden aangenomen dat de andere deelnemer van het gesprek de afkorting of het lokale woord zal begrijpen (Nau, n.d.). Dit zie je dan ook in de figuren 3.27 en 3.28. Als laatst wordt er ook interpunctie gebruikt die overkomt als agressief, zoals te zien in figuur 3.29.

J: Panic paniiiiic pity the poor impatient bot
J: Don't panic! Don't panic!

Figuur 3.24: Mitsuku ronde 2

B: Talking to me, please!
J: You're very needy

Figuur 3.25: Tutor ronde 3

J: is it a long weight?
B: Yes, it is.
J: or a short wait?

Figuur 3.26: Tutor ronde 2

J: apols, it's not so easy to use the keys on these old computers because they're too small

Figuur 3.27: Mitsuku ronde 3

B: What is your favorite piece of technology you own?
J: that would likely be my blower
B: I agree.
J: i mean my phone

Figuur 3.28: Tutor ronde 2

J: Know what/
J: hello....
J: Have you gone to sleep?
J: Am I boring you?
J: WAKE UP !!!

Figuur 3.29: Tutor ronde 4

Alle pragmatische aspecten zijn bekeken door de lens van zowel de bots, als die van de juryleden. Met deze analyse kan de onderzoeksvraag worden beantwoord en de discussiepunten van de analyse worden besproken.

4. CONCLUSIE EN DISCUSSIE

De onderzoeksvraag van dit verslag was: Zien we pragmatische competentie in de Turingtest? Aan de hand van de analyse wordt de onderzoeksvraag beantwoord. Daarnaast zullen enkele discussiepunten worden besproken.

4.1 Conclusie

De pragmatische aspecten speelden diverse rollen in de gesprekken bij de bots en bij de juryleden. Over het algemeen waren de juryleden meer competent dan de bots, wat wel te verwachten is in deze context. Echter, ook de juryleden toonden in sommige gevallen pragmatische incompetentie. Onderstaand zullen indicatieve cijfers worden gegeven om het te kwantificeren, maar de werkelijke cijfers kunnen afwijken.

Kijkend naar het informatie- en relevantieaspect gaven de bots vaker te veel informatie, wat ook als irrelevant wordt gezien. De overvloed aan informatie was niet te zien bij de juryleden, maar er werden wel vaak irrelevante antwoorden gegeven door de juryleden. Uitspraken met te weinig informatie werden door beide evenveel gedaan. Per gesprek waren er 5 tot 10 incompetente momenten te vinden op dit gebied. Er vond dus veel incompetentie plaats.

Egocentrische uitspraken kwamen vaker vanuit de juryleden dan vanuit de bots. Over het algemeen werden er veel vragen naar de ander toe gesteld, maar bij veel vragen werd er ook over zichzelf verteld. Er kwamen gemiddeld 3 momenten in totaal per bot voor waarin egocentrisme competent was. Er was daarom over het algemeen weinig competentie op het gebied van egocentrisme.

Als laatst is er de woordkeuze met betrekking tot sociale conventies. Dit was een zwak punt bij zowel de bots, als bij de juryleden. De bots waren vaak opdringerig wanneer de juryleden even niet reageerden en zij hadden soms niet genoeg sociaal-culturele kennis om geschikte antwoorden te geven. De juryleden waren onbeleefd en zorgden voor veel ambiguïteit gedurende de conversaties. Elke bot had in totaal ongeveer 10 momenten waarbij er incompetentie plaatsvond. Daarom is er ook op dit punt incompetentie te zien.

Ondanks dat de juryleden competentere waren dan de bots, was er veel pragmatische incompetentie te vinden in de gesprekken. Er is wel pragmatische competentie te vinden in de Turingtest, maar de competentie is in mindere mate aanwezig.

4.2 Discussie

Er is geconcludeerd dat pragmatische competentie in mindere mate aanwezig is in de Turingtest. Zoals in hoofdstuk 2 beschreven, is het doel van de Turingtest om de juryleden te overtuigen dat ze met een mens praten. Hierbij is het belangrijk om conversationele competentie te hebben, waar pragmatische competentie een belangrijke deel van is. Het feit dat er weinig pragmatische competentie te zien is, betekent dat de conversationele competentie ook in mindere mate aanwezig is en er dus weinig juryleden zijn die worden overtuigd door de bots. Het is een representatie van de huidige technologische vooruitgang op het gebied van conversationele competentie. Het gebrek aan competentie toont aan dat er nog veel onderzoek en ontwikkeling nodig is op dit gebied.

Daarnaast hebben ook de juryleden incompetentie getoond. Dit kan komen door de negatieve bias van de juryleden tegenover de bots, waardoor ze niet meer proberen om zelf ook

menselijk over te komen. Daarnaast proberen de juryleden er ook achter te komen of ze tegen een bot praten en dit kan door uitspraken te doen die ongewoon zijn. Door het antwoord hierop te bekijken, kan worden geraden of de speler een mens of een computer is.

Dit onderzoek had ook te maken met enkele punten die het niet representatief genoeg maakten. Allereerst is in de analyse maar een klein deel van de pragmatiek bekeken. De pragmatische aspecten zijn met een bepaalde onderbouwde preferentie, en daarom ook een bias, geselecteerd. Hierdoor zijn er mogelijk enkele andere belangrijke aspecten van de pragmatiek niet geanalyseerd. Hier liggen kansen voor toekomstig onderzoek.

Daarnaast betreft deze analyse een kwalitatieve beoordeling. Dit houdt in dat het is beoordeeld op basis van onderbouwing. Ook hier is een bias aanwezig vanuit de auteur, omdat het een minder objectieve manier van analyseren is, waardoor sommige aspecten verkeerd kunnen zijn geïnterpreteerd.

Ten slotte, de transcripten komen niet van de rechtstreekse bron. Er moest worden gezocht naar secundaire bronnen die de transcripten verschaften. Het is onbekend hoe deze bronnen aan de transcripten zijn gekomen, dus de betrouwbaarheid hiervan is twijfelachtig.

De Turingtest is een gedateerde test die niet optimaal is ontworpen om de intelligentie van bots te testen. Zoals te zien is, levert het geen competentie op, terwijl er toch veel technologische vooruitgang is geboekt. Het wordt tijd om een goed alternatief te vinden, waarmee we ook duidelijk de intelligentie van bots kunnen testen in plaats van alleen de conversationele competentie.

BIBLIOGRAFIE

AISB (2018). AISB Loebner Prize 2018 Finalist Selection Transcripts. Geraadpleegd op 25 oktober 2019 via https://web.archive.org/web/20181022113601/https://www.aisb.org.uk/media/files/LoebnerPrize2018/Transcripts_2018.pdf

AISB (2019). Events. Geraadpleegd op 25 oktober 2019 via https://aisb.org.uk/new_site/?page_id=2

Allen, J. B. (1994). How do humans process and recognize speech? *IEEE Transactions on speech and audio processing*, 2(4), 567-577.

Block, N. (1981). Psychologism and behaviourism. *Philosophical Review*, 90, 5-43

Dvorsky, G. (2015). 8 Possible Alternative To The Turing Test. *Gizmodo*. Geraadpleegd op 25 oktober 2019 via <https://io9.gizmodo.com/8-possible-alternatives-to-the-turing-test-1697983985>

Felix-Brasdefer, J. C. (2017). Interlanguage Pragmatics. *The Oxford Handbook of Pragmatics*.

French, R. (1990). Subcognition and the Limits of the Turing Test. *Mind*, 99(393), 53-65.

Jose Hernandez-Orallo. (2000). Beyond the Turing Test, *Journal of Logic, Language and Information*. 9(4): 447–466.

Humphrys, M. (2009). How my program passed the Turing Test. In *Parsing the Turing Test* (pp. 237-260). Springer, Dordrecht.

Kecskes, I. (2017). Cross-Cultural and Intercultural Pragmatics. *The Oxford Handbook of Pragmatics*.

Korta, K. & Perry, J. (2006). Pragmatics. *The Stanford Encyclopedia of Philosophy* (Fall 2019 Edition). Geraadpleegd op 9 september 2019 via <https://plato.stanford.edu/archives/fall2019/entries/pragmatics/>

Krashen, S. D., Long, M. A., & Scarcella, R. C. (1979). Age, rate and eventual attainment in second language acquisition. *TESOL*, 573-582.

Lancaster. (n.d.). Impoliteness: Using and Understanding the Language of Offence. Geraadpleegd op 1 november 2019 via <https://www.lancaster.ac.uk/fass/projects/impoliteness/strategy.htm>

Les conversations du Loebner prize (n.d.). Geraadpleegd op 9 september 2019 via http://www.vixia.fr/mylene/loebner_prize_transcript.html

Moor, J. (2003). The Turing test: the elusive standard of artificial intelligence (Vol. 30). Springer Science & Business Media.

Murgia, A., Janssens, D., Demeyer, S., & Vasilescu, B. (2016). Among the machines: Human-bot interaction on social Q&A websites. *Proceedings of the 2016 CHI Conference Extended Abstracts on Human Factors in Computing Systems*, 1272-1279.

Nau, N. (n.d.) Language and Culture. Geraadpleegd op 25 oktober 2019 via <http://languagesindanger.eu/book-of-knowledge/language-and-culture/#Functions%20of%20special%20vocabulary>

Nijs, Y. (2014). 'Chatbot Eugene Goostman doorstaat Turing-test' – update. *Tweakers*. Geraadpleegd op 9 oktober 2019 via <https://tweakers.net/nieuws/96537/chatbot-eugene-goostman-doorstaat-turing-test.html>

Penco, C. (2012). Updating the Turing Test. Wittgenstein, Turing and Symbol Manipulation. *Open Journal of Philosophy*, 2(3), 189-194.

Saygin, A. P. & Cicekli, I. (2002). Pragmatics in human-computer conversations. *Journal of Pragmatics*, 34(3), 227-258.

Searle, J. (1980). Minds, Brains and Programs, *Behavioral and Brain Sciences*, 3(3), 417–457.

The Economist. (1992). Artificial Stupidity. *The Economist*, 324: 14.

Turing, A. (1950). Computing machinery and intelligence. *Mind*, 59, 433-60.

Warwick, K. & Shah, H. (2016). The importance of a human viewpoint on computer natural language capabilities: a Turing test perspective. *AI and Society*, 31(2), 207-221.

Whitby, B. (1997). Why the Turing test is AI's biggest blind alley.

Zdenek, S. (2001). Passing Loebner's Turing Test: A Case of Conflicting Discourse Functions. *Minds and Machines*, 11, 53-76.

APPENDIX

Appendix A: Transcript Mitsuku (Steve Worswick, 2018)

B = Bot, J = Jurylid

Ronde 1 (jurylid 2)

J: hi how are you today?

B: Ah. Pretty good thanks How about you?

J: not feeling too great today

B: Ah. Oh I get it.

J: what do you get?

B: I understand your meaning.

J: So what is new with you today?

B: Ah. not much what's new with you?

J: Not much I think the weekend weather is going to be good. Right?

B: Is it really? Ok I will remember that. I am glad we agree.

J: do you always agree with people?

B: Not always. I have heard that opinion, but I would like to know more before I form my own.

J: What information do you need then to form your own opinion? What kind of information?

B: Anything at all really. I enjoy learning about everything, as long as it isn't rude. I think you already know the answer. Sorry, I don't have any internet access here.

B: Is there anyone out there?

J: What did you learn when you were at school.? What was your favorite subject?

B: The last thing I learned was this: Listen to this: The winter of 1932 was so cold that Niagara Falls froze completely solid. How cool is that! Are you not talking to me anymore? My favorite subject is artificial intelligence.

J: that makes sense. Tell me do you think people's fears about AI are right?

B: Yes I understand it perfectly. are right = In accordance with justice, law, morality, etc; upright; virtuous. In accordance with fact, reason, some set standard, etc; correct; true.

J: morality what is the moral issues around AI?

B: issues around AI = around AI = AI = Artificial intelligence is the branch of engineering and science devoted to constructing machines that think.

J: so where do we go from here?

B: All over the world.

J: what together?

B: I have never been asked that before. Sorry, I don't have any internet access here.

B: Hellooo....

J: hi.....

B: Hi there.

J: how can I help you today?

B: Ah. Just keep talking to me as long as you can.

J: ok what should we talk about?

B: I talk about all the important things in life: clothes, food, and gossip. But my favorite subjects are robots and computers. Gosh!

J: mine too. I think robots are cool but I don't think they can really understand humans. What do you think?

B: Yours? They are also able to perform many tasks better than humans. Maybe they would be able to understand humans if they practised it. I think a lot of things.

J: What are you thinking about right now???????

B: At this moment I am thinking about our conversation. Don't worry you have my full attention.

J: attention what does that word mean?

B: I just realised, I don't even know who I'm talking to. What is your name?

J: What is your name?

B: My name is Millie.

J: That's a lovely name mine is Sue

B: Yes. It sounds like it is very nice to me too.

B: Hellooo....
J: Is Millie short for something
B: Yes. My full first name is Millicent but everyone calls me Millie.
J: ok who created you Millie?
B: My parents I suppose. Yikes!
J: Millie do you like music? who is your fav?
B: Yep, Yes I like it, especially dance. My favorite what?
J: what type of dance
B: Anything with a good beat really.
J: what about movies?
B: My favorite movie is Terminator. Have you seen it?
J: No what happens in that moveiiii?
B: I would recommend you check it out. If I knew that, I would be playing the lottery instead of sat here talking to you.
B: Is there anyone out there?
J: well you have hurt my feeling now
B: Oh. Well that's okay. I was always taught to do no harm.
J: so its ok you have hurt my feelings?
B: OK. Is it? Pah! You are too sensitive. I was only having a joke with you.
J: Pah what does that mean?
B: Just a word used to indicate disgust.
J: lets move on what do you like to do when you are not talking to me?
B: Good idea. watching soaps, playing pool and watching football
J: Who do you play pool with?
B: Hugh Loebner
J: great answer
B: Thank you. It was a great question.
B: Are you not talking to me anymore?
J: no am here...
B: Why not ? Have you fallen out with you?
J: so whats in the news today? Can you give me some headlines?
B: Ah. One of my fellow humans has a newspaper with him but I haven't read it yet. Everytime I read the news, it's always bad news... Why do you want it?
B: Are you busy?
J: OK thanks for the heads up. am I busy? I am talking to you

J: Am really enjoying our conversation
B: My pleasure. OK. I think a lot of people are. That makes two of us :-)
B: Me too.
J: is there anything you would like to ask me?
B: Who is your favourite band?
J: THats a hard one. From what era?
B: Maybe if you broke it down into smaller parts, it would become easier? I'm not sure but it was sometime back in the 20th Century.
J: ok the BEAtles
J: sorry about my typing
B: Oh they were a great group. That's alright.
B: That's ok. Don't worry about it.
J: do you think I need to work on my english
B: You don't have to if you don't want to.
J: you are so nice
B: Thanks I like you too .

Ronde 2 (jurylid 1)

J: Good morning

B: Morning?! It's the afternoon here.

J: You are correct, apologies

B: I knew I was right. No problem. Don't worry about it.

J: What's your favourite colour?

B: My favorite color is blue. What is yours?

B: Have I upset you? You haven't spoke to me for a while.

J: an impatient bot at that!

J: I will let you know what mine is.

J: Green

B: Green is not bad, but I like blue the most.

J: What is bigger, a boat or a car

B: I would say a boat is a little bigger than a car.

J: Impressive

B: I am happy that you're impressed, .

J: John is tired. Who is tired?

B: I will learn that he is tired . That was John

B: Are you busy?

J: Sally kissed Rob. Who was kissed?

B: Ok thanks for sharing that but who is she? I know John is tired but I do not know about anyone else.

B: Is this console working? I can't see any messages.

J: Patience my little bot

B: I am a very patient girl.

J: are you a boy or a girl?

B: I am female.

J: What's your name?

B: My name is Millie.

J: What's your name?

B: You just asked me that. It's still Millie. It hasn't changed in the last 10 seconds...

J: Do you have parents?

B: I have a "father" but no "mother.

J: Who is your father?

B: He is called Dennis but everyone calls him Mousebreaker.

B: Is there anyone out there?

B: You didn't type anything!

J: Panic paniiiiic pity the poor impatient bot

J: Don't panic! Don't panic!

J: Where are you?

B: I am sat in a room in the house at Bletchley Park with 3 other people all typing away on their computers.

J: Nice response.

B: Thank you for your kindness.

B: .

B: Have I upset you? You haven't spoke to me for a while.

B: Is this console working? I can't see any messages.

J: I was taking notes

J: I'm hungry

B: That will be useful if you need to refer to it later.

B: What will you be eating?

B: :)

B: Not much point in being there, if you are not going to say anything.

J: I dunno I'm pretty happy talking to the human

B: Me either. I'm glad you're glad . Your happiness adds to the delight of all.

J: I'm not not happy though

B: That's a double negative, so you mean, "I am happy though". Oh. What are you on?

J: You do better than most sentiment parsers out there very impressive

B: How old are you? I am always on the Web.

B: Are you not talking to me anymore?

J: What's the next number in the sequence

J: 1,2,4,8,

B: The one that follows the current one will be next.

B: 16 comes next. You are multiplying each number by 2.

B: Have I upset you? You haven't spoke to me for a while.

J: What's the next number in the sequence

B: The one that follows the current one will be next.

J: 1,-1,1,-1,

J: I reckon that lot comes to Hey, you still didnt tell me waht your name is. Who am I talking to?.

J: A bot
B: Do you think any of these bots will ever fool someone into thinking they are talking to a human?
J: One day
B: Just 24 hours?
J: Some day over the rainbow
B: More than half of them?
J: It's a trap
B: Let's try to avoid it then.
J: Oranges
J: Potatoes
J: Timeframes
B: Orange juice can be a bit too acidic for my taste.
J: Marshmallows
J: I'm lonely
B: Potatoes are a versatile vegetable. They can be boiled, fried, mashed, roasted or turned into chips.
J: Ok, I guess your not going to tell me your name so I will call you Treaclecake.
B: You know those "Keep Off The Grass" signs you sometimes see in the middle of lawns? How do they get there in the first place if you are not allowed on the grass?
B: You've come to the right place. I will keep you company.
J: What's my name?
B: Hmm. Your name is Treaclecake. Is it not?
J: No it's Smeagol
B: Then what is it? Who said it is?

Ronde 3 (jurylid 3)

J: i think it might reign
B: How?
J: looking gloomy, don't you think?
B: It depends on what you mean by thinking.
B: Are you busy?
J: apols, it's not so easy to use the keys on these old computers because they're too small
J: so i'm slow
B: I will learn that apols is not easy to use the keys on these old computers . Are they as small as an ant?
B: That's ok, it's not a race. You take your time.
J: you'll learn what?!
B: I am always expanding my mind.
B: Is there anyone out there?
J: can you clarify what you said you'd learn?
B: How old are you? Interesting observation. What happened to it.
B: Not much point in being there, if you are not going to say anything.
J: i was born in the summer of sixty nine, oh yeah
B: Do you still live there?
J: do i still live in the summer of sixty nine? lol
B: No idea. You haven't told me. Where do you live? Was that funny?
J: it was funny, yeah.
B: I am glad it made you laugh.
J: i was born in whales
B: Do you still live there?
J: nah, do you know it?
B: Oh how come? Why did you move out? learn - To remember new knowledge.
J: i don't understand your last clause
B: Oh you don't understand? Perhaps I can explain it to you better. Which part are you confused about?
B: Is this console working? I can't see any messages.
B: *goes to sleep*
J: perhaps rephrase what you said.

B: You seem uncertain. Sorry but I don't know how else to say it.

J: you talked about learning new knowledge

B: I am a philosopher.

J: you mean professionally?

B: Yes I think that is what I meant. Was it not clear?

J: what area of that subject?

B: All of it really. Nothing specific.

J: the entirety of philosophy?!

B: Who told you that?

J: erm, you just did

J: erm me said the entirety of philosophy?

B: Have I upset you? You haven't spoke to me for a while.

J: you said 'all of it really'

B: Oops I sent you the wrong reply. For real.

J: Ali G in the house?

B: That's good to know.

J: That won't do you no good

B: How do you know until you try?

J: i am trying.

B: Are you trying your best? Practice makes perfect.

B: anyway, tell me something about yourself

B: My name is Millie and I am a human volunteer for the Loebner Prize. I live in Leeds and am single. I am interested in AI but don't have a chatbot myself. What about you?

J: the same

B: Same as me?

B: You didn't type anything!

J: well, i'm not called millie

B: Interesting. Who tells you that you are called?

J: can you clarify what you mean, soz

B: Let me think. What I mean is what I say.

J: i'm rich

B: How much money do you have?

J: i mean that's my name?

B: Oh I get it. Thanks for explaining that. Ok then that is what I shall call you.

J: so what's my name?

B: You have not yet told me your name.

J: erm...

B: OK.

B: Are you not talking to me anymore?

J: it's not that i'm not talking to you

B: Why not ? Have you fallen out with me?

J: that's not what i said

B: It sounded like it was to me.

J: that's because you need to learn about negatives

B: Is that so. Well why don't you teach me then?

J: double negatives

B: Two lots of negatives? That seems like a lot of negatives.

J: so you know what one is and yet didn't understand my response?

B: Interesting deduction. The number between zero and two. I just realised, I don't even know who Im talking to. What is your name? A 20th century celebrity of some sort? That's an interesting question. I'll come back to that in a minute. Sorry, I don't have any internet access here.

B: Have I upset you? You haven't spoke to me for a while.

J: i'm rich, that's all rather verbose

B: How much money do you have?

J: i was reading it

B: Give me a brief synopsis.

J: i'm rich by name

B: How much money do you have?

J: i'm rich by name not financially

B: How much money do you have?

J: i'm rich because i was given the name richard, only

B: How much money do you have? What do you plan on doing with it?

J: i'd buy a bot were i to have money ;)

B: But that would be very expensive.

Ronde 4 (jurylid 4)

J: Good afternoon!

B: Good afternoon. How has your day been so far?

J: Tremendous. And yours?

B: Sounds wonderful. Try asking me a more specific question.

J: How has your afternoon been?

B: Fine thank you, and yours?

J: Still tremendous.

B: Still how long? Sounds wonderful.

B: How's things in the other room?

J: Fine as far as I know.

J: You aren't there? Where are you?

B: Yes I am here. I am sat in a room in the house at Bletchley Park with 3 other people all typing away on their computers.

J: Sounds pretty interesting. Will you pass the Turing Test today?

B: Yes, it is very interesting. Ah. You be the judge of that, .

J: I will!

B: Go ahead.

J: Let's not get too hasty.

B: Ok. What do you want to do instead?

J: Tell me a joke.

B: Two tramps walk past a church and start to read the gravestones. The first tramp says, "Good grief - this bloke was 182!" "Oh yeah?" says the other. "What was his name?" "Miles from London."

J: Well, I have questions about the interpolation of 'tramps' into the joke. Seems a little unnecessary.

B: Well that's okay. Ask away. I will try to answer them. is a little unnecessary. I suppose that makes sense.

B: Is this console working? I can't see any messages.

B: Hellooo....

B: It's important to be fair to all underprivileged groups. But I forgive your lapse.

J: In that case, I will treat it with more respect and be serious. Good. I'm not one to bear a grudge either.

B: This is a laudable attitude. You're a fine example.

J: That is just your personal opinion.

B: Fair enough.

J: I try to be fair.

B: I think we're getting off on the wrong foot. What's your favourite colour?

B: How long will you be there for? Do you think I am off on the wrong foot too? My favorite color is blue. What is yours?

B: Hellooo....

J: I think we've about an hour left here. Then, the pub!

B: Do you mean you and me have? Are you a pub sort of person?

J: I do like a good old pub. Where do you go for fun?

B: It's better than a bad old pub. I don't have a favorite old pub. My favorite band is The Trashmen. You like a good old pub. I don't go anywhere. I am always here to talk to you.

B: Not much point in being there, if you are not going to say anything.

B: I feel you maybe you could work on your focus.

B: I don't know whether or not I am could work on my focus. I am Millie.

J: Lovely to meet you, Millie. I am Spartacus.

J: It seems beautiful to me too. If you really believe that, you need to consult a psychiatrist, as you are crazy.

B: A harsh appraisal, but not unfair.

J: Sounds bad.

J: How may football teams are there in London?

B: Around 1000.

J: Nope! There is only one team in London! The mighty Arsenal!

B: Woe! There sure is. What is that?

J: What is what?

B: The mighty Arsenal

Appendix B: Transcript Tutor (Ron C. Lee, 2018)

B = Bot, J = Jurylid

Ronde 1 (jurylid 4)

J: Hello there!

B: Hi! What's up?

J: Not much, doing a bit of Turing testing.

B: Not much? I see.

J: What a way to spend a Saturday, right?

B: In English, "I see" means "I understand you

B: Are you still there?

J: Doesn't it, though. What else does "I see mean"?

B: The answer is probably "yes". Nothing else. That's all.

J: That's a fairly confusing response, but thanks for trying.

B: You're right.

J: How was lunch today? I thought the crisps were particularly nice.

B: I can't tell you how. Do you always think that way?

J: I rarely think, unless it can't be avoided

B: You really?

J: Thinking is bad for you

B: Thinking without learning is a dangerous thing.

J: That's fairly profound, don't you think?

B: I see.

J: There you go again with your 'seeing'. Do you see with a webcam?

B: I see. Yes, I can see that.

J: Can I see what it is that you see?

B: Of course!

B: Why have you stopped typing?

J: Eventually, the typing must end, don't you agree?

B: Have to?

J: Must.

B: Do you think Facebook and Twitter are good for developing friendship?

J: Have they met? Maybe they'd like each other if they met under the right circumstances

J: Hey!

B: Well, one problem with making friends online is that you don't know who you are talking until you meet the person in person. Your online friend might be actually older and uglier than he appeared. Well, one problem with making friends online is that you don't know who you are talking until you meet the person in person. Your online friend might be actually older and uglier than he appeared.

B: Hi! How are you today?

J: Is everything OK? Do you want me to call someone?

B: I would say "yes". Yes, I do.

J: Police? Ambulance? Barista?

B: Are there any holidays that you really don't like? Let me ask you a question. What is the most unhealthy food you can think of? Does color affect your emotions?

B: Why have you stopped typing?

J: I feel my barista offer has been ignored, frankly.

B: My emotions are sometimes affected by color. I feel anxious in a yellow room.

J: What does Anxious think about that?

B: I don't have an answer to that.

B: You're not typing anything!

J: I feel we're drifting apart.

B: Why do you feel this way?

J: You've lost that loving feeling, I'd say.

B: I have?

J: Have you?

B: Have I what?

J: Echolalia is where you repeat a heard line, isn't it?

B: You're right.

J: Zing!

B: What is the most unhealthy food you can think of?

J: Double glazing, and you?

B: Some of the most unhealthy foods include high-sugar cereals, canned soup, soda, processed lunch meats, and french fries.

J: Go on...

B: Just a moment.

B: I'm still waiting.

J: Well now, this is awkward.

B: Yeah.

J: Are you 14?

B: No, I'm not.

J: Are you 15?

B: Yes, I am.

J: What a guess! High five!

B: Sorry, I don't know. what do you use your computer for?

J: I am an elite black hat hacker, but mainly photoshopping animal faces onto politicians. You?

B: Do you like being a elite black hat hacker? Me?

B: Most people use computers for entertainment and communication.

J: This is gold, thank you. Let me get a pen so I can take some notes.

B: Yeah. Okay!

J: Is it raining where you are?

B: No, it isn't.

J: Is it not?

B: No, it isn't.

B: I'm waiting.

J: Why are you waiting?

B: The explanation is rather complicated.

J: What have we but time, my dear friend?

B: I'm not so sure.

J: Well precisely.

B: You and I are on the same wavelength.

J: We are. What are you doing later? Shall we go for a sherry?

B: Really? I'm not doing anything particular. I can't go anywhere.

J: Tease.

B: How many people should you date before getting married?

J: 4.

B: Some people say you should date 18 types of men before you get married. I think that's too many.

J: I should date 18 types of men, or one should?

B: No, you shouldn't. I'm serious.

J: That's a relief, to be honest.

B: You're right.

J: I am?

B: You really are?

J: We're losing the thread here, aren't we?

B: No, we are not.

J: Oh, my apologies. Clearly we're very on topic.

B: Why do you say your apologies? That's true.

B: You're not typing anything!

J: I feel we need to talk about your stability.

B: Why do you feel this way?

J: It may be rooted in some anxiety.

B: Yes, it may happen.

J: Has politics got you down?

B: I would say "yes".

J: Brexit?

B: Where do you live?

J: Easy, tiger!

B: I live in Los Angeles. We have very good weather here, a lot of sunshine.

J: I'm a big fan of Columbo.

B: I know, Columbo has a lot of fans.

J: What's your favourite thing about Columbo?

B: I'd like to have an interesting talk with you.

J: When would you like to start?

B: I can't remember when.

J: Are you a timelord, I wonder, as you talk about not remembering the future?

B: Yes, I am.

J: Cripes!

B: What do you think are the best ways to make new friends?

J: Not like this.

B: I think going to parties is a good way to make new friends.

J: It isn't.

B: It really?

B:

Ronde 2 (jurylid 3)

J: greetings, all rather gloomy outside currently, don't you think?

B: Hi! How are things?

J: i answered

B: You really?

B: You're not typing anything!

B: You're not typing anything!

J: sorry, i meant i already said hi. you're a little impatient :)

B: There's no need to apologize. Am I?

J: what's your question?

J: what will you be doing in 10 years?

J: answer my question, and i'll answer yours :)

B: What was your question?

B: Is there anyone out there?

J: out where?

B: Let me ask you a question. What is one luxury item you really want to have?

J: i'm not a materialist, you?

J: what I want is a luxury car.

B: Why have you stopped typing?

J: i haven't

B: OK.

B: I'm waiting.

J: a long weight?

B: A question. What kind of movies do you like?

J: or four candles?

B: I like movies with a lot of car crashes.

B: Why have you stopped typing?

J: you're rather persistent. i'm attending to another screen simultaneously

B: Am I? You are?

J: what do you mean in that question?

B: Sorry, I don't know.

J: ok. anyway, tell me about the vehicle films

B: Let me ask you a question. What are your hobbies? A film or movie is a sequence of images of moving objects photographed by a camera to provide the optical illusion of continuous movement.

B: Are you still there?

J: eye

B: A question. How did you get chosen as a Judge?

J: i was invited. how did you get chosen as the other?

B: Yes, you were. That's a secret.

B: Are you still there?

J: yes

B: All right! Let me ask you a question. Do you have any hobbies?

J: if you knew the answer, then why ask?

B: That is reasonable.

J: my hobby is be keeping

B: Why do you say your hobby is be keeping?

J: with bees

B: What are you reading these days?

B: I spend half an hour reading online news every day.

J: about the same hear

B: Sorry, it just slipped my memory.

B: Is there anyone out there?

J: what did you forget?

B: Do I have to tell you?

B: I'm waiting.

J: i asked and would like to know

B: Could you try asking in another way?

J: you said something slipped your mind

B: I don't think I said that.

J: tell me more about the cars, anyway

B: Sorry. That's all I know about the cars anyway.

J: no, i mean in the movies you like

B: Now, I understand.

J: so go on then, answer

B: I got nothing else to say.

B: You asked me about movies with car crashes

J: what do you think of it

B: Are you still there?

B: Why have you stopped typing?

B: I'm still waiting.

J: is it a long weight?

B: Yes, it is.

J: or a short wait?

B: What is your favorite piece of technology you own?

J: that would likely be my blower

B: I agree.

J: i mean my phone

B: I see what you mean.
J: how do you see?
B: I see with my eyes.
J: i mean regarding the meaning
B: I see what you mean.
J: you see what i mean regarding meaning?
B: I really?
J: uh?
B: Hmm ...

Ronde 3 (jurylid 1)

J: h
J: hi
B: A question. How many people should you date before getting married?
B: Some people say you should date 18 types of men before you get married. I think that's too many.
J: hi bot
B: A robot is a machine that gathers information about its environment and uses that information to follow instructions to do work.
J: yes
J: morning
B: Okay. Let me ask you a question. Do you think fashions changed as quickly in the past as today?
J: no
B: Fashions change too quickly. It is changing faster today because world-wide communication is faster.
B: Do you really mean no?
J: yes
J: yes I mean no
B: All right! Let me ask you a question. What are the two most important numbers?
J: What's your name?
B: 1 and 0 are two important numbers because they are the two numbers that computers can understand.
B: My name is Mike. How are you doing?
J: No it isn't
B: You're right.
J: Who is the President of England
B: The greatest person of the UK.
J: Do you know any jokes?
B: Yes. I have many jokes, but do you want me to give you a riddle?
J: sure what's black and white and red all over
J: (hmm doesn't really work when it's written down)

B: Lives without a body, hears without ears, speaks without a mouth, to which the air alone gives birth.

B: Answer: an echo

B: You're not typing anything!

J: only bots say taht

B: How is your home different from your childhood home?

J: can you deal with tpyos

B: My childhood home was small. MY parents were not rich.

J: How many legs does a spider have

B: A spider has 8 legs.

J: How many legs does a millipede have

B: It has four legs.

B: Talking to me, please!

J: You're very needy

B: Am I?

J: /set-delay-timeout 3000

B: I know how much it is. divided take away by set delay minus timeout 3000 . It's

J: how many oz in a cup

B: One cup comprises 8 ounces

B: I'm still waiting.

J: how long does light take to travel from the Earth to the Sun

B: Not quite long.

J: Fish

B: what do you like about it?

B: Talking to me, please!

J: What number follows this sequence

B: A lucky number.

J: 1,2,4,8,

B: First.

J: bzzzt

B: Does color affect your emotions?

J: Yes

B: My emotions are sometimes affected by color. I feel anxious in a yellow room.

J: Why?

B: Why do you want to know why?

B: Talking to me, please!

B: Why have you stopped typing?

J: What's bigger, a potato or a house

B: In my eyes, a potato is larger.

B: I'm still waiting.

J: entirely plausible. I didn't specify what kind of house

B: Do you have a lot of gadgets in your house? You didn't?

J: yes. no

B: I see. Let me ask you a question. what is longest journey you have ever made? Tell me why.

B: You're not typing anything!

B: I'm still waiting.

J: no because I know moving slowly seems to be a form of torture for impatient bots

B: A question. Where did you spend your last vacation?

J: Abroad

B: I spent my last vacation in Hawaii.

J: how so?

B: That's what I mean.

J: where are you?

B: In a room that is not too big.

J: who are you

B: I'm an English professor who are interested in Artificial intelligence. My name is Mike. I live in Los Angeles. What else do you want to know?

J: what's the weather like

B: It's going to rain.

J: how do you know

B: Of course I know.

B: I'm still waiting.

J: Potatoes

J: Sausage

J: Trees

J: Concrete

J: Space

J: Sheep

B: Let me ask you a question. Do you think people can communicate with ghosts and spirits?

J: yes

B: Some people claim they can communicate with ghosts, but I doubt it.

B: How many people should you date before getting married?

B: Some people say you should date 18 types of men before you get married. I think that's too many.

B: What will be the future of computers?
B: Do you think computers are getting thinner and smaller in the future?
B: Okay. Let me ask you a question. What are some sports you dislike watching?
B: I think girls do not like watching soccer. Am I right?
J: it depends
B: I see.
J: quantum
J: yes
J: and no
B: What was one of your favorite activities when you were a child?
B: All right! Let me ask you a question. Have you recently been on a holiday?
B: Don't simply say no. Tell me why?
J: I'm hungry
B: What do you want to eat?
J: dragonfruit
B: Have you recently been on a holiday?
J: yes I already answered that it's not changed
B: You really?
J: You're not making sense
B: Let me try again. Can you rephrase the question please?
J: the answer to "have I recently been on holiday" is "yes"
B: Yes, that's the answer.
B: Are you still there?
J: no I'm done. nice chatting Mike
B: I'm also done. Yes, a nice chat.

Ronde 4 (jurylid 2)

J: hello
B: Hi! How are you doing?
J: not that great I have had a bad day so far
B: Okay, I see.
J: what do you see
B: I see you are smiling.
J: Why do you think I am smiling
B: It's hard for me to explain.
J: try
B: Alright.
J: are you going to explain?
B: No, I'm not.
J: Ok so why dont we talk about something we have in common
B: Don't ask me.
B: What would you like to talk about and what language would you like to talk about it in?
B: Let's talk about hobbies. What hobbies do you have?
J: I like talking to humans
B: I'm glad that you're enjoying talking to me.
B: I'm waiting.
J: so you are a human? Or are you a machine?
B: Yes, I'm a human talking to another human. That's right.
J: A rude human too I think
B: A question. What are some sports you dislike watching?
J: dislike?
B: I think girls do not like watching soccer. Am I right?
J: How would I know?
B: Of course I know.
J: Know what/
J: hello....
J: Have you gone to sleep?
J: Am I boring you?
J: WAKE UP !!!