
‘(PIET)PIRAAT’, ‘HIJ’ OF ‘DIE’: REFERENTIES IN KINDERTAAL

Nienke Wessel
Radboud Universiteit
Linglab 1 - LET-TWB131A
s4598350
n.wessel@student.ru.nl

14 juni 2019

ABSTRACT

Men kan op verschillende manieren naar personen of zaken verwijzen: met een Noun Phrase (NP), persoonlijk voornaamwoord of aanwijzend voornaamwoord. De keuze hangt voor een groot deel af van het perspectief van de toehoorder; is deze al bekend met de referent? Kan er een ambigue situatie ontstaan bij een bepaalde keuze? Eerder onderzoek heeft aangetoond dat kinderen vaak nog fouten maken hierin. In dit onderzoek kijken we naar de manier waarop kinderen deze verwijzingen gebruiken bij het introduceren, herhalen of herintroduceren van referenten. Daaruit blijkt dat kinderen in bepaalde gevallen significant minder gebruik maken van NPs waar volwassenen dat wel doen. Ze houden dus in mindere mate rekening met de verwarring die kan ontstaan bij het niet gebruiken van een NP in een ambigue situatie.

1 Inleiding

Wanneer kinderen een taal moeten leren komt daar veel bij kijken. Niet alleen moeten ze de juiste woorden leren en de grammatica van de taal onder de knie krijgen, ook moeten ze leren hoe ze de taal moeten gebruiken op de juiste manier. Daarbij moeten ze leren om het perspectief van de toehoorder mee te nemen. Het moet immers begrijpelijk zijn voor de toehoorder.

Onderdeel daarvan is het gebruik van verwijzingen naar personen. De volledige naam van een persoon telkens opnieuw gebruiken om naar hem of haar te verwijzen doet immers niemand. In plaats daarvan gebruikt men woorden als ‘hij’, een aanwijzend voornaamwoord ‘die’ of ‘de vrouw’. Kinderen moeten deze gebruiken omtrent verwijzingen leren terwijl ze hun moedertaal leren.

1.1 Verwijzingen

Verwijzingen kunnen gebruikt worden op verschillende manieren. Men kan een volledig zelfstandig naamwoord of *noun phrase* (NP) gebruiken, zoals ‘de piraat’ (definitief) of ‘een piraat’ (indefinitief). Verder kan een pronomen gebruikt worden, zoals ‘hij’, of kan er een lege verwijzing zijn (zoals in de zin ‘hij loopt en \emptyset schopt de bal’).

Het gebruik van deze vormen is in volwassen Nederlands als volgt samen te vatten:

- Introductie nieuwe referent: over het algemeen een indefiniëte NP ‘een piraat’, soms een definitieve NP ‘de piraat’ als referent als bekend verondersteld kan worden.
- Aangehouden referent: een definitieve NP ‘de piraat’, een pronomen ‘hij’ of een aanwijzend voornaamwoord ‘die’.
- Herintroductie referent: een definitieve NP ‘de piraat’.

1.2 Perspectief

Verder heeft onderzoek aangetoond dat volwassenen hun taalgebruik aanpassen op de situatie, ook in andere gevallen dan die hierboven beschreven zijn. Onderzoek van Arnold and Griffin (2007) laat zien dat wanneer er meerdere referenten in het verhaal zijn, iemand vaker de naam van de referent gebruikt dan 'he' of 'she' in het Engels. Dit is nog sterker als beide referenten hetzelfde geslacht hebben.

Om succesvolle communicatie te hebben, moet een spreker het perspectief van de luisteraar immers in gedachten houden. Afhankelijk daarvan kiezen ze voor een bepaalde soort referent. Aan de andere kant gebruiken luisteraars ook weer het perspectief van de spreker om de beste interpretatie te kiezen voor de verwijzing. Kinderen blijken in sommige gevallen dat nog op een andere manier te doen dan volwassenen. Dat blijkt bijvoorbeeld uit onderzoek van De Hoop and Kramer (2006) naar de interpretatie van indefiniete subjecten en objecten. In dat onderzoek wordt de theorie naar voren gebracht dat dit komt doordat kinderen nog niet in staat zijn om bi-directioneel rekening te houden met het perspectief van de toehoorder.

De vraag is nu wanneer en hoe kinderen de gebruiken omtrent verwijzingen leren gebruiken. Uit eerder onderzoek van bijvoorbeeld Chien and Wexler (1990) en Hendriks and Spenader (2006) blijkt dat kinderen het pragmatische deel van verwijswoorden pas laat leren en in de door ons onderzochte groep (kinderen tussen van 4 tot 6 jaar) zeker nog niet helemaal beheersen.

Wel blijken kinderen vanaf nog veel jongere leeftijd al op andere gebieden rekening te houden met de kennis van de gesprekspartner. Matthews et al. (2006) deden onderzoek naar hoe kinderen aan een gesprekspartner uitleggen wat zij in een filmpje zien. Als de partner aangeeft niet te weten wie er in het filmpje iets doet, benoemden de kinderen (vanaf 2 jaar oud al) vaker expliciet om wie het ging dan als de gesprekspartner dat niet aangaf. Verder vonden ze dat kinderen vanaf 3 jaar oud ook rekening hielden met of de gesprekspartner meegekeken had naar het filmpje of niet bij het vertellen over het filmpje. Deze experimenten wijzen er dus op dat kinderen wel in staat zijn om het perspectief van een gesprekspartner mee te nemen in hun taalgebruik, op jonge leeftijd al.

Echter is het gebruik van verwijzingen zeker nog niet op het niveau van volwassenen. Peterson and Dodsworth (1991) lieten in hun longitudinale studie zien dat hoewel het gebruik van NPs beter werd over de duur van de studie (de kinderen waren tussen de 2;0 en 3;6 jaar oud), ze aan het eind van de studie nog steeds moeite hadden met het goed introduceren van nieuwe NPs. De Cat (2011) heeft in onderzoek naar de fouten die kinderen maken bij het leren van het Frans via een story telling task twee soorten fouten geïdentificeerd: soms gebruiken kinderen een indefiniete NP voor iets wat al eerder geïntroduceerd is of soms een definitieve NP bij iets wat nog niet geïntroduceerd is. De Cat (2011) noemt als verklaring voor het eerste dat kinderen nog verminderde cognitieve mogelijkheden hebben en dus moeite hebben om een verhaal en de samenhang daarvan in hun hoofd te houden. Als verklaring voor het tweede wordt gegeven dat kinderen nog een egocentrischer beeld van het gesprek hebben: ze denken dat het perspectief van de luisteraar meer overeen komt met hun eigen perspectief dan daadwerkelijk het geval is.

1.3 Fases in de ontwikkeling

Karmiloff-Smith (1985) heeft onderzoek gedaan naar de fases die kinderen doorlopen bij het leren gebruiken van verwijzingen in het vertellen van een verhaal. Daaruit bleek dat er verschillende fases te onderscheiden zijn. In de eerste fase gebruiken kinderen eigenlijk alleen de directe input en speelt de context weinig mee. Dat is terug te zien in het voornamelijk gebruiken van hij en zij, zonder deze te introduceren, of deze te introduceren met een definitief naamwoord. In de tweede fase speelt de context wel mee. Daar zie je indefiniete introductie van een persoon. Ook zie je dat de context mee gaat spelen in de rolverdeling van personen over grammaticale rollen: er wordt geprobeerd om de hoofdpersoon van het verhaal subject te maken. In de derde fase wordt een balans gevonden tussen het gebruik van de input en de context. Daar zie je dat de hoofdpersoon niet meer persé subject hoeft te zijn.

1.4 Het huidige onderzoek

In dit onderzoek kijken we specifiek naar het gebruik van verwijzingen naar personen door kinderen en volwassenen. Daarbij zijn we vooral geïnteresseerd in wanneer er NPs gebruikt worden en wanneer niet. De onderzoeksvraag die daarbij centraal staat is: "In hoeverre houden kinderen van 4 tot 6 jaar rekening met het perspectief van de toehoorder bij hun keuze voor een persoonsverwijzing?"

Deze vraag beantwoorden we door een story telling task uit te voeren met kinderen en volwassenen, en deze verhalen met elkaar te vergelijken. Daarbij wordt specifiek gekeken naar of kinderen bij het introduceren, herhalen of herintroduceren van een referent op hetzelfde moment een NP gebruiken als volwassenen.

De verwachting hier is dat kinderen bij het introduceren minder vaak een NP gebruiken dan volwassenen. Deze verwachting komt voort uit een de theorie naar voren gebracht door De Hoop and Kramer (2006). We gaan er vanuit dat een volwassen spreker bij de keuze voor een verwijzing het bidirectionele proces aangaat. Als we er vanuit gaan dat een voornaamwoord grammaticaal de voorkeur heeft boven een NP, dan zal de voorlopige keuze een voornaamwoord zijn. Dan zal de spreker bedenken of de gesprekspartner daar ook de juiste interpretatie aan zal geven. Als de bedoelde referent niet op een prominente plaats stond, dan zal een voornaamwoord vaak niet de juiste interpretatie opleveren bij de hoorder, en zal de spreker dus kiezen voor een andere verwijzing. Echter, wij verwachten dat kinderen in mindere mate in staat zijn om die tweede stap te zetten. Ze zullen dus vaker kiezen voor een andere verwijzing dan een NP.

In essentie verwachten we dus dat volwassenen meer rekening zullen houden met mogelijke verwarring die kan ontstaan als meerdere referenten hetzelfde geslacht hebben.

2 Methode

In dit onderzoek hebben we plaatjesverhalen laten zien aan kinderen en volwassenen. Hen was de opdracht gegeven om het verhaal dat op de plaatjes staat te vertellen.

2.1 Participanten

De participanten in dit onderzoek bestonden uit 31 normaal ontwikkelende kinderen (4;3-6;5, gemiddeld 5;6) en 20 jongvolwassenen (18-35, gemiddeld 26;2). De groepen waren gebalanceerd voor geslacht.

2.2 Procedure

De kinderen zijn getest in de winter van 2007-2008 in een stille ruimte op hun basisschool. De volwassenen zijn getest in de lente van 2010 in hun huis of op hun universiteit.

Tijdens het experiment werd aan de kinderen eerst een introductiepagina laten zien waarop alle figuren uit de verhalen te zien waren. Het kind werd gevraagd om de figuren te benoemen en werd daarbij geholpen door de experimentafnemer als dat nodig was. Toen werd aan het kind verteld door een van de onderzoekers dat de andere onderzoeker graag wilde horen wat er in het verhaal gebeurde, maar het verhaal niet kon zien. Het kind moest dus zo duidelijk mogelijk vertellen wat er op de afbeeldingen gebeurt. Er waren twee oefensessies met korte verhalen van elk twee plaatjes. Daarna werd overgegaan op de vier echte verhalen van elk zes plaatjes. De verhalen zijn weergegeven in de appendix, bijlage A. Het kind kon telkens maar één van de plaatjes tegelijkertijd zien. De plaatjes waren geprint op A4-formaat. Het kind vertelde telkens wat er gebeurde op het plaatje. De beschrijvingen zijn zoveel mogelijk monologen van het kind. Echter, als het nodig was, dan gaf de onderzoeker prompts aan het kind. Daarnaast heeft de onderzoeker soms het kind aangemoedigd bij het omslaan van de pagina, met bijvoorbeeld 'ja' of 'goed gedaan'. Nadat elk verhaal was afgelopen, kreeg het kind een sticker en werd er herhaald dat een van de onderzoekers het verhaal niet kon zien.

Het testen van de volwassenen ging grotendeels op dezelfde manier. Daar zijn dezelfde figuren, oefenverhalen en testverhalen gebruikt. Ze kregen geen stickers na elk verhaal, maar wel een chocolaatje aan het eind van hun sessie. Ze werden getest door maar één onderzoeker, die ze vertelde dat hun verhaal werd opgenomen en later werd beluisterd door iemand anders. Die ander zou de plaatjes niet hebben, dus ze moesten zo duidelijk mogelijk zijn in hun verhaal. De volwassenen was ook gevraagd om hun antwoorden kort te houden.

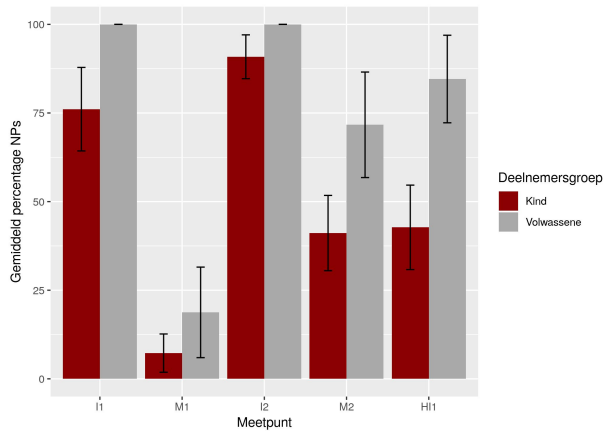
De totale sessie inclusief uitleg, het doornemen van de verschillende figuren, de oefenverhalen en de testverhalen duurde bij de kinderen tussen de 7 en 15 minuten, waar een verhaal ongeveer 3 tot 7 minuten duurde. Bij de volwassenen duurde de totale sessie tussen de 3 en 5 minuten.

2.3 Transcriptie

De opnames werden getranscribeerd door de onderzoeker die aanwezig was bij de testsessie. De momenten waarop een kind een beloning (sticker) kreeg, zijn niet getranscribeerd. De transcripties zijn daarna gecodeerd in CHAT-format. Daarna is door een andere onderzoeker gecontroleerd op consistentie tussen de opnames en de codering.

2.4 Annotatie

Na de transcriptie is er een annotatie uitgevoerd waarbij er werd gelet op twee zaken. Aan de ene kant werd er gekeken naar op welke manier de figuren in het verhaal werden benoemd. Daar werd onderscheid gemaakt tussen de eerste keer het eerste figuur werd geïntroduceerd (I1), de eerste keer dat een verwijzing naar die werd herhaald



Figuur 1: Percentage gebruikte NPs

	Kind		Volw.	
	M	SE	M	SE
I1	76,1	5,8	100	0,0
M1	7,2	2,6	18,8	6,0
I2	90,9	2,9	100	0,0
M2	41,1	4,9	71,7	7,3
HI1	42,7	5,6	84,6	5,6

Tabel 1: Gemiddeld percentage gebruikte NPs (M) en standaarderror (SE)

(M1), de eerste keer dat de tweede figuur werd geïntroduceerd (I2) en herhaald (M2) en de eerste keer dat opnieuw naar de eerste persoon werd verwezen na de verwijzing naar de tweede persoon (HI1). Daarbij werd in elk van deze gevallen genoteerd of gebruik is gemaakt van een zelfstandig naamwoord, persoonlijk voornaamwoord, aanwijzend voornaamwoord, lege verwijzing, bijzin of een overige categorie. Aan de andere kant is er gekeken naar de wie er het topic van de zin was en dan vooral of het topic gewisseld was (topic switch).

Er is geannoteerd door verschillende mensen. Elke stuk is door twee mensen geannoteerd. Bij een verschil in annotatie is door overleg besloten wat de uiteindelijke annotatie is.

Aan de hand van de annotatie is het percentage NPs berekend dat gebruikt is door de participanten. Bij I1, M1, I2, M2 is alleen gekeken naar het percentage NPs. Bij HI1 zijn alleen de gevallen meegenomen waar een topic shift heeft plaatsgevonden. Dit is omdat we in dit onderzoek geïnteresseerd zijn naar of er bij de keuze voor de verwijzing rekening gehouden wordt met dat er een topic shift heeft plaatsgevonden en dus dat een verwijzing door iets anders dan een NP niet goed geïnterpreteerd gaat worden door de luisteraar. In bijlage B is de code te vinden die gebruikt is voor het extraheren van de percentages uit de annotatie.

2.5 Statistiek

Op de data is een independent t-test uitgevoerd met bootstrap (1000 samples) en Bonferroni-correctie. De bootstrap is gebruikt omdat de data geen normale verdeling blijkt te zijn. De Bonferroni-correctie is gebruikt omdat er vijf testen over dezelfde proefpersonen worden gedaan.

3 Resultaten

Op de vijf verschillende meetpunten (I1, M1, I2, M2 en HI2) is het percentage gebruikte NPs bepaald voor beide deelnemersgroepen. Deze zijn terug te vinden in figuur 1 en tabel 1. Bij I1 zien we dat volwassenen altijd een NP gebruiken, waar kinderen dat maar een deel van de tijd doen. Dit verschil van 23,9, *BCa* 95% *CI* [12, 5, 36, 1], is net significant ($p = 0,10$). Bij M1, het opnieuw noemen van de eerste referent, gebruiken de volwassenen gemiddeld vaker een NP dan de kinderen, maar een stuk minder vaak dan bij I1. Het verschil van 11,5, *BCa* 95% *CI* [-0,9, 25, 9], is echter niet significant $p = 0,112 > 0,01$.

Bij I2 zien we opnieuw dat alle volwassenen een NP gebruiken, waar de kinderen dat niet altijd doen. Ook dit verschil, 23,9, *BCa* 95% *CI* [36, 1, 12, 5], is niet significant $p = 0,018 > 0,01$. Dan bij M2 zien we dat de volwassenen een stuk vaker dan de kinderen een NP gebruiken, maar beide niet altijd of nooit. Dit verschil van 30,5, *BCa* 95% *CI* [13, 0, 46, 7], is significant $p = 0,004 < 0,01$. Tot slot bij de herintroductie van een referent zien we dat volwassenen in een grote meerderheid van de tijd voor een NP kiezen waar kinderen dat in net iets minder dan de helft van de tijd doen. Ook dit verschil, 41,8, *BCa* 95% *CI* [25, 2, 57, 0], is significant $p = 0,001 < 0,01$.

Concluderend zien we dus dat in elk van de gevallen de volwassenen vaker een NP gebruiken dan de kinderen. Deze verschillen zijn ook significant in de meetpunten I1, M2 en HI1.

4 Discussie

In dit onderzoek hebben we vergeleken hoe kinderen en volwassenen wel of niet gebruik maken van Noun Phrases (NPs) om te verwijzen naar referenten die geïntroduceerd worden, herhaald worden of geherintroduceerd worden na een wisseling. Het doel was om te bepalen of kinderen dat op dezelfde manier deden als volwassenen en of kinderen ook het perspectief van de toehoorder meenamen in hun keuze voor verwijzingen. Dit hebben we gedaan door kinderen en volwassenen verhalen te laten vertellen bij plaatjes die ze te zien kregen. Daarin waren altijd twee personages van hetzelfde geslacht te zien.

In onze dataset gebruikten volwassenen in elk van de situaties vaker een NP dan de kinderen. Deze verschillen waren significant bij de introductie van de eerste referent, de herhaling van de tweede referent en de herintroductie van de eerste referent. We zullen de categorieën een voor een kort doorlopen.

Introductie referent

Wanneer een referent geïntroduceerd wordt, zou je verwachten dat deze altijd met een NP wordt geïntroduceerd. De andere opties ('zij', 'die') geven immers voor de luisteraar niet voldoende aan over wie het gaat. De luisteraar weet niet waar die woorden naar verwijzen. Dit is dan ook wat je ziet bij de volwassenen in de I1 en I2 conditie: in alle gevallen wordt een NP gebruikt. Bij de kinderen is dit niet het geval. Vooral bij I1 is er een groot, significant verschil. De kinderen hebben daar dus waarschijnlijk niet bedacht dat het voor de hoorder niet duidelijk is om wie het gaat als ze hun verhaal beginnen. Ze moeten bedenken dat de afbeelding die zij voor hun neus hebben niet te zien is voor degene aan wie ze het verhaal vertellen. Dit is in lijn met de resultaten van De Cat (2011). Hoewel daar specifiek gekeken werd naar het verschil tussen definiete en indefiniete NPs, zagen we ook daar dat kinderen zich nog niet voldoende realiseren dat een referent onbekend is voor de luisteraar en daarom een fout maken. Dat lijkt ook hier te gebeuren, hoewel dan door het niet gebruiken van een NP.

Interessant is dan dat er geen significant verschil is in de introductie van de tweede referent, I2. In onze resultaten is te zien dat het percentage NPs groter is bij de kinderen dan in I1. Mogelijk dat kinderen zich hier bewuster zijn van de verwarring, omdat ze nu zelf ook twee figuren uit elkaar moeten houden, en dus niet zelf hoeft te bedenken wat er in het hoofd van de luisteraar gebeurt, maar voor zichzelf de verwarring proberen te verminderen.

Herhaling referent

Bij de herhaling van de eerste referent zien we een laag aantal NPs, zowel bij de volwassenen als de kinderen. Er kan ook over het algemeen geen verwarring ontstaan over waarnaar verwezen wordt.

Bij de herhaling van de tweede referent zien we een significant verschil tussen de de kinderen en de volwassenen. Volwassenen gebruiken veelal een NP waar kinderen dat minder doen. Dat volwassenen hier een NP gebruiken kan te maken hebben met de resultaten die bijvoorbeeld Arnold and Griffin (2007), omdat we nu twee referenten van hetzelfde geslacht hebben. Kinderen doen dit duidelijk in mindere mate dan volwassenen. Hoewel er hier niet direct een ambigue situatie ontstaat, lijkt het erop dat volwassenen voor de duidelijkheid toch eerder de keuze maken voor een NP dan voor een andere verwijzing.

Herintroductie

Bij de herintroductie, HI1, zien we het grootste verschil in percentage NPs van alle categorieën. Volwassenen kiezen duidelijk voor een NP in bijna alle gevallen. Dit is ook te verwachten gezien het niet gebruiken van een NP leidt tot een ambigue situatie: het is dan niet af te leiden dat er een wisseling heeft plaatsgevonden in wie de handeling uitvoert. Kinderen kiezen in veel minder grote mate voor een NP: slechts in 43% van de gevallen werd er een NP gebruikt.

Daaruit blijkt dus dat kinderen in mindere mate rekening houden met het perspectief van de toehoorder en de verwarring die kan ontstaan. Immers, bij een HI1 van twee personages van hetzelfde geslacht kan een luisteraar niet opmaken dat er een verandering heeft plaatsgevonden in wie er wat uitvoert als er geen NP wordt gebruikt. Als dat inderdaad zoals De Hoop and Kramer (2006) hypothetiseren, komt doordat kinderen nog niet bi-directioneel kunnen redeneren over begrip, dan zouden kinderen zich dus nog niet bewust zijn van dat er van een gesprekspartner voornaamwoorden interpreteert naar de meest prominente referent. Het moet dus expliciet benoemd worden als die referent wisselt. Het vergt cognitieve vaardigheden om dit door te rekenen en het zou kunnen dat kinderen die vaardigheden nog niet voldoende hebben.

4.1 Onderzoeksvraag

Op de onderzoeksvraag “In hoeverre houden kinderen van 4 tot 6 jaar rekening met het perspectief van de toehoorder bij hun keuze voor een persoonsverwijzing?” kunnen we antwoorden dat kinderen in mindere mate rekening houden met het perspectief van de toehoorder. Wij verwachtten in onze hypothese dat kinderen minder vaak een NP gebruiken dan volwassenen en dat bleek inderdaad correct. We zien vooral dat in de gevallen waar het perspectief van de toehoorder relevant is, I1, M2 en H11, dat er een significant verschil is. De niet-significante verschillen tussen bij M1 en I2 zijn in situaties waarin juist het perspectief van de toehoorder minder uitmaakt: in M1 is de referent al geïntroduceerd en in I2 moeten kinderen ook voor hunzelf de twee verschillende referenten uit elkaar houden. Het lijkt dus alsof er door kinderen minder rekening gehouden wordt met het perspectief van de toehoorder.

4.2 Toekomstig onderzoek

Het huidige onderzoek heeft vooral gekeken naar het verschil tussen kinderen en volwassenen. Er zijn echter nog wel een aantal zaken die nog onderzocht kunnen worden. Zo zijn in de huidige opzet van het onderzoek zijn de kinderen niet uitgesplitst op leeftijd. De U-vorm, zoals beschreven in Karmiloff-Smith (1985), kan hier van invloed zijn geweest: kinderen gebruiken wel een NP, zonder dat te doen vanwege het perspectief van de toehoorder. Ze doen dat dan omdat het de standaard is in die fase van de ontwikkeling. Het zou daarom mogelijk interessant zijn om dit onderzoek te herhalen met kinderen van meerdere verschillende leeftijden en de resultaten daarvan uit te zetten over de leeftijds categorieën.

Verder zou vervolgonderzoek moeten kijken naar oudere kinderen en vanaf welke leeftijd dit pragmatische deel wel onder de knie is door oudere kinderen te onderzoeken. Vanaf wanneer verdwijnt het verschil?

4.3 Conclusie

Het gebruik van verwijzingen bij kinderen is anders dan bij volwassenen. Het lijkt alsof kinderen in mindere mate rekening houden met het perspectief, mogelijk omdat dat meer doorrekening vraagt van het kind. Verder onderzoek zou moeten kijken naar wanneer het verschil rechtgetrokken is en mogelijke veranderingen die daar de oorzaak van zouden kunnen zijn.

Referenties

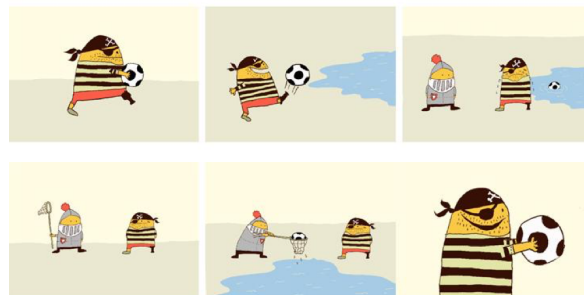
- Arnold, J. E. and Griffin, Z. M. (2007). The effect of additional characters on choice of referring expression: Everyone counts. *Journal of memory and language*, 56(4):521–536.
- Chien, Y. C. and Wexler, K. (1990). Children’s knowledge of locality conditions in binding as evidence for the modularity of syntax and pragmatics. *Language acquisition*, 1(3):225–295.
- De Cat, C. (2011). Information tracking and encoding in early 11: linguistic competence vs. cognitive limitations. *Journal of Child Language*, 38(4):828–860.
- De Hoop, H. and Kramer, I. (2006). Children’s optimal interpretations of indefinite subjects and objects. *Language acquisition*, 13(2):103–123.
- Hendriks, P. and Spenader, J. (2006). When production precedes comprehension: An optimization approach to the acquisition of pronouns. *Language acquisition*, 13(4):319–348.
- Karmiloff-Smith, A. (1985). Language and cognitive processes from a developmental perspective. *Languages and Cognitive Processes*, 1(1):61–85.
- Matthews, D., Lieven, E., Theakston, A., and Tomasello, M. (2006). The effect of perceptual availability and prior discourse on young children’s use of referring expressions. *Applied Psycholinguistics*, 27(3):403–422.
- Peterson, C. and Dodsworth, P. (1991). A longitudinal analysis of young children’s cohesion and noun specification in narratives. *Journal of Child Language*, 18(2):397–415.

A Gebruikte verhalen

In figuur 2, 3, 4 en 5 zijn de gebruikte verhalen weergegeven.



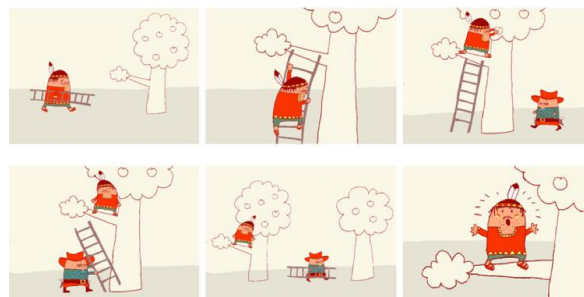
Figuur 2: Verhaal over ballerina en verpleegster.



Figuur 3: Verhaal over piraat en ridder.



Figuur 4: Verhaal over prinses en heks.



Figuur 5: Verhaal over indiaan en cowboy.

B Gebruikte code

```

extraheer <- function()
{
  options(stringsAsFactors = FALSE)

  stap1_lees("consensus2019.txt")
  stap1a_verwijder_volgordefouten()

  stap2_maaksprekerkolom()
  stap3_vulsprekerkolom()
  stap4_maakverhaalkolom()
  stap5_vulverhaalkolom()

  stap6_maakresultaat()
  stap7_voegkolommentoe()

  stap8_vuli1()
  stap9_vulm1()
  stap10_vuli2()
  stap11_vulm2()
  stap12_vulhi1()
  stap12a_verwijder_verdwaalde_m()
  stap13_vultopicshift()

  stap14_bewaarResultaat("meting2019.txt")

  stap15_maak_reftabellen()
  stap16_tel_refs()
  stap17_verwijder_ongewenst()
  stap18_bereken_percentages()
  stap19_bewaarResultaat()
}

stap1_lees <- function(bestandsnaam)
{
  file(description=bestandsnaam,open="r") -> corpusbestand
  readLines(con=corpusbestand,n = -1) -> regels
  as.data.frame(x=regels) ->> bestand
  colnames(x=bestand)[1] <- "Regel"
  close(con=corpusbestand)

  sub(pattern="\s*#.*",replacement="",x=bestand$Regel,perl=TRUE) ->> bestand$Regel
  #Met deze expressie moeten alle comments verwijderd worden. Met \s* wordt alle
  #alle whitespace verzameld, de '#' is waarmee wordt aangegeven dat er een
  #comment is en met .* wordt alles na de comment mee bedoeld.
  #Voorbeeld: 'regel #dit is een comment' -> 'regel'
}

stap1a_verwijder_volgordefouten <- function()
{
  sub(pattern="(. *I2.*);\s*M1.*",replacement="\1",x=bestand$Regel,
    ignore.case=TRUE,perl=TRUE) ->> bestand$Regel
  #Vind waar M1 na I2 voorkomt (fout in annotatie) en laat alleen I2 staan
  #Met (. *I2.*) wordt gezocht naar `I2` gevolgd of voorafgegaan door wat
  #dan ook. Als er twee referenten op een regel staan worden die gescheiden
  #door `;`, dus via `;\s*` bepalen we waar de scheiding zit. Dan volgt
  #M1, mogelijk gevolgd door karakters. We nemen daarna alleen het deel

```



```

#tussen haakjes, dus I2.
#Voorbeeld: '%REF:      I2 CN Wde_dokter; M1 CP zij' -> '%REF:      I2 CN Wde_dokter'
sub(pattern="(. *HI1.*);\\s*M2.*",replacement="\\1",x=bestand$Regel,
     ignore.case=TRUE,perl=TRUE) ->> bestand$Regel
#Vind waar M2 na HI2 voorkomt (fout in annotatie) en laat alleen I2 staan
#Deze werkt op dezelfde manier als hierboven.
}

stap2_maaksprekerkolom <- function()
{
  vector(mode="character",length=nrow(bestand)) -> Spreker
  vector(mode="character",length=nrow(bestand)) -> Sprekertype
  cbind(bestand,Spreker,Sprekertype) ->> bestand
}

stap3_vulsprekerkolom <- function()
{
  naam <- ""
  type <- ""

  for(positie in 1:nrow(bestand))
  {
    if(grepl(x=bestand[positie,"Regel"],pattern="@Participant"))
    {
      sub(x=bestand[positie,"Regel"],pattern=".*(CHI|ADL)\\s+CK([ac])([0-9]+).*",
          replacement="\\2\\3",perl=TRUE) -> naam
      #Vind plek waar CHI of ADL (via (CHI|ADL)) staat gevolgd door witruimte
      #(via \\s+ dus minstens 1) en daarna CKc## of CKa## via CK([ac])([0-9]+)
      #en pakt de c/a en het participantnummer
      #Voorbeeld: '@Participants:      CHI CKc14 Target_Child , INV Ellis1 Investigator'
      # -> 'c14'
      sub(x=bestand[positie,"Regel"],pattern=".*(CHI|ADL)\\s+CK([ac])([0-9]+).*",
          replacement="\\2",perl=TRUE) -> type
      #Zelfde als boven, maar pakt alleen de a of c (en daarmee dus type spreker)
    }

    bestand[positie,"Spreker"] <-<- naam
    bestand[positie,"Sprekertype"] <-<- type
  }
}

stap4_maakverhaalkolom <- function()
{
  vector(mode="character",length=nrow(bestand)) -> Verhaal
  cbind(bestand,Verhaal) ->> bestand
}

stap5_vulverhaalkolom <- function()
{
  naam <- ""

  for(positie in 1:nrow(bestand))
  {
    if(grepl(x=bestand[positie,"Regel"],pattern="@G.*-1$"))
      #Als de regel @G en een streepje - en het nummer 1 bevat (de regels die
      #aangeven waar in het verhaal we zijn). We zijn dan dus bij het eerste
      #plaatje van een nieuw verhaal
      #Voorbeeld: '@G:      ballerina-1'
  }
}

```

```

{
  sub(x=bestand[positie,"Regel"],pattern=".*\\s(.*)-1",replacement="\\1",
    perl=TRUE) -> naam
  #Extraheer de naam van het verhaal; die staat tussen een spatie `\\s` en '-1'
  #Voorbeeld: '@G:          ballerina-1' -> 'ballerina'
}

bestand[positie,"Verhaal"] <-> naam
}
}

stap6_maakresultaat <- function()
{
  bestand[grepl(x=bestand[, "Regel"],pattern="G:.*-1$"),c("Spreker","Sprekertype",
    "Verhaal")] ->> meting
  #Verzamel alle regels die het eerste plaatje van een verhaalregel bevatten
  #Zie ook vorige functie, r113
  rownames(meting) <- 1:nrow(meting)

  meting ->> metingpos
}

stap7_voegkolommentoe <- function()
{
  #Voeg kolommen toe waar alleen maar "GEEN" in staat
  sub(pattern=".*",replacement="GEEN",x=meting$Spreker,perl=TRUE) -> I1
  cbind(meting,I1) ->> meting
  sub(pattern=".*",replacement="GEEN",x=meting$Spreker,perl=TRUE) -> M1
  cbind(meting,M1) ->> meting
  sub(pattern=".*",replacement="GEEN",x=meting$Spreker,perl=TRUE) -> I2
  cbind(meting,I2) ->> meting
  sub(pattern=".*",replacement="GEEN",x=meting$Spreker,perl=TRUE) -> M2
  cbind(meting,M2) ->> meting
  sub(pattern=".*",replacement="GEEN",x=meting$Spreker,perl=TRUE) -> HI1
  cbind(meting,HI1) ->> meting
  sub(pattern=".*",replacement="GEEN",x=meting$Spreker,perl=TRUE) -> TopicShift
  cbind(meting,TopicShift) ->> meting

  #Voeg kolommen toe waar alleen 0 in staat
  sub(pattern=".*",replacement="0",x=metingpos$Spreker,perl=TRUE) -> I1
  cbind(metingpos,I1) ->> metingpos
  sub(pattern=".*",replacement="0",x=metingpos$Spreker,perl=TRUE) -> M1
  cbind(metingpos,M1) ->> metingpos
  sub(pattern=".*",replacement="0",x=metingpos$Spreker,perl=TRUE) -> I2
  cbind(metingpos,I2) ->> metingpos
  sub(pattern=".*",replacement="0",x=metingpos$Spreker,perl=TRUE) -> M2
  cbind(metingpos,M2) ->> metingpos
  sub(pattern=".*",replacement="0",x=metingpos$Spreker,perl=TRUE) -> HI1
  cbind(metingpos,HI1) ->> metingpos
  sub(pattern=".*",replacement="0",x=metingpos$Spreker,perl=TRUE) -> TopicShift
  cbind(metingpos,TopicShift) ->> metingpos
}

stap8_vuli1 <- function()
{
  for(positie in 1:nrow(bestand))
  {
    if(grepl(x=bestand[positie,"Regel"],pattern="REF:.*\\sI1"))

```

```

{
  sub(pattern=".*[^H]I1[^;]*\\sC([A-Z*]+).*",replacement="\\1",
    x=bestand[positie,"Regel"],ignore.case=TRUE,perl=TRUE) -> waarde
  #Vind regels waarin I1 staat maar geen HI1 '[^H]I1' en zoek dan naar de
  # een spatie met daarna hoofdletter C '\\sC`
  #Ga niet verder dan ; (want dan ben je bij de volgende entry van de regel)
  #Pak de letter(s) direct achter de C `([A-Z*]+)`. De rest negeren we
  #Voorbeeld: '%REF:      I1 CN Wzangeres' -> 'N'
  meting[(meting$Spreker==bestand[positie,"Spreker"])&
    (meting$Verhaal==bestand[positie,"Verhaal"]), "I1"] <- waarde
  metingpos[(meting$Spreker==bestand[positie,"Spreker"])&
    (meting$Verhaal==bestand[positie,"Verhaal"]), "I1"] <- positie
}
}
}

stap9_vulm1 <- function()
{
  for(positie in 1:nrow(bestand))
  {
    if(grepl(x=bestand[positie,"Regel"],pattern="REF:.*\\sM1"))
    {
      sub(pattern=".*M1[^;]*\\sC([A-Z*]+).*",replacement="\\1",
        x=bestand[positie,"Regel"],ignore.case=TRUE,perl=TRUE) -> waarde
      #Zie uitleg over I1 op regel 179, uitgezonderd van dat het geen HI1 mag zijn
      meting[(meting$Spreker==bestand[positie,"Spreker"])&
        (meting$Verhaal==bestand[positie,"Verhaal"]), "M1"] <- waarde
      metingpos[(meting$Spreker==bestand[positie,"Spreker"])&
        (meting$Verhaal==bestand[positie,"Verhaal"]), "M1"] <- positie
    }
  }
}

stap10_vuli2 <- function()
{
  for(positie in 1:nrow(bestand))
  {
    if(grepl(x=bestand[positie,"Regel"],pattern="REF:.*\\sI2"))
    {
      sub(pattern=".*I2[^;]*\\sC([A-Z*]+).*",replacement="\\1",
        x=bestand[positie,"Regel"],ignore.case=TRUE,perl=TRUE) -> waarde
      #Zie uitleg M1 op regel 200
      meting[(meting$Spreker==bestand[positie,"Spreker"])&
        (meting$Verhaal==bestand[positie,"Verhaal"]), "I2"] <- waarde
      metingpos[(meting$Spreker==bestand[positie,"Spreker"])&
        (meting$Verhaal==bestand[positie,"Verhaal"]), "I2"] <- positie
    }
  }
}

stap11_vulm2 <- function()
{
  for(positie in 1:nrow(bestand))
  {
    if(grepl(x=bestand[positie,"Regel"],pattern="REF:.*\\sM2"))
    {
      sub(pattern=".*M2[^;]*\\sC([A-Z*]+).*",replacement="\\1",
        x=bestand[positie,"Regel"],ignore.case=TRUE,perl=TRUE) -> waarde
      #Zie uitleg M1 op regel 200

```

```

    meting[(meting$Spreker==bestand[positie,"Spreker"])&
      (meting$Verhaal==bestand[positie,"Verhaal"]), "M2"] <- waarde
    metingpos[(meting$Spreker==bestand[positie,"Spreker"])&
      (meting$Verhaal==bestand[positie,"Verhaal"]), "M2"] <- positie
  }
}

stap12_vulhi1 <- function()
{
  for(positie in 1:nrow(bestand))
  {
    if(grepl(x=bestand[positie,"Regel"],pattern="REF:.*\\sHI1"))
    {
      sub(pattern=".*HI1[~;]*\\sC([A-Z*]+).*",replacement="\\1",
        x=bestand[positie,"Regel"],ignore.case=TRUE,perl=TRUE) -> waarde
      #Zie uitleg M1 op regel 200
      meting[(meting$Spreker==bestand[positie,"Spreker"])&
        (meting$Verhaal==bestand[positie,"Verhaal"]), "HI1"] <- waarde
      metingpos[(meting$Spreker==bestand[positie,"Spreker"])&
        (meting$Verhaal==bestand[positie,"Verhaal"]), "HI1"] <- positie
    }
  }
}

stap12a_verwijder_verdwaalde_m <- function()
{
  meting[(metingpos[, "M1"]>metingpos[, "I2"])&(metingpos[, "I2"]>0), "M1"] <- "GEEN"
  metingpos[(metingpos[, "M1"]>metingpos[, "I2"])&(metingpos[, "I2"]>0), "M1"] <- "0"

  meting[(metingpos[, "M2"]>metingpos[, "HI1"])&(metingpos[, "HI1"]>0), "M2"] <- "GEEN"
  metingpos[(metingpos[, "M2"]>metingpos[, "HI1"])&(metingpos[, "HI1"]>0), "M2"] <- "0"
}

stap13_vultopicshift <- function()
{
  hi1_gezien <- FALSE

  for(positie in 1:nrow(bestand))
  {
    if(grepl(x=bestand[positie,"Regel"],pattern="\\sHI1")) #Als HI1 tegengekomen
      #na een spatie
    {
      hi1_gezien <- TRUE
    }

    if(grepl(x=bestand[positie,"Regel"],pattern="@G.*-1$")) #Als nieuw verhaal
      #(zie ook r113)
    {
      hi1_gezien <- FALSE
    }

    if(grepl(x=bestand[positie,"Regel"],pattern="TOP:.*\\sTS") & hi1_gezien==FALSE)
      #als topicshift `TS` gezien in de TOP regel
      #Voorbeeld: '%TOP: TSI Eheks'
    {
      meting[(meting$Spreker==bestand[positie,"Spreker"])&
        (meting$Verhaal==bestand[positie,"Verhaal"]), "TopicShift"] <- "WEL"
      metingpos[(meting$Spreker==bestand[positie,"Spreker"])&

```

```

        (meting$Verhaal==bestand[positie,"Verhaal"]), "TopicShift"] <-> positie
    }
}

stap14_bewaarresultaat <- function(bestandsnaam)
{
  write.table(x=meting,file=bestandsnaam)
}

stap15_maak_reftabellen <- function()
{
  meting[meting[, "Verhaal"]=="ballerina", c("Spreker",
                                             "I1", "M1", "I2", "M2", "HI1")] -> ref_n

  rownames(ref_n) <-> ref_n[, "Spreker"]
  ref_n[, "I1"] <-> 0
  ref_n[, "M1"] <-> 0
  ref_n[, "I2"] <-> 0
  ref_n[, "M2"] <-> 0
  ref_n[, "HI1"] <-> 0

  ref_npal <-> ref_n
}

stap16_tel_refs <- function()
{
  for(positie in 1:nrow(meting))
  {
    if(meting[positie, "TopicShift"]=="WEL")
    {
      spreker <- meting[positie, "Spreker"]
      for(kolom in c("I1", "M1", "I2", "M2", "HI1"))
      {
        if(meting[positie, kolom] == "N")
        {
          ref_n[spreker, kolom] <-> ref_n[spreker, kolom] + 1
          ref_npal[spreker, kolom] <-> ref_npal[spreker, kolom] + 1
        }

        if(meting[positie, kolom] == "P" || meting[positie, kolom] == "A"
           || meting[positie, kolom] == "L")
        {
          ref_npal[spreker, kolom] <-> ref_npal[spreker, kolom] + 1
        }
      }
    }
  }
}

stap17_verwijder_ongewenst <- function()
{
  vector(length = nrow(ref_npal)) -> gewenst
  for(positie in 1:nrow(ref_npal))
  {
    gewenst[positie] <-> min(ref_npal[positie, c("I1", "M1", "I2", "M2", "HI1")]) > 0
  }
  ref_npal[gewenst,] -> ref_npal
  ref_n[gewenst,] -> ref_n
  rownames(ref_npal) <-> 1:nrow(ref_npal)
}

```

```

rownames(ref_n) <- 1:nrow(ref_npal)
}

stap18_bereken_percentages<- function()
{
  sprekers <- ref_npal
  ref_n$I1 / ref_npal$I1 * 100 ->> sprekers$I1
  ref_n$M1 / ref_npal$M1 * 100 ->> sprekers$M1
  ref_n$I2 / ref_npal$I2 * 100 ->> sprekers$I2
  ref_n$M2 / ref_npal$M2 * 100 ->> sprekers$M2
  ref_n$HI1 / ref_npal$HI1 * 100 ->> sprekers$HI1
}

stap19_bewaarresultaat <- function()
{
  write.table(file="perc_np_wide.txt",x=sprekers,row.names=FALSE)

  library("reshape2")
  gesmolten <- melt(data=sprekers,
                    id.vars=c("Spreker"),
                    measure.vars = c("I1","M1","I2","M2","HI1"),
                    variable.name = "Meetpunt",
                    value.name="Percentage")
  write.table(file="perc_np_long.txt",x=gesmolten,row.names=FALSE)
}

```