Modelling characters' mental depth in stories told by children aged 4-10

Dijk, B.M.A. van; Duijn, M.J. van

Modelling Characters' Mental Depth in Stories Told by Children Aged 4-10

Bram van Dijk (b.m.a.van.dijk@liacs.leidenuniv.nl)
Max van Duijn (m.j.van.duijn@liacs.leidenuniv.nl)
Leiden Institute for Advanced Computer Science (LIACS), Niels Bohrweg 2
Leiden, 2333 CA The Netherlands

Abstract

From age 3-4, children are generally capable of telling stories about a topic free of choice. Over the years their stories become more sophisticated in content and structure, reflecting various aspects of cognitive development. Here we focus on children's ability to construe characters with increasing levels of mental depth, arguably reflecting socio-cognitive capacities including Theory of Mind. Within our sample of 51 stories told by children aged 4-10, characters range from flat "actors" performing simple actions, to "agents" having basic perceptive, emotional, and intentional capacities, to fully-blown "persons" with complex inner lives. We argue for the underexplored potential of computationally extracted *story-internal* factors (e.g. lexical/syntactic complexity) in explaining variance in character depth, as opposed to *story-external* factors (e.g. age, socioeconomic status) on which existing work has focused. We show that especially lexical richness explains variance in character depth, and this effect is larger than and not moderated by age.

Keywords: children's stories; cognitive development; Theory of Mind; fictional minds; NLP

Introduction

From early childhood on children tell stories to themselves and others as part of their daily play activities (Sutton-Smith, 1981; Cremin et al., 2016). Such storytelling has been described as a kind of *cognitive play* that—besides being the source of a lot of fun—forms a natural crossroads of various key areas in child development (Paley, 1990; Bergen, 2002; Smith & Roopnarine, 2019). Telling stories involves language skills at the phonological, lexical, syntactic, and pragmatic levels (Southwood & Russell, 2004). It draws further on cognitive abilities such as memorizing, planning, organizing knowledge of the world (McKeough & Genereux, 2003), and empathizing with others, in particular to work out how characters should behave, speak, feel, and think in ways that are relatable and interesting for an audience (Nicolopoulou, 1993; Zunshine, 2006; Van Duijn, Sluiter, & Verhagen, 2015).

In this paper we are interested in the representation of mental activities of characters and the place this has in child development. Existing theoretical work has linked children's ability to render character minds to the mastery of socio-cognitive skills, in particular *mindreading* or *Theory of Mind* (ToM).[1] Empirical research has shown that the complexity of characters and their mental activities that children can deal with tends to increase with age (e.g. Nicolopoulou & Richner, 2007; Nicolopoulou & Ünlütabak, 2017). For this paper, we recorded and transcribed 51 oral stories elicited from Dutch children of different ages and backgrounds, during storytelling workshops integrated in their daily school and daycare environments. A total of 268 characters were represented in these stories, each of which we assessed in terms of its mental depth. To give two brief opening examples of what we looked at (translated from stories we collected earlier as part of a pilot):

(1) *they sit neatly in a row but the other [puppy] always enters later* (told by a child aged 4y;1m)
(2) *they sat down as always until he was not looking [...] then they went inside the school director's office and secretly took the key* (9y;11m)

Characters presented in the fragments (1) and (2) fall at the lower and higher ends of the scale of mental depth that we will introduce in more detail below respectively. Fragment (1) introduces characters with arguably different perspectives on the staged setting: some are already inside, while another one enters later. However, there is no fleshing out of mental activity by any of these characters at all, and this is representative of the rest of the story, which revolves around movements (coming in, going out) and actions (eating) only. This is very different in (2), where the implied protagonists' *awareness* of what the school director does not *see*, and hence *knows*, is central in the story's plot.

In line with existing work, we observe an overall increase of character mental depth with the age of the children telling the stories in our sample. However, it is our aim to understand in more detail which factors drive children's ability to render more complex characters. To this end, we develop a framework using computational techniques and statistical

[1] For a general overview of literature on mindreading/ToM, i.e. the ability to take others' perspectives and reason about their behavior in terms of emotional and intentional states, see Apperly (2010). For an overview of theoretical work linking ToM with children's stories see Nicolopoulou (2016) and Zunshine (2019).

modelling for mapping out relationships between, on one side, the mental depth of story characters and, on the other side, multiple *story-external* variables (e.g. age, socioeconomics) and automatically parsed *story-internal* variables (e.g. vocabulary, syntax). Our results show that in particular the *lexical richness* a story exhibits can be used as a reliable predictor of character depth: it explains a larger proportion of the variance compared to age, and is not moderated by age. We discuss the role of lexical richness and other variables in understanding children's ability to deal with characters of different levels of mental complexity, both within our current sample and in larger, more diverse samples in the future.

Concepts & Hypotheses

Narrative and Storytelling

Narrative plays a key role in human communication. On a daily basis adults and children alike use stories to share their perceptions and imaginations with others, typically in causally, temporally, and logically structured ways. Classic definitions of narrative often emphasize criteria such as goal-directedness, causality, or the unfolding of series of actions over time (e.g. Duijnmeijer, De Jong, & Scheper, 2012). However, in this research we cast the definitional net a bit wider and argue that children's stories could also be descriptions of situations, events, or characters in which goals, causal relations, or a clear temporal development are not immediately present. What we take as our central criterion here to demarcate stories from other speech phenomena is *mediatedness* or *transcendence*, marked by a departure from the actual speaker and their immediate here-and-now (cf. Nicolopoulou & Richner, 2007; Zeman, 2018). For example, children merely describing their situation during the storytelling workshop in which we collected our data would *not* be telling a story (e.g. "I am sitting on a chair in the group circle…"), whereas children describing a real or imagined situation set elsewhere would be, even if that situation is not worked out any further with additional characters and events ("Yesterday I had a silent disco…").

Children's Stories and Development

In this paper we focus on two of the developmental trajectories that naturally intersect in stories that children tell, social cognition and language, against the background of their more general development, which we approximate via age and education level of the parents/caretakers. Following a large body of research (for an overview see Tompkins, Farrar, & Montgomery, 2019; Milligan, Astington, & Dack, 2007), we expect these trajectories to be interrelated and it is our longer-term aim to contribute to further understanding of this interrelatedness by studying stories that children tell. Here we develop a framework for mapping out features within such stories that we assume to be manifestations of developmental progression on the linguistic and social-cognitive levels. Our hypotheses at this stage concern the co-occurrence of and relationships between these features within the stories; testing whether this is indeed indicative of the development of the children who tell them is outside the scope of this paper.

Social cognition Firstly, we are interested in socio-cognitive sophistication of the stories, which we operationalize as the mental depth that characters exhibit. Using a slightly adapted version of the typology introduced by Nicolopoulou & Richner (2007) we rate each character's mental activity on a nine-level scale. These levels fall under three main categories: *actors* undergoing (level I) or performing (II) simple actions, *agents* having basic perceptive, expressive, emotional, and intentional capacities (III-V), and *persons* capable of coordinating beliefs, desires, expectations, and so on, with different imagined realities (VI-VII) and/or other characters' cognitive states (VIII-IX; see Methods and Table 1 below for more details).

Language Secondly, we are interested in linguistic qualities of the stories, which we operationalize on two levels: vocabulary and syntax. As a measure of vocabulary sophistication (a.k.a. lexical richness) we assessed the vocabulary of each story by computing the probability of the occurrence of each word that a child used approximated by frequencies in a benchmark lexicon. This metric builds on the idea that the difficulty of words from the perspective of a language learner is strongly negatively correlated with how frequently it occurs (Vermeer, 2001). Thus, using less frequent words means using less probable words, and this we take to indicate a more sophisticated vocabulary. The idea is that a richer vocabulary functions as a communicative and mental toolbox that enables a child to render both the physical and social world better. This toolbox can be especially helpful when engaging in demanding tasks such as telling a story, where there is a sustained pressure for finding the right words to get the desired message across to an audience (Curenton & Justice, 2014).

As a measure of syntactic complexity, we calculated the average distance between syntactically dependent words. It is well-established that language structures which employ longer dependency distances between head words and dependent words are more difficult to process (Gibson, 1998, 2000; Gildea & Temperly, 2010). An example of this difference is given by King & Just (1991) in terms of subject-extracted relative clauses (3) and object-extracted relative clauses (4):

(3) The reporter who attacked the senator admitted the error.
(4) The reporter who the senator attacked admitted the error.

In both sentences the verb "attacked" is dependent on the pronoun "who". In (4) these dependents are not adjacent, but have two words in between, which makes that sentence more challenging to process.

Average dependency distance seems to capture language skills more generally. For example, it can be used to distinguish English text written by natives from that written by L2 learners (Oya, 2015) and speech from individuals with mild cognitive impairments from speech produced by typically developed speakers (Roark, Mitchell, & Hollingshead, 2007). Our idea here is that children capable of handling more complex syntactic structures, as indicated by their stories exhibiting higher average dependency distances, have more powerful formats available for representing events in the social and mental worlds, in discourse as well as in their own strands of reasoning (cf. De Villiers & De Villiers, 2014).

Hypotheses Firstly, we hypothesize that stories exhibiting a more sophisticated vocabulary contain characters with higher levels of mental depth. Secondly, we hypothesize that stories with larger syntactic dependency lengths contain characters with higher levels of mental depth.

Background variables Existing work has shown that the mental depth of characters in stories that children tell increases with their age (Nicolopoulou & Richner, 2007), which is why we include it in the model. Parent education functions as a proxy for socioeconomic status in our model; there is evidence that children from parents who have a higher socioeconomic status perform better on ToM tasks (Shatz et al., 2003).

Methods

Data Collection & Transcription

For our data collection we offered storytelling sessions to various institutions in the medium-sized Dutch cities Leiden, Tilburg and Utrecht. Three schools (two in Leiden, one in Utrecht), one daycare (Leiden) and one community center (Tilburg) were willing to cooperate. Around 200 children in total participated in sessions held between September 2019 and June 2020. We were able to include 98 stories told by 54 children (M_{age}(SD) = 6.81(1.66), range = 4.17-10.1; 30 females, 2 unknown) in our database after receiving consent forms from their parents. In order to maximize independence between observations we use only the first story told by each child, and due to missing information on the consent forms an additional 3 stories dropped out, resulting in a subset of 51 stories for this paper. Our experiment and data management protocols were assessed and approved by the Ethical Committee of the Leiden University Faculty of Science (file no. 2020 – 002). Our storytelling sessions were held in group circle settings. After briefly exploring some general features of stories interactively (e.g. "What is a story?", "What do we find in stories?") and narrating a short standard exemplary

Table 1: Annotation scheme for character depth. All examples are literal quotes from our dataset, followed by a somewhat liberal/idiomatic English gloss, followed by the unique ID of the story from which it was taken.

ACTORS	I	Passive actor: does nothing or undergoes something passively *er was een boerderij met schapen* (there was a farm with sheep; 060301) *toen ze vos gingen pakken* (when they went to catch fox; 060301)
	II	Active actor: coupled to action/description without clear intention/goal-directedness *hij is verkleed met echte laarzen* (he was dressed up with real boots; 010701)
AGENTS	IIIa	Implicit intention in action *wij gingen autorijden* (we went to drive the car; 010501)
	IVa	Action in response to/contrast with situation or event *er kwam een baby [...] en toen ging de vader weg* (a baby came and then the father left; 021302) *ze gingen [...] treetje voor treetje maar het ging mis* (they went step by step but it went wrong; 030303)
	Va	Complex/explicit intention/goal-directedness in action *kwam [hij ons] ophalen en toen naar huis brengen* (came he pick us up and then bring home; 010501)
	IIIb	Simple/implicit emotion, perception, attention *ik keek naar die* (I looked at that one; 060701)
	IVb	Emotion in response to situation/event or action in response to emotion/perception *ik ging [...] naar bowling [...] en het was heel leuk* (I went bowling and it was a lot of fun; 020501)
	Vb	Complex/explicit emotion, perception, attention, utterance *in de auto was het heel stom, want we moesten lang wachten* (in the car it was very stupid, since we had to wait long; 010501) *Papa zei er is niets aan de hand* (Daddy said all is fine; 010504)
PERSONS	VI	Explicit intentional state (desire, belief, etc.), implicitly coordinated with actions/events *de moedertrol wou de andere trol redden* (the mother troll wanted to save the other troll; 011301)
	VII	Explicit intentional state, explicitly coordinated (equation, contrast, etc.) with non-intentional aspects of the storyworld *die wou ik doen maar toen ging zij hem helemaal afmaken* (that one I wanted to do but then she went to fully complete it; 010503)
	VIII	Explicit intentional state, coordinated with the intentional state of another character or past/future self *ze wou heel graag buiten spelen, maar het mocht niet van haar moeder* (she very much wanted to play outside, but her mother did not allow it; 010201)
	IX	Multiple-order intentionality *vader en moeder wouden niet dat ze naar een bos ging [want ze] wisten dat in de in het bos een kluizenaar woonde. En dat wist Amelie ook dat ze dat wisten* (father and mother did not want that she would go to the woods since they knew that a hermit was living in the woods. And Amelie knew too that they knew this; 021301)

fantasy story, we invited children to tell a story about a topic free of choice. Voice recordings were made after informing children about this. Afterwards the recordings were pseudonymized and transcribed by the authors and research assistants twice: first orthographically (including "noise" such as false starts, wrong conjugations, broken-off words, etc.), and second normalized, thus without noisy elements, to enhance compatibility with computational language processing tools. All transcripts were double-checked for consistency with the audio files. In addition to the story data, personal data such as age of the children and parental education levels were collected through consent forms. Transcripts, data, and code are available via OSF: https://osf.io/k52e8/?view_only=3f659438af7b43549d9ae42d0ec29310.

Manual Annotation of Character Depth

We have loaded all anonymized transcriptions in the open online content annotation tool CATMA (version 6.1.3; Gius et al., 2020), where we created a tag set for character depth. Tags within this set were based on the typology introduced by Nicolopoulou & Richner (2007). A few adaptations were made, however, in terms of the three main levels (actor, agent, and person) our tag set remained compatible with the original typology. See Table 1 for descriptions and examples of the tags we have used to assign a level of character depth to each character. Our workflow included a first stage in which we, the authors of this paper, discussed the first 10 stories openly, followed by a second stage in which the remaining 41 stories were annotated by each of us independently. In a third stage, all tags that differed were discussed until consensus was reached. Finally, the annotations were considered fixed and downloaded from CATMA in TEI-XML format.

We extracted the *maximum character depth by level* with a Python script. This variable represents the highest level of character depth reached in a story on a scale from 0 to 9, corresponding with the levels in the topology set out in Table 1 when discarding subcategories indicated by letters (e.g. IVa and IVb both count as 4), where 0 indicates the theoretical option of no characters being presented in a story (which did not occur in our current dataset), value 1 corresponds with level I in Table 1, and so on.

Parsing & Linguistic Variable Extraction

Vocabulary Probability Our approach was to take the *textual vocabulary* of a representative reference corpus, which consists of all the lemmas constituting the vocabulary of the corpus (Fengxiang, Yang, & Wang, 2016). We use this benchmark to compute the probability of each story vocabulary, treating it as a subset of the textual vocabulary. Lemma probabilities were approximated by relative frequency counts in the reference corpus.

We obtained lemmas for each story by parsing normalized story transcripts with the memory-based Frog parser (Van den Bosch et al., 2007). We used as reference corpus the "free text" lexicon (FTL) of the BasiScript corpus (Tellings et al. 2018), which consists of essays of primary school children with minimal teacher intervention, thus staying close to the free story paradigm. We removed interpunction and named entities from the FTL, which yielded a total number of token instances N of 3699822, and a vocabulary V of 46570 lemmas.

The estimated probability of some lemma l_i occurring in story S is given by $P(l_i) = \frac{(c_i+1)\frac{N}{N+V}}{N}$, with c_i being the count of token instances of l_i in the FTL, adjusted for words not occurring in the FTL. We used Laplacian smoothing since the FTL includes many typical fantasy constructions such as "trollensnot" (troll snot) with count 1, but not the similar construction "eenhoornsnot" (unicorn snot) which exists in our stories. We calculated the probability L of the vocabulary of S with $L = \frac{1}{S}\sum_{i=1}^{n} P(l_i)$, the fraction being a normalizing factor, and converted them to permille for convenient interpretation. The interpretation of the probability can be phrased as follows: if one draws a lemma from the FTL, how likely is it that it belongs to the story vocabulary? For sophisticated vocabularies this probability will be lower.

Dependency Distance We used the Alpino parser (Van Noord, 2006) to extract all dependencies per utterance. The dependency distance of the ith dependency relation DD_i is typically set to 1 for adjacent words, 2 if one extra word occurs in between the dependents, and so on. Overall dependency distance MD_{sent} for a sentence with n words is given by $MD_{sent} = \frac{1}{n-1}\sum_{i=1}^{n-1} |DD_i|$ (Wang & Liu, 2017). Then, for a story consisting of multiple utterances, $MD_{story} = \frac{1}{u}\sum_{i=1}^{n} MD_i$ where u is the total number of utterances in a story.

Results

Bivariate explorations

Prior to constructing the linear model that we use for assessing our hypotheses, we explore various correlations between a subset of the variables outlined above.

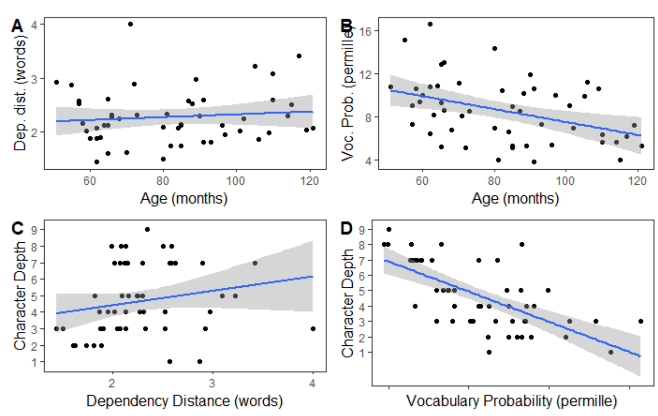

Figure 1: Explorative correlation plots

Firstly, it appears that Dependency Distance correlates weakly with Age (Fig. 1A, Pearson's $r = 0.1$) and Vocabulary Probability correlates moderately with Age (Fig 1B, Pearson's $r = -0.4$). It makes sense that as children grow older, both their vocabularies and syntactic dependencies are becoming increasingly complex. Secondly, Dependency Distance correlates weakly with Character Depth (Fig. 1C, Pearson's $r = 0.2$) and Vocabulary Probability correlates strongly with Character Depth (Fig. 1D, Pearson's $r = -0.67$), indicating that the relations between the parsed linguistic variables and our dependent variable Character Depth are in the expected direction, albeit in quite different gradations. In the next section we scrutinize these bivariate explorations using a multiple linear regression model.

Hypothesis testing

We consider a multiple regression model most appropriate for the analysis; due to the limited number of observations per institution in our dataset, a mixed-effects model did not converge properly. Our model includes Dependency Distance, Vocabulary Probability, Age, Education Parents, and interactions between Dependency Distance and Age and Vocabulary Probability and Age as predictors of Character Depth. The model accounts for about 53% of the variance in Character Depth ($R^2 = .525$, $F[6, 44] = 8.132$, $p < 0.001$). Standardized coefficients sorted on size are given in Table 2.

In line with our first hypothesis, we see that the simple effect of Vocabulary Probability has the largest negative and significant slope. This indicates that as the vocabulary of a story becomes less probable, i.e. the lexical richness of that story goes up by our measure, characters tend to become more complex in terms of their mental depth, with other effects fixed at mean level.

In addition, we observe in Table 2 a positive and significant simple effect of Age, which means that as children get older, the characters they use in their stories tend to get more complex in terms of mental depth, with other effects fixed at mean level. However, this effect is only a bit over half the size of that of Vocabulary Probability ($\beta = 0.582$ versus $\beta = -1.117$).

We learn more about the relation between Vocabulary Probability and Age by looking at the small non-significant interaction effect Vocabulary Probability * Age in Table 2. It

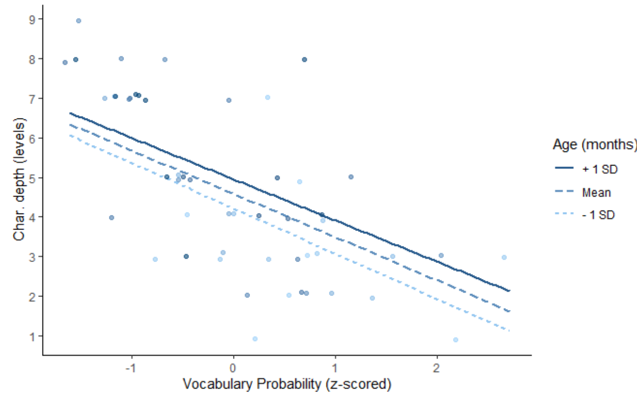

Figure 2: Interaction plot of Vocabulary Probability * Age

indicates that the effect of vocabulary is not moderated by age, in other words, is not significantly different for children at different age levels. This is visible in Figure 2, where three lines indicate predictions of character depth for various levels of age, but have similar slopes.

With respect to our second hypothesis, we observe in Table 2 that the simple effect Dependency Distance and the interaction Dependency Distance * Age are both small and non-significant. Thus, contrary to our expectation, this model suggests that the distance between syntactically dependent words does not explain the observed variation in the levels of characters' mental depth, nor can we say that age plays a moderating role here. Finally, we can see in Table 2 that the main effect of Education Parents is positive and a bit smaller than age, but non-significant, suggesting that parental years of education do not reliably predict levels of characters' mental depth either.

Although we saw in the bivariate explorations section that there is a moderate correlation between Vocabulary Probability and Age, ($r = -.4$), we have no indications that these and other predictors pose multicollinearity issues for the estimates in our model, since all computed variance inflation factors are below 1.44. We thus find some tentative evidence for the idea that in our model, Vocabulary Probability and Age have independent effects.

Table 2: Model terms sorted on standardized coefficient size.

Term	*Estimate*	*SE*	*t*	*p*	*5% CI*	*95% CI*
(Intercept)	4.716	0.250	18.892	<.001	4.212	5.219
Vocabulary Probability	-1.117*	0.289	-3.860	<.001	-1.701	-0.534
Age	0.582*	0.265	2.193	0.036	0.047	1.117
Education Parents	0.425	0.265	1.602	0.116	-0.110	0.961
Age * Vocabulary Probability	0.161	0.292	0.551	0.584	-0.428	0.750
Age * Dependency Distance	0.146	0.232	0.628	0.533	-0.322	0.614
Dependency Distance	-0.016	0.242	-0.067	0.947	-0.110	0.473

($R^2 = .525$, $F[6, 44] = 8.132$, $p < 0.001$) * significant at the .05 level, 2-tailed Character Depth (M = 4.67, SD = 2.1)

General Discussion

Our central finding is that lexical richness is a key story-internal factor for predicting a story's socio-cognitive sophistication, as manifested in the mental depth of characters. This finding has multiple implications and possible interpretations. Firstly, it seems to follow that rich vocabularies are particularly helpful in organizing and describing the storyworld, including its social and mental aspects. In theory, this could be entirely independent of actual socio-cognitive skills possessed by the child telling the story: it could be merely a matter of being able or not to find the right words for fleshing out a character in terms of its emotional and intentional states. However, with existing research in mind (e.g. Milligan et al., 2007; De Villiers & De Villiers, 2014) it appears more likely that our observed effect extends beyond the realm of the stories as such, and that possessing a richer vocabulary not only enhances a child's communication about the social world, but also supports its understanding of and ability to reason about socio-cognitive matters. Here it is particularly salient that the effect is larger than and not moderated by age. This adds a new perspective to the debate about the period in which children start to invoke others' mental states in their language (for an overview see Nicolopoulou, 2016). Rather than disclosing a "Rubicon" moment for ToM-language use in children, we propose a methodology that is able to show *what it is* about certain aspects of language development, such as having access to a richer lexicon, that engenders fleshing out mental activity in more detail, regardless of what age a child has. To substantiate such an interpretation, further research is needed focused on establishing firmer links between patterns observed inside stories and development as it takes place within the children that tell them. Here we see a role for collaborative work involving both (computational) linguists/narratologists and developmental psychologists.

For syntactic complexity the picture is quite different; we see no significant evidence for its contribution to character depth in our sample. Although in our bivariate explorations we saw a hint of the relation we hypothesized, in our model it was probably trumped by other effects. A reason for this could be that speech employs overall lower dependency distances compared to written text, which for children may even be stronger the case. If dependency effects are thus generally smaller, we must revisit this prediction with more data and maybe also compare and evaluate different metrics of syntactic complexity, such as clause length and words per finite verb.

A general remark about our methodology is that the use of computational language processing tools makes operationalizing "narrative sophistication", as we have done (and as is also proposed by Nicolopoulou in Cremin et al., 2016), a lot easier, more reproducible, and more scalable. With larger datasets we might in the future be able to use story-internal variables to approximate children's narratological and linguistic capacities, as well as related cognitive skills, when no external information about the storytellers is available, or when collection and storage of sensitive data from children or parents is to be minimized. In addition to (and to provide a more solid foundation for) such computational approaches, we see multiple directions in which research may go that aims to deepen our understanding of the relationships between socio-cognitive development and narrative/linguistic competence. A possibility would be to include stories from a more diverse population, for example by involving atypically developing children, and/or collect additional data about each individual storyteller's performance on relevant standardized tasks (e.g. those used by Wellman & Liu, 2004). Another exciting possibility would be to compare our sample to story corpuses in other languages, ideally differing substantially from Dutch in their syntactic and semantic structuring. Such extensions could help to further bootstrap patterns within the stories on trends in individual development, and shed light on directionality and causality of the interactions.

Finally, insufficient returned consent forms and other factors diminished the number of children per session we could include, which constrained this study to a fixed-effects model. Using more advanced random effect modelling we could most likely make better estimates of the relevant relationships, since such models would be able to take session-bound dependencies between for instance vocabularies into account. With this perspective in mind, we emphasize that a first improvement for our future research will be to focus on more participants per workshop session. Currently, the prospects for our story corpus are looking good: recent data collections in Spring 2021 yielded about 200 additional stories to be analyzed. The goal for the rest of this year is to compile a corpus of at least 500 stories, consisting of around 8 hours of high-quality child speech recordings and 50000 tokens, that is open to researchers with all kinds of backgrounds and interests. A huge bonus so far is that children love our storytelling workshop, and are happy to see us come each time.

Acknowledgments

We thank Isabelle Blok, Yasemin Tunbul, Nikita Ham, Iris Jansen, Bram Loth, and Ties Westerbeek for their help with collecting and processing the data, and we thank Barend Beekhuizen for helpful comments. The research reported in this paper was supported by the Elise Mathilde Fonds, Leiden University Fund, and the Dutch Research Council (NWO).

References

Apperly, I. (2010). *Mindreaders. The cognitive basis for Theory of Mind*. New York: Psychology Press.

Bergen, D. (2002). The role of pretend play in children's cognitive development. *Early Childhood Research & Practice, 4*(1), 1-13.

Cremin, T., Flewitt, R., Mardell, B., & Swam, J. (Eds.) (2016). *Storytelling in early childhood: enriching language, literacy and classroom culture*. NY: Routledge.

Curenton, S.M., & Justice, L.M. (2004). Children's preliteracy skills: Influence of mothers' education and

beliefs about shared-reading interactions. *Early Education and Development, 19*(2)*,* 261-283.
De Villiers, J.G. & De Villiers, P.A. (2014). The role of language in Theory of Mind development. *Top Lang Disorders*, *34*(4), 313-328.
Duijnmeijer, I., De Jong, D., & Scheper, A. (2012). Narrative abilities, memory and attention in children with a specific language impairment. *International Journal of Language & Communication Disorders 47*(5), 542-55.
Fenxiang, F., Yang, Y., & Wang, Y. (2016). The probability distribution of textual vocabulary in the English language. *Journal of Quantitative Linguistics, 23*, 49-70.
Gibson, E. (1998). Linguistic complexity: locality of syntactic dependencies. *Cognition, 68*, 1–76.
Gildea, D. & Temperly, D. (2010). Do grammars minimize dependency length? *Cognitive Science, 34*, 286-310.
Gius, E. Meister, J., Petris, M., Meister, M., Bruck, C., Jacke, J., Schumacher, M., Flüh, M., & Horstmann, J. (2020). *CATMA 6.1*. Zenodo.
King, J., & Just, M. (1991). Individual differences in syntactic processing: The role of working memory. *Journal of Memory and Language, 30*, 580–602.
McKeough, A., & Genereux, R. (2003). Transformation in narrative thought during adolescence: The structure and content of story compositions. *Journal of Educational Psychology, 93*(3), 537-52.
Milligan, K., Astington, J., & Dack, L. (2007). Language and Theory of Mind: Meta-analysis of the relation between language ability and false-belief understanding. *Child Development, 78*(2), 622–646.
Nicolopoulou, A. (1993). Play, cognitive development, and the social world: Piaget., Vygotsky, and beyond. *Human Development, 36*, 1-23.
Nicolopoulou, A. (2016). Young children's pretend play and storytelling as modes of narrative activity. In Douglas, S. & Stirling, L. (Eds.) *Children's play, pretense, and story* (pp. 7-27). New York: Routlegde.
Nicolopoulou, A. & Richner, E. (2007). From actors to agents to persons: The development of character representation in young children's narratives. *Child Development 78*(2), 412-429.
Nicolopoulou, A. & Ünlütabak, B. (2017). Narrativity and mindreading revisited. Children's understanding of theory of mind in a storybook and in standard false belief tasks. In Ketrez, F. et al. (Eds). *Social environment and cognition in language development. studies in honor of Ayhan Aksu-Koç* (pp. 151-166). Amsterdam: Benjamins.
Oya, M. (2011). Syntactic dependency distance as sentence complexity measure. In *Proceedings of the 16th International Conference of Pan-Pacific Association of Applied Linguistics* (pp. 313-316). Chigasaki: PAAL Japan.
Paley, V. (1990). *The boy who would be a helicopter. The uses of storytelling in the classroom.* Cambridge MA: Harvard UP.
Roark, B., Mitchell, M., & Hollingshead, K. (2007). Syntactic complexity measures for detecting mild cognitive impairment. In Cohen, K.B. & Demner-Fushman, D. (Eds.) *BioNLP: Biological, Translational, and Clinical Language Processing* (pp. 1-8). Madison: ACL.
Shatz, M., Diesendruck, G., Martinez-Beck, I., & Akar, D. (2003). The influence of language and socioeconomic status on children's understanding of false belief. *Developmental Psychology, 39*, 717-729.
Smith, P., & Roopnarine, J. (2019). *The Cambridge handbook of play. Developmental and disciplinary perspectives*. Cambridge: Cambridge UP.
Southwood, F., & Russell, A. (2004). Comparison of conversation, freeplay and story generation as methods of language sample elicitation. *Journal of Speech, Language and Hearing Research, 47*(2), 366-76.
Sutton-Smith, B. (1981). *The folkstories of children*. Philadelphia: U of Pennsylvania P.
Tellings, A., Oostdijk, N., Monster, I., Grootjen, F., & Van den Bosch, A. (2018). BasiScript. A corpus of contemporary Dutch texts written by primary school children. *International Journal of Corpus Linguistics, 23*, 494-508.
Tompkins, V., Farrar, M.J., & Montgomery, E.M. (2019). Speaking Your Mind: Language and Narrative in Young Children's Theory of Mind Development. In Benson, J.B. (Ed.) *Advances in Child Development and Behavior* (pp. 109-140). Amsterdam: Elsevier.
Van den Bosch, A., Busser, G.J., Canisius, S., & Daelemans, W. (2007). An efficient memory-based morphosyntactic tagger and parser for Dutch. *Proceedings of the 17th Meeting of Computational Linguistics in the Netherlands* (pp. 191-206). Utrecht: LOT Occasional Series.
Van Duijn, M.J., I. Sluiter, & A. Verhagen. (2015). When narrative takes over: The representation of embedded mind states in Shakespeare's *Othello. Language and Literature*, *24*(2): 148-166.
Van Noord, G. (2006). At last parsing is now operational. *TALN06. Verbum Ex Machina. Actes de la 13e conference sur le traitement automatique des langues naturelles* (pp. 20-42). Leuven: Leuven University Press.
Vermeer, A. (2001). Breadth and depth of vocabulary in relation to L1/L2 acquisition and frequency of input. *Applied psycholinguistics, 22*, 217-234.
Wang, Y., & Liu, H. (2017). The effects of genre on dependency distance and dependency direction. *Language Sciences, 59*, 135-147.
Wellman, H. M. & Liu, D. (2004). Scaling of theory-of-mind tasks. *Child Development*, *75*(2): 523–541.
Zeman, S. (2018). What is a Narration – and why does it matter? In Hübl, A. & Steinbach, M. (Eds.), *Linguistic Foundations of Narration in Spoken and Sign Languages* (pp. 173-206). Amsterdam: John Benjamins.
Zunshine, L. (2006). *Why we read fiction. Theory of Mind and the novel*. Columbus: The Ohio State UP.
Zunshine, L. (2019). What Mary Poppins knew: Theory of Mind, children's literature, history. *Narrative, 27*(1), 1-29.