UD-English-CHILDES: A Collected Resource of Gold and Silver Universal Dependencies Trees for Child Language Interactions

Xiulin Yang[♆] **Zhuoxuan Ju**[♆] **Lanni Bu**[♆] **Zoey Liu**[❀] **Nathan Schneider**[♆]
[♆]Georgetown University
[❀]University of Florida
{xy236, zj153, lb1437, nathan.schneider}@georgetown.edu
liu.ying@ufl.edu

Abstract

CHILDES is a widely used resource of transcribed child and child-directed speech. This paper introduces UD-English-CHILDES, the first officially released Universal Dependencies (UD) treebank. It is derived from previously dependency-annotated CHILDES data, which we harmonize to follow unified annotation principles. The gold-standard trees encompass utterances sampled from 11 children and their caregivers, totaling over 48K sentences (236K tokens). We validate these gold-standard annotations under the UD v2 framework and provide an additional 1M silver-standard sentences, offering a consistent resource for computational and linguistic research.

1 Introduction

The Child Language Data Exchange System (CHILDES) (MacWhinney, 2000) has long been a key resource for research in language acquisition, computational modeling of child language, and the evaluation of Natural Language Processing (NLP) tools. However, many analyses rely on different grammatical assumptions (e.g., Pearl and Sprouse, 2013; Szubert et al., 2024; Liu and Prud'hommeaux, 2021; Gretz et al., 2015; Sagae et al., 2007), and therefore adopt divergent annotation frameworks or standards. While most existing annotations use syntactic dependencies—in part due to the relative simplicity of annotation and parsing and the growing adoption of the Universal Dependencies (UD) framework (Nivre et al., 2016, 2020)—annotation practices remain inconsistent across datasets. This is largely due to the lack of a unified guideline for annotating children's speech, which presents unique challenges not fully addressed by existing UD documentation.

As UD treebanks have become valuable resources in both NLP (e.g., Jumelet et al., 2025; Opitz et al., 2025) and language acquisition research (e.g., Clark et al., 2023; Hahn et al., 2020), there have been increasing efforts to parse CHILDES data using tools such as `stanza` (Liu and MacWhinney, 2024). However, the resulting annotation quality is often inconsistent and cannot be guaranteed. In this paper, we compile, harmonize, and manually correct major UD-style annotations of CHILDES data into a consistent, unified UD format, resulting in a gold-standard treebank of 48K sentences and 236K tokens (including, e.g., the tree in Figure 1). In addition, we construct a larger silver-standard treebank of 1M sentences and 6M tokens produced by `stanza`[1] and report parser accuracy estimates. We publicly release both datasets.[2]

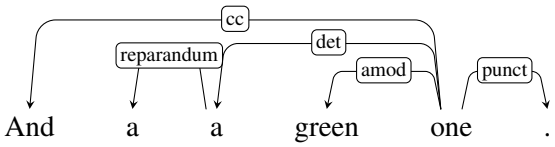

Figure 1: UD tree for a child utterance from Lily (Providence corpus, sentID=16916280)

2 Related Work

2.1 CHILDES Corpora

CHILDES is a collection of child–adult conversations recorded in naturalistic or laboratory settings. It has played a central role in both language acquisition research and the development of NLP tools. In addition to specialized corpora—such as clinical datasets (Gillam and Pearson, 2004), naturalistic family interactions (Gleason, 1980), and controlled laboratory studies (Newman et al.,

[1] `stanza 1.9.2 (combined model)`

[2] Official gold UD release: https://github.com/UniversalDependencies/UD_English-CHILDES Note: Due to a postprocessing error, the gold UD release from the `main` branch is missing approximately 10K sentences. For complete access to the data, please use the `dev` branch. The `main` branch will be updated in the next official release scheduled for November 2025.
Silver release: https://github.com/xiulinyang/UD-CHILDES.

Corpus	Children	Speakers	Trees	UPOS	Feats	Utterances		Tokens	
						Gold	Silver	Gold	Silver
S+24	Adam	Adults	Gold	Gold	Convrtd	17,233	0	91,114	0
LP21	Eve	All	Gold	Silver	Silver	2,207	0	8,497	0
LP23	10 Children	All	Gold	Silver	Silver	34,530	0	168,284	0
UD-English-CHILDES	11 Children	All	Gold/Silver	Gold/Silver	N/A	**48,183**	**1,197,471**	**236,941**	**6,892,314**

Table 1: Overview of CHILDES-based UD treebanks compiled in this paper and our newly-released UD-English-CHILDES treebank. Source corpus labels (S+24, LP21, LP23) are defined in §3. Note that there is overlap in the Adam data: S+24 figures are counts from the original dataset; for our version, these were filtered to avoid duplicates and merged with corresponding LP23 utterances. The heading **Gold** refers to the subset of utterances for which trees and UPOS have been manually corrected according to the UD v2 framework; **Silver** refers to the subset with fully automatic annotations from `stanza`.

Child	Corpus	Child age range	Gold sents	Gold toks	Silver sents	Silver toks
Laura	Braunwald (Braunwald, 1971)	1;3-7;0 (1;3-7;0)	4,622	21,079	41,862	205,427
Adam	Brown (Brown, 1973)	1;6-5;2 (1;6-5;2)	16,736	84,643	93,315	452,348
Eve	Brown	1;6-5;1 (1;6-5;2)	2,207	8,497	108,044	532,319
Abe	Kuczaj (Kuczaj, 1977)	2;4-5;0 (2;4-5;0)	4,167	22,437	38,630	230,489
Sarah	Brown	1;6-5;2 (1;6-5;2)	5,347	23,233	104,926	517,654
Lily	Providence (Demuth et al., 2006)	0;11-4;0 (0;11-4;0)	1,499	6,337	79,573	422,245
Naima	Providence	1;3-3;11 (0;11-4;0)	2,534	14,360	236,350	1,422,543
Violet	Providence	0;11-4;0 (0;11-4;0)	721	1,857	32,801	164,975
Thomas	Thomas (Lieven et al., 2009)	2;0-4;11 (2;0-4;11)	4,240	20,333	313,550	2,039,132
Emma	Weist (Weist and Zevenbergen, 2008)	2;2-4;10 (2;1-5;0)	2,423	13,730	74,825	474,460
Roman	Weist	2;2-4;9 (2;1-5;0)	3,653	20,557	73,595	467,633

Table 2: Detailed statistics for each child, including counts of gold and silver annotations and their corresponding age ranges in months. Ages in the silver corpus are shown in parentheses. For source corpus URLs see Appendix A.

2016)—CHILDES supports a wide range of approaches to developmental linguistics. Many of its corpora inform foundational theories of language acquisition, particularly the poverty of the stimulus hypothesis (Chomsky, 1976). Researchers frequently use child-directed speech from CHILDES to quantify the distribution of linguistic structures that are central to these theories, such as wanna contraction (Getz, 2019), anaphoric one (Foraker et al., 2009; Pearl and Mis, 2011), auxiliary fronting (Perfors et al., 2011), and syntactic islands (Pearl and Mis, 2011). It has also been used in computational models of language acquisition (e.g., Abend et al., 2017).

CHILDES has also emerged as a valuable resource for NLP tool benchmarking and language model pretraining. Following the work of Huang (2016), studies such as Liu and Prud'hommeaux (2023) have highlighted the challenges faced by UD parsers when applied to child-directed speech, showing substantial performance gaps compared to adult data. CHILDES also supports recent research on pretraining dynamics (Feng et al., 2024) and the development of efficient language models, including in initiatives like the BabyLM Challenge (Choshen et al., 2024; Charpentier et al., 2025).

2.2 Spoken Language Treebanks

Overview The development of UD project has fostered the development of spoken language annotations across a wide variety of languages, such as Beja (Kahane et al., 2021) and Japanese (Omura et al., 2023), as documented in Dobrovoljc (2022). For English, the GUM corpus (Zeldes, 2017) incorporates several spoken genres.

CHILDES Dependency Treebanks Early dependency parsing research on English CHILDES data utilized a custom inventory of grammatical relations (GR; Sagae et al., 2004, 2005). These gradually evolved to address CHILDES-specific challenges (Sagae et al., 2007), and were applied to the entire English CHILDES corpus using a supervised parser (Sagae et al., 2010).

More recently, UD-style annotations have been introduced to CHILDES. Liu and MacWhinney (2024) release an automatically parsed version of the English CHILDES corpus, annotated with UD trees using `stanza`. Liu and Prud'hommeaux (2021) used a semi-automatic method to convert previous GR-based annotations into UD trees, focusing on child-produced speech (ages 18–27 months) from the Eve data within the Brown corpus (Brown, 1973). Subsequently, Szubert et al. (2024)

```
# sent_id = 22497 (normalized sentence ID across corpora; used to avoid
                collisions since some corpora share identical sentence IDs)
# original_sent_id = 946255 (original sentence ID from the corpus, as assigned
                           in childsr)
# childes_toks = who's that (original token string from childsr)
# child_name = Abe
# corpus_name = Kuczaj
# gold_annotation = True
# speaker_age = 43.72369042485472 (child's age in months)
# speaker_gender = male (child's gender)
# speaker_role = Father (speaker role in conversation)
# type = question (sentence type annotation)
# text = Who's that?
1-2   Who's _     _     _     _     _     _     _     _
1     Who   who   PRON  WP    _     0     root  0:root      _
2     's    be    AUX   VBZ   _     1     cop   1:cop _
3     that  that  PRON  DT    _     1     nsubj 1:nsubj     SpaceAfter=No
4     ?     ?     PUNCT ?     _     1     punct 1:punct     _
```

Figure 2: Example of a gold-annotated CoNLL-U sentence from the CHILDES-Providence corpus, with added parenthetical explanations of sentence-level metadata. Enhanced UD (EUD) relations are added deterministically by the script at https://github.com/amir-zeldes/gum/blob/master/_build/utils/eng_enhance.ini.

developed gold-standard UD annotations by automatically transforming GR annotations and manually correcting them. Their dataset includes child-directed speech from the Adam data of the Brown corpus and the Hebrew Hagar corpus (Berman, 1990), addressing spoken-language-specific phenomena such as repetitions and non-standard vocabulary, as well as a mapping to semantics.

Building upon these efforts, Liu and Prud'hommeaux (2023) significantly expanded UD annotations to cover utterances from 10 children aged 18–66 months (Adam from the Brown corpus as well as 9 children from other corpora), incorporating both child and caregiver speech. Their work tackles complex spoken-language features, including speech repairs and restarts.

Although Liu and Prud'hommeaux (2021, 2023) provide manually corrected UD trees, their annotations are inconsistent with the UD v2 framework, lack Universal Part-of-Speech (UPOS) tags, and have not been independently verified. Szubert et al. (2024) offer verified data, but they follow the UD v1 annotation guidelines. To date, there is no official UD release for CHILDES speech data.

3 Annotations

3.1 Data Source & Statistics

This work leverages three existing UD treebanks: Szubert et al. (2024) (henceforth **S+24**), Liu and Prud'hommeaux (2021) (**LP21**), and Liu and Prud'hommeaux (2023) (**LP23**), summarized in Table 1. As these treebanks were already annotated, our human annotation efforts focused primarily on correcting errors and harmonizing annotations across corpora. We present post-compilation statistics in Tables 1 and 2. Table 1 summarizes the full corpus and its source contributions, and Table 2 provides per-child statistics.

In the official UD release, we divide the corpus based on the children's names and genders. The training and dev splits (90% and 10%, respectively) are constructed from the data of Adam, Lily, Naima, Sarah, Roman, Laura, and Abe. The corpora of Eve, Violet, Emma, and Thomas are reserved for the test split. Details are reported in Table 3.

3.2 Annotation Pipeline

Following Liu and Prud'hommeaux (2023), we collect CHILDES corpora using the R package `childesr` (Sanchez et al., 2019).[3] Sentence normalization can be found in the paper. As the data from LP21 and LP23 are only parsed but not tagged yet, sentences with existing dependency annotations are identified and automatically tagged with UPOS using `stanza` (Qi et al., 2020), while unannotated sentences are assigned both UPOS and dependency trees. Our current work focuses on correcting previously human-annotated data. To ensure conformity with UD guidelines, we run all processed sentences through the UD validation tool[4] and manually fix those that fail validation. The correction work is performed by three linguistics graduate students trained in UD annotation. In

[3] https://langcog.github.io/childes-db-website/

[4] https://github.com/UniversalDependencies/tools/blob/master/validate.py

Split	Children	Corpus	Gold Sents
Train	Adam, Lily, Naima, Sarah, Roman, Laura, Abe	Brown, Providence, Weist, Kuczaj, Braunwald	34,732
Dev	Adam, Lily, Naima, Sarah, Roman, Laura, Abe	Brown, Providence, Weist, Kuczaj, Braunwald	3,860
Test	Eve, Violet, Emma, Thomas	Brown, Providence, Weist, Thomas	9,591

Table 3: Data splits for the official UD_English-CHILDES with associated children, corpora, and gold-standard sentence counts.

total, we made approximately 8,000 corrections.

Many of the errors stem from mismatches between UPOS tags and dependency labels (as LP21 and LP23 used automatic UPOS tagging). In addition, we address format issues such as multiword tokens, spacing mismatches (e.g., `SpaceAfter`), and deprecated dependency relations not supported by current UD guidelines (e.g., `compound:svc`, `obl:about_like`, `nmod:over_under`). The 5 most common linguistic issues were as follows:

`advmod` tagged as `ADP` This error commonly appears with phrasal verbs such as *get up* and *take over*. The original annotation assigns `advmod` as the dependency relation to phrasal verbs with POS tag `ADP`. We revise these to `compound:prt`, in accordance with the UD treatment of phrasal particles.

Auxiliaries tagged as `VERB` or `PART` Auxiliaries such as *be* and *have* are frequently misclassified as main verbs or particles. In some cases, lemmas are also mislabeled—most notably, the lemma of contracted forms like *'s* is incorrectly assigned as *'s* rather than the appropriate auxiliary *be*. We correct both the POS and lemma annotations in these cases.

Lexical items tagged as `PUNCT` The `stanza` parser often mislabels disfluent word fragments in spontaneous speech as punctuation marks (e.g., *OK/INTJ Adam/PROPN ride/VERB dat/**PUNCT** ./PUNCT*. We reassign these tokens appropriate UPOS labels based on context and speaker intent, often as interjections.

Determiner misrecognition Ambiguous or reduced forms of determiners—such as *de* —are frequently misidentified as proper nouns (`PROPN`). We manually review these cases and reannotate them as `DET` when appropriate.

Function word heads with dependents In previous treebanks, words appearing in functional relations such as `case`, `mark`, and `aux` have been assigned children, which violates UD's constraint that these words should be leaf nodes. We reassign the erroneous dependents to the appropriate content heads, ensuring the structure conforms to UD's projectivity and function word constraints.

3.3 Harmonization

Each treebank follows its own annotation guidelines, which are largely based on UD but not fully compliant. We performed a series of normalization steps to harmonize them into a consistent format. Our unified format is primarily based on LP23, with several adaptations described below.

Metadata In our normalized CoNLL-U files, we include the following metadata fields with an example provided in Figure 2: `sent_id` (normalized sentence IDs); `original_sent_id` (utterance ID retrieved via the `childesr` R package); `childes_toks` (tokenized utterance); `corpus_name` (original corpus name); `gold_annotation` (indicates whether the sentence is manually annotated); `speaker_gender`, `speaker_role`, and `speaker_age` (speaker/child metadata); `text` (the text aligned with the tree), and `type` (sentence type). Table 4 summarizes the distribution of the main sentence types and compares them with those in the UD 2.15 release of GUM (Zeldes, 2017), a multi-genre English corpus. Notably, questions occur in the CHILDES conversations at a much higher rate—they are nearly half (45%) as frequent as declarative utterances, as opposed to 9% in GUM.

Type	CHILDES			GUM
	CDS	CS	Overall	Overall
declarative	16,112	15,884	31,996	7,695 (`decl`)
question	2,882	11,413	14,295	716 (`q`, `wh`)
imperative	509	288	797	1,326 (`imp`, `intj`)
emphatic				
others	601	494	1095	2,409

Table 4: Sentence type counts in gold CHILDES and GUM corpora. **Question** includes *question*, *self interruption question*, *trail off question*, and *interruption question*. **Others** encompasses less frequent categories: *trail off*, *interruption*, *self interruption*, and *quotation next line*.

Punctuation To bring the transcripts in line with written English conventions, we capitalize the first word of each utterance and infer sentence-final

Metrics	Children's speech	Parents' speech	Overall
LAS	81.2	86.3	84.2
UAS	87.2	91.0	89.5

Table 5: LAS and UAS scores for children's speech, parents' speech, and overall performance.

punctuation at the end of each sentence based on the sentence type provided in the metadata.[5]

Reparandum Each of the three treebanks defines its own subtypes for the `reparandum` and `parataxis` relations. For example, S+24 includes labels such as `parataxis:repeat` not present in the current UD guidelines. Similarly, LP21 and LP23 annotate `reparandum` with subtypes such as `restart` and `repetition` to mark special utterance features of children's speech. To ensure consistency across treebanks, we move all such subrelation information to the `MISC` column.

Others As S+24 was annotated using the UD guidelines version 1.0, we convert the annotation using UD version 2.0 with a script[6] and manual annotation. For example, we shifted the head-dependent direction of `flat` in the annotations.

Since S+24 and LP23 overlap in the Adam corpus, we merged the annotations from these two treebanks. 3375 sentences are repetitive in S+24. We removed these sentences from our corpus.[7]

To ensure a more linguistically plausible analysis, we also diverged from Liu and Prud'hommeaux (2023) in our treatment of interjections. Instead of annotating utterances consisting solely of interjections (e.g., *Ha ha ha ha*) as `conj`, we used the `flat` relation.

3.4 Silver Data Assessment

To create silver-standard annotations, we apply `stanza` to the utterances that were not sampled by the previous treebanks (but were from the same CHILDES datasets, i.e. conversations involving the 11 children in Table 2). To estimate the quality of these silver annotations, we evaluate the parser's performance on the gold-standard data. We report Labeled and Unlabeled Attachment Scores (LAS/UAS) in Table 5. The parser achieves an overall LAS of 83.3. Performance is higher on parents' speech (86.3 LAS) than on children's speech (81.2 LAS), likely due to the greater syntactic regularity and lower frequency of disfluencies in adult utterances. The overall high-quality data can be more easily verified by human annotators than annotated from scratch. It also provides valuable training data for improving parsers on spoken language.

[5]The original data transcribes various kinds of prosodic information such as pauses. At present we do not retain this information or attempt to infer corresponding punctuation like commas and parentheses.

[6]https://github.com/UniversalDependencies/tools/tree/master/v2-conversion

[7]883 sentences from S+24 could not be merged because S+24 and LP21 are using different data sources, and were therefore removed from our treebank as well.

4 Conclusion & Future Work

In this paper, we present the first harmonized UD treebanks for CHILDES, covering 11 corpora and over 48k sentences from both child-directed and child-produced speech. The three datasets we compiled do not preserve conversational structure, and as a result, the finalized gold-standard treebank lacks coherent dialogue sequencing. Preserving such structure would require additional manual annotation to make sure all sentences are gold. However, since our annotations include the `original_sent_id` field, reconstructing the conversation structure is straightforward. Furthermore, morphological features have not been annotated or independently verified. Future work will focus on further corrections to the silver-standard data and the continued expansion of the treebanks. We welcome collaboration on this ongoing effort.

Acknowledgments

We acknowledge Ida Szubert, Omri Abend, Samuel Gibbon, Louis Mahon, Sharon Goldwater, Mark Steedman, and Emily Prud'hommeaux for their contributions to the original UD treebanking efforts. We also thank Brian MacWhinney for helpful discussions and anonymous reviewers for their suggestions.

References

Omri Abend, Tom Kwiatkowski, Nathaniel J. Smith, Sharon Goldwater, and Mark Steedman. 2017. Bootstrapping language acquisition. *Cognition*, 164:116–143.

Ruth A. Berman. 1990. On acquiring an (S)VO language: subjectless sentences in children's Hebrew. *Linguistics*, 28(6):1135–1166.

Susan R Braunwald. 1971. Mother-child communication: the function of maternal-language input. *Word*, 27(1-3):28–50.

Roger Brown. 1973. *A first language: The early stages*. Harvard University Press.

Lucas Charpentier, Leshem Choshen, Ryan Cotterell, Mustafa Omer Gul, Michael Hu, Jaap Jumelet, Tal Linzen, Jing Liu, Aaron Mueller, Candace Ross, Raj Sanjay Shah, Alex Warstadt, Ethan Wilcox, and Adina Williams. 2025. BabyLM turns 3: Call for papers for the 2025 BabyLM workshop. *arXiv preprint arXiv:2502.10645*.

Noam Chomsky. 1976. *Reflections on language*. Temple Smith London.

Leshem Choshen, Ryan Cotterell, Michael Y Hu, Tal Linzen, Aaron Mueller, Candace Ross, Alex Warstadt, Ethan Wilcox, Adina Williams, and Chengxu Zhuang. 2024. [Call for papers] the 2nd BabyLM challenge: Sample-efficient pretraining on a developmentally plausible corpus. *arXiv preprint arXiv:2404.06214*.

Thomas Hikaru Clark, Clara Meister, Tiago Pimentel, Michael Hahn, Ryan Cotterell, Richard Futrell, and Roger Levy. 2023. A cross-linguistic pressure for Uniform Information Density in word order. *Transactions of the Association for Computational Linguistics*, 11:1048–1065.

Katherine Demuth, Jennifer Culbertson, and Jennifer Alter. 2006. Word-minimality, epenthesis and coda licensing in the early acquisition of English. *Language and speech*, 49(2):137–173.

Kaja Dobrovoljc. 2022. Spoken language treebanks in Universal Dependencies: an overview. In *Proceedings of the Thirteenth Language Resources and Evaluation Conference*, pages 1798–1806, Marseille, France. European Language Resources Association.

Steven Y. Feng, Noah Goodman, and Michael Frank. 2024. Is child-directed speech effective training data for language models? In *Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing*, pages 22055–22071, Miami, Florida, USA. Association for Computational Linguistics.

Stephani Foraker, Terry Regier, Naveen Khetarpal, Amy Perfors, and Joshua Tenenbaum. 2009. Indirect evidence and the poverty of the stimulus: The case of anaphoric *one*. *Cognitive Science*, 33(2):287–300.

Heidi R Getz. 2019. Acquiring *wanna*: Beyond Universal Grammar. *Language Acquisition*, 26(2):119–143.

Ronald Bradley Gillam and Nils A Pearson. 2004. *Test of narrative language*. Pro-ed Austin, TX.

Jean Berko Gleason. 1980. The acquisition of social speech routines and politeness formulas. In *Language*, pages 21–27. Elsevier.

Shai Gretz, Alon Itai, Brian MacWhinney, Bracha Nir, and Shuly Wintner. 2015. Parsing Hebrew CHILDES transcripts. *Language Resources and Evaluation*, 49:107–145.

Michael Hahn, Dan Jurafsky, and Richard Futrell. 2020. Universals of word order reflect optimization of grammars for efficient communication. *Proceedings of the National Academy of Sciences*, 117(5):2347–2353.

Rui Huang. 2016. An evaluation of POS taggers for the CHILDES corpus. *CUNY Academic Works*.

Jaap Jumelet, Leonie Weissweiler, and Arianna Bisazza. 2025. MultiBLiMP 1.0: A massively multilingual benchmark of linguistic minimal pairs. *arXiv preprint arXiv:2504.02768*.

Sylvain Kahane, Martine Vanhove, Rayan Ziane, and Bruno Guillaume. 2021. A morph-based and a word-based treebank for Beja. In *Proceedings of the 20th International Workshop on Treebanks and Linguistic Theories (TLT, SyntaxFest 2021)*, pages 48–60, Sofia, Bulgaria. Association for Computational Linguistics.

Stan Kuczaj. 1977. The acquisition of regular and irregular past tense forms. *Journal of Verbal Learning and Verbal Behavior*, 16(5):589–600.

Elena Lieven, Dorothé Salomo, and Michael Tomasello. 2009. Two-year-old children's production of multiword utterances: A usage-based analysis. *Cognitive Linguistics*, 20(3):481–507.

Houjun Liu and Brian MacWhinney. 2024. Morphosyntactic analysis for CHILDES. *Language Development Research*, 4(1).

Zoey Liu and Emily Prud'hommeaux. 2021. Dependency parsing evaluation for low-resource spontaneous speech. In *Proceedings of the Second Workshop on Domain Adaptation for NLP*, pages 156–165, Kyiv, Ukraine. Association for Computational Linguistics.

Zoey Liu and Emily Prud'hommeaux. 2023. Data-driven parsing evaluation for child-parent interactions. *Transactions of the Association for Computational Linguistics*, 11:1734–1753.

Brian MacWhinney. 2000. *The CHILDES project: Tools for analyzing talk, Volume I: Transcription format and programs*. Psychology Press.

Rochelle S Newman, Meredith L Rowe, and Nan Bernstein Ratner. 2016. Input and uptake at 7 months predicts toddler vocabulary: the role of child-directed speech and infant processing skills in language development. *Journal of Child Language*, 43(5):1158–1173.

Joakim Nivre, Marie-Catherine de Marneffe, Filip Ginter, Yoav Goldberg, Jan Hajič, Christopher D. Manning, Ryan McDonald, Slav Petrov, Sampo Pyysalo, Natalia Silveira, Reut Tsarfaty, and Daniel Zeman. 2016. Universal Dependencies v1: A multilingual treebank collection. In *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC'16)*, pages 1659–1666, Portorož, Slovenia. European Language Resources Association (ELRA).

Joakim Nivre, Marie-Catherine De Marneffe, Filip Ginter, Jan Hajič, Christopher D Manning, Sampo Pyysalo, Sebastian Schuster, Francis Tyers, and Daniel Zeman. 2020. Universal Dependencies v2: An evergrowing multilingual treebank collection. *arXiv preprint arXiv:2004.10643*.

Mai Omura, Hiroshi Matsuda, Masayuki Asahara, and Aya Wakasa. 2023. UD_Japanese-CEJC: Dependency relation annotation on corpus of everyday Japanese conversation. In *Proceedings of the 24th Annual Meeting of the Special Interest Group on Discourse and Dialogue*, pages 324–335, Prague, Czechia. Association for Computational Linguistics.

Juri Opitz, Shira Wein, and Nathan Schneider. 2025. Natural language processing RELIES on linguistics. *Computational Linguistics*, pages 1–23.

Lisa Pearl and Benjamin Mis. 2011. How far can indirect evidence take us? Anaphoric *one* revisited. In *Proceedings of the Annual Meeting of the Cognitive Science Society*, volume 33.

Lisa Pearl and Jon Sprouse. 2013. Syntactic islands and learning biases: Combining experimental syntax and computational modeling to investigate the language acquisition problem. *Language Acquisition*, 20(1):23–68.

Amy Perfors, Joshua B. Tenenbaum, and Terry Regier. 2011. The learnability of abstract syntactic principles. *Cognition*, 118(3):306–338.

Peng Qi, Yuhao Zhang, Yuhui Zhang, Jason Bolton, and Christopher D. Manning. 2020. Stanza: A Python Natural Language Processing Toolkit for many human languages. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics: System Demonstrations*, pages 101–108, Online. Association for Computational Linguistics.

Kenji Sagae, Eric Davis, Alon Lavie, Brian MacWhinney, and Shuly Wintner. 2007. High-accuracy annotation and parsing of CHILDES transcripts. In *Proceedings of the Workshop on Cognitive Aspects of Computational Language Acquisition*, pages 25–32, Prague, Czech Republic. Association for Computational Linguistics.

Kenji Sagae, Eric Davis, Alon Lavie, Brian MacWhinney, and Shuly Wintner. 2010. Morphosyntactic annotation of CHILDES transcripts. *Journal of Child Language*, 37(3):705–729.

Kenji Sagae, Alon Lavie, and Brian MacWhinney. 2005. Automatic measurement of syntactic development in child language. In *Proceedings of the 43rd Annual Meeting of the Association for Computational Linguistics (ACL'05)*, pages 197–204, Ann Arbor, Michigan. Association for Computational Linguistics.

Kenji Sagae, Brian MacWhinney, and Alon Lavie. 2004. Adding syntactic annotations to transcripts of parent-child dialogs. In *Proceedings of the Fourth International Conference on Language Resources and Evaluation (LREC'04)*, Lisbon, Portugal. European Language Resources Association (ELRA).

Alessandro Sanchez, Stephan C. Meylan, Mika Braginsky, Kyle E. MacDonald, Daniel Yurovsky, and Michael C. Frank. 2019. childes-db: A flexible and reproducible interface to the child language data exchange system. *Behavior Research Methods*, 51(4):1928–1941.

Ida Szubert, Omri Abend, Nathan Schneider, Samuel Gibbon, Louis Mahon, Sharon Goldwater, and Mark Steedman. 2024. Cross-linguistically consistent semantic and syntactic annotation of child-directed speech. *Language Resources and Evaluation*.

Richard M Weist and Andrea A Zevenbergen. 2008. Autobiographical memory and past time reference. *Language Learning and Development*, 4(4):291–308.

Amir Zeldes. 2017. The GUM corpus: creating multilayer resources in the classroom. *Language Resources and Evaluation*, 51(3):581–612.

A Sources of the Coprora

In this work, we include the sources from the following corpora:

- `https://childes.talkbank.org/access/Eng-NA/Braunwald.html`
- `https://childes.talkbank.org/access/Eng-NA/Brown.html`
- `https://childes.talkbank.org/access/Eng-NA/Kuczaj.html`
- `https://phon.talkbank.org/access/Eng-NA/Providence.html`
- `https://childes.talkbank.org/access/Eng-UK/Thomas.html`
- `https://childes.talkbank.org/access/Eng-NA/Weist.html`