Column ID	Column Description
id	Reference ID
file_name	File Name of the Transcript
lang	Language of the Transcript Taken from Transcript's Preamble
lang_google	Language of the Transcript Detected By Google Translation API
lang_comp	An auxiliary column comparing if language detection are the same between google and preamble
narrator	Non-Child Speaker (The Narrator) determined by the most "Proliferate" Speaker in the Transcript
syntax_integrity	TRUE = There is no syntax error found in the transcript by CLAN standard.
mlu	Mean Length of Utterance By Words
ttr	Type Token Ratio by selecting 100 Random Words
vs_mean_dis	An Example for Illustration: I eat a pizza. So I also drink a lot. Mean Levenshtein edit distance between utterances in variation sets, with gap = 0 $\text{vs\_mean\_dis} = \frac{LevenshteinEditDistance[Utterance_n, Utterance_{n+1}]}{Max[Length[Utterance_n], Length[Utterance_{n+1}]]} = \frac{4}{Max[4,6]} = \frac{1}{3}$
vs_mean_dis_ori	Mean Levenshtein edit distance without taking out words from Stop List, with gap = 0 which consists of: what, my, this, where, and, is, you, the
vs_token_p_g0	Proportion(p) of words (token) in Varset, with gap = 0. $\text{vs\_token\_p} = \frac{\text{Tokens Repeated}}{\text{Total Token Count}} = \frac{2 \times 2}{10} = 0.4$
vs_type_p_g0	Proportion(p) of unique words (type) in Varset, with gap = 0. $\text{vs\_type\_p} = \frac{\text{Types Repeated}}{\text{Total Type Count}} = \frac{2 \times 2}{8} = 0.5$
vs_ttr_g0	A Type Token Ratio of words in varsets, with gap = 0. $\text{vs\_ttr} = \frac{\text{Tokens in Varset}}{\text{Types in Varset}} = \frac{2}{2} = 1$
vs_utter_p_g0	An example for illustration: B, A, A, B, B, A, A, A (where A is in varset, B is not) The proportion(p) of utterances in Varset across the whole transcripts, with gap = 0. $\text{vs\_utter\_p} = \frac{N[\text{Utterances in Varset}]}{\text{Total } N[\text{Utterance}]} = \frac{5}{8}$
vs_mean_len_g0	The mean length of utterances in Varset across the whole transcripts, with gap = 0. $\text{vs\_utter\_p} = \frac{\sum Length[\text{Varset Sequence}_n]}{\text{Total } N[\text{Varset Sequences}]} = \frac{2+3}{2} = 2.5$
vs_token_p_g1	Proportion(p) of words (token) in Varset, with gap $\leq$ 1.
vs_type_p_g1	Proportion(p) of unique words (type) in Varset, with gap $\leq$ 1.
vs_ttr_g1	A Type Token Ratio of words in varsets, with gap $\leq$ 1.
vs_utter_p_g1	The proportion(p) of utterances in Varset across the whole transcripts, with gap $\leq$ 1.
vs_mean_len_g1	The mean length of utterances in Varset across the whole transcripts, with gap $\leq$ 1.
lsbq_dob	Date of birth of the child, retrieved from "LSBQ".
lsbq_dot	The Day of Testing, retrieved from "LSBQ".
chi_age	Age of Child on The Day of Testing, in days.
chi_l1	Child's Understanding in L1, i.e. English. (0~1)
adu_edu	Adult's (The narrator) Education Attainment in Text
adu_l1	Adult (The narrator) Relative Proficiency in L1, i.e. English. (0~1)
adu_l2	Adult (The narrator) Relative Proficiency in L2. (0~1)
adu_bi_sub	Adult (The narrator) Bilingual Index by Subtraction. (0~1)
adu_bi_div	Adult (The narrator) Bilingual Index by Division. (0~1)
adu_bi_edi	Adult (The narrator) Bilingual Index by Edinburgh's Formula (L1-L2)/(L1+L2). (-1~1)
hh_l1	Household (HH) Use of L1, i.e. English. (0~1)
hh_l2	Household (HH) Use of L2. (0~1)
hh_bi_sub	Household (HH) Bilingual Index by Subtraction. (0~1)
hh_bi_div	Household (HH) Bilingual Index by Division. (0~1)
hh_bi_edi	Household (HH) Bilingual Index by Edinburgh's Formula (L1-L2)/(L1+L2). (-1~1)
nl_adu_name	Name of Child, retrieved from "Narration Mastersheet" (nl = narration log).
nl_chi_name	Name of Narrator, retrieved from "Narration Mastersheet".
nl_narrator	Whether the narrator is the father or mother to the child, retrieved from "Narration Mastersheet".
nl_dob	Date of birth of the child, retrieved from "Narration Mastersheet".
nl_dot	The Day of Testing, retrieved from "Narration Mastersheet".
nl_lang	Language of the transcript, retrieved from "Narration Mastersheet".
nl_location	Location of Testing, retrived from "Narration Mastersheet".
nl_chi_age	AAge of Child on The Day of Testing, retrieved from "Narration Mastersheet".