

Column ID	Column Description
id	Reference ID
file_name	File Name of 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	<p>An Example for Illustration:</p> <p>I eat a pizza.</p> <p>So I also drink a lot.</p> <p>Mean Levenshtein edit distance between utterances in variation sets, with gap = 0</p> $vs_mean_dis = \frac{LevenshteinEditDistance[Utterance_n, Utterance_{n+1}]}{Max[Length[Utterance_n], Length[Utterance_{n+1}]} = \frac{4}{Max[4, 6]} = \frac{1}{3}$
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	<p>Proportion(p) of words (token) in Varset, with gap = 0.</p> $vs_token_p = \frac{Tokens\ Repeated}{Total\ Token\ Count} = \frac{2 \times 2}{10} = 0.4$
vs_type_p_g0	<p>Proportion(p) of unique words (type) in Varset, with gap = 0.</p> $vs_type_p = \frac{Types\ Repeated}{Total\ Type\ Count} = \frac{2 \times 2}{8} = 0.5$
vs_ttr_g0	<p>A Type Token Ratio of words in varsets, with gap = 0.</p> $vs_ttr = \frac{Tokens\ in\ Varset}{Types\ in\ Varset} = \frac{2}{2} = 1$
vs_utter_p_g0	<p>An example for illustration:</p> <p>B, A, A, B, B, A, A, A</p> <p>(where A is in varset, B is not)</p> <p>The proportion(p) of utterances in Varset across the whole transcripts, with gap = 0.</p> $vs_utter_p = \frac{N[Utterances\ in\ Varset]}{Total\ N[Utterance]} = \frac{5}{8}$
vs_mean_len_g0	<p>The mean length of utterances in Varset across the whole transcripts, with gap = 0.</p> $vs_utter_p = \frac{\sum Length[Varset\ Sequence_n]}{Total\ N[Varset\ Sequences]} = \frac{2+3}{2} = 2.5$
vs_token_p_g1	Proportion(p) of words (token) in Varset, with gap ≤ 1.
vs_type_p_g1	Proportion(p) of unique words (type) in Varset, with gap ≤ 1.
vs_ttr_g1	A Type Token Ratio of words in varsets, with gap ≤ 1.
vs_utter_p_g1	The proportion(p) of utterances in Varset across the whole transcripts, with gap ≤ 1.
vs_mean_len_g1	The mean length of utterances in Varset across the whole transcripts, with gap ≤ 1.