

Rules Table V2.0

Aa Pattern Coding	≡ Meaning	≡ regex/steps to decode	≡ Comments	≡ Implemented
<u>"+.?"</u>	trailing off of a question	@pause@	Trailing off occurs when speakers shift attention away from what they are saying, can't find the words to finish their thought, or sometimes even forgetting what they were going to say.	
<u>"+. ."</u>	trailing off	" " "	Trailing off occurs when speakers shift attention away from what they are saying, can't find the words to finish their thought, or sometimes even forgetting what they were going to say.	
<u>"+/."</u>	interruption by another speaker	replace interruptor's line with @ipause@ and concatenate the before-after		
<u>"+/?"</u>	interruption of a question by another speaker	?		
<u>"+//."</u>	self-interruption	remove?	This symbol is used for an utterance that is incomplete because a speaker interrupted him/herself	
<u>"+//?"</u>	self-interruption of a question			

Aa Pattern Coding	≡ Meaning	≡ regex/steps to decode	≡ Comments	≡ Implemented
<u>"+,,"</u>	self completion (after an interruption)	newline and code removed	removed	
<u>"++"</u>	other completion (after an interruption)	newline and code removed		
<u>"+"./."</u>	quotation follows on next line	\+"/\.\ [\\+ gram\\]n\\+"` removed \+"/\.\ [\\+ gram\\]n` removed \+"/\.\ [\\+ gram\\] n\\+"` removed \+"/\.\ [\\+ gram\\] n` removed same for jar, es. and exc in place of gram \+"/\.\ n\\+"` removed \+"/\.\n\\+"` removed \+"/\.\ n` removed \+"/\.\n` removed		
<u>"+"""</u>	quoted utterance occurs on this line (use at beginning of utterance as link, not a terminator)	newline and code removed		
<u>"+"."."</u>	material quoted on this line was preceded by quoted material on previous line	treated as followed, and removed		

Aa Pattern Coding	≡ Meaning	≡ regex/steps to decode	≡ Comments	≡ Implemented
<u>"<"</u>	lazy overlap marking (at beginning of utterance that overlapped the the previous utterance)	newline and code removed	The “lazy overlap” symbol shows that the one speaker (PAR) spoke while the the other speaker (INV) was speaking, but it doesn’t show at what point in the Investigator’s sentence the Participant started talking	
@n	neologism (e.g., sakov@n)	can be masked? ** have it connected not seperated	more neologism may mean fluent aphasia avg token → any position wala word prediction Threshold value	
<u>exclamations</u>	common ones: ah, aw, haha, ow, oy, sh, ugh, uhoh	standard ways ? will have to check from model Exclamation preprocessing literature review	can replace it with exclaim	
<u>interjections</u>	common ones: mhm, uhuh, hm, uhuh	standard ways ?	has a correlation with type of aphasia	
<u>fillers</u>	common ones: &-um, &-uh	standard ways ?	why is it separated with interjections?	

Aa Pattern Coding	≡ Meaning	≡ regex/steps to decode	≡ Comments	≡ Implemented
<u>underscores</u>	for combinations such as you_know, a_lot, how_come, kind_of, of_course, as_well, all_of_a_sudden, oh_my_gosh, as_well_as	replaced with space caution: actions/gestures contain underscores	REASON == argument is that this should be treated as one word	
<u>letters</u>	@l	it is a letter		
<u>letter sequence</u>	abcdefg@k	round of the closest meaningful word scrabble api, given rearranged word find the closest meaning full words use edit distance on an exhaustive dictionary	again two versions	Done
<u>xxx</u>	unintelligible speech, not treated as a word		M: 0 and 1 version N: this is what couldn't be clearly heard so can convert → avg token or <UNK> token	
<u>www</u>	untranscribed material (e.g., looking through pictures, talking with spouse), must be followed by %exp tier (see below)		<action> and removed Search : Aphasia towards visual actions	

Aa Pattern Coding	≡ Meaning	≡ regex/steps to decode	≡ Comments	≡ Implemented
<u>%exp:</u>	talking to the cameraman, lines not transcribed		www and corrected	
<u>&=text</u>	simple local event and gestures (e.g., &=laughs, &=sighs, &=ges:fishing, &=head: shake, &=ges:wave)	again these are just actions	treat it like exp [main thing here but can be grouped to exp and gpx]	Done
<u>&+text</u>	phonological fragment (&+sh &+w we came home)	<frag>	first letter/sound of the following word two versions: removed and not removed Limitation of LMs	
<u>%gpx:</u>	gestural and proxemic activity that extends through the whole sentence	removed three instances	statistically insignificant / can put it into beside a sentence and treat it like exp	
<u>[: text]</u>	target/intended word for errors (e.g., tried [: cried])		correct word make 2 forms - incorrect wala and correct form	Done
<u>[*]</u>	error (e.g., paraphasia -- w&k@u [: wet] [*])	right word	page number 18	

Aa Pattern Coding	≡ Meaning	≡ regex/steps to decode	≡ Comments	≡ Implemented
[/]	retracing without correction (e.g., simple repetition) put repeated items between <> unless only one word was repeated	removed	might use it as a tag having some meaning 0 and 1 versions for prediction use 1 version be aware <> is reserved already	
[//]	retracing with correction (e.g., simple word or grammar change) put changed items between <> unless only one word was changed	removed	same as above	
<u>word [x N]</u>	word repeated N times	pending	for example : red [x 4] → red red red red → red <red red red>	
(.)	pause between words	@pause@	what is the timecode correlation answer → no specific relation	Done
(.)	long pause between words	@pauser@	"	Done
(. . .)	extra long pause between words	@pausest@	"	Done
(2:13.12)	pause of 2 minutes and 13.12 seconds	not present		
()	shortenings -- e.g., runnin(g) for running, (be)cause for because	connected the word	even mor ignores the parantheses	Done

Aa Pattern Coding	≡ Meaning	≡ regex/steps to decode	≡ Comments	≡ Implemented
<u><4,5-digit-numbers></u> <u><4-5-digit-numbers></u>	Timecodes	subtracting and get time duration also for pauses	- subtract for each word and keep them - in case of no data → -1 [will use some distribution/proportion]	
<u>mor_gra</u>	tags	send them parallelly with the sentence and keep in mind that this needs to be removed along with its corresponding word in the original sentence.		
<u>INV</u>	Interviewer's speech	for replacing with avg token since it has context		