

Symbol	Description	Source	Updates due to ELAN compatibility, python code and otter use
<S001> <S002> <S003>	Begin each turn with speaker code No line space after each speaker	CorCenCC	Use speaker tagging function in Otter; Include <> as part of speaker tags
<S?>	where it's not possible to identify the speaker (NB not assigning gender)		Unnecessary (so far) with video
<SS>	when a number of speakers talk at once		
<b>punctuation</b>			
	remove all except to indicate questions		Python removing punctuation and capitals after full stops
<b>Overlaps and interruptions</b>			
=x  x+ =x +x	when one speaker interrupts another use = at the beginning of the interruption  Where the interrupted speaker carries on, use + at end of the interrupted turn and beginning of continued turn  <\$001> I think I would like+ <\$002> =Right <\$001> +to teach	CANCODE	<b>As col 1</b>  Allows us to look for interruptions, overlaps, backchannels and continued turns
<b>unfinished speech (false starts)</b>			
x=	= incomplete word at the end of the incomplete word	CANCODE	
<b>Interruption</b>			

/=x	to indicate an interruption at the start of the turn, closed up to the first word	CANCODE	See above for overlaps R
<b>repetition</b>			
	Adding in orthographically but not tagging	CorCenCC	Adding in orthographically but not tagged
<b>contextual info</b> transcriber's notes			
[]	[] for non verbal vocalisations [laugh] [cough] [breath out] [breath in] @tech noise/@ [any other info]	SO = screen orientation E= Extra	ELAN does not recognise angled brackets <>, so square brackets are used for NV vocalisations e.g. [tech noise] [signal loss] [whisper counting]  We will create new tier for tech noise once text transcription brought into ELAN
<b>guesses</b>			
	<G> one more </G> with guessed words inserted !unclear/!	CANCODE	See note above re angled brackets – guesses now within exclamation mark !guess/! And unclear marked same with square brackets ![unclear]/!
<b>unintelligible</b>			
!	![unclear]/! <unclear> for one word <unclear2> etc for more <unclear?> if uncertain how many words	CorCenCC	Unclear now marked as above with square brackets: ![unclear]/!

<b>pauses</b>			
... [pause]	For pauses of any length? ... for discernible short pause [pause] for silence of 3 to 10 seconds [silence] for prolonged silences of longer than 10 seconds		[pause] for all pauses
Screen sharing			[screen share start] [screen share stop] [slide change] [screen change] [screen scroll]
<b>Fillers vocalisation</b>		BNC – Love (2020, p.119)	
	<ol style="list-style-type: none"> <li>1. Short filler um uh am? (depending on following sound)</li> <li>2. Longer vowel filler um: the: me: or no lengthening?</li> <li>3. Mm (backchannel)</li> </ol> <p>BNC: •er = a long or short “er” or “uh” vowel, as in “bird”; there may or may not be an “R” sound at the end; usually=uncertainty;  •erm=as for “er” but ends as a nasal sound;  ah = has the vowel found in “father”</p>		<p><b>These are representations for all:</b></p> <p><b>Fillers</b> uh (hesitation) um (hesitation)</p> <p><b>Responses</b> oh (surprise/upset) mm (any length of backchannel) uh? (uncertainty request repetition)</p>

	<p>or a similar vowel; usually=realisation, frustration or pain;</p> <ul style="list-style-type: none"> <li>•oh=has the vowel found in “road” or a similar vowel; usually=mild surprise or upset;</li> <li>•eh=has the vowel in “bed” or the vowel in “made” or something similar, without an “R” or “M” sound at the end; usually=uncer-tainty, or “please say again?”;</li> <li>•mm=has a nasally “M” or “N” sound from start to end; usually=agreement;</li> <li>•huh=like an “er” but with a clear “H” sound at the start; usually=surprise;</li> <li>•uhu=two shortened “uh” or “er”-type vowels with an “H” sound between them, usually=disagreement; OR, a sound like the word “ahah!”; usually=success or realisation;</li> </ul>		
<b>capitalisation</b>			
	<p>Names of people, places, subjects, departments, institutions, courses, films, books, etc. should be typed with initial capitals. Check anonymisation parameters below.</p>	CorCenCC	

<b>anonymisation</b>			
<anon>	<anon> firstname1 </anon> <anon> surname1 </anon> <anon> institution1 </anon> Numbering for each recording begins with 1 and include list of names with each transcription or retain running list of names for entire corpus / all recordings?		New coding system added to data log document – simplified to e.g. [anon_FN1] [anon_\$003]