

Annotation scheme for pauses and hesitations

(18 Sept 2008)

	Type of coding phenomena	TEI-based XML coding	'readable' Version/ Example	referring to section in TEI P5 guidelines
A	TEXT STRUCTURE			
A1	sections which have been annotated (divisions)	<div start="1:31" end="1:59" > [annotated section] </div>	n/a	n/a
A2	speakers/ utterances (main speaker: u1, embedded speaker: u2)	<u1 who="S2"> [text] </u1>	S2:	8.3.1
	OVERLAP CODING			8.4.2 synchronisation & overlap 16.5 Synchronisation
A3a	MAIN SPEAKER <u1>	<anchor synch="[TS]"/> TEXT </anchor>	Colour coding, e.g.:	
A3b	embedded speaker <u2>	<u2 who="SX"> <o synch "TS"> TEXT </o> </u2>	S2: there's no point in writing every quarter [S1: yeah] still haven't got it if you know what I mean.	
B	PAUSE PHENOMENA timed			
B1	silent pause	<pause type="silent" dur="0.3"/>	<0.3>	8.3.2
B2	filled pause/vocal (er, em, sigh, laughter etc.)	<vocal dur="0.3"> em </vocal>	<em 0.3>	8.3.3 semi-lexical phenomena
B3	syllable stretch	them <vocal type="stretch" dur="0.7"/>	them <::0.7>	n/a
C	FURTHER HESITATION PHENOMENA not timed			
C1	false start/restart/repetition	so <del type="falseREstart"> I </vocal> I got	so <rep> I </rep> I got	8.4.5 based on TEI (deletion)
D	MWU			
D1	MWU candidate	<mwu> I think </mwu>	I think [bold]	n/a

Part of the project on

Second language speech fluency and the role of pauses in automatically extracted MWUs

(<https://www.nottingham.ac.uk/english/research/cral/doku.php?id=projects:slsf>)