This file describes the coding of the TextGrid segments of the Kiel-Corpus of read speech, spontaneous speech, and the Lindenstrassen-Corpus. This text uses the symbols < and > to enclose literal strings from the TextGrid- and s1h-files.

Information about the TextGrid files:
Each TextGrid file has four interval-tiers:

Tier-nr.	Name	Description
1	original	Segment labels as they appear in the ".s1h" files after the "hend"-mark
2	segment	Segment labels without certain meta-information (see below)
3	realized	Labels of actually realized segments, incl. stress and function word information
4	basic	Like tier 3, but without stress and function-word markers

All tiers have the same number of segments and a segment number refers in all four tiers to the same stretch of time in the signal (e.g., the 5th segment in tier 1 is the 5th segment in tiers 2, 3, and 4 as well and belongs to the same signal part; only the labels in the tiers differ and can be 'empty').

Coding of meta-information and 'empty' segments:
Praat does not allow segments with the length '0 samples' and segments on one tier cannot overlap, i.e., the segments must form a strictly consecutive 'chain' (but segment names can be 'empty', i.e. have no labels). The Kiel-Corpus on the other hand does contain certain meta-information like e.g. 'beginning of sentence' (<c:>), 'function word' (<+>), and information about deleted (in relation to a canonical transcription) segments (e.g. <d->) which all have a length of '0' samples. To be able to code this information into one tier (instead of using separate tiers e.g. for "function word", "sentence", etc.) these labels with the length '0' are converted to receive the length '1 sample' and this sample is subtracted from the next (real) interval associated with a segment. This procedure encodes all information from the Kiel-Corpus and it leaves the numbering inside the four tiers of one TextGrid the same by introducing a small length-reduction (62.5μs for 16 kHz and 90.9μs for 11 kHz recordings per sample) for the next segment. (Sometimes there are more than one 1-sample segments introduced in front of a real segment; note that even 10 such 1-segment segments are together shorter than 1 ms. Praat allows segment length shorter than the length of 1 sample, but the script to generate the segments did not use this option.)

Description of the tiers:

Tier 1, original:
This tier shows the original labels from the Kiel ".s1h" files.
These are the same labels as in the ".s1h" text-files beyond the <hend> mark. The only difference is that the symbols for single and double quotes (<'> and <">), indicating primary and secondary stress or focus (<">), have been replaced with <\'1> and <\'1\'1> respectively (single and double-quotes are meta-symbols in Praat and cannot be used as parts of labels; <\'1> and <\'1\'1> appear as single and double quotes within Praat's Edit window).

Tier 2, segment:
Essentially, this tier shows the canonical transcription and the realized sound, i.e. a label like <n-m> indicates that an [n] has been realized as [m]; or <n-> expresses that an [n] has been deleted, whereas <-n> shows that an [n] has been inserted in relation to the canonical transcription.
Technically, tier 2 copies the labels from tier 1, but the following symbols have been **removed**:

##	beginning of word
$#	compound boundary
$	word-internal segment
#…	interverbal sounds, pauses or punctuation
.,;?!	punctuation
MA	'obsolete mark'
%	unsure sign
Q-	deleted glottal stop
-q	glottalization
[cghlnpqrsvwz]:	nonverbal marks (e.g. hesitation, cough, indicators etc.)
:k	nonverbal sign (click, external noise)
~	nasal mark
=/+	false starts, also in the variants =/-, /+, /-, etc. (transcription errors?)
_	(underline) edge of intraword interruption
+	function word marker
-hp	transcription error?
=6	transcription error?
ma	transcription error?

Tier 3, realized:
These are the labels of the realized (i.e. transcribed) segments. Additionally, (i) the function word marker (<+>) and hesitation marker (<v:>), which exist in tier 1 (the original transcription) only on the last (or first) sound of a word have been extended throughout the word in front of each segment. The symbol <+> is used to indicate segments that are part of function words and <§> mark segments that are part of a hesitations. (ii) The VOT, which is transcribed as a <-h> after a plosive has been recoded as the plosive symbol with an attached <h> (i.e., a symbol sequence like <d -h a n+> in tier 2 becomes <+d +dh +a +n> in tier 3. (iii) Furthermore, the focus-mark (two single quotes after a dollar-sign in tier 1: <$'>) has been transformed to a secondary stress-mark on the next vowel (i.e. <$'' O> in tier 1 becomes <''O> in tier 3).

Tier 4, basic:
This tier has essentially only the segment labels as in tier 3, but without any stress, hesitation, and function-word marks.

Example:
Beginning of “g071a000.s1h” with the Text “Ja, guten Tag, dann fange ich einfach mal an” realized as [j'aː g'uːn th'ax dh@n f'aŋ ɪç 'aɪnf"ax ma 'an]. All quotes < ' > (single and double) are coded as < '\1 > in the TextGrid-files, but are written here as < ' >, as they appear in Praat.

Sample # in “.s1h”	Sample # in TextGrid	Time (sec) in TextGrid	Tier 1 original	Tier 2 segment	Tier 3 realized	Tier 4 basic
5935	5935	0.371	#c:			
5935	5936	0.371	#-s:			
6265	6265	0.392	#-h:			
12695	12695	0.793	##j	j	j	j
14161	14161	0.885	$'a:	'a:	'a:	a:
15406	15406	0.963	#,			
15406	15407	0.963	##g	g	g	g
16222	16222	1.014	$'u:	'u:	'u:	u:
16728	16728	1.046	$t-n	t-n	n	n
18048	18048	1.128	$@-	@-		
18048	18049	1.128	$n-	n-		
18048	18050	1.128	##t	t	t	t
18282	18282	1.143	$-h	-h	th	th
18682	18682	1.168	$'a:-'a	'a:-'a	'a	a
19646	19646	1.228	$k-x	k-x	x	x
20264	20264	1.267	#.			
20264	20265	1.267	#c:			
20264	20266	1.267	##d	d	+d	d
21567	21567	1.348	$-h	-h	+dh	dh
21823	21823	1.364	$a-@	a-@	+@	@
22292	22292	1.393	$n+	n	+n	n
23549	23549	1.472	##f	f	f	f
25393	25393	1.587	$'a	'a	'a	a
26256	26256	1.641	$N	N	N	N

Sample # in ".s1h"	Sample # in TextGrid	Time (sec) in TextGrid	Tier 1 original	Tier 2 segment	Tier 3 realized	Tier 4 basic
27308	27308	1.707	$@-	@-		
27308	27309	1.707	##Q-	Q-		
27308	27310	1.707	$I	I	+I	I
27715	27715	1.732	$C+	C	+C	C
28310	28310	1.769	##Q-	Q-		
28310	28311	1.769	$-q			
28310	28312	1.770	$'aI	'aI	'aI	aI
30071	30071	1.879	$n	n	n	n
30870	30870	1.929	$#f	f	f	f
31768	31768	1.986	$"a	"a	"a	a
32228	32228	2.014	$x	x	x	x
32840	32840	2.053	##m	m	+m	m
34090	34090	2.131	$a:	a:	+a:	a:
35508	35508	2.219	$l-+	l-		
35508	35509	2.219	##Q-	Q-		
35508	35510	2.219	$-q			
35508	35511	2.219	$'a	'a	'a	a
37298	37298	2.331	$n	n	n	n