

## Unicode Characters for Discourse Transcription (Selected)

For transcribing spoken discourse at higher delicacy/granularity levels involving intonation symbols (in the Discourse Transcription system, DT2), it may be preferable to use a Unicode font such as Arial Unicode MS. Unicode is an explicit system for representing characters which can help to avoid incompatibilities in transferring transcription data from one computer format or platform to another, and in using special characters in general. Tables 1 and 2 below present Unicode information for various characters used in transcription, some of which are not present in a standard keyboard layout. The Unicode numbers are given in hexadecimal notation. In addition, for characters which are included in the ANSI or ASCII standard, the decimal equivalent is given. For characters which are used with the same or similar meaning in the International Phonetic Alphabet (IPA), the standard IPA number is given.

To access Unicode characters in, for example, Microsoft Word, use the “Insert/Symbol” function, or type the 4-character Unicode number followed by Alt-X. In OpenOffice.org, from the menu select “Insert/Special Character”.

**Table 1. Standard ASCII & ANSI characters**

Sign	Function (DT2)	Unicode Name	Uni- code # (hex)	ANSI -ASCII # (dec)	IP A #	Unicode Subset	
:	lag/lengthening	COLON	003A	58		Basic Latin	1.
;	speaker/agent attribution	SEMI-COLON	003B	59		Basic Latin	2.
@	laugh pulse	COMMERCIAL AT	0040	64		Basic Latin	3.
%	glottal	PERCENT SIGN	0025	37		Basic Latin	4.
!	booster	EXCLAMATION MARK	0021	33		Basic Latin	5.
ı	low booster	INVERTED EXCLAMATION MARK	00A1	161		Latin-1	6.
¿	appeal onset/plateau	INVERTED QUESTION MARK	00BF	191		Latin-1	7.
°	piano	DEGREE SIGN	00B0	176		Latin-1	8.
¤	whisper	CURRENCY SIGN	00A4	164		Latin-1	9.
±	non-interactive source	PLUS-MINUS SIGN	00B1	177		Latin-1	10.
÷	chant/stylization	DIVISION SIGN	00F7	247		Latin-1	11.
*	prominence/stress	ASTERISK	002A	42	612	Basic Latin	12.
^	primary accent	CIRCUMFLEX ACCENT	005E	94	597	Basic Latin	13.
`	secondary accent	GRAVE ACCENT	0060	96	515	Basic Latin	14.
©	co-action prefix	COPYRIGHT SIGN	00A9	169		Latin-1	15.
<	markup label start	LESS-THAN SIGN	003C	60		Basic Latin	16.
>	markup label end	GREATER-THAN SIGN	003E	62		Basic Latin	17.

*Du Bois: Unicode Characters*

<b>Sign</b>	<b>Function (DT2)</b>	<b>Unicode Name</b>	<b>Uni- code # (hex)</b>	<b>ANSI -ASCII # (dec)</b>	<b>IP A #</b>	<b>Unicode Subset</b>	
/	head onset (intonation)	SOLIDUS	002F	47	903	Basic Latin	18.
\	tail (intonation)	REVERSE SOLIDUS	005C	92		Basic Latin	19.
[	overlap start	LEFT SQUARE BRACKET	005B	91		Basic Latin	20.
]	overlap end	RIGHT SQUARE BRACKET	005D	93		Basic Latin	21.
{	phonetics start	LEFT CURLY BRACKET	007B	123		Basic Latin	22.
}	phonetics end	RIGHT CURLY BRACKET	007D	125		Basic Latin	23.
«	vox start (alternate)	LEFT-POINTING DOUBLE ANGLE QUOTATION MARK	00AB	171		Latin-1	24.
»	vox end (alternate)	RIGHT-POINTING DOUBLE ANGLE QUOTATION MARK	00BB	187		Latin-1	25.
#	unintelligible	NUMBER SIGN	0023	35		Basic Latin	26.
×	false start, reparandum	MULTIPLICATION SIGN	00D7	215		Latin-1	27.
~	pseudograph	TILDE	007E	126		Basic Latin	28.
¬	constructed example	NOT SIGN	00AC	172		Latin-1	29.
\$	comment line	DOLLAR SIGN	0024	36		Basic Latin	30.
§	prosodic sentence	SECTION SIGN	00A7	167		Latin-1	31.
¶	paratone	PILCROW SIGN [= PARAGRAPH SIGN]	00B6	182		Latin-1	32.
	accent unit	VERTICAL LINE	007C	124	507	Basic Latin	33.
⋮	accent unit (alternate)	BROKEN BAR	00A6	166		Latin-1	34.
+	clitic boundary	PLUS SIGN	002B	43		Basic Latin	35.
—	linking [cf. IPA 509]	LOW LINE	005F	95		Basic Latin	36.
-	morpheme boundary	HYPHEN-MINUS	002D	45		Basic Latin	37.
-	word truncation	EN DASH	2013	150		General Punctuation	38.
—	intonation unit truncation	EM DASH	2014	151		General Punctuation	39.
†	reset, minor (alternate)	DAGGER	2020	134		General Punctuation	40.
‡	reset, major (alternate)	DOUBLE DAGGER	2021	135		General Punctuation	41.

**Table 2. Non-ANSI characters**

Sign	Function (DT2)	Unicode Name	Uni- code # (hex)	IPA #	Unicode Subset	
Ɱ	top	MODIFIER LETTER EXTRA-HIGH TONE BAR	02E5	519	Spacing Modifier Letters	42.
Ɐ	high	MODIFIER LETTER HIGH TONE BAR	02E6	520	Spacing Modifier Letters	43.
Ɒ	mid	MODIFIER LETTER MID TONE BAR	02E7	521	Spacing Modifier Letters	44.
ⱱ	low	MODIFIER LETTER LOW TONE BAR	02E8	522	Spacing Modifier Letters	45.
Ⱳ	bottom	MODIFIER LETTER EXTRA-LOW TONE BAR	02E9	523	Spacing Modifier Letters	46.
↗	high [IPA = 'upstep']	UPWARDS ARROW	2191	518	Arrows	47.
↘	low [IPA = 'downstep']	DOWNWARDS ARROW	2193	517	Arrows	48.
↕	top	UPWARDS TWO HEADED ARROW	219F		Arrows	49.
↕	bottom	DOWNWARDS TWO HEADED ARROW	21A1		Arrows	50.
↗	downstepped high	DOWNWARDS ARROW FROM BAR	21A7		Arrows	51.
↘	upstepped low	UPWARDS ARROW FROM BAR	21A5		Arrows	52.
↵	pitch reset, minor	ASSERTS	22A6		Mathematical Operators	53.
↵	pitch reset, major	MODELS	22A9		Mathematical Operators	54.
ⴞ	amplitude reset, minor	DOWN TACK	22A4		Mathematical Operators	55.
ⴞ	amplitude reset, major	INTERCALATE	22BA		Mathematical Operators	56.
↗	upkey	NORTH EAST DOUBLE ARROW	21D7		Arrows	57.
↘	downkey	SOUTH EAST DOUBLE ARROW	21D8		Arrows	58.
⇒	midkey	RIGHTWARDS DOUBLE ARROW	21D2		Arrows	59.
↗	high-key	UPWARDS DOUBLE ARROW	21D1		Arrows	60.
↘	low-key	DOWNWARDS DOUBLE ARROW	21D3		Arrows	61.
↗	rise [IPA = 'global rise']	NORTH EAST ARROW	2197	510	Arrows	62.
↘	fall [IPA = 'global fall']	SOUTH EAST ARROW	2198	511	Arrows	63.
→	level	RIGHTWARDS ARROW	2192		Arrows	64.
→	truncated level	RIGHTWARDS ARROW WITH STROKE	219B		Arrows	65.
↗↘	rise-fall [sequence]	[combination: rise + fall]	[comb.]		Arrows	66.
↘↗	fall-rise [sequence]	[combination: fall + rise]	[comb.]		Arrows	67.
↻	rise-fall [alternate]	CLOCKWISE TOP SEMICIRCLE ARROW	21B7		Arrows	68.
↻	fall-rise [alternate]	BOTTOM ARC ANTICLOCKWISE ARROW	293B		Arrows	69.
ˆ	rise	MODIFIER LETTER ACUTE ACCENT	02CA	513	Spacing Modifier Letters	70.
˘	fall	MODIFIER LETTER GRAVE ACCENT	02CB	515	Spacing Modifier Letters	71.

Sign	Function (DT2)	Unicode Name	Uni- code # (hex)	IPA #	Unicode Subset	
-	level	MODIFIER LETTER MACRON	02C9	514	Spacing Modifier Letters	72.
^	rise-fall	MODIFIER LETTER CIRCUMFLEX ACCENT	02C6	525	Spacing Modifier Letters	73.
ˇ	fall-rise	CARON	02C7	524	Spacing Modifier Letters	74.
ˈ	primary stress	MODIFIER LETTER VERTICAL LINE	02C8	501	Spacing Modifier Letters	75.
ˌ	secondary stress	MODIFIER LETTER LOW VERTICAL LINE	02CC	502	Spacing Modifier Letters	76.
	intonation unit	DOUBLE VERTICAL LINE	2016	508	General Punctuation	77.
..	hold, micro-pause	TWO DOT LEADER	2025		General Punctuation	78.
...	pause, untimed	HORIZONTAL ELLIPSIS	2026		General Punctuation	79.
aː	phonemic long [IPA=half-long]	MODIFIER LETTER HALF TRIANGULAR COLON	02D1	504	Spacing Modifier Letters	80.
a:	long (alternate) [IPA = long]	MODIFIER LETTER TRIANGULAR COLON	02D0	503	Spacing Modifier Letters	81.
͡	linking (alternate, IPA)	UNDERTIE	203F	509	General Punctuation	82.
h̥	in-breath	DOUBLE-STRUCK CAPITAL H	210D		Letterlike Symbols	83.
h̄	exhalation	SCRIPT CAPITAL H	210B		Letterlike Symbols	84.
☺	smile voice with speech	WHITE SMILING FACE	263A		Miscellaneous Symbols	85.
⊗	co-action locus, general	CIRCLED LATIN CAPITAL LETTER X	24CD		Enclosed Alphanumerics	86.
⊙	co-action gesture	CIRCLED LATIN CAPITAL LETTER G	24BC		Enclosed Alphanumerics	87.
Ⓟ	co-action beat	CIRCLED LATIN CAPITAL LETTER B	24B7		Enclosed Alphanumerics	88.
①	co-action type 1	CIRCLED DIGIT ONE	2460		Enclosed Alphanumerics	89.
②	co-action type 2	CIRCLED DIGIT TWO	2461		Enclosed Alphanumerics	90.
③	co-action type 3	CIRCLED DIGIT THREE	2462		Enclosed Alphanumerics	91.
④	co-action type 4	CIRCLED DIGIT FOUR	2463		Enclosed Alphanumerics	92.
┌	gaze start	LEFT CEILING	2308		Miscellaneous Technical	93.
┐	gaze end	RIGHT CEILING	2309		Miscellaneous Technical	94.
└	co-action start	LEFT FLOOR	230A		Miscellaneous Technical	95.
┘	co-action end	RIGHT FLOOR	230B		Miscellaneous Technical	96.
◐	gaze arrive	CIRCLE WITH LEFT HALF BLACK	25D0		Geometric Shapes	97.
◑	gaze depart	CIRCLE WITH RIGHT HALF BLACK	25D1		Geometric Shapes	98.
◕	gaze 1	FISHEYE	25C9		Geometric Shapes	99.
◔	gaze 2	BULLSEYE	25CE		Geometric Shapes	

[rev. 6-Jun-2006]