Filename: Bravo-File-Format.press

This memo describes the file format for documents produced by Bravo (**not** **BravoX). It is applicable to Bravo versions 6.0 through 8.6, although much of the discussion applies to other versions of Bravo as well.**

INTRODUCTION

From time to time, someone will ask about documentation concerning the file format generated by Bravo. As far as I know (I haven't researched the subject), two principal papers exist, and both are difficult to find. They were written several years ago by Greg Kusnick, a member of the Bravo programming team headed by Charles Simonyi.

At the time, Greg was addressing a single recipient, Dan DeSantis. Dan & I were developing a program to translate Bravo documents into a form accepted by the ill-fated Diamond editor. Greg's first paper, written about September 1976, described the essentials of Bravo's file format. The second paper, written in October 1976, concentrated on the format for tab settings.

This memo is largely a re-hash of Kusnick's work, with a few additions bringing it up to date. For the most part the additions result from deductions I made after experimenting with Bravo 7.5. Not being a Bravo expert, I may have overlooked or misunderstood some things. If such matters are brought to my attention, I'll be happy to revise the memo.

Bill Maybury May 30, 1981

OVERVIEW

Each Bravo document is stored as a *single* textual file. There are also support files, for fonts and user profile (in User.Cm), but these are considered independent of the document. Note that the same document may have significantly different renderings if different support files are used. (E.g. Greek letters can become arrowheads if the user profile is altered a little...or plain English can become Greek gibberish.)

Bravo files have no End-of-File marker; Bravo relies on the file system for length information. The maximum length for a Bravo file is 65,536 bytes. Actually, a limit of 60K is more realistic; beyond that Bravo is likely to complain. A length of zero is okay.

Bravo is quite willing to accept a file containing simple text, i.e. no control information. Such files represent "vanilla" documents having *no* paragraphs (in the Bravo sense). At present, Laurel and most Mesa Utilities produce vanilla documents. When rendered by Bravo, these documents have default formatting, subject to User.Cm.

Except for vanilla documents, Bravo files contain a sequence of paragraphs, where each paragraph consists of its text, if any, followed by a "trailer" specifying control information. The trailer begins on a special character, control-z (^z), optionally followed by formatting data. The trailer terminates on a carriage return (CR).

A trailer's formatting data indicates two things: how the paragraph as a whole is to look and how its runs of characters are to look. The first part is sometimes called the "par-look" section while the second is called the "char-look" section. The char-look section begins on a special character, the backslash (\). Both sections are optional.

The following BNF description summarizes the overall makeup of a Bravo document file:

BravoDoc	::=	VanillaDoc **|** SeqOfPars	
VanillaDoc	::=	{**character**}^{0}	*-- i.e. zero or more ASCII characters*
SeqOfPars	::=	{Paragraph}^{1}	*-- i.e. one or more Paragraphs*
Paragraph	::=	{**character**}^{0} Trailer	
Trailer	::=	**^z** ParLooks^{0} CharLooks^{0} **CR**	*-- i.e. ParLooks & CharLooks are optional*
ParLooks	::=	{parLook}^{1}	*-- par-look details are discussed next*
CharLooks	::=	\ {charLook}^{1}	*-- char-look details are discussed later*

PARAGRAPH LOOKS (PAR-LOOKS)

Except for tab-settings, the presence of a given par-look strictly depends on whether the corresponding property differs from the norm or not. Thus, the "z" (right-margin) par-look doesn't occur for a paragraph having the default right margin,. And, the "d" (line1-margin) par-look doesn't occur for a paragraph whose first line has the same left margin as the rest of its lines. But, the "c" (centered) par-look occurs for *every* paragraph which is centered.

Tab-settings propagate (unless overridden by new tab-settings). That is, if a given tab-setting is specified for paragraph n, the setting remains in effect for paragraphs n+1, n+2, etc. Tab-setting par-looks are always parenthesized and appear after any other par-looks.

The other par-looks begin on a small letter. The letter alone is sufficient for the Boolean par-looks (doc-profile, justified, centered, hardcopy). The others have a CARDINAL value after their letter. In the case of margins, the value is in mica units; otherwise, it's in point units. (There are 2540 micas or 72 points per inch.) The table below covers all par-looks which may precede a tab-setting (if any). They are given in Bravo's preferred order.

Table of par-looks (except for tab-settings):

par-look	letter	units	remarks
right-margin	z	micas	default = 527 pts unless overridden by User.Cm
left-margin	l	micas	default = 85 pts unless overridden by User.Cm
line1-margin	d	micas	1st line's left margin differs from that of other lines
vertical-tab	y	pts	default = -1 (i.e. *no* vertical tabbing)
line-leading	x	pts	default = 1 pt unless overridden by User.Cm
par-leading	e	pts	default = 0 pts unless overridden by User.Cm
doc-profile	q	----	
justified	j	----	
centered	c	----	
hardcopy	w	----	
keep	k	pts	default = 0 pts unless overridden by User.Cm

Bravo understands two types of tabs: named & plain. Consequently, there are two forms of par-looks for tab-settings: tab-stop & tab-interval. The two are mutually exclusive.

> However, if extra tabs occur in a paragraph having tab-stops, a default interval is automatically applied by Bravo. This default may be specified in User.Cm; if it isn't, Bravo assumes a 60 pt interval.

The tab-interval par-look is simply a parenthesized CARDINAL value:

(interval)

where **interval** is in mica units. This par-look is generated when the user executes Bravo's "Look TAB =" command. It also results by default when applicable. For example, the tab-interval par-look is almost always present in a document's initial paragraph (not if that paragraph uses tab-stops, of course). The **interval** effectively partitions paragraph lines into equal-sized units starting at the paragraph's left margin. Consider the example,

(2540)

which specifies a tab-interval of one inch. If the paragraph's left margin is at 85 pts, there will be effective tab stops every inch (72 pts) to the right, i.e. at:

157 pts, 229 pts, 301 pts, 373 pts, etc.

A tab-stop par-look consists of a parenthesized pair of CARDINAL values separated by a comma; no blank follows the comma:

(name,tabStop)

TabStop is in mica units. The **name** is a value in the range [0..13], corresponding to *n* IN {1, 2, 3, 4, 5, 6, 7, 8, 9, a, b, c, d, e}, respectively, for "Look TAB *n*" commands. There may be a sequence of tab-stops; in which case they occur in **name** order. Consider this example:

(0,5080)(1,12700)(10,7620)

Tab-1 is set 2 inches from the left edge of the *page's* (not paragraph's) left margin; tab-2 is set at 5 inches, and tab-b is set at 3 inches, i.e. between tabs 1 & 2. Bravo accepts useless settings, by the way, i.e. tab-stops outside the paragraph's margins.

> Tab-stops can be cleared as well as set. The special **tabStop**, 65535, denotes this. For example, the par-look (0,65536) clears tab-1.

CHARACTER LOOKS (CHAR-LOOKS)

Char-looks describe the properties of a run of text in a paragraph. Here's an oversimplified rule:

> *Char-looks contain a property-change description followed by a run-length.*

The property-change description tells how properties of the new run differ from those of the preceding run (if any) in the paragraph. (If there is no preceding run in the paragraph, the description tells how properties differ from defaults.) The run-length tells how many characters occur in the new run.

So much for oversimplification; there are two exceptions to the above rule:

> A paragraph's *first* char-look may or may not have a property-change description. If only the run-length is present, Bravo applies default properties to the run. A paragraph's *last* char-look (possibly its only one) has no run-length. Bravo applies the property-change description to all remaining characters in the paragraph.

One obvious technicality should be mentioned: If a paragaraph has *no* char-looks, all of its characters have default properties.

Property-change descriptions consist of one or more "looks" (character-property specifications). Looks have two basic forms. The most common form is merely a letter, e.g. i for italic, b for bold, and u for underlined. Note that these are small letters; their capitals reverse the property.

The other basic form consists of a small letter immediately followed by a CARDINAL. (A blank is appended for separation in cases where run-length follows the look.) This form of look is used for specifying changes in font-number, offset, color, or tab name. For example, switching to font 1 results in the look f1.

The following table covers all character looks. It is given in what seems to be Bravo's preferred order. (The ordering of looks in a char-look is probably unimportant, however.)

Table of character looks:

look	letter	default	remarks
font-number	f	0	values IN [0..9]
offset	o	0	values IN [0..255], discussed next
tab-or-color	t	0	values IN [0..14], discussed later
underlined	u	FALSE	(U means non-underlined)
bold	b	FALSE	(B means non-bold)
italic	i	FALSE	(I means non-italic)
graphic	g	FALSE	(G means non-graphic)
visible	v	FALSE	(V means non-visible)
overstrike	s	FALSE	(S means non-overstrike) the look is no longer in use
vanished	n	FALSE	(N means non-vanished) the look is no longer in use

Offset values are interpreted as 8-bit INTEGERs. For example, the look o249 represents a negative offset of 7 points.

The tab-or-color look has a dual use, representing color or a named tab. Color is simple:

t0 for Black (used to revert back to uncolored text)
t1 for Cyan
t2 for Green
t3 for Magenta
t4 for Red
t5 for Violet
t6 for Yellow

(Regarding highlighting: If a character is "underlined" and also has a color other than Black, Bravo assumes the character is to be drawn non-underlined but on a colored background.)

Analysis is needed to distinguish named-tab looks from color looks. (Note that t1 may mean "named-tab-1" or "Cyan color.") For TAB characters, named-tab is the reasonable choice. For other characters, color is the reasonable choice...if they have looks t0, t1, t2, t3, t4, t5, or t6. If they have looks t7, t8, t9, t10, t11, t12, t13, or t14, the characters are named "tabs." (Peculiarly enough, Bravo allows *any* character to have the named-tab property.)

Except for t0, named-tab looks correspond in straightforward fashion to the tabs they name:

look	tab	look	tab	look	tab	look	tab
		t4	4	t8	8	t12	c
t1	1	t5	5	t9	9	t13	d
t2	2	t6	6	t10	a	t14	e
t3	3	t7	7	t11	b		

Named-tab look t0 un-names the next "tab." I.e. it causes the run to revert to un-named or plain tab property. Typically, a t0 look appears shortly after each named-tab look.

EXAMPLES

The five examples below contrast the normal appearance of a paragraph with its file-format appearance...as rendered by Bravo's Unformatted-Get command. (To improve readability, the paragraph trailer's control-z character has been replaced by ^z, and a CR has been inserted before the trailer's carriage-return.) The examples are fairly self-explanatory. The first one illustrates an "empty" italicized paragraph having no par-looks.

- -

^z\i CR

- -

*** CENTERED & BOLD ***

*** CENTERED & BOLD *** ^zc\b CR

- -

*** NAMED-TAB-1 at 4": , COLOR: Cyan ***

*** NAMED-TAB-1 at 4": , COLOR: Cyan *** ^z(0,10160)\22t1 1t0 8t1 5t0 CR

- -

*** *ITALICS* UNDERLINED FONT-1 *ALL-OF-THOSE* *MIX* & MATCH OFFSET 9 pts ***

*** ITALICS UNDERLINED FONT-1 ALL-OF-THOSE MIX & MATCH OFFSET 9 pts
*** ^z\4i7I3u10U3f1 6f0 3f1ui12f0UI3ui3I3f1 5f0U3o9 12o0 CR

- -

*** RIGHT-MARGIN 7" LEFT-MARGIN 2" LINE-1-MARGIN 1.5" LINE-LEADING 4 pts PARAGRAPH-LEADING 19 pts JUSTIFIED KEEP 8 pts TAB-INTERVAL 1"
TAB: TAB: TAB: TAB: ***

*** RIGHT-MARGIN 7" LEFT-MARGIN 2" LINE-1-MARGIN 1.5" LINE-LEADING 4 pts
PARAGRAPH-LEADING 19 pts JUSTIFIED KEEP 8 pts TAB-INTERVAL 1"
TAB: TAB: TAB: TAB: *** ^zz17780l5080d3810x4e19jk8(2540) CR