

Unicode Technical Note #37**DUPLOYAN SHORTHAND RENDERING MODEL**

Version	1
Authors	Van Anderson
Date	2014-07-01
This Version	http://www.unicode.org/notes/tn37/tn37-1.html
Previous Version	n/a
PDF Version	http://www.unicode.org/notes/tn37/utn37-1-duployan.pdf
Latest Version	http://www.unicode.org/notes/tn37/

Summary

The Duployan (historically "Duployéan") shorthands and Chinook script are used as a secondary shorthand for writing French, English, German, Spanish, Romanian, and as an alternate primary script for several first nations' languages of interior British Columbia, including the Chinook Jargon, Okanagan, Lilloet, Shushwap, and North Thompson. This Technical Note describes the rendering model and other details of script usage.

Many tables in this document depend on the availability of specialized glyphs, which are contained in the Duployan Proposal font. This font is embedded via the CSS stylesheet, but a PDF version of this note is also supplied for environments that don't support that. See also [Section 11, Resources](#) for a TrueType version of the font.

Status

*This document is a **Unicode Technical Note**. Sole responsibility for its contents rests with the author(s). Publication does not imply any endorsement by the Unicode Consortium. This document is not subject to the [Unicode Patent Policy](#).*

For information on Unicode Technical Notes including criteria for acceptance, see <http://www.unicode.org/notes/>.

Contents

1. [Historical Overview of the Duployan and adaptations](#)
2. [Typology](#)
3. [Script Structure](#)
4. [Collation](#)
5. [Input](#)
6. [Principles of the Duployan Shorthands and Chinook scripts](#)
7. [Confusability and Usage](#)
8. [Parsing of Shorthand Overlap Sequences](#)
9. [References](#)

10. Documentation

11. Resources

- Other References
- Modifications

1. Historical Overview of the Duployan and adaptations

The Duployan shorthands and Chinook script are used as a secondary shorthand for writing French, English, German, Spanish, Romanian, and as an alternate primary script for several first nations' languages of interior British Columbia, including the Chinook Jargon, Okanagan, Lillooet, Shushwap, and North Thompson. The original Duployan shorthand was invented by Emile Duployé, published in 1860, as a stenographic shorthand for French. It was one of the two most commonly used French shorthands, being more popular in the south of France, and adjacent French speaking areas of other countries. Adapted Duployan shorthands were also developed for English, German, Spanish, and Romanian. The basic inventory of consonant and vowel signs - all in the first two columns of the allocation - have been augmented over the years to provide more efficient shorthands for these languages and to adapt it to the phonologies of these languages and the languages using Chinook writing. There currently exists no encoding - PUA or otherwise - for the representation of the Duployan or Chinook. Indeed, the submission of the Duployan Shorthands and Chinook script to the Unicode Consortium has necessitated the creation, from scratch, of the first Duployan/Chinook font, and the allocation is based solely on the internal logic of the script and affinity of usage among characters.

The Chinook script was an adaptation and augmentation of the Duployan shorthand by fr. Jean Marie Raphael LeJeune, used for writing the Chinook Jargon and other languages of 19th c. interior British Columbia. Its original use and greatest surviving attestation is from the run of the *Kamloops Wawa*, a (mostly) Chinook Jargon newsletter of the Catholic diocese of Kamloops, British Columbia, published 1891-1923. At the time, the Chinook Jargon pidgin was widely spoken from SE Alaska to northern California, from the Pacific to the Rockies, and sporadically outside this area. Although the Chinook Jargon was the lingua franca in many communities of the Pacific Northwest, it was generally a spoken, rather than written language. Most attempts at documentation used the Latin script to approximate Jargon words with English or French phonology, and indeed, dictionaries of the Chinook Jargon are still readily available in these Latinate orthographies. In contrast, the archives of the *Kamloops Wawa*, written in Chinook, includes a considerable dictionary, but also constitutes an unparalleled 3+ decade corpus of Chinook Jargon usage during the height of its spread and utility. The Chinook Script makes use of the basic Duployan inventory, with the addition of several derived letterforms and compound letters.

In 1984, the "Students' Practical Encyclopedia" (*Enciclopedia practică a copiilor*) was published in Romania, containing the "Curs de Stenografie" by Margareta Sfințescu. This shorthand was an adaptation of the Duployan for Romanian, using a few of the Chinook and Duployan shorthand compound letters as basic letterforms, and several basic vowel forms with diacritics. It also makes use of a "doubling mark" to indicate a general duplication of a word or phonemic form.

The Pernin shorthand was first published by Helen M. Pernin as "Pernin's Universal Phonography" no later than 1882. There is an alternate version of the Pernin shorthand published as "Pernin's Practical Reporter", that has different affixes. The next year, John Mathew Sloan published the competing Sloan-Duployéan method, which was expanded in 1918, when Denis R. Perrault published the Perrault-Duployan system. All three of the above, being the main English adaptations of Duployan, enjoyed some popularity, but never attained the reach of Pitman or Gregg shorthands. All three systems share many characters with Chinook and each other. The most significant anomalies of these systems are the invariant vowel signs in Pernin, the quarter-circle combined consonants, found in each system but with differing values, the extensive use of vowel diacritics in Sloan, and heavy shading of letters - as the

voiced consonants in Pitman-based systems are - to indicate "r" flavored letters in Sloan.

Unsupported orthographies. Currently, materials are unavailable to attempt including Carl Brandt's English Duployan adaptation or George Galloway's extension of the Sloan-Duployéan in the current encoding. Similarly, documentation of the adaptations of Duployan to German and Spanish are unavailable, so complete support for these orthographies is probably not offered in the current allocation. Allocation space has been set aside to reasonably accommodate extensions for some of these extensions of the Duployan script.

2. Typology

Duployan is, at its core, an alphabetic (consonant & vowel) stenographic (simple line & curve) writing system (cf. Pitman shorthand, a stenographic abjad). It classifies under the geometric shorthands, in that the model letterforms are generally based on circles and lines (cf. Gregg's elliptical shorthand). In general, there is a visual and functional distinction between consonants, which are based either on lines or large semi-circles and have invariable orientation, ie consonants do not rotate to match with surrounding letters; and the vowels, which are generally based on circles, quarter arcs, and small semi-circles, and generally reshape and orient contextually. It is an LTR script, proceeding down the page in lines like most modern Western scripts, although individual letters may be written right-to-left.

3. Script Structure

The core repertoire of the Duployan writing contains several classes of letters, differentiated primarily by visual form and stroke direction, and nominally by phonetic value. Letter classes include the line consonants (P, T, F, K, & L-type) and arc consonants (M, N, J, & S-type), circle vowels (A, O, & W-vowels), nasal vowels, and orienting vowels (U/Eu, I/E). In addition, the Chinook writing contains spacing letters, compound consonants, and a logograph. The extended Duployan shorthand includes four other letter classes - the complex letters (multisyllabic symbols with consonant forms), and high, low, and connecting terminals for common word endings. The Romanian stenography, Pernin, Perrault, and Sloan orthographies add a few letters or letter forms, ideographs, and several combined letters. Most "core" letters have related variant forms, including the addition of ancillary dots and crosses, size variants, and the compounding of vowels.

Since the Duployan was originally developed as a shorthand system, strings of letters are joined together cursively into words in Duployan, Romanian, Pernin, Perrault, and Sloan, or nominally syllabic units in Chinook - usually with a single circle vowel for each unit. The original Duployan and its offshoots all encourage overlapping for initialisms and abbreviations and many prescribe overlaps and raised or lowered text height for some morphemes or phonemes.

4. Collation

Information on collation of Duployan scripts is generally ambiguous and arbitrary. Many dictionaries and primers simply cite in that language's Latin alphabetical order with no attempt made at native collation. [Other sources](#) group words by novel alphabetization, no more or less canonical than any other. The Romanian "Curs de Stenografie" does make an effort at native collation, starting with vowels, and then in the general order of the consonants in this allocation. The collation algorithm prescribed herein is based on principles derived from the Chinook, but results in a similar order as the Romanian.

The most logical collation, given the structure of the script, is to collate by general shape, which places primacy on the consonants which, being invariant, tend to determine the shape of a word. Vowels have their own order, and clusters of one or more vowels should be collated as if they were a single vowel. Initial vowel clusters are ordered before the first consonant, medial and final clusters after the last.

Collation starts with consonants - initial vowels (ie no consonant) << H << P << T << F << K << L << M << N << J << S << combined consonants << medial/final vowels - then Affixes - attached << high << low - and finally signs. Secondary weight is given to diacritics, marks, and the bold R letters in the Sloan

orthography - all characters which do not change the basic shape of the word form. Tertiary weight is given to the joiners, spaces, and format controls, some of which can indicate semantic content, but often indicate presentation form.

All variant and compound consonants are collated directly after their base letters, with voiced consonants and their variants after the last unvoiced variants. The vowels collate similarly - O, A, I, U, Ou, Ow, Nasals - with variants collated after their base letters.

This collation order, based on the numeric values of letters in Chinook, corresponds significantly with the order of words in the Romanian "Curs de Stenografie", except that F/V comes before K/G instead of after, and A comes after O instead of before.

Collation table The ornamental horizontal rules in this document show the general collation order in simplified form.

Primary collation: Initial vowel cluster < H < X < P < B < P N < T < TH < SLOAN DH < D < DH < D S < F < V < F N < K < KK < G < SLOAN J < K M < L < HL < LH < R < RH < R S < M < M N < M WITH DOT < M S < M N S < N < N M < N WITH DOT < N S < N M S < J < J M < J N < J WITH DOT < J WITH DOTS INSIDE AND ABOVE < J S < J M S < J N S < J S WITH DOT < S < S J < S WITH DOT < S WITH DOT BELOW < S S < S J S < S T < S T R < S P < S P R < T S < T R S < W < WH < W R < S N < S M < K R S < G R S < S K < S K R < medial/final vowel cluster.

< ATTACHED SECANT < ATTACHED TANGENT < ATTACHED TAIL < ATTACHED I HOOK < ATTACHED E HOOK < ATTACHED TANGENT HOOK < ATTACHED LTR SECANT < LOW VERTICAL SECANT < MID VERTICAL SECANT < HIGH VERTICAL SECANT < LEFT HORIZONTAL SECANT < MID HORIZONTAL SECANT < RIGHT HORIZONTAL SECANT < HIGH ACUTE ARC < HIGH TIGHT ACUTE < HIGH GRAVE ARC < HIGH LONG GRAVE < HIGH DOT < HIGH CIRCLE < HIGH LINE < HIGH WAVE < HIGH VERTICAL < LOW ACUTE ARC < LOW TIGHT ACUTE < LOW GRAVE ARC < LOW LONG GRAVE < LOW DOT < LOW CIRCLE < LOW LINE < LOW WAVE < LOW VERTICAL < LOW ARROW < O WITH CROSS < word/affix signs from outside the Duployan block - Less Than, Greater Than, Multiplication, Plus Sign, etc.

Vowel order: O < WO < AOU < A < WA < OA < Sloan OW < I < E < WI < WEI < Romanian I < Sloan EH < Sloan EE < Short I < EE < IE < UI < YE < Long I < U < EU < XW < U N < LONG U < UH < OOH < Sloan U < OU < OW < WOW < Romanian U < Vocalic M < Nasal I < Nasal U < Nasal O < Nasal A < Pernin AN < Pernin AM < Sloan AN < Sloan EN < Sloan ON;

Secondary collation: No marks < Combining R < Double Mark < Diacritics.

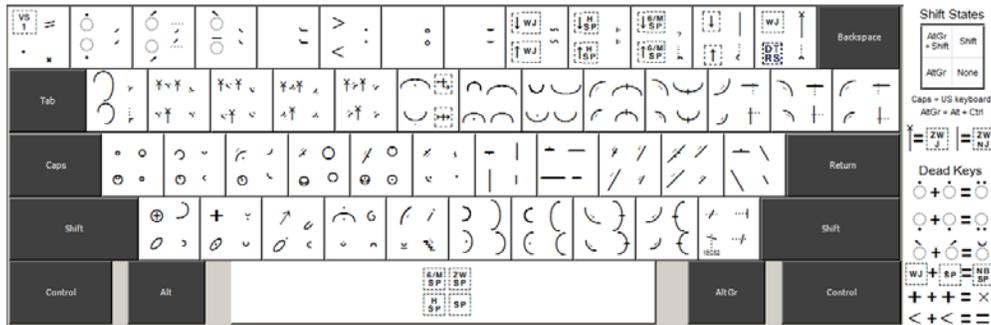
Tertiary collation: No format < Variation Selectors < SP/NBSP < ZWNJ < ZWJ; < shorthand formats Letter Overlap < Continuing Overlap < Down < Up; < following ZWSP < HSP < 6/MSP < THSP < 4/MSP < 3/MSP < NSP < MSP.

Irrelevant: All punctuation, including CHINOOK FULL STOP.

5. Input

A Basic Duployan keyboard layout has been devised for inputting Duployan text. This places the most common characters in the easiest to reach key positions. Keys are also defined for the basic nasal vowels with inherent joiners, which are the necessary encoding form in many orthographies. A character map, MSKLC file, and installation files are attached in keyboard.zip.

This keyboard layout should be considered informative as a base layout for the complete Duployan and Chinook. Other Duployan keyboards should not be constrained by this layout; specifically, a layout for a particular orthography should place the characters necessary for that language on the most convenient keys, regardless of the general layout, and should not necessarily provide for access to all Duployan characters or alternate forms of the nasal vowels.



6. Principles of the Duployan Shorthands and Chinook scripts

Rendering Duployan Characters.

Duployan characters, like characters in most shorthand scripts, can cursively connect, combine, and change shape depending on their context. Its appearance is affected by the presence of adjacent characters, ligaturing, the font used to render the character, and the application or system environment. These variables can cause the appearance of Duployan and Chinook characters to differ from their nominal glyphs (used in the code charts). Duployan and Chinook characters are default joining to each other, except for the high and low affixes and where otherwise noted in the code chart. Characters marked as non-joining, and any characters from other blocks are non-joining to Duployan and Chinook characters by default. Exceptions are Zero Width Joiner (U+200D), by definition, and the Shorthand Format Controls (U+1BCF0-U+1BCF3), which are tied to Duployan as Script_Extensions, and alter the joining characteristics of adjacent stenographic characters. Defined width spacing characters (U+2000-U+200B) preserve the height of the cursive stroke, so they should be treated as simple joining characters, with a blank glyph image.

Invariant letters.

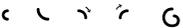
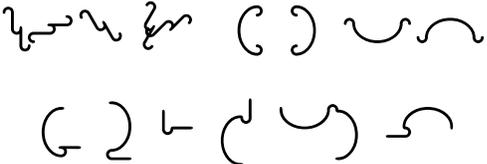
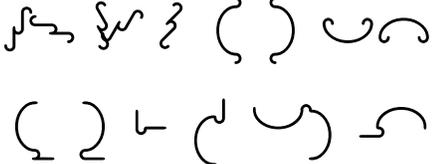
The majority of characters in the Duployan shorthands and Chinook scripts are invariant letters. They have a static shape, orientation, and stroke direction, and the set of invariant letters is almost completely contiguous with the consonants. Each invariant has a size - as many as three; a shape - line, quarter-circle, semicircle; a static orientation - N/S, E/W, NE/SW, NW/SE; an inherent stroke direction - generally LTR or TopToBottom; and many have derived and compound variants with markings (crosses or dots). They will usually cursively connect - the end of first character's stroke is the beginning of the second's - but will also overlap with a following character when shorthand formats are used. A few invariant letters and all of the high and low affixes are classified as *non-joining characters*, that interact typographically with adjacent characters like a word or text break, and only have a stroke direction when overridden by ZWJ (U+200D).

It can be assumed in the following that similar characters, like D, D-S, TH, and DH have the same cursive, overlapping, and other connecting properties as the character on which it is based, ie T. Likewise, variations of N - N-S, N-M, N-M-S, and Ng - connect like an N, and so on. The invariant letters can be generally classified as *P type* (line with N-S stroke direction), *T type* (line, W-E), *F type* (line, NW-SE), *K type* (line, NE-SW), *L type* (line, SW-NE), *M type* (N-E-S semicircle), *N type* (N-W-S semicircle), *J type* (W-N-E semicircle), and *S types* (W-S-E semicircle), and combined consonants (all quarter-circles, see code chart). Furthermore, the P,T,F,K, and L collectively constitute the *Line consonants*, and the M,N,J, and S types, as well as the combined consonants, are *arc consonants*.

Orienting characters.

Many vowels have a consistent shape, but rotate to align with the preceding character and will mirror to allow the following character to attach without crossing the vowel or preceding character. When adjacent one non-joining and one joining character, these orienting vowels will rotate to align with the

adjacent joining character, and mirror right/up or left/down based on their identity as a *primary orienting* or *secondary orienting* vowel. Likewise, when adjacent two similar type characters, or if the following character allows mirroring either way, they will align with the preceding character and mirror according to their orientation. Directional affinities are preserved, even when preceded by a non-joining and followed by a joining character. Primary orientation indicates an affinity for a stroke direction towards the right, and up when lacking a right/left distinction. Conversely, secondary orientation is left/down. Many orienting vowels come in pairs, with opposite orientations but the same basic shape. Except for 'I' and 'E', orienting characters can be bracketed by ZWJ/ZWNJ (U+200D/U+200C) to make a joining or non-joining invariant version. 'I' and 'E' have related invariant characters encoded separately.

Table 1: Comparison of Primary and Secondary Orienting Vowels	
Primary (right/up) Orienting Vowels-	Secondary (left/down) Orienting Vowels -
	
	

Related to the orienting vowels and invariant letters are the attached affixes. Many of these, noted in the charts with "dots [to] show position on and relative orientation to base glyph", act as spacing or non-spacing marks that do not effect joining of adjacent characters, but do rotate to match the angle of the base character. Some, noted in the charts with "dots [to] show position on base glyph", are non-spacing invariant marks.

Circle vowels.

The most commonly encountered vowel letters are the circle vowels. These vowels connect to preceding and following characters, with the adjacent characters entering the circle vowel at a tangent, and most (except Ou U+1BC0E) exiting the vowel shape at a tangent. The circle vowels often take partial contextual forms, with the adjacent characters implicitly completing the circle by crossing tangents.

Circle vowels followed but not preceded by a joining letter have a clockwise stroke direction into line consonants and will lie inside the arc of an arc consonant. Circle vowels preceded but not followed by a joining character will again sit inside the arc of an arc consonant, as if followed by a T-type if following a line consonant, and above the end of a T-type consonant. Circle vowels adjacent two line consonants will lie outside the angle created by the intersection of the two lines. When adjacent same type line consonants, they will again lie as if followed by a T-type. When adjacent an arc consonant and another invariant, the circle vowel will follow the angle rule as given above, but when the adjacent characters do not present an angle, the circle vowel will lie in the same position as if the following joining character were not there.

Many sequences of successive circle vowels default ligature forms. Where a ligature is not available, or when overridden by an intervening ZWJ + ZWNJ + ZWJ (U+200D + U+200C + U+200D), successive circle vowels not preceded by a joining character will connect at the vertical tangent on the shared side. If a preceding joiner character is present, cursively connected circle vowels will sit on opposite sides of the end of the previous character, with the following character determining the position of the second

character, as with a primary orienting vowel (see Table 5 and additions to Figure 16-3, below).

The Duployan Letter Sloan Ow (U+1BC75) and, in the Pernin and Sloan orthographies, discretionary ligatures of circle vowels, are classified as *reverse circle vowels*. These reverse circle vowels are opposite a regular circle vowel, ie they have a withershins stroke direction, will lie outside of arc vowels, inside the angle of two line consonants, &c. Reverse circle vowels are not known to interact typographically with other vowel characters.

Table 2: Circle Vowels and Reverse Circle Vowels		
Circle Vowels		Reverse Circle Vowels

Nasal vowels

Nasal vowels are the only Duployan characters that are positioned contextually. A fully implemented typeface will allow for three different renderings of the four basic nasal vowels (U+1BC18-U+1BC1B). When adjacent two joining characters, the nasal vowels will render as a diacritic placed outside the angle of the adjacent characters, shadowing the position of circle vowels adjacent two characters, explained above. When still preceded by a joining character, but followed by ZWJ (U+200D) + a joining character or by any non-joining character, the nasal vowel will render as a primary or secondary orienting vowel in relation to the preceding joining character. It will either join with the following character, if a ZWJ intervenes, or be unjoined with ZWNJ or a non-joining character. Likewise, when following a joining character + ZWJ/ZWNJ (U+200D/U+200C), the nasal vowel will render as a primary or secondary orienting vowel in relation to the *following* joining character. The Duployan Letter Vocalic M (U+1BC16) is always a primary orienting vowel, cutting backward in relation to the preceding character, and does not position diacritically. A nasal vowel not preceded by a joining character, and followed by a ZWJ + joining character will still orient in relation to the following joining character, allowing for consistent use of Nasal Vowels + ZWJ in orthographies that do not use diacritic positioning of nasals.

When bracketed by Zero Width Joiners, nasal vowels will render as combining invariant characters as per the nominal glyph images. ZWNJ (U+200C) can be used when the orienting or invariant nasal vowel is not to be connected to an adjacent joining character. The Pernin and Sloan nasal vowels (U+1BC7A-U+1BC7E) are always invariant, and the Vocalic M (U+1BC16) never. The orthography of the Romanian stenography uses the two U arc vowels (U+1BC0D, U+1BC1D) as nasals, however the Romanian stenography uses nasals as orienting vowels (+ZWJ), and no marking is needed for proper rendering.

Table 3: Nasal Vowels		
		F + An + T

\ + ʀ + $\begin{matrix} \text{[ZW]} \\ \text{[J]} \end{matrix}$ + -	→ \	F + An _j + T
\ + ʀ + $\begin{matrix} \text{[ZW]} \\ \text{[NJ]} \end{matrix}$ + -	→ \-	F + An _{nj} + T
\ + ʀ + $\begin{matrix} \text{[ZW]} \\ \text{[NJ]} \end{matrix}$ + -	→ \-)	F + On _{nj} + T
- + $\begin{matrix} \text{[ZW]} \\ \text{[J]} \end{matrix}$ + ʀ + -	→ -	T + _j An + T
- + $\begin{matrix} \text{[ZW]} \\ \text{[J]} \end{matrix}$ + ʀ + -	→ -	T + _j On + T
\ + $\begin{matrix} \text{[ZW]} \\ \text{[J]} \end{matrix}$ + ʀ + $\begin{matrix} \text{[ZW]} \\ \text{[J]} \end{matrix}$ + -	→ \	F + An _i + T

P.S. The logic behind the prescribed use of ZWJ/ZWNJ is that it deprives the surrounding context from the nasal vowel, specifying only whether the adjacent characters will join, as there are no known ligatures of nasal vowels. The joiner controls could be replaced by any non-joining character and result in the same rendering of the nasal.

Compound vowels.

The default rendering of compound vowel sequences (or vowel clusters) depends on the nature of the vowels involved. Most orthographies prefer ligation to simple compounding of circle vowels. However, compounding that visually preserves each member is regularly encountered in sequences involving orienting vowels combined with a circle vowel or any number of other orienting vowels. As a rule, circle vowels act as if an adjacent orienting vowel were a line consonant whose orientation is determined by any joining characters adjacent the vowel cluster. The entire sequence should be rendered as if it were an orienting vowel, although a circle vowel between a joining character and orienting vowel will sit opposite the orienting vowel, touching at the intersection of the orienting vowel and joining character. These vowel clusters have primary or secondary orientation determined usually by the first character of the sequence, but the last character when not preceded by a joining character. When the vowel cluster is not adjacent any joining characters, default rendering is along a horizontal mid-line, as with clusters of circle vowels (see circle vowels, above).

Table 4: Compound vowels

o + c	→ ȝ	A + I
o + v	→ æ	A + E
o + c + -	→ ȥ	A + I + T
o + v + -	→ ȣ	A + E + T
c + o + -	→ ȣ	I + A + T
l + o + c	→ ȡ	P + A + I
l + o + c + -	→ ȡ̄ OR ȡ̅	P + A + I + T
l + o + c + (→ ȡ̄̅ OR ȡ̅̅	P + A + I + M
v + c	→ ȣ	E + I
c + v	→ ȣ	I + E
o + c	→ ȝ	O + I

Ligatures, Allographs, and Standard Variants.

Ligaturing behaviour is fairly limited in the Duployan orthographies, especially in comparison with other cursive scripts like Arabic and Devanagari. As with Arabic, Devanagari, and other complex scripts, ligatures can be expressly requested by use of Zero Width Joiner (U+200D). Zero Width Non Joiner (U+200C) should break a ligature into its component characters, and the sequence ZWJ + ZWNJ + ZWJ (U+200D + U+200C + U+200D) would break a default ligature and render the characters by default joining behaviour (see circle vowels, above).

Discretionary Features.

All discretionary contextual/ligature forms can be requested in plain text by using ZWJ (U+200D).

The Pernin orthography makes use of a contextual form for repeated consonants, reducing the second consonant to a small blot (in writing, caused by increasing pen or pencil pressure) at the end of the previous character's stroke. This applies to both identical and similar consonants, with the first consonant represented by its full form, eg. T+Dot = T+T or T+D or T+Th &c.

Pernin also prescribes a ligature form of a circle vowel preceding the Pernin R (Duployan letter L, U+1BC05), unless it is followed by another circle vowel. The ligature form is an identically sized reverse circle vowel (see Circle Vowels, above). Similarly, in the Sloan orthography, an initial circle vowel preceded by an R (U+1BC15) will render as a reverse circle vowel.

Standard Variants.

All standard variants are requested in plain text Duployan by using Variation Selector 1 (U+FE00).

Pernin prescribes a "slight upward tick inclining to the left" for an L (U+1BC05, Pernin R) following R (U+1BC15, Pernin L), and one "to the right" for an R after L. This upward tick can also sometimes be found, generally at word end, following other consonants. These ticks are a standard variation sequence of the Duployan Letter L and the Duployan Letter R, encoded as L/R + VS1 (U+FE00).

The Duployan Letter W (U+1BC70) is the most variable letter among Duployan scripts. In the Sloan and Perrault orthographies, it is a full quarter arc, written NE-SW, 12 o'clock to 9 o'clock. On the other hand, in Pernin, it is closer to a one-sixth arc, starting closer to the 11 o'clock position, though still roughly the same length arc (larger diameter) than the Sloan/Perrault variety. Following K and G (U+1BC04, U+1BC14), the Duployan Letter W takes the form of a hook - Perrault tending a bit more wave-like than Pernin. Sloan prescribes other characters for K/G+W, and does not have a hook-form of W. The Pernin variant of W can be accessed in plain text by the use of the variation sequence W + VS1 (U+1BC70 + U+FE00), and the hook form as a default ligature/contextual form. As with all default ligatures, the unligated, joined sequence of K/G + W can be requested in plain text with a medial ZWJ + ZWNJ + ZWJ (U+200D + U+200C + U+200C).

In Chinook usage, the letters M, N, J, and S (U+1BC06 - U+1BC09) can be used as numbers (see numbers below). When they do so, they are smaller than the normal sized "letter" forms. These variants can be specified, again, by the variation sequence M/N/J/S + VS1 (U+1BC06/7/8/9 + U+FE00).

Other default ligatures and contextual forms.

Default features are unmarked in plain text. Unligated forms of these character sequences can be requested with the joining sequence ZWJ + ZWNJ + ZWJ (U+200D + U+200C + U+200C).

Most orthographies have some means of indicating the junction of two same type line consonants. Usually, this comes in the form of a slight (\leq line width) jog at the intersection, or sometimes a short cross-tick at the intersection of the characters or an angle change of L/R characters. For the purposes of plain text, the jog is considered the unligated form of the character sequence, and is the neutral default rendering. An implementation can prescribe the cross tick, or other indicator as a default rendering. ZWJ should always request the Pernin dotted form, above, and never the tick, angle, or jog.

The Romanian orthography prescribes contextual forms for the Romanian U character (U+1BC1F) and its compounds. The nominal form given in the code charts is for non-medial contexts. When medial, it takes the form of Duployan Letter Ow (U+1BC0F). Positional ligatures include the sequence O + Romanian U (U+1BC0A + U+1BC1F), when initial or final, taking the form of an elongated, oval shaped, plain circle vowel. Medially, A or O + Romanian U (U+1BC0A/U+1BC0B + U+1BC1F) exhibits the default joining behaviour of sequential circle vowels, sitting on opposite sides of the end of the previous character - Romanian U again appearing in its medial "Ow" form. Following other vowels, Romanian U appears in diminished form, as a sort of tail.

Romanian also prescribes a ligated form of the vowel sequence O + A (U+1BC0A + U+1BC0B) that is visually identical to the letter Wa (U+1BC5B).

Lastly, the Duployan thick letter selector (U+1BC7F, DTLS) does not have a visual form of its own, but causes the previous character to be rendered as a thick variant, representing the addition of an 'R' sound to a Sloan letter. The Duployan Letter R (U+1BC15) can not substitute a ligature behavior for the DTLS, as the added 'R' sound can occur in the middle of a compound letter.

Table 5: Ligatures, Allographs, & Alternates		
Discretionary features		
+	→	B + p
+ ° + / + -	→ ⌋	B + Ar + T
/ + o +	→ ρ	rO + P
Standard variants		
/ + / + VS1	→ /	R + L variant
- + / + VS1	→ ⌋	T + R variant
⌋ + VS1	→ ⌋	W variant
⌋ + VS1	→ ⌋	M variant
⌋ + VS1	→ ⌋	N variant

◡ + VS1	→ ◡	J variant
◤ + VS1	→ ◤	S variant
Default features		
// + Ꞑ	→ ʞ	K + W
+ ʞ + -	→ ʞ̣	P + Rom U + T
+ 0 + ʞ	→ ʞ̇	B + O + Rom U
+ 0 + ʞ + -	→ ʞ̣̇	B + O + Rom U + D
+ • + ʞ	→ ʞ̣̇	B + A + Rom U
+ • + ʞ + -	→ ʞ̣̣̇	B + A + Rom U + D
+ ˙ + • + ʞ	→ ʞ̣̣̇̇	B + I + A + Rom U
- + -	→ - -	T + D
+ 0 + •	→ ʞ̇	B + O + A
\ + DT RS + •	→ ʞ̣̇	F _R + A

Additions to Figure 16-1. Prevention of Joining

	+	◡			→ ◡
1BC11		1BC08			
	+	ZW NJ	+	◡	→ ◡
1BC11		200C		1BC08	

Additions to Figure 16-2. Exhibition of Joining Glyphs in Isolation

		᠒			→ ᠒
		1BC1F			
ZW J	+	᠒	+	ZW J	→ ᠒
200D		1BC1F		200D	

Additions to Figure 16-3. Effect of Intervening Joiners

Character Sequences				As Is	ZW NJ	ZW J ZW NJ ZW J	ZW J
	◦	/	-	᠒	᠒	᠒	᠒
1BC11	1BC0B	1BC05	1BC02				
	○	◦	-	᠒	᠒ ^e	᠒	᠒
1BC11	1BC0A	1BC0B	1BC02				
	/	/		/ or / or /	//	/	/
	1BC15	1BC05					

Joined text.

The most common form of character interaction is that of the cursive connection. The termination of a character stroke leads directly into the beginning of the next character. Vowel signs follow the dynamic shaping discussed above, but fundamentally are the same as other joining characters, joining at a tangent to adjacent characters. Non-joining characters - any character from other scripts, and those

found in Duployan - have a small intervening space, as with standard alphabetic writing.

Unjoined text.

The Duployan script has a cursive conjoining property that, like Arabic, is effected by the use of the Zero Width Non-Joiner (ZWNJ, U+200C). ZWNJ encodes a break within a word, turning an otherwise joining character into a non-joining character, and resetting the cursive stroke height to neutral. This break is usually found only at nominally syllabic boundaries in Chinook texts, and where a separated letter or letters indicates an affix in the Duployan shorthands. This break is smaller than a word space, in some instances involving negative kerning, and is not a word break. ZWNJ and Zero Width Joiner (U+200D) will also change the positioning of the nasal vowels (see nasal vowels, above).

Overlapping text.

The use of overlapping letters to indicate abbreviations and initialisms is found in many systems of shorthand. As such, the current proposal allocates a block of shorthand format characters, which encode non-default text flow in any shorthand. Included are two overlap control characters: the first (U+1BCF0) indicating a single *letter overlap*, with the text continuing to flow as if that overlapping character did not exist, and the second (U+1BCF1) indicating a *continuing overlap* where the text flow proceeds from the overlapping character. In Duployan, this behaviour is limited to consonants, circle vowels, and orienting vowels overlapping consonants.

The overlapping behavior in Duployan shorthands and Chinook is fairly straightforward: for two line consonants, two arc consonants, or a vowel overlapping any consonant, the two characters overlap at approximately 3/5 along the stroke of the first consonant and 2/5 along the stroke of a second consonant or the middle of a vowel. For overlaps of arc and line consonants, the arc consonant is split into the first and second half of the arc, an arc overlapping a line taking place in the first half, line over arc in the second. The line consonant, again at the 3/5 / 2/5 point, will meet the arc at a perpendicular angle, or as close as possible, never beyond the middle of the arc, nor past the end.

It is unknown if or how M type and N type or J type and S type arc consonants would overlap each other until such a time as examples of this occurrence are documented. Default rendering should indicate the overlap in some way, either preserving control characters, or through an offset. Same type line consonants also will not overlap, necessitating similar default rendering; L-type and K-type consonants will not overlap each other, as well, due to their similar angle.

As indicated above, the flow of text continues either with the first character in the case of U+1BCF0, or with the second in the case of U+1BCF1. An overlapping letter can also take another overlapping letter before returning to the original text flow. Also, in the Romanian shorthand, long line consonants (U+1BC11-U+1BC15) can take two overlapping characters, indicated by two Letter-Overlap control characters (U+1BCF0 + U+1BCF0) followed by the two overlapping characters. With double overlaps, the first overlapping character overlaps at approximately 1/3 of the stroke length of the base character, the second at ~ 2/3. See Parsing of Shorthand Overlap Sequences, below.

Down step.

The Romanian shorthand prescribes that a certain set of word endings be indicated by letters following not in the default direction of text flow - to the right, but below the word. Likewise, the Sloan-Duployéan and Pernin methods prescribe contracted word endings, wherein the next word is started low, to signal a dropped sound at the end of the previous word. As such, a shorthand format has been defined (U+1BCF2) that indicates a following character should be rendered below the previous character, with any subsequent joined characters proceeding relative to the lowered glyph. At word boundaries, this causes the next word (or stenographic period) to be lowered. Because the lowering is a part of the previous word, the lowered word boundary should be indicated by the shorthand format down step, followed by a width defined space (U+2002-U+200b) and the next word, or period (U+2E3C?). Note that the step format control is found directly after the preceding word, as it encodes a phoneme missing from

the end. When Cross' Eclectic shorthand is encoded, a space will come before the step format control as the change in alignment represents a missing initial phoneme.

Up step.

The Sloan-Duployéan and Pernin methods also prescribe contracted word endings, where the next word is started high, signaling the dropped sound. A shorthand format has been defined (U+1BCF3) to indicate a following word (or stenographic period) to be raised. Even though the up control is only found at word boundaries, this boundary form is still indicated by the shorthand format up step, followed by a width defined space and the next word, or period.

Aligned text.

The last form of contracted words in Sloan-Duployéan and Pernin are non-stepping, with the two words even. As with distinctions in spacing with the Step formats, distinctions in spacing of aligned text are encoded with defined-width space characters (ZWSP, U+200B; HSP, U+200A; 6/MSP, U+2006; 4/MSP, U+2005; 3/MSP, U+2004; ENSP, U+2003; EMSP, U+2002) or the non-breaking counterparts thereof (hence, Word Joiner, U+2060, as well). Note that Thin Space (U+2009), is not used, due to its common equivalence to the Six-Per-Em Space (U+2006). The regular space characters (U+0020 and U+00A0) cause the following word to start at a neutral baseline, and cannot be used for aligned or stepped word boundaries. If different sized spaces are needed unaligned, again, the above space characters can be used, preceded by ZWNJ. Note that the natural letterspacing of unjoined characters is retained with step format controls, so a ZWSP (U+200B) will not cause the adjacent characters to touch, and will, in fact, appear identical to a ZWNJ (U+200C), except that alignment will be preserved.

Table 6: Text flow		
Joined Text		
+ ◡ + .	→ ◡ .	PJH
— + / + ‘	→ — / ‘	DKX
Unjoined Text		
ZW + ◡ ZW + .	→ ◡ .	P.J.H
— ZW + / ZW + ‘	→ — / ‘	D.K.X
Letter Overlaps		
— / / H + ◡	→ — / / ◡	Line _x S

\cup $\boxed{+}$ + $- \setminus //$	\rightarrow $\cancel{\cup}$	$S_x \text{Line}$
$ $ $\boxed{+}$ + $/$	\rightarrow $\cancel{ }$	$B_x R$
$-$ $\boxed{+}$ + $/$	\rightarrow $\cancel{-}$	$D_x G$
\setminus $\boxed{+}$ + $-$	\rightarrow $\cancel{\setminus}$	$V_x D$
$/$ $\boxed{+}$ + $ $	\rightarrow $\cancel{/}$	$G_x B$
$/$ $\boxed{+}$ + \setminus	\rightarrow $\cancel{/}$	$R_x V$
$($ $\boxed{+}$ + $($	\rightarrow $\cancel{(}$	$M_x M$
$($ $\boxed{+}$ + \cup	\rightarrow $\cancel{(}$	$M_x S$
$($ $\boxed{+}$ + $)$	\rightarrow $\cancel{(}$	$M_x J$
\setminus + \circ + $-$ $\boxed{+}$ + \setminus + \circ + $-$	\rightarrow $\cancel{\setminus}$	$KAT_{xK} AT$
Continuing & Double Overlaps		
\setminus + \circ + $-$ $\boxed{+}$ + \setminus + \circ + $-$	\rightarrow $\cancel{\setminus}$	$KAT_{x+} KAT$
\cup $\boxed{+}$ + $ $ $\boxed{+}$ + \cup	\rightarrow $\cancel{\cup}$	$S_x B_x J$
$-$ $\boxed{+}$ $\boxed{+}$ + \circ + $/$ + \cup	\rightarrow $\cancel{-}$	$D_{xA+K} Un$
Under affixes		

(+ ˘ + [↓] + ◡ + ◦	→	Min _{-SA}
Under word		
— + [↓] + [ZW SP] + -	→	D _{-T}
Over word		
+ [↑] + [ZW SP] + c	→	B ⁺ _{-le}

Combining diacritical marks on vowels.

Several Duployan orthographies use combining diacritical marks to distinguish vowels. These diacritics include acute, grave, breve, macron, under macron, over dot, under dot, diaeresis, under diaeresis, &c. They can appear on orienting vowels, circle vowels, and nasal vowels (On, and An). Although there are several vowel letters with marks included in the allocation, these are not decomposable as a combining sequence, as the diacritic marks change position along with their "base" orienting vowel. Combining diacritics indicate vowels with diacritics that consistently appear above or below the base character, no matter the adjacent joining characters.

Affixes.

Except for Chinook, every Duployan orthography makes extensive use of a set of marks - often similar, in appearance, to diacritics - and letters to symbolize lexical affixes. The unattached high and low Duployan affixes (U+1BC2A-2F, U+1BC3A-3E, U+1BC50, U+1BC83, U+1BC93) act much like spacing characters - the marks are written next to the word root, and will be either higher or lower than the adjacent letter.

The attached affixes (U+1BC40, x41, x44, etc.) touch or cross the first or last letter of a word (again for prefixes or suffixes), with the location of crossing (and touching if not evident) symbolized by a dotted line in the charts. The character names list specifies if the character rotates to complement the angle of the base letter, or is invariant. An attached affix always attaches to a letter, never to an affix. Since affixes are encoded logically, and unattached affixes can logically occur between a root and an attached affix, the displayed order of affixes may be different from the encoded order.

Third, some orthographies use letters or sequences of letters to indicate affixes, some of which appear similar to the high or low affix signs. As a rule, signs that are similar to a letter, but unmotivated - that is, they don't symbolize a sound of the affix - or if a high and low pair is found in the orthography, they are symbolized by affix signs, not letters. Signs that are motivated and aren't paired high/low should be represented by a letter, often separated by ZWNJ (U+200C) from the root, whether the affix usually appears lower or higher than the adjacent character or not. Some letter affixes are encoded with the shorthand format Continuing Overlap (U+1BCF1). For consistency, the shorthand format Letter Overlap (U+1BCF0) should not be used to combine an affix to a root - even if the root is a single character.

In the Sloan orthography, successive high and low affixes or letters-as-affix and high/low affixes are written joined together. These compound affixes always position like the first high or low affix in the compound. It is encoded as affix 1 + ZWJ + affix 2, whether it is affix sign + affix sign or letter + affix sign. Letter + letter affixes do not need to be joined by ZWJ, as they are already joining characters. As with other affixes, if the compound ends with a letter-affix, it must also be followed by ZWNJ if it does

not cursively connect with the word root. Likewise, some high/low affix signs can be used as an attached affix, again encoded with ZWJ (U+200D).

Table 7: Diacritics and affixes		
Diacritics and Precomposed Vowels		
+ ◌ + ◌ + ◌	→ ◌̣	P + E + Underdot
+ ◌	→ ◌̣	P + Rom I
- + ◌ + ◌ + ◌	→ ◌̣̣	T + E + Underdot
- + ◌	→ ◌̣̣	T + Rom I
Affix Signs and Letters		
' + - + ' + '	→ ' - '	arc D arc
- + ' + ,	→ - ' ,	D ^{arc} _{arc}
- + - + [ZWJ] + ,	→ - ◌̣̣	D ^{line+} _{arc}
◌ + ' + ◌	→ ' ◌	/ ^{arc} S → ^{arc} /S
- + ◌ + / + [ZWJ] + (→ - ρ (DOK-M
/ + - + [ZWJ] + /	→ L /	KT-R
- + [ZWJ] + " + [ZWJ] + ◌	→ ◌̣̣	T +vert+ I

Numbers.

The Duployan orthographies each have a distinct means of expressing numbers. Some number systems must utilize formatting requiring markup to represent all aspects of the number system, and as of this time, there is no expectation that a full transcription of all number forms should be representable in plain text. The Chinook number system uses Duployan characters and markup to indicate numbers. The Romanian shorthand and French Duployan use regular European/Arabic numerals in conjunction

with Duployan characters, combining marks, and markup to indicate magnitude and aspect. Sloan and Pernin use markup and non-Duployan characters in conjunction with regular European/Arabic numerals.

Chinook numbers.

The Chinook number characters are 1 P, 2 T, 3 F, 4 K, 5 R, 6 M, 7 N, 8 J, 9 S, 0 O, 10 A, 100 Wa, and 1000-enclosing circle handled with markup. The numbers can be indicated Hanzi-style with P-S combining with O, A, or Wa to indicate value, although an O, A, or Wa must be preceded by a P to indicate a single hundred or ten, unlike Hanzi numerals. P-S connects to O, A, and Wa the same as in running text. O is used unconnected to indicate a zero or connected for the tens with a following digit zero, while A is used when connecting the tens to a non-zero ones digit. The enclosing circle for thousands surrounds the entire group of up to five characters (P-S + Wa + P-S + A/O + P-S), and can nest inside itself to indicate millions - a separate circle surrounding a following thousands group. Chinook numbers can also be indicated Indian/Arabic style, with the digits 0 9 (O-S) having place value. This is especially common when writing years or when numbering items, as opposed to enumerating them. The digits generally connect cursively, the same as in Hanzi-style Chinook numbers. For most Chinook writers, the numeral forms of M, N, J, and S are about half-size normal, and are requested in plain text by M/N/J/S + VS1 (U+FE00).

Romanian numbers.

The Romanian number system uses the European/Arabic numerals to indicate numbers 0-99, with marks to indicate further powers of ten: an overdot (U+0307) for hundreds, a preceding Middle Dot (U+00B7) for thousands, a dot below (U+0323) for millions, and a following Middle Dot for thousand millions. As with most systems using marks to indicate magnitude, these marks can be used in conjunction, e.g. a dot above and dot below for hundred millions. Multiplicative forms (with the prefix *ân-*) use the character A Nasal (U+1BC1B) before a number, percentages with Combining Ring Above (U+030A), and grade with the degree sign (U+00B0). Ordinals are symbolized by a following T (U+1BC02), while fractions are written numerator over denominator, with no solidus or line. This representation of fractions constitutes a presentation form of already encoded fraction signs or can be explicitly expressed using markup, never with the shorthand format down step (U+1BCF2).

Pernin numbers.

The Pernin number system uses the European/Arabic numerals to write numbers, although periods (U+002E) can be used instead of zeros. An underline (by markup) indicates ordinals (first, second...), while an overline (again) indicates the numerical adverbs (once, twice...). The Pernin system suggests, however, that "when large numbers are to be written ... it is better to indicate ... us[ing] a corresponding shorthand contraction for thousand, million, etc.", such contractions left to the individual.

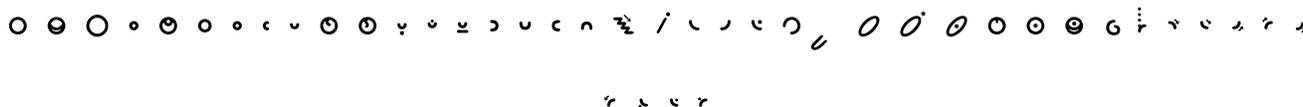
Sloan numbers.

The Sloan number system uses the European/Arabic numerals to write numbers, and can be used for ordinals, iteratives, &c. e.g. 2: two, twice, second, secondly. The shorthand aspect in the Sloan system is the use of an overline, strikethrough, and underline (all represented with markup) for magnitude as follows: Overline: hundreds; Strikethrough: thousands; Underline: millions. Again, these can be used in conjunction with each other to indicate, for example hundred millions with an overline and underline.

French Duployan numbers.

The French Duployan number system, like the Romanian, uses the European/Arabic numerals with Duployan letters and affixes indicating magnitude and aspect. Magnitude is indicated as follows: Hundreds with an S (U+1BC09) after the number; Thousands with the Duployan Affix High Dot (U+1BC2C) following the number; Millions with the Duployan Affix Low Grave Arc (U+1BC3B) following; and Thousand Millions (Milliards) with a following R (U+1BC15) like a large solidus. As above, these

indicators of magnitude can be combined, e.g. an S and high dot indicating hundred thousands. For ordinals, the Duployan Affix Low Dot (U+1BC3C) is used following any indications of magnitude; Adverbs with the Duployan High Acute Arc (U+1BC2A); Approximates (*dizaine, douzaine, &c*) with the Duployan High Grave Arc (U+1BC2B); Adverbials with the Duployan High Circle (U+1BC2D); Percents with the Duployan Low Circle (u+x3D), doubled for Per mill. Manuscripts will indicate the numbers 4 and 6 with an underline to distinguish these number forms from the words "quittance" and "mot" to which the regular number forms show affinity; This distinction should be handled with markup or by typeface choice.



7. Confusability and Usage

Given the complex shaping engine required to render Duployan text, there can be ambiguity as to which character or character sequence should be used to represent a given form. The full names list supplied can be consulted for known ambiguities, but this is not an exhaustive list. For dotted letters vs. diacritics, the determining factor is always whether the dot moves in relation to the letter contextually, as explained in diacritics, above. The dotted consonants should always be used and never decomposed; e.g. HL (U+1BC68) ≠ H (U+1BC00) + L (U+1BC05) and S with dot below (U+1BC49) ≠ S (U+1BC09) + Dot Below (U+0323). Other confusables are in the affixes, and the rule (as given above) is that an affix that is motivated uses the letters, generally unjoined to the word, e.g. Pernin Inter- = In (U+18) + T (U+1BC02) + ZWNJ, Magn- = M (U+1BC06) + ZWNJ, and Multi- = M (U+1BC06) + Continuing Overlap (U+1BCF1). When there is a positional distinction (high vs. low), the affix signs should always be used.

Romanian word signs

For the most part, the extensive list of Romanian word signs are unambiguous. The Duployan Letter Ow (U+1BC0F) should only be used in Romanian text as an overlapping character or as a word sign. In running text, the Ow shape represents the medial form of the Duployan Letter Romanian U (U+1BC1F). In numeric contexts, the Degree Sign (U+00B0) and Combining Ring Above (U+030A) should be used instead of the High Circle Affix (U+1BC2D) for indicating percentages and grade of Romanian numbers. Likewise, the Combining dots (U+0307 & U+0323), Combining Diaereses (U+0308 & U+0324), and Middle Dot (U+00B7) should be used to indicate powers of ten instead of the Dot affixes (U+1BC2C & U+1BC3C) and letter H (U+1BC00).

Proper Names

Most Duployan shorthands prescribe that proper names be marked, as there are no majescole letters. Universally, they prescribe an underline, which should not be encoded in plain text, but handled through markup.

Stenographic Full Stop (Period)

The Stenographic Full Stop, U+2E3C, is used with shorthand/stenography systems in place of the normal period. Oftentimes, these systems will make use of a dot for a letter, word, or affix symbol, and the crossed period is used to avoid ambiguity. Due to its script=common attribute, and its unsuitability to any SMP blocks, this punctuation mark should be placed in the BMP.

8. Parsing of Shorthand Overlap sequences

Parsing as a tree. Even though the handling of Duployan characters with Shorthand Format Overlap

Controls is fairly simple, it is based on a more robust model with a few simple rules analysable as an N-ary tree: 1) each Overlap Control (branch) has as its base (parent node) the most recent character in the text stream 2) each Overlap Control must take a single shorthand character as its "argument" (child node), 3) the argument of each Overlap Control is allocated by a preorder insertion, where the number of branches of a particular node (character) is defined by the number of consecutive Overlap Controls directly following the character in the text stream. 4) for a Continuing Overlap to be valid, its base (parent node) must be the original base character, or the argument (child node) of another valid Continuing Overlap.

As a rule, nasal vowels, affix signs, and the letters H, X, the H-modified consonants from 1BC60-1BC6F, and all vowel characters are not bases for an overlapping character. The letters H, X, I, E, and affix signs do not overlap other letters. The voiced (medium length) line consonants can take two characters overlapping in the Romanian orthography.

Parsing as a stream. The structure of a shorthand overlap sequence can also be analysed as a simple stream 1) each Overlap Control has as its base the most recent character in the text stream, 2) each Overlap Control must take a single shorthand character as its argument, 3) after the the initial base, each character binds to one Overlap control - the first unbound Overlap control in the most recent group of overlap controls with an unbound member. 4) for a Continuing Overlap to be valid, its base must be the original base character, or a character bound to another valid Continuing Overlap.

Example. The example given below is for demonstration purposes only, counterfactually presuming that a Duployan character can take three Overlaps or be a third overlapping character. Known textual examples contain just a few overlaps associated with a single parent base character. Each character and overlap control is numbered identically in the text stream, parsing structures, and output image, and is color matched between the parsing structures and output image.

Character 1 is the highest base character. If there are no Continuing overlaps, the next non-overlapping character will cursively connect to this character.

Characters 2, 3, & 4 are Overlap Format Controls, with Character 1 as their base. In this case, Character 2 is a Continuing Overlap, and is the first (leftmost) overlap of Character 1, while 3 and 4 are the middle and rightmost overlaps

Character 5 is the first character overlapping Character 1. Since Character 2, of which this character is the argument, is a valid Continuing Overlap, the next non-overlapping character will cursively connect to this character, if it is not the base for another Continuing Overlap.

Character 6 is a Continuing Overlap, with Character 5 as its base.

Character 7 is the character overlapping Character 5. Since Character 6 was a valid Continuing Overlap, the next non-overlapping character will cursively connect to this character.

Character 8 is the second character overlapping Character 1. It is the argument for Character 3.

Characters 9 and 10 are Letter Overlaps with Character 8 as their base.

Character 11 is the first character overlapping Character 8, and is the argument of Character 9.

Character 12 is the second character overlapping Character 8, and is the argument for Character 10.

Character 13 is the third character overlapping Character 1, and is the argument of Character 4.

Since there are no remaining unbound Overlap controls, **Character 14** is not an overlapping character. Since the cursive connection was passed into the overlaps by the Continuing Overlap Format Controls (Characters 2 and 6), this character cursively connects to Character 7, instead of Character 1, and the rest of the word would continue from it.

Example Text Stream, Parsing Examples, and Sequence Rendering

number	1	2	3	4	5	6	7	8	9	10	11	12	13	14
--------	---	---	---	---	---	---	---	---	---	----	----	----	----	----

glyph	—	⊞	⊞	⊞	/	⊞	\	/	⊞	⊞	\	-		/
code point	U+1BC22	xF1	xF0	xF0	x14	xF1	x03	x15	xF0	xF0	x03	x02	x01	x14

Parsing of Shorthand Steps.

In contrast with overlaps, the Shorthand Step Format controls have a simple grammar: ZWNJ (U+200C), space & NBSP (U+0020 & U+00A0) and all non shorthand characters will return a text stream to a neutral baseline. Spacing characters (except for space and NBSP), including ZWSP, ZWJ, and WJ (U+200B, U+200D, & U+2060), and all gc=Mn will preserve the current baseline (ie, the height of the cursive stroke) and advance the text stream, if appropriate. Shorthand Step Format controls always act in relation to the current stroke height, whether it is neutral or has been altered by preceding characters.

Shorthand steps should only be used between two letters, or adjacent to a spacing character.

Given that future shorthands will need to be encoded with varying step heights, and the needs of those shorthands should take precedence, this proposal does not define whether multiple instances of an Up or Down step is legal. Until such time as a determinative shorthand is encoded, a second (or more) Up or Down Step Format control should be interpreted as raising or lowering the stroke height a second (or more) time.

Shorthand Control level of implementation.

Different shorthands will have need for differing levels of implementation of the Shorthand Format Controls. Support of arbitrarily complex Overlap sequences shall not be required for Unicode conformance; therefore, the block description for each encoded shorthand should include specifications for the width of overlaps (maximum number of overlaps assignable to a single character), the depth (how many overlaps can "stack" on each other), and the breadth (how many overlaps are assignable to an already overlapping character). If the depth is only 1, then the breadth is, by default, 0. For example, the Duployan Shorthands and Chinook, as a whole, have a width of 2 (on the medium line consonants in Romanian), a depth of 2 (Chinook abbreviations), and a breadth of 1. Implementations for specific orthographies are: French Duployan: 1x1; Romanian: 2x1; Chinook: 1x2(1); Sloan: 1x1; Pernin: 1x1; Perrault: 0.

and for a list of technical reports.

[Versions] Versions of the Unicode Standard

<http://www.unicode.org/versions/>

For details on the precise contents of each version of the Unicode Standard, and how to cite them.

Modifications

The following summarizes modifications from the previous version of this document.

1 Initial version

Copyright © 2012-2014 Van Anderson and Unicode, Inc. All Rights Reserved. The Unicode Consortium and [authors] make no expressed or implied warranty of any kind, and assume no liability for errors or omissions. No liability is assumed for incidental and consequential damages in connection with or arising out of the use of the information or programs contained or accompanying this technical note. The Unicode [Terms of Use](#) apply.

Unicode and the Unicode logo are trademarks of Unicode, Inc., and are registered in some jurisdictions.