LANGUAGE DOCUMENTATION & CONSERVATION

Vol. 16 (2022), pp. 364-398
http://nflrc.hawaii.edu/ldc/
http://hdl.handle.net/10125/74684
Revised Version Received: 12 Jun 2022

Type right: Examining the underlying causes of common typeface and font errors for Indigenous orthographies, and a possible path forward

Julia Schillo
Simon Fraser University

Mark Turin
University of British Columbia

Despite considerable typographical innovations over the past twenty years that have enabled and facilitated typing capabilities for many Indigenous language orthographies, typographical errors continue to disproportionately affect Indigenous languages. These include errors in glyph shapes, which impact legibility, and issues with glyph positioning, which impact readability. In this article, the glottalization accent mark is used to demonstrate how such errors manifest in various widely used typefaces. Through a case study of the glottalization accent mark, we identify the root causes of common typographical errors, stemming from the Unicode Standard, which provides the code structure for digital typing, and from the typeface design methodology used to create most of the typefaces available to Indigenous language communities. Many Unicode characters used by Indigenous orthographies lack rigorous and precise semantic definitions, leading to inconsistencies in glyphs created through a language-agnostic typeface design process that does not require designers to be familiar with the specific orthographies for which they design glyphs. To address these issues, we recommend that Unicode revisit the character semantics of Indigenous orthographic elements to create more robust semantic definitions and that typeface designers use a community-partnered design methodology that engages with the goals of language reclamation and revitalization.

1. Introduction[1] The past twenty years have seen considerable typographical innovation, particularly in terms of making typing more universal and increasingly accessible for language communities across the globe (Haralambous 2007: 53–54; Hermes et al. 2016). Despite these positive developments, common and widespread issues for typing in many North American Indigenous languages persist. While language speakers can identify when text in their language displays incorrectly, the

[1] Acknowledgements: We are grateful to Bridget Chase, Cris Hernandez, and Aidan Pine for their insightful comments and suggestions, and to the two anonymous peer reviewers whose knowledge and specific suggestions have helped to improve and refine this contribution considerably. We also thank Gvủí/Rory Housty, Marianne Ignace, and Aiyana Twigg for generously taking the time to share their experiences of typing in Indigenous languages. Our thanks to the editors of LD&C for their responsiveness and support throughout the process, and to Rui Yamawaki and Cedar Lay for such careful and professional copy-editing and layout design. Needless to say, the responsibility for any remaining errors lies solely with the authors.

E-ISSN 1934-5275

underlying causes of incorrect formatting and representation remain elusive to most users.

This article examines the sources of some of the most common errors that occur when typing in many of the languages Indigenous to North America. We begin by situating this article within its linguistic and orthographic context. Next, we describe a number of the key typographical problems that language users encounter. After this, we give a brief overview of the mechanics and theory that make typing possible, which will then allow us to analyze the root causes of the typographical issues that we describe. We conclude by proposing ways to strengthen typeface design and character coding moving forward.

1.1 Linguistic and orthographic context While widespread typographical issues persist for many Indigenous languages in North America (and beyond), this contribution focuses primarily on the orthographic needs of Indigenous languages spoken in the Pacific Northwest.

The Pacific Northwest is the northwestern region of North America commonly defined as British Columbia, Washington State, and Oregon.[2] The Pacific Northwest is a region of extreme linguistic diversity, with thirty-four languages spoken in British Columbia alone (First Peoples' Cultural Council 2018a: 1). Language communities in this region are generally small, with memberships numbering between a few hundred and a few thousand in total. While all of the languages spoken in this region are endangered, with very small numbers of fluent speakers and most speakers belonging to the grandparent generation (First Peoples' Cultural Council 2018a: 1), a number of innovative language revitalization projects are underway that are working to build the next generation of language speakers.[3]

The languages spoken in the region exhibit a large inventory of phonemic consonants, with many complex consonant clusters. These phonemic inventories often include many post-velar consonants, glottalized (ejective) and nonglottalized consonants, as well as many affricates.[4] Although Pacific Northwest language communities are extremely diverse, their orthographies exhibit some shared features. Most orthographies used by Indigenous language communities in the Pacific Northwest are Latin-based and are tasked with representing these expansive phonemic inventories, which have relatively few sounds in common with the European languages for which the Latin script was initially developed.

Most orthographies for Pacific Northwest languages are based specifically on the North American Phonetic Alphabet (NAPA),[5] a system of phonetic notation

[2] Some definitions of the Pacific Northwest also extend to include neighboring regions of Alaska, the Yukon, Idaho, and California.

[3] For more information about the status of languages and revitalization work in British Columbia specifically, see First Peoples' Cultural Council (2018b).

[4] For examples from the Salish language family, see Davis (2019: 457) and from the Athabaskan language family, see Hargus (2010).

[5] Also known as the Americanist Phonetic Alphabet.

that is widely used by linguists for the transcription of North American Indigenous languages. NAPA was developed to promote phonetic precision and orthographic transparency, with each letter corresponding to one sound, as opposed to other orthographies in which letter combinations are used to realize individual sounds. As a consequence of this approach, NAPA has a large inventory of letters and diacritical marks to represent the many diverse phonemes of North American Indigenous languages (Pullum & Ladusaw 1996: xxii–xxiv). While NAPA is primarily Latin-based, it also includes some Greek-based letters such as the lambda and gamma, as well as letters specific to phonetic alphabets such as a belted-L <ɬ> and a glottal stop <ʔ>. In addition, NAPA makes extensive use of diacritical marks such as carons (hačeks), glottalization accent marks, and underdots, as well as superscript letters (Pullum & Ladusaw 1996: 301–302). A partial NAPA chart demonstrating its orthographic elements can be found in Figure 1.[6] These orthographic features are reflected in many Pacific Northwest Indigenous orthographies, which also include Greek-based letters, phonetic alphabet letters, a plethora of diacritical marks, and superscript letters.

	Bilabial	Labio-dental	Inter-dental	Dental	Alveolar	Palato-alveolar	Palatal	Velar	Uvular	Pharyngeal	Glottal
Stop	p b			t̪ d̪	t d		k̠ g̠	k g	q ɢ		ʔ
Affricate					ȼ dz	č ǰ					
Lateral Affricate					ƛ ʎ						
Fricative	ɸ β	f v	θ ð	s̪ z̪	s z	š ž	x̠ ɣ̠	x ɣ	χ ʁ	ħ ʕ	h ɦ
Lateral Fricative					ɬ						
Nasal	m	m̥			n		ñ	ŋ	ɴ		
Ejectives	pʼ tʼ kʼ										

Figure 1. A partial NAPA chart based on the chart provided in Pullum & Ladusaw (1996: 301–302), including the diacritical mark used for ejectives, also known as glottalized consonants.

Because NAPA was primarily intended to be used by academic linguists in service of their technical needs, readability and legibility are not its principal goals, and even less so any accompanying aesthetic sensibilities. This leads to specific challenges for NAPA-based orthographies, which are used in day-to-day typographical contexts and thus need to facilitate and support fluent and natural reading. Given that many orthographies of Pacific Northwest languages employ a large number of the specialized letters and diacritical marks found in NAPA, the typographical options available to speakers and writers of these languages are constrained to a reduced set of available typefaces that actually include all of the necessary characters.

We also draw the reader's attention to the large number of orthographies in use

[6] Some sounds that are less relevant to this article are omitted, such as implosives and retroflex consonants.

across the region. For each Indigenous language, there may be multiple competing or coexisting orthographies that correspond to or are favored by different subsections of a community who speak and write a language.[7] Some orthographies are also in a state of semi-standardization, meaning that although rules of standardization have been developed, they are not yet accepted or employed by all language users. There are many reasons why the application of standardized rules may not be stable or consistent, including variation in pronunciation within a community, confusion or disagreement about such standardized forms, and/or technical limitations, which include lacking easy access to an appropriate keyboard for input on different digital devices or a lack of comfort with the tools needed to represent a language in digital spaces.

1.2 Typographical terminology A few notes on terminology are in order, beginning with the distinction between a font and a typeface. While the term 'font' is commonly used by anyone typing using a digital keyboard, it is necessary to disambiguate between a font and a typeface. A *typeface* is a group of glyphs of orthographic elements (e.g., letters, punctuation, and diacritical marks) that are designed to be displayed together. An example of a recognizable typeface is Times New Roman (Morison & Lardent 1932). A *font* refers to a specific variation of a typeface and is a term most relevant to typesetting and web design. Examples of specific fonts would be Times New Roman Regular, Times New Roman Italic, Times New Roman Bold, and Times New Roman Bold Italic. Thus, what is commonly referred to as the 'Times New Roman font' is actually a family of related fonts that comprises the Times New Roman typeface. It is important to distinguish between different fonts because each font is downloaded and installed onto a computer as a separate file, even if these are then clustered together as a font family in the operating system.[8] In this article, we primarily use the term *typeface* because the aspects of type design on which we focus relate to the typeface as a whole and are not specific to individual fonts. A more in-depth discussion of typography is offered in §3.

1.3 References from speakers In preparation for this article, we contacted a number of colleagues who are speakers of Indigenous languages from the Pacific Northwest to inquire about their experiences with typing. In the following sections of this contribution, we will make reference to their comments. The speakers who generously gave us their thoughts and described their experiences include:

Ġvùí/Rory Housty: a Haíłzaqvḷa[9] learner and teacher from the Heiltsuk Nation who works as an educator in the adult immersion courses for the Heiltsuk Tribal Council – Haíłzaqvḷa Revitalization Program and who maintains Haíłzaqvḷa language spaces on social media, including Twitter (@rhousty) and TikTok (@gvuwi).

---

[7] For examples, see FirstVoices (n.d.b).

[8] While font family and typeface are very similar terms, *typeface* is a term that applies to any medium, while *font family* is a term specific to digital environments.

[9] The English name for this language is Heiltsuk.

Marianne Ignace: a professor in the Department of Linguistics and the Department of Indigenous Studies and the director of the Indigenous Languages Program and First Nations Language Centre at Simon Fraser University. She is active in research, teaching, and community-based course delivery of Secwepemctsín, X̲aad Kil, and Sm'algya̲x.[10]

Aiyana Twigg: a learner of Ktunaxa[11] and a recent graduate of the University of British Columbia, as well as a recipient of the 2022 Lieutenant Governor's medal for inclusion, democracy, and reconciliation, who is from the Ktunaxa and Blackfoot Nations.

2. Typographical issues for Indigenous orthographies Despite considerable typographical developments and technical innovations, typing in many Indigenous languages is still much more difficult than typing in dominant, majority, colonial languages such as English, French, and Spanish. The most widespread and severe cause of this difficulty is a continued scarcity of typefaces that contain the full range of necessary glyphs for all of the orthographic elements of an Indigenous language. This typographical scarcity is due to the lack of overlap between characters in many Indigenous orthographies and the characters found in the orthographies of dominant languages for which most typefaces have been developed. Finding typefaces that accommodate the needs of many of the Indigenous languages of the Pacific Northwest is no simple task and generally yields a small number of options. Ignace notes that she and her students are confined to using a small number of typefaces, namely Times New Roman, Arial (Nicholas & Saunders 1982), and Calibri (De Groot 2007) (Marianne Ignace, pers. comm.). Twigg also notes that she often encounters tofu (a small white box indicating that a character is missing from a typeface) when typing in Ktunaxa (Aiyana Twigg, pers. comm.).

Moreover, even among the constrained selection of typefaces that include all of the characters in a specific Indigenous orthography, there are still aspects of certain typefaces that make them unappealing to speakers, leading to typefaces being underused or even going unused. There are two primary factors leading to lack of use: First, elements of glyph design may be in conflict with a community's preferred visual representation for their language's orthography and/or interfere with legibility. Second, the placement of certain glyphs can be incorrect and/or interfere with readability.[12]

Housty, Ignace, and Twigg all flagged such issues. Twigg explains that she primary relies on Arial and Times New Roman because of the challenges that she encounters with glyph design and placement in other typefaces:

[10] The English names for these languages are Shuswap, Haida, and Tsimshian, respectively.

[11] The English name for this language is Kutenai.

[12] Legibility describes the ease with which a reader can identify characters (e.g., letters or diacritical marks) based on the design of their glyphs. Readability describes the ease with which a reader can read a block of text, based on elements of typesetting, such as the amount of space between characters.

> Another problem that I sometimes find is that certain fonts do not support Ktunaxa well. Arial, and Times New Roman are fonts that generally do not have issues when writing Ktunaxa. The size is good (not too big or small), the letters are spaced correctly, and the apostrophe above certain glottal letters is represented properly. However, there have been many other fonts where I have tried to type Ktunaxa […] but the opposite occurs (large letters, spacing is not right). Additionally, the apostrophe above the glottal is not above the letter as it should be, but instead like a regular apostrophe ([for example]: instead of q̓ it looks like q'). This is a detrimental problem, considering that just the apostrophe alone is another version of writing the glottal (ʔ) in Ktunaxa (sometimes people write ki'su'k kyukyit, instead of kiʔsuʔk kyukyit). (Aiyana Twigg, pers. comm.)

Housty and Ignace both echo this complaint, with Housty noting that various diacritical marks move to unexpected positions when using certain typefaces (Ġvǔí/ Rory Housty, pers. comm.), and Ignace describing that when positioned under a lowercase <g>, the underline diacritic used in X̲aad Kil[13] is not sufficiently distinct in many typefaces. In fact, the lowercase <g> + underline diacritic combination <g̲> is unreliably represented in so many typefaces that speakers have resorted to typing it as a capital letter – yielding <G̲> (Marianne Ignace, pers. comm.). Similarly, the diacritic used in Secwepemctsín to indicate that a consonant is glottalized sometimes formats too high above the letter that it modifies (Marianne Ignace, pers. comm.).

For typeface/font users typing in dominant, majority, colonial languages that are well-resourced, glyph design and formatting rarely, if ever, cross their mind. If glyph design does catch their attention, it is only because they have a personal like or dislike of the aesthetics of a certain typeface, and they have the freedom to choose from thousands of other typographical options that better suit their tastes. By contrast, speakers of Indigenous languages with complex orthographic systems are usually very cognizant of typographical challenges and the ever-present likelihood of issues in design and formatting. Indigenous language speakers are all too familiar with how typefaces can inhibit legibility and readability, although they are not always able to identify the root cause behind these issues.

In the following subsections, we offer examples of some of the common typographical issues facing many Indigenous orthographies used in the Pacific Northwest. We provide examples of some of the complaints raised by users looking to type in Indigenous languages and identify aspects of typefaces that can lead to their underutilization.

For reasons of comparison and consistency, all examples that we provide are illustrated using the glottalization accent mark, which looks similar to a quotation mark positioned above or to the upper-right of the letter that it modifies.[14] While any

[13] The <g̲> letter represents a pharyngealized stop in X̲aad Kil.

[14] While many different orthographic mechanisms are used to represent glottalization, the accent mark is the most common in orthographies in the Pacific Northwest.

number of characters could have been chosen to illustrate our point, the glottalization accent mark is the cause of some of the most widespread typographical errors – indeed, it was identified by both Ignace and Twigg as a common source of concern – which in turn can be traced back to a number of sources (described in §4), making it a particularly suitable candidate for exemplifying a variety of issues and causes. For a discussion of typographical issues for Indigenous languages that investigates other typographical elements, see Schillo & Turin (2020).

In addition, the glottalization accent mark is a feature common to many orthographies of Indigenous languages spoken in the Pacific Northwest. Our orthographic examples are drawn from alphabet charts provided on various public, community-run language sites hosted by FirstVoices, an online platform for dictionaries and other language resources for First Nations languages maintained by the First Peoples' Cultural Council, an organization supporting Indigenous language and culture revitalization in British Columbia (First Peoples' Cultural Council n.d.). We are working from the assumption that letters on the FirstVoices alphabet charts were provided by the community language teams who maintain the language sites and thus make use of each orthography's preferred Unicode points. This detail is relevant for the analysis offered in §4. An overview of Unicode is given in §3.1. The typefaces used for the examples in this article are ones produced and disseminated by Apple, the First Peoples' Cultural Council, Google, Microsoft, and SIL International. An explanation of these choices is provided in §3.2.

2.1 Issue 1: Glyph design Elements of glyph design/shape may conflict with a community's preferred visual representation for characters in their orthography. While any typeface user may have typefaces they find more or less aesthetically appealing, the issues discussed in this section transcend aesthetic preferences. Instead, we focus on design elements that interfere with what language communities consider to be integral aspects of the shape and form of a letter/diacritical mark.

In the case of the glottalization accent mark, its shape varies across different typefaces, with glyphs being either a curved, vertical shape or a straight, diagonal shape (see Figure 2 for examples of both). In our experience, the shape most widely considered by speakers with whom we have worked to be 'correct' or true to what is intended is the curved version, which more closely resembles the handwritten accent mark. Some language communities and speakers find both versions acceptable, while others report that the straight shape is less suitable for representing this accent mark.

Figure 2. A straight, diagonal glottalization accent mark in Noto Sans (Google 2013a) and two versions of curved, vertical glottalization accent marks in Aboriginal Sans (Harvey 2004) and Arial (Nicholas & Saunders 1982).

A key reason for the preference for a curved glottalization accent mark is that it is perceived to be more visually discrete, while the straight version more closely resembles an acute accent. This distinction is important for languages whose orthographies have both a glottalization accent mark and an acute accent, such as Halq'eméylem, Ucwalmícwts, nsyilxcən, nłeʔkepmxcin, and Secwepemctsín[15] (First-Voices n.d.c–g), to name a few. Haíłzaqvḷa offers a strikingly good example of a case in which visual distinctions between glottalization and acute accent marks are critical: not only are both accent marks present in the Haíłzaqvḷa orthography, but both appear on letters that would otherwise be identical (Heiltsuk Cultural Education Centre 1986) (see Figure 3).

Figure 3. Haíłzaqvḷa letters that include acute accents (blue) and glottalization accent marks (red). This example is illustrated using Noto Sans (Google 2013a) to demonstrate the visual similarities when using a straight, diagonal glottalization accent mark.

[15] The English names for these languages are Halkomelem, Lillooet, Okanagan, Thompson, and Shuswap, respectively.

2.2 Issue 2: Placement of glyphs As outlined earlier, orthographies of Indigenous languages of the Pacific Northwest are notable for their extensive diacritic inventories, with some letters containing multiple diacritical marks. Many typefaces show persistent and widespread problems with diacritical marks formatting in the wrong position relative to the character that they modify. The most severe form of this issue is when glyphs collide with one another (see Figure 4 for an example), profoundly and negatively impacting readability.

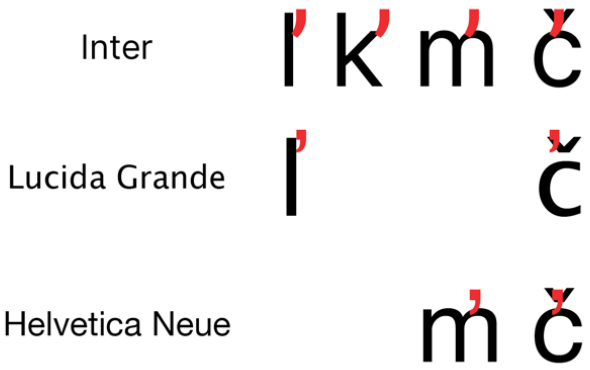

Figure 4. Small letter (lowercase) characters with a glottalization accent mark colliding and one instance of the glottalization accent mark colliding with both a small letter and another diacritical mark. These examples are illustrated using the letters <l̓>, <k̓>, <m̓>, and <č̓>, and with the typefaces Inter (Andersson n.d.), Lucida Grande (Bigelow & Holmes 2000), and Helvetica Neue (D. Stempel AG 1983).

While Figure 4 demonstrates an extreme – albeit, not uncommon – example of incorrect placement of the glottalization accent mark, more subtle, if less visually catastrophic, errors are also very common. The two primary positions in which glottalization accent marks are meant to occur are either directly above the letter they modify or above and to the right of the letter (demonstrated in Figure 5). For the remainder of this article, we will refer to the diacritical mark's position directly above the letter as the *stacked* placement and refer to the position above and to the right as the *upper-right* placement. A community's orthography may make use of only one of these positions or a combination of the two in order to tailor the placement of the glottalization accent mark to the different letters it modifies.

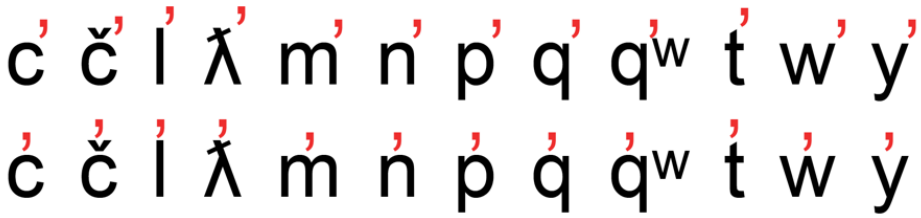

Figure 5. Upper-right positioning (above) versus stacked positioning (below) of the glottalization accent mark. This example uses Arial (Nicholas & Saunders 1982).

Issues arise when the accent mark is positioned[16] in a different place to that which was intended by speakers. For example, an orthography may only accept the stacked placement, but the glottalization accent mark incorrectly positions to the upper-right when using some typefaces – as Twigg identified with her example of q̇ (Aiyana Twigg, pers. comm.). In some cases, the accent mark may even appear in an entirely incorrect position, such as to the left-hand side of a character as shown in Figure 6.

Figure 6. Glottalization accent mark incorrectly positioned to the left of the small letter L in Helvetica Neue (D. Stempel AG 1983).

3. Typographical context To understand the causes of the errors described in §2 – as well as similar issues, such as those described by Housty, Ignace, and Twigg – a foundational understanding of typography and the mechanisms involved in typing is needed. The following subsections offer an overview of the tools and technical structures that support and enable typing, beginning with Unicode (§3.1), followed by typeface design (§3.2), and ending with keyboards (§3.3).

3.1 Encoding and Unicode In order to use a typeface digitally, a coding structure must be in place to support it. With the advent of digital technologies, various methods of character encoding emerged. In fact, so many different methods of encoding were developed that standards became necessary to make typing processes more

[16] The term *position* is used to describe the placement of a diacritical mark in relation to the letter that it modifies.

universal and mutually compatible between computers and operating systems. One of the earliest and most popular encoding standards to take hold was the American Standard Code for Information Interchange (ASCII) in the early 1960s (Haralambous 2007: 29). While ASCII worked well for the English language and orthography, it was constrained by its 7-bit encoding, which limited the number of characters that could be encoded to only 128 (Pine & Turin 2018: 27). As a result, ASCII could only be used for a limited number of languages, and some characters had to be merged, such as the apostrophe, the left quotation mark, and the right quotation mark, all of which were combined into one character (Kuhn 2007).

In the early 1990s, a new standard emerged with the goal of unifying encodings and defining and supporting the characters of global orthographies: the Unicode Standard. Prior to Unicode, characters had different – and sometimes conflicting – encodings in different systems. Unicode did away with this variation, instead assigning each character a unique code point and thus unifying them (Unicode Consortium 2021: xxi). Unicode was also developed on an encoding foundation far more expansive than its predecessors, allowing for many more characters to be included. Unicode's stated goal is to include all of the characters of every orthography in the world (Unicode Consortium 2021: xxi). Thanks to Unicode, which is now the dominant global standard, typing has become more accessible and universal (Haralambous 2007: 53–54; Pine & Turin 2018: 27), and its adoption has been critical for the digital mobilization of Indigenous languages (Carpenter et al. 2021: 145). Housty highlights that Unicode-compliant fonts have been pivotal in making Haíłzaqvḷa more accessible online and on different software and says, "The Unicode font ensures that someone won't need the Haíłzaqv font installed so I don't have to worry about converting something into a pdf before sending. I'm also grateful the Unicode font has allowed us to have an online dictionary" (Ġvùí/Rory Housty, pers. comm.).

3.1.1 Unicode principles The methodology for coding characters used by the Unicode Consortium – the group which maintains the Unicode Standard – is guided by ten principles, four of which are particularly relevant for this article: the principles of *unification*; *dynamic composition*; *semantics*; and *characters, not glyphs* (Unicode, Inc. 2019).

The principle of *unification* dictates that "characters that are equivalent in form are given a single code" (Unicode, Inc. 2019). This means that two orthographic elements that look and function the same can be coded using the same code point. For example, even though the English <j> corresponds to a different phoneme than the German <j>, only one code point is needed because both <j>s look the same and function the same in relation to surrounding text. This principle aims to reduce the total number of code points to only those that are strictly necessary and to avoid duplicates.

We must note that the principle of unification has two interpretations: one that is focused on encodings and another that is more theoretical and orthography-based. The example that we provided regarding the English and German <j> reflects a theoretical standpoint that focuses on the orthographic element itself in relation to the language communities that it serves. For the Unicode Consortium, this principle is

also concerned with unifying prior encodings of the character that were developed under different standards. This is relevant to note because from the outset, Unicode was tasked with unifying characters encoded by many different standards and systems. Richard Gillam (2003: 72) provides an example of two encoding standards that had encoded the same character differently: "Think about ASCII and EBCDIC in American English, for example. The capital letter A encoded by ASCII (as 0x41) is the same capital letter A that is encoded by EBCDIC (as 0xC1), so it makes little sense to have these two source values map to different codes in Unicode." In this sense, the principle of unification dictates that "Unicode unifies character codes from its various source encodings, whenever they can be demonstrated beyond reasonable doubt to refer to the same character" (Gillam 2003: 73).

In this contribution, we focus on the theoretical reading of unification, and we concentrate our analysis on orthographic elements in isolation from and unencumbered by any prior encoding methods. Our choice is motivated by the fact that prior encoding systems were scarce for many Pacific Northwest Indigenous orthographies. Orthographies were instead primarily created and represented using the technology of the time, namely typewriters (Carpenter et al. 2021: 125).[17] For representing orthographies digitally, communities made use of hybrid strategies such as photographing handwritten text and sharing these image files since they did not always have easy access to keyboard input systems or typefaces. In some cases, community members used unrelated characters to substitute those that were not available through ASCII, such as using an <@> to stand in for a <ə> (Pine & Turin 2018: 28). For this reason, prior encoding methods are not always the best sources to reference when determining the full character inventories of Indigenous orthographies. Instead, a familiarity with specific orthographic systems, particularly in their handwritten forms, can often be more reliable and appropriate.

The principle of *dynamic composition* determines that accented forms of orthographic elements can be split into separately coded characters, which can then be combined as *combining characters* (Unicode, Inc. 2019). Like the principle of unification, this principle helps to reduce the total number of coded characters. A single letter in an orthography may comprise multiple components, and Unicode codes each component separately so that users can mix and match orthographic elements. For example, German has three letters that include an umlaut (known typographically as the 'diaeresis' diacritic): <ü>, <ö>, and <ä>. Instead of separately coding each of those three letters, as well as the three letters with no diaeresis (<u>, <o>, and <a>) – which would have generated a total of six coded characters – Unicode instead codes the diaeresis separately as a combining character, which can then be paired with the three letters, thus requiring a total of only four characters rather than six. The diaeresis can then be used productively to combine with other letters as well, for example: <ï> and <ÿ>.

The principle of *semantics* dictates that characters be "well defined, and their definitions clearly show what they are and what they are not" (Haralambous 2007:

[17] For example, characters for the SENĆOŦEN orthography were created by overlapping letters with punctuation marks on an English language typewriter (W̱SÁNEĆ Leadership Council 2019).

59). In creating code points, Unicode is engaged in a process of documenting and defining elements of global scripts and orthographies. This principle requires that necessary and unnecessary aspects of orthographic elements are determined in order to define them.

To understand the Unicode principle requiring that *characters, not glyphs*, be coded, we must first define both terms. A *character* is the core form of an orthographic feature, while a *glyph* is a manifestation of a character. To equate this distinction to linguistic concepts, a character corresponds to the underlying form of an orthographic element, while a glyph is a surface form of a character, and there can be an infinite number of surface forms. Unicode seeks only to code characters.

By way of illustration, each typeface contains a glyph that corresponds to a character. For example, the small letter (lowercase) M character is represented by a distinct glyph in each typeface. The small letter M looks different in Times New Roman and in Arial because each of these typefaces have their own distinct small letter M glyphs. However, each of their glyphs are understood to correspond to the same small letter M character and are thus recognizable as carrying the same meaning. The Unicode principle of coding characters, not glyphs, requires that a single code point be associated with small letter M, and not two separate code points for the glyph of Times New Roman and the glyph of Arial. Like the principle of unification, the principle of coding characters, not glyphs, is intended to minimize the total number of code points to only those that are necessary.

But how does Unicode differentiate between a glyph and a character? How does Unicode know, for example, that a narrow small letter M and a wide small letter M are both glyphs and not two separate characters? This is where the principle of semantics comes into play, and it becomes Unicode's task to clarify the necessary, core elements of a character. Is a specific width a core element of the small letter M character? No, it is not; different widths do not interfere with its recognizability as the same orthographic element. What about the inclusion of serifs: would a small letter M with and without serifs constitute two separate characters? Also no, the inclusion or exclusion of serifs is not character features of Latin-based characters. Do the arches of the small letter M need to be rounded, or could they instead be square-shaped? This depends somewhat on context. Square-shaped arches are indeed less common in designs for a small letter M and may look out of place in some typefaces. However, if this is part of a more expressive typeface that has other elements with similar square shapes in places where it is less standard to do so, then a small letter M would be recognizable as such within the framework of that typeface. What about the number of arches: could a small letter M be made with only one arch? Definitely not, as this would be confused with the small letter N. Thus, two separate character points must be coded for the small letter M and the small letter N. These are the kinds of considerations that must be taken into account when determining the features of a character as compared to a glyph. See Figure 7 for a visual representation of these examples.

Feature	Example glyph	Corresponds to small letter M character
narrow	m Arial	✓
wide	m Montserrat	✓
serifs	m Times New Roman	✓
square arches	m Elevon	✓
one arch	n Arial	✗

Figure 7. Examples in Arial (Nicholas & Saunders 1982), Montserrat (Ulanovsky et al. 2017), Times New Roman (Morison & Lardent 1932), and Elevon (Dalton Maag n.d.).

However, the distinction between a character and a glyph is not always as clear-cut as differentiating between a small letter M and a small letter N. Relating the process to linguistic concepts, we would do well to remember that many phonemes cannot be determined solely through minimal pairs. As such, it can be difficult to adequately define a character, and to disambiguate situations in which two separate characters are required. For example, in the English orthography an <a> and an <ɑ> are both entirely acceptable ways of representing the same letter (a small letter A). However, in NAPA, the <a> and <ɑ> represent different letters that correspond to different vowels (Pullum & Ladusaw 1996: 298). As such, distinguishing aspects of a character from aspects of a glyph is an intensive undertaking, and one that Unicode generally approaches on a case-by-case basis. At the same time, this case-by-case approach leaves room for human error dependent on one's familiarity with an orthography or on the demand for certain characters, which creates increased pressure to ensure that their semantic definitions are correct. As a result, characters and their semantics tend to be very well-defined for dominant, colonial, Western languages

such as English, but are more vulnerable and uncertain for minority, Indigenous, and non-Western orthographies. While certainly not infallible, Unicode must still be commended as one of the only coding standards that has attempted to define and distinguish between characters and glyphs (Haralambous 2007: 54).

3.1.2 Organization of the Unicode Standard The Unicode Standard is organized by script – the overarching writing system under which orthographies develop – hence why English <j> and German <j> are coded as a single character. Characters in different scripts, even characters that look the same, are coded as separate points. For example, the Latin capital letter B <B> and the Cyrillic capital letter Ve <B> are assigned two different points despite appearing visually identical. Unicode maintains a list of *confusable characters* (or *confusables*) – characters that are likely to be confused with other characters due to visual similarities – to help differentiate between such characters as Latin capital letter B and Cyrillic capital letter Ve (Unicode, Inc. n.d.).

Once characters are disambiguated, they are assigned a code point, which indicates their identity. The Unicode Consortium maintains a character database, which contains a full list of their coded characters as well as each character's associated semantic information, which is composed of the character's identity and normative properties (Haralambous 2007: 95). Yannis Haralambous (2007: 95) explains that "Unicode defines the *identity* of a character as the combination of its description and its representative glyph."[18] A character's description is communicated through the name of the code point, together with any additional information about its identity and use that Unicode may provide in a bulleted list, while the representative glyph is shown as an accompanying visual example that Unicode provides for reference. Using <j> as an example, its description is 'Latin small letter J' and its representative glyph is an image of a sample glyph of a small letter J (Unicode, Inc. 2020d; Unicode, Inc. 2021: 6). Haralambous (2007: 95) continues, "The *semantics* of a character are given by its character identity and its *normative properties*."[19] Normative properties communicate aspects of a character's function, such as to which category and script it belongs, among many other defining functions (Haralambous 2007: 95–120).[20]

Characters are assigned to different categories based on their function. For example, *letters* are a category of phonologically meaningful orthographic elements that occupy their own horizontal space – that is, they occupy space along the horizontal baseline on which the text appears (Haralambous 2007: 98). *Nonspacing diacritical marks*, or *combining diacritical marks*, are characters – such as accent marks – that do not appear in isolation; that is, they always modify another character, most often a letter (Haralambous 2007: 99). These diacritical marks are considered nonspacing because they do not occupy their own horizontal space and instead format within the space of the letter that they modify. *Punctuation* are orthographic

[18] Italics maintained from the original quote.

[19] Italics maintained from the original quote.

[20] The Unicode character database is publicly available. See Unicode, Inc. (2021).

elements that are not phonologically meaningful but are instead used to clarify word and sentence structure.

3.2 Typeface design and typefaces for Indigenous languages While some typefaces have been designed with a specific Indigenous language in mind, these are few in number, and such tailor-made typefaces do not exist for most North American Indigenous languages/orthographies. Instead, the majority of typefaces that are used by Indigenous language communities are not designed for these languages – nor with any specific language in mind. They are created with the intention that they can be used for as wide a range of languages as possible by including glyphs for as many characters as possible. The design methodology for these typefaces hinges on referencing Unicode characters and their semantics, as it would otherwise be very difficult for a designer to be familiar with each character for which they design a glyph. Such a design process – which we refer to as a *language-agnostic design process* – is able to rapidly generate glyphs, including glyphs used by languages that may otherwise be overlooked. Additionally, language-agnostic typefaces often include glyphs for multiple different scripts.

The primary producers and distributors of these typefaces include large Web and technology corporations, such as Apple, Google, and Microsoft, as well as Indigenous language-focused organizations such as the First Peoples' Cultural Council and SIL International. Corporations are often interested in these typefaces because they facilitate Web searches, email correspondence, and texting in most of the languages that their user base will speak and write in. While Indigenous language communities may not be the intended primary users, the typefaces produced and distributed by these corporations still include the necessary glyphs to type in many Indigenous languages and are widely accessible, making them appealing to users typing in Indigenous languages. Typefaces produced and distributed by the First Peoples' Cultural Council and SIL International differ slightly, in that their primary audience is Indigenous language users. Thus, the extended glyph sets that make up their typefaces are often tailored to the characters used by Indigenous orthographies that are prevalent within the geographic range covered by the organization.

Throughout this article, analyses and figure examples are all demonstrated using language-agnostic typefaces produced and disseminated by Apple, the First Peoples' Cultural Council, Google, Microsoft, and SIL International, representing some of the most accessible and widely used language-agnostic typefaces in circulation.

3.3 Keyboards Any computer user will be familiar with the physical keyboard, which includes keys that can be pressed to produce a character. Additionally, users who have typed on a computer in more than one language, using more than one orthography, will also be familiar with the idea of a virtual keyboard. A *virtual keyboard* comprises "the table of correspondences between (physical) keys and the (virtual) characters that they generate" (Haralambous 2007: 164). For example, on a standard US layout QWERTY keyboard, the <z> character is located to the right of the SHIFT key. However, in an InScript Nepali keyboard, the character in the same position on the physical keyboard is <श> (Typing Baba n.d.). While these different

characters are typed using the same key on the physical keyboard, the underlying virtual keyboard communicates that the physical key corresponds to different Unicode points, thus generating different characters.

Multiple virtual keyboards can be downloaded onto one computer, and users can toggle between them to type in different languages. In addition to the characters that correspond to the keys of the physical keyboard, more characters can be made available by coding them to correspond to certain key combinations, such as SHIFT + <insert key> or ALT/COMMAND + SHIFT + <insert key>.[21] The same process is at play on a standard US layout QWERTY keyboard, where the key to the right of the SHIFT key generates a small letter Z, while the same key pressed simultaneously as SHIFT generates a capital letter Z. Similarly, by pressing SHIFT + 7, one can generate an <&>. As with the coding of characters themselves, keyboard standards are developed to ensure that users employ the same characters when typing in the standard orthography of their language (Hermes et al. 2016: 280).

4. Identifying the sources of the problems Understanding the mechanisms involved in typing, we can now identify what causes the issues demonstrated in §2. The issues described in §2 relate to inconsistencies that users cannot anticipate other than through extensive experience. As described, the shape of the glottalization accent mark may be inconsistent, as well as its position relative to the character it modifies, although the source of these inconsistencies is not always apparent to users. Indeed, there is no single source to which one can attribute these errors. Instead, we have identified a series of underlying causes that point to all levels of the typing process, including problems originating in the Unicode Standard itself, issues stemming from typeface design, and ones that relate to character input on the user end, which in turn are often impacted by the choice of software. The following subsections offer further details about these various causes.

4.1 Stemming from Unicode: Ambiguity of Unicode points While the glottalization accent mark is recognized by speakers as a single accent mark, there are in fact multiple Unicode character points that can be, and are, used to type it. Three of the most common points used to represent the glottalization accent mark are:

U+0313 – Combining comma above

- This nonspacing diacritical mark is positioned directly above the letter it modifies.
- The character's description notes that it is used in NAPA to signify glottalization.
- For the character's description and representative glyph, see Unicode, Inc. (2021: 33–34). For the character's normative properties, see Unicode, Inc. (2020b).

[21] For an example of one such keyboard for Indigenous languages of the Pacific Northwest, see First Nations and Endangered Languages Program (n.d.).

U+0315 – Combining comma above right

- This nonspacing diacritical mark is positioned to the upper-right of the letter it modifies.
- The character's description does not mention its use to signify glottalization.
- For the character's description and representative glyph, see Unicode, Inc. (2021: 33–34). For the character's normative properties, see Unicode, Inc. (2020c).

U+02BC – Modifier letter apostrophe

- This is a modifier letter, meaning that while it does not appear in isolation, it does occupy its own horizontal space next to the letter it modifies (as opposed to diacritical marks that are nonspacing).
- This character is categorized as a letter, differentiating it from the apostrophe used for punctuation, which has a separate Unicode point.
- The character's description notes that it is used to represent glottal stops and glottalization and that it is a letter used in the alphabets of many languages.
- For the character's description and representative glyph, see Unicode, Inc. (2021: 29–30). For the character's normative properties, see Unicode, Inc. (2020i).

The representative glyphs of all three characters are curved. See Figure 8 for examples of each character.

As explained in §3.1.2, descriptions associated with Unicode points are part of their semantic information, which is composed of their description, a representative glyph, and their normative properties. It is relevant to note that none of the descriptions of these three points explicitly state that the characters are curved. However, each of their names draw a visual correlation to punctuation marks: commas and apostrophes.

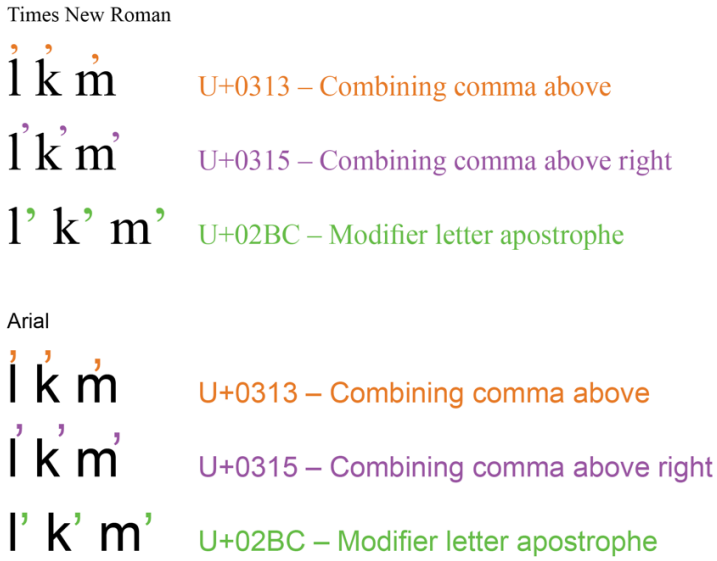

Figure 8. Examples of each of these three Unicode points in both Times New Roman (Morison & Lardent 1932) and Arial (Nicholas & Saunders 1982). *Combining comma above* is shown in orange, *Combining comma above right* in purple, and *Modifier letter apostrophe* in green.

Character descriptions based on surface resemblance to punctuation marks can have significant consequences for the design of resulting glyphs. Although the character descriptions of these points, with the exception of *Combining comma above right*, mention that these characters are used to represent glottalization, it is unlikely that most typeface designers are familiar with the phonological process of glottalization and what its representation entails across various orthographies. Instead, most typeface designers primarily reference the character names, which derive from punctuation-based terminology. Punctuation marks, including commas and apostrophes, are for the most part more forgiving in terms of their expected shape because they do not correspond to pronounceable elements of written language. By contrast, letters and diacritical marks carry more rigid expectations of their predicted forms.[22] For example, while it is acceptable to make a comma straight or curved, it is not customary to make an acute accent straight or curved. Although in many orthographies it is not acceptable for a glottalization accent mark to be either straight or curved – keep-

[22] This is not to say that type cannot bend standards, but rather that successful adaptations are usually more intentional than what we describe in this article and are undertaken with a greater understanding of how characters are used within an orthographic context.

ing in mind that this requirement does vary by language community – its association with commas and apostrophes, stemming from Unicode's character descriptions for these points, leads to an incorrect assumption on the part of some typeface designers that the corresponding glyphs are open to similar levels of flexibility as those seen in the associated punctuation marks. Even the curved representative glyphs that Unicode provides for these points are not enough to prevent misconceptions from arising.

4.2 Stemming from typeface design As a consequence of the confusion related to Unicode's character descriptions, there exists considerable variation in the shapes of the resulting glyphs for *Combining comma above*, *Combining comma above right*, and *Modifier letter apostrophe*. Additionally – although the source of this inconsistency is less clear – it is regrettably quite common for diacritical marks to be placed in an incorrect position relative to the letters that they modify. These inconsistencies exist on both an inter-typeface and intra-typeface level (inconsistencies between different typefaces and inconsistencies within a typeface, respectively), adding an extra layer of confusion for users who view the characters as *syno-characters*[23] – that is, different characters that correspond to the same element within an orthography – and are thus not expecting such variation between them. §4.2.1 and §4.2.2 examine some of the inconsistencies that stem from typeface design through the lens of inter- and intra-typeface glyph comparison.

4.2.1 Inter-typeface inconsistencies While some level of inter-typeface differences in the shape and design of glyphs is to be expected, significant differences in glyph positioning are unpredictable and troublesome. Although *Combining comma above* and *Combining comma above right* clearly describe the position in which the diacritical mark is meant to appear (above the letter it modifies, and above and to the right of the letter it modifies, respectively), and Unicode lists both *Combining comma above* and *Combining comma above right* as confusables for each other (Unicode, Inc. n.d.), in many typefaces the glyphs do not position correctly.

In Figure 6, we offer an example of the glottalization accent mark appearing to the left of the small letter L. To provide more context, that example uses *Combining comma above* and the error appears on the small letter L because it is so narrow in comparison to other characters. This same error occurs in a number of other typefaces when using *Combining comma above*, such as Inter (Andersson n.d.), Lucida Grande (Bigelow & Holmes 2000), and Aboriginal Sans (Harvey 2004). Additionally, for Helvetica Neue (used for the example in Figure 6) (D. Stempel AG 1983), this issue is present for both *Combining comma above* and *Combining comma above right*. This in turn relates to a wider issue of there being no differentiation in the placement of the two diacritical marks. Although the two Unicode points were coded

[23] We have adapted the term *syno-character* from *synoglyph*, which is used to describe glyphs that look different but correspond to the same character, such as the English <a> and <ɑ>. In this case, we instead use the term *syno-character* because Unicode has coded multiple characters in order to facilitate different positionings, although they correspond to the same orthographic element.

separately with the specific intent of providing two different positions for the characters, typefaces such as Helvetica Neue as well as Noto Serif (Google 2013b) do not make this distinction. As a result, it is impossible for the glottalization accent mark to appear in the stacked position when using either of these typefaces. See Figure 9 for examples of these inconsistencies.

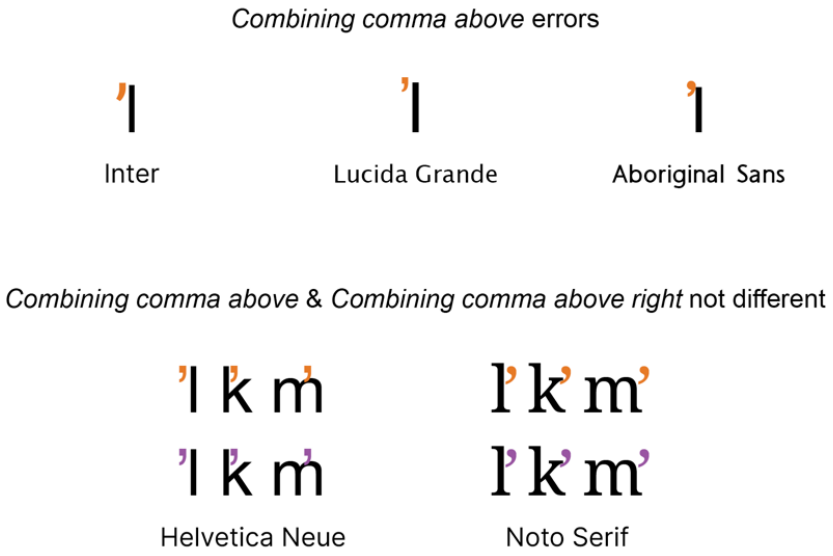

Figure 9. Above: Combining comma above incorrectly formatting to the left of the small letter L in the typefaces Inter (Andersson n.d.), Lucida Grande (Bigelow & Holmes 2000), and Aboriginal Sans (Harvey 2004). Below: The combining diacritical marks of Helvetica Neue (D. Stempel AG 1983) and Noto Serif (Google 2013b) incorrectly appear in the same placement. *Combining comma above* is shown in orange, and *Combining comma above right* is shown in purple.

The source of these errors is less transparent, as the Unicode character descriptions leave little room for misinterpretation with the intended position clearly articulated in each character's names and with Unicode listing them as confusables. Most likely, these errors are oversights that stem from the language-agnostic design process in which typeface designers generate so many glyphs that they are not always able to thoroughly check all of their work.

4.2.2 Intra-typeface inconsistencies Within a typeface, a great deal of variation may exist in the shape and design of the glottalization accent mark, depending on which point is used. Ideally, *Combining comma above*, *Combining comma above right*, and *Modifier letter apostrophe* would all be the same shape. This way, lan-

guage communities whose keyboards make use of a combination of these points would still be able to achieve a unified visual style for this diacritical mark, with the position of the glottalization accent mark tailored to the various letters that it modifies. For many typefaces, this is indeed the case, with each of the corresponding points looking the same, and only formatting in different positions, as in Arial, Arimo (Matteson 2013a), Noto Sans (Google 2013a), Noto Serif, and Gentium Plus (Gaultney 2010).

However, in other typefaces, the glyphs for *Combining comma above*, *Combining comma above right*, and *Modifier letter apostrophe* are differently shaped and differently sized. Often, *Modifier letter apostrophe* is the outlier, with a design that diverges from the two combining diacritical points. For example, the combining diacritical marks in Tahoma (Carter 1994) are both curved, while *Modifier letter apostrophe* is straight. It is also common for *Modifier letter apostrophe* to be shaped roughly the same as the diacritical marks, but larger, as seen in Aboriginal Sans, Aboriginal Serif (Harvey 2007), Doulos SIL (SIL International 1992), Charis SIL (SIL International 1997), Times New Roman, and Tinos (Matteson 2013b). In some cases, all three points have their own distinct designs, such as in Lato (Dziedzic 2010), where *Modifier letter apostrophe* is curved, *Combining comma above* is shaped like a quotation mark, and *Combining comma above right* is straight. Similarly, in Andika (Gaultney & Olsen n.d.), *Modifier letter apostrophe* shares its curved shape with *Combining comma above*, although *Modifier letter apostrophe* is bigger, while *Combining comma above right* is straight. Examples of these intra-typeface inconsistencies are given in Figure 10.

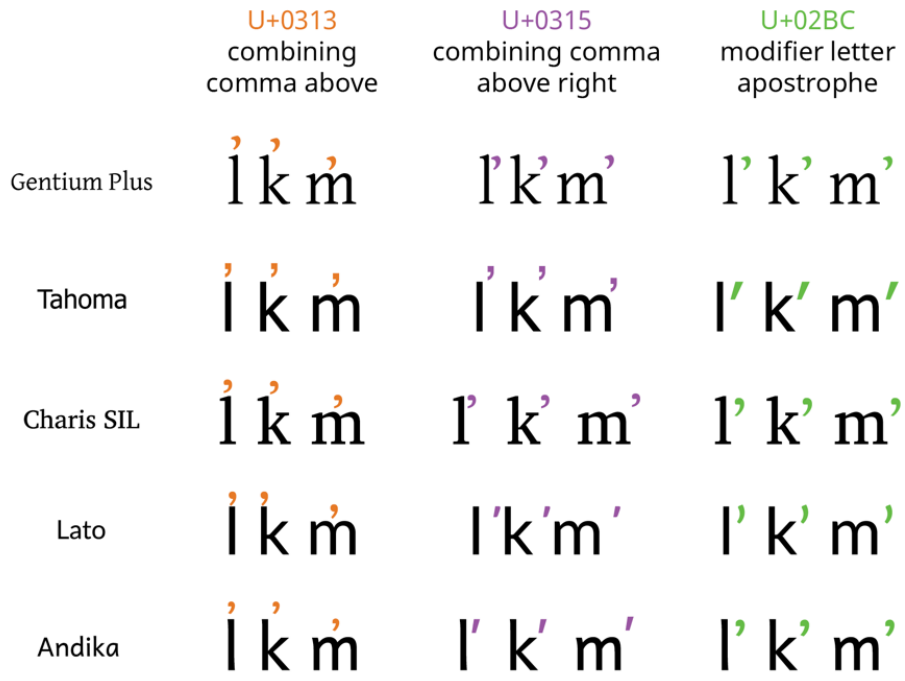

Figure 10. Table demonstrating the different shapes and sizes of the glyphs corresponding to the three Unicode points used for the glottalization accent mark. Gentium Plus (Gaultney 2010) demonstrates the appearance of the points when they have a unified design, while the other four typefaces illustrate the variation that is possible within a typeface. Examples are in Tahoma (Carter 1994), Charis SIL (SIL International 1997), Lato (Dziedzic 2010), and Andika (Gaultney & Olsen n.d.) typefaces.

These intra-typeface inconsistencies are readily apparent in orthographies with glottalization accent marks in both the stacked and upper-right positions (depending on the letter), which therefore need to use a combination of points to realize the accent mark in its various positions. For example, the dii?diitidq[24] orthography uses all three characters to represent the glottalization accent mark (FirstVoices n.d.a). Figure 11 uses Arial to demonstrate the visual unity achieved in a typeface with a unified design for the various points and uses Lato to demonstrate the disunity in a typeface in which each point has a distinctive design.

[24] The English name for this language is Ditidaht or Nitinaht.

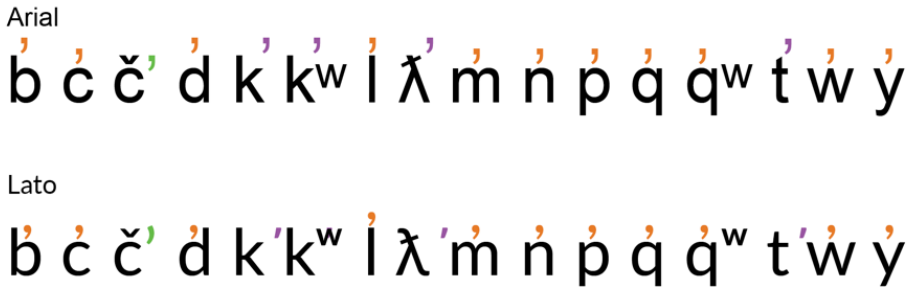

Figure 11. Glottalized consonants in the dii?diitidq orthography. Examples of *Combining comma above* are displayed in orange, *Combining comma above right* in purple, and *Modifier letter apostrophe* in green. The top row uses the Arial typeface (Nicholas & Saunders 1982), while the bottom row uses the Lato typeface (Dziedzic 2010).

4.3 Stemming from users: Punctuation use and software It is common for speakers to substitute punctuation marks to represent the glottalization accent mark when technological constraints or capabilities require it. For example, if a speaker does not have the necessary keyboard downloaded or is unsure how to use it,[25] they may instead approximate the glottalization accent mark using a punctuation mark.[26] *U+2019 – Right single quotation mark* (Unicode, Inc. 2020j; Unicode, Inc. 2021: 214–215) and *U+0027 – Apostrophe* (Unicode, Inc. 2020a; Unicode, Inc. 2021: 2, 4) are the punctuation marks most commonly used to represent the glottalization accent mark.

Users who approximate the glottalization accent mark by using punctuation marks may experience additional inconsistencies caused by their chosen software platform. Of the two punctuation marks commonly employed to approximate the glottalization accent mark, *Right single quotation mark* (commonly called 'curly quote') is most often a curved shape, while *Apostrophe* (commonly called 'straight quote') is often straight and vertical. *Apostrophe* corresponds to the key directly to the right of the semicolon on a standard US layout QWERTY keyboard. This is a remnant of the ASCII era, during which curly quotes and apostrophes were merged. Unicode bases its initial 128 points off of the characters encoded by ASCII (Pine & Turin 2018: 27), an intentional choice to ease the transition from one standardized system to the other and also to ensure backwards compatibility. As a result, key-

[25] A more in-depth discussion about keyboard issues for Indigenous languages is outside the scope of this article and is a topic that deserves a separate paper in its own right. It is worth noting, however, that both Housty and Twigg identified keyboard issues as an obstacle to their abilities to type in their own languages (Ġvùí/Rory Housty, pers. comm.; Aiyana Twigg, pers. comm.).

[26] Similar approximations for other orthographical elements in Heiltsuk are discussed in Pine & Turin (2018: 28).

boards still overapply the use of the *Apostrophe* point as a holdover from the ASCII era (Kuhn 2007). Users who type the *Apostrophe* – which is more easily accessed on a US layout QWERTY keyboard – have a higher likelihood of yielding a straight, vertical glyph.

Although *Apostrophe* is the character used by a US layout QWERTY keyboard, the practice of using *Apostrophe* as a stand-in for quotation marks is frowned upon by typographers (Fleishman 2016). This has led Microsoft Word to include an editor function known as 'smart quotes,' which automatically switches *Apostrophe* to *Right single quotation mark* or *Left single quotation mark* when typed (Microsoft n.d.).[27] A consequence of such software determinism is that, depending on the platform on which one is typing, users may be shown either a straight apostrophe or curly quote when pressing the very same key combination on their keyboard. This adds yet another complication that can lead to the glottalization accent mark being inconsistently shaped.

The root cause of this issue is the use of punctuation to approximate diacritical marks. This practice will likely become less common given the growing desire among many language communities to be able to type in their languages and to use an appropriate keyboard to do so (Chase & Fortier 2021).[28]

5. Discussion As demonstrated in §2 and §4, a great deal of (unwelcome) variation in typographical representations can exist for a single orthographic element, stemming from issues at all levels and stages of the typing process. From confusing and underdefined character semantics in the Unicode Standard, to errors on the part of the typeface designer, to variation stemming from approximation of characters on the user's end, many factors combine to contribute to inconsistencies when typing in Indigenous languages.

It is clear that there is considerable confusion surrounding the character semantics that the Unicode Standard provides for the three points used to represent the glottalization accent mark. We have identified multiple indicators of this confusion, with the shape of the accent mark varying at both inter- and intra-typeface levels, including one of its most common glyph shapes not being perceived as correct by many of the language communities whose orthographies utilize this character. Additionally, glyphs for the glottalization accent mark do not always format in the correct position, despite Unicode providing multiple diacritical marks with the explicit purpose of offering two options for positioning. We argue that confusion about the character's shape stems from an underdefinition of character semantics by the Unicode Consortium combined with a lack of familiarity with the characters on the part of many typeface designers. On the other hand, issues with glyph positioning are primarily the result of an overburdened typeface design process that increases the risk of mistakes being neither identified nor addressed before implementation.

---

[27] While this editor function can be turned off, it is generally preset to be on during software installs.

[28] Using letters and diacritical marks instead of punctuation not only helps to ensure visual consistency but also aids with searchability and recognizability of characters in digital files and applications.

The underdefinition and resulting vulnerabilities of Unicode's character semantics are ultimately the consequence of an imbalance in the application of their guiding principles. The Unicode Consortium has stated that it is impossible to fully satisfy all of their principles simultaneously and that compromises must be made (Haralambous 2007: 58). The Unicode principles we investigate in this article demonstrate the kinds of conflicts that can emerge as a result. The principles of unification and dynamic composition aim to reduce the total number of characters that are coded. The principle of semantics aims to define aspects of a character as thoroughly as possible. In order to define characters rigorously, more context must be given for a character – for example, by coding letter and diacritic combinations as one point, which would result in fewer positioning errors. However, such an approach would violate the principle of dynamic composition. Additionally, semantics would be strengthened by giving greater weight to differences that could be interpreted as defining aspects of a character, such as the need for a curved glottalization accent mark. However, doing so would increase the risk of violating the principle of unification and coding duplicate characters.

The descriptions of the three Unicode characters discussed – *Combining comma above*, *Combining comma above right*, and *Modifier letter apostrophe* – also do not make reference to the shape that the character must be, instead equating them to punctuation marks. This underspecification leads to the variation in glyph shapes seen in many of the typefaces examined in this article. However, when investigating this variation, it becomes apparent that there is no consensus about the 'correct' shape of the glottalization accent mark among the language communities who use it. For some communities, the curved and straight versions of the glottalization accent mark are synoglyphs and thus equally acceptable, while for others, only the curved version is acceptable. That the straight-shaped glottalization accent mark is so unwelcome for some orthographies, thus making it unusable, is a strong indicator that this shape violates core semantics of the character in these cases. Such differences in conceptualizations about the character shape of the glottalization accent mark indicate that there is no singular glottalization accent mark character, and rather that there are two characters with different semantic requirements. While these considerations clarify a great deal about the semantics of the glottalization accent mark, they point to the need for more coded characters, which in turn would put additional pressure on the principle of unification.

We contend that there is a trend of Unicode prioritizing the principles of unification and dynamic composition to the detriment of the principle of semantics, specifically in the case of orthographic elements of Indigenous, non-Western, and minority languages. That is to say, for Indigenous, non-Western, and minority languages, the Unicode Consortium places stronger emphasis on reducing orthographic elements to their core character components in order to code fewer characters. In the process, important semantic aspects that would necessitate the creation of additional character points are overlooked.

To return to the diaeresis example given in §3.1.1, while the diaeresis is coded as a separate, combining diacritical mark, *U+00A8 – Diaeresis* (Unicode, Inc. 2020e; Unicode, Inc. 2021: 8–9), many of the most common orthographic letters that utilize

the diaeresis – specifically those used in the orthographies of European languages – are also coded as one unit. These are known as *precomposed characters* (Unicode, Inc. 2019). While it is not necessary to have separate, precomposed characters for <ä>, <ö>, and <ü>, as all three can be achieved through a letter + diacritical mark combination, all three have them nonetheless: *U+00E4 – Latin small letter A with diaeresis* (Unicode, Inc. 2020f; Unicode, Inc. 2021: 8, 11), *U+00F6 – Latin small letter O with diaeresis* (Unicode, Inc. 2020g; Unicode, Inc. 2021: 8, 12), and *U+00FC – Latin small letter U with diaeresis* (Unicode, Inc. 2020h; Unicode, Inc. 2021: 8, 12). Precomposed characters help to contextualize the use of the diaeresis and ensure that its positioning is well-defined. This in turn leads to less potential for errors in the typeface design process and to glyphs in which the positioning of the diaeresis is intentional and precise. It is regrettable that precomposed characters do not exist for glottalized consonants and for other similar letter + diacritical mark combinations in Indigenous orthographies.

Although members of the Unicode Consortium acknowledge that compromises must be made in order for their system to function, it is our contention that the consequences of such compromises are much more impactful for and evident in characters used by Indigenous orthographies than they are for the orthographies of dominant languages. Compromises made to the semantic integrity of characters in Indigenous languages are more evident precisely because equivalent compromises have not been made for the characters of dominant, European languages, which are much less prone to issues and mistakes akin to those we describe in §2. Even if similar compromises had been made for dominant languages, they would still not have the same impact that they have for Indigenous languages due to the level of familiarity that typeface designers have with dominant languages – a familiarity that enables them to make more informed decisions regarding glyph designs. Unicode runs the risk of reproducing and replicating the same inequalities of resourcing and attention that negatively impact underresourced and historically marginalized languages on a wider societal scale. Simply put, in our analysis, it is clear that not all languages and orthographies are treated equally.

Also underlying many of the issues we describe in this article is the language-agnostic design process used by many typeface designers. This design process does not require typeface designers to have any contextual familiarity with uses of the characters for which they design glyphs, instead relying on the character semantics provided by Unicode when designing glyphs. While this language-agnostic design process has produced many of the available typefaces currently used by Indigenous language communities, this methodology also has unanticipated consequences. By depending almost entirely on Unicode points and their semantics, glyph design is particularly vulnerable for characters whose semantics are underdefined or ambiguous. Because designers of such typefaces are not always familiar with all of the characters for which they are designing glyphs and are instead working in the highly decontextualized space of the Unicode character database, they are not able to identify instances of underdefined or ambiguous character semantics. As a result, unintentional errors can and do emerge. Similarly, because such a large number of glyphs are produced through this language-agnostic design methodology, there is a greater risk of er-

rors being overlooked during the process, such as the positioning errors described in §4.2.1. Thus, although these typefaces are often the only options available to many Indigenous languages, they are not tailored to fit specific languages and their orthographies and do not always accommodate the specific orthographic needs of a community. Juliet Shen (2010: 22), who uses the term *pan-indigenous* to describe what we call language-agnostic typefaces, explains, "Because of the sheer volume of characters contained in these pan-indigenous fonts (numbering in the thousands) […] not all scripts are well served by these sets."

This begs the question: If a language-agnostic design process is so prone to errors, why is it still so popular? The simple answer is that the language-agnostic design process is intended to quickly fill gaps, and generally, it does this quite well. By designing glyphs for characters that are not part of dominant language orthographies, language-agnostic typefaces have provided typographical capabilities and options to language communities whose needs have been – and continue to be – overlooked. Similarly, language-agnostic typefaces are necessary for major operating systems and social media platforms, as they allow users to type in a larger number of languages. We have much to be grateful for from language-agnostic typefaces, as they have been a driving force in offering typographical options and capabilities in an otherwise very constrained operating space. Having recognized this, however, we repeat that their purpose is to fill gaps, not to produce typefaces that demonstrate a nuanced understanding of the use of characters in a specific Indigenous orthography – specifically those characters that Unicode leaves underdefined and that may have conflicting semantics in different orthographies. As a consequence, there will always be errors and shortcomings in how this design methodology is able to serve underresourced orthographies.

For typeface users from Indigenous language communities, the inconsistencies that we identify in this contribution make typing unpredictable and unreliable. For communities who already have a highly restricted array of typographical options available to them due to a paucity of typefaces containing the full range of necessary glyphs, it is particularly demotivating when a typeface that *does* contain all of the necessary glyphs and *should* work is still unable to correctly accommodate the needs of their orthography due to issues with glyph shape and positioning. Because of these constraints and limitations, communities often make compromises of their own and may overlook typographical issues and errors. Such compromises are not the result of any actual ambivalence on the part of language communities, but rather a consequence of the widespread, persistent, and seemingly unavoidable inconsistencies and errors in the typefaces to which they have access.

6. Recommendations Looking toward the future of typography for Indigenous languages, we offer a number of suggestions for both the Unicode Consortium and for typeface designers.

First, we believe that the semantics of some of the characters already coded in the Unicode Standard are not comprehensive enough and that new code points must be added to fill these gaps. While a number of the characters used in Pacific North-

west Indigenous orthographies do not yet have any associated Unicode points,[29] even those that have already been coded may lack integrity in terms of their documented character semantics. For example, although *Combining comma above*, *Combining comma above right*, and *Modifier letter apostrophe* were likely meant to provide a comprehensive array of code points for the glottalization accent mark, their semantics were not clear enough, leaving shortcomings that must be corrected through the introduction of new points.

In the case of the glottalization accent mark, we recommend that for each of the existing code points used to represent this character, a corresponding point be added, which has a description that more clearly specifies that it must be curved. Because some communities have a strong preference for a glottalization accent mark that has a curved shape, while other communities accept both a curved shape and a straight shape, the current points do not adequately address both of these preferences. The current points are usable for communities that accept both shapes, but for communities for whom the curved shape is a necessary character feature, new points should be added, which have descriptions that clearly indicate that the glottalization accent mark must be curved. Such a handling would entail that the glottalization accent mark will be split into two characters: one that is necessarily curved and one that is optionally curved.

Second, we recommend that character descriptions for phonologically meaningful orthographic elements no longer be based on a phenotypic similarity to punctuation marks, as this can generate misinterpretation and misunderstanding. Instead, we argue that character descriptions along the following lines would be more appropriate, descriptive, and unambiguous: "Combining curved glottalization accent above" and "Combining curved glottalization accent above right."

We acknowledge that adding new points and revisiting character semantics will likely be a lengthy process, including both careful reevaluation and subsequent implementation of any new Unicode points. Community keyboards would need to be rewritten and redistributed, and new glyphs would need to be added to typefaces and updated fonts distributed. While this is a worthwhile endeavor, our third suggestion pertains to a shorter-term solution based in typeface design. This solution requires that typeface designers reorient their methodology away from the language-agnostic, decontextualized design process and instead toward a community-partnered, orthography-specific process.

Although language-agnostic typefaces have been pivotal in providing typographical options to minority language communities, they carry within them inherent flaws, which will inevitably continue to produce errors and discontent. The language-agnostic methodology is more prone to errors and oversights. However, even if a designer is able to comprehensively research every character and thor-

[29] As a case in point, a number of Latinized Greek-derived letters are not yet coded, such as a Latin theta and a Latin lambda. At the time of writing, speakers must instead approximate these letters using their Greek-script equivalents, leading to potentially inconsistent stroke modulation and glyph shapes on account of the aesthetic differences of the two scripts (see Tiro Typeworks 2021). In some cases, speakers also approximate characters using unrelated but visually similar characters, such as the Heiltsuk uppercase lambda, which is represented using the *Turned sans-serif capital Y* – an upside-down uppercase Y (Heiltsuk Cultural Education Centre 1986).

oughly check for errors in their work, there will still be unavoidable pitfalls to the language-agnostic design process until the Unicode Standard becomes much more robust in its service of Indigenous orthographic features. As an alternative, a community-partnered, orthography-specific methodology would help to recontextualize characters, particularly ones for which different communities have conflicting and divergent needs that are not yet accommodated by the Unicode Standard, needs that cannot be simultaneously met by a language-agnostic typeface.

Such a community-partnered, orthography-specific design process requires typeface designers and community representatives to work closely and collaboratively to determine the Unicode points being used for a specific community's keyboard and then to define their own character semantics for those points. For example, in the case of the glottalization accent mark, community representatives are best positioned to determine and explain to designers whether or not they require a curved character, after which the typeface designer can create glyphs with these clarified character semantics in mind. Such a design process – one that is undertaken *with* a language community rather than *for* a language community – would enable designers to create more precise, usable, and culturally grounded typefaces that will better serve the language communities and orthographies to which they are tailored.

While such an orthography-specific approach would help to address some of the shortcomings of the Unicode Standard, it is not a permanent solution. To ensure interoperability, the Unicode Standard must be expanded to include points for all necessary characters.

Additionally, we must clarify that community-partnered, orthography-specific typefaces are not intended to replace language-agnostic typefaces, as they each serve different purposes. Language-agnostic typefaces are still necessary for allowing multilingual use across major operating systems, social media platforms, and various software tools. However, language-agnostic typefaces are currently the *only* typefaces available to many Indigenous languages, which is not ideal as they are more prone to errors and are not visually tailored to represent the culture of any specific language community.

Community-partnered, orthography-specific typefaces speak to the goals of language revitalization. Through community consultation, glyphs can be designed in ways that are more culturally grounded, allowing typefaces to articulate the aesthetic sensibilities and visual culture of a community. Similarly, a community-partnered methodology would help to generate a sense of community ownership in the process and in any resulting typefaces. For these reasons, reorienting toward a community-partnered, orthography-specific design process has the potential to be generative and more inclusive.[30]

7. Conclusion Typography for Indigenous languages has seen significant developments in the past two decades (Hermes et al. 2016). Characters can be coded using the Unicode Standard, which has taken significant steps to include orthographic ele-

[30] For an example of a typeface designed in a community-partnered, orthography-specific way and a description of the methodology used for the project, see Shen (2010).

ments used by Indigenous languages. A multitude of typefaces have been developed and disseminated that include glyphs for the various characters used in Indigenous orthographies, most of which are created using a language-agnostic design process. Despite many positive developments, there is still much work to be done.

Many of the Unicode characters that are primarily used by Indigenous orthographies continue to lack rigorous and precise semantic definitions. This in turn leads to unwanted variation in glyphs produced by a language-agnostic typeface design process, as typeface designers are not sufficiently familiar with many of the characters for which they are designing glyphs and, at the same time, are not aware of Unicode's semantic shortcomings for the relevant characters. When typeface designers who create language-agnostic typefaces work with very extensive character inventories for languages that they neither speak nor read, it is of no surprise that errors often creep into the design process.

In order to address these issues, we urge the Unicode Consortium to consider revisiting the character semantics of various Indigenous orthographic elements and explore creating new points in respectful dialogue and close conversation with the communities who seek to type in their languages. Additionally, we encourage typeface designers to move away from a language-agnostic design process and move toward one that is community-partnered and orthography-specific. Such a recalibration would facilitate communities and designers working together to clarify aspects of character semantics that remain ambiguous in the Unicode Standard and to create culturally grounded typefaces. Through such steps, we anticipate that innovative and highly functional typefaces that better serve the needs of Indigenous language communities and their orthographies can be developed.

References

Andersson, Rasmus. n.d. *Inter* [Typeface]. Google Fonts. (https://fonts.google.com/specimen/Inter) (Accessed 2022-05-28.)

Bigelow, Charles & Kris Holmes. 2000. *Lucida Grande* [Typeface]. Bigelow & Holmes.

Carpenter, Jennifer, Annie Guerin, Michelle Kaczmarek, Gerry Lawson, Kim Lawson, Lisa P. Nathan, & Mark Turin. 2021. Locally contingent and community-dependent: Tools and technologies for Indigenous language mobilization. In Link, Adrianna, Abigail Shelton, & Patrick Spero (eds.), *Indigenous languages and the promise of archives*, 125–155. Lincoln: University of Nebraska Press.

Carter, Matthew. 1994. *Tahoma* [Typeface]. Microsoft. (https://docs.microsoft.com/en-us/typography/font-list/tahoma) (Accessed 2022-05-28.)

Chase, Bridget & Kyra Fortier. 2021. Networks of support: How online resources are built, maintained and adapted for community language revitalization needs. (Presented at the 7th International Conference on Language Documentation & Conservation, Honolulu, 4–7 March.) (https://www.youtube.com/watch?v=SWEkPLXOtoc) (Accessed 2021-09-10.)
D. Stempel AG. 1983. *Helvetica Neue* [Typeface].
Dalton Maag. n.d. *Elevon* [Typeface]. Adobe Fonts. (https://fonts.adobe.com/fonts/elevon) (Accessed 2022-05-28.)
Davis, Henry. 2019. Salish languages. In Siddiqi, Daniel, Michael Barrie, Carrie Gillon, Jason D. Haugen, & Éric Mathieu (eds.), *The Routledge handbook of North American languages*, 452–472. New York: Routledge.
De Groot, Luc(as). 2007. *Calibri* [Typeface]. Microsoft. (https://docs.microsoft.com/en-us/typography/font-list/calibri) (Accessed 2022-05-31.)
Dziedzic, Łukasz. 2010. *Lato* [Typeface]. Google Fonts. (https://fonts.google.com/specimen/Lato) (Accessed 2022-05-28.)
First Nations and Endangered Languages Program. n.d. *First Nations Unicode font*. Vancouver, BC: University of British Columbia. (https://fnel.arts.ubc.ca/resources/font/) (Accessed 2021-09-27.)
First Peoples' Cultural Council. 2018a. *Fact sheet: Report on the status of B.C. First Nations languages 2018, third edition*. Brentwood Bay: First Peoples' Cultural Council. (https://fpcc.ca/wp-content/uploads/2020/05/FPCC-Fact-Sheet-Language-Report-2018.pdf) (Accessed 2021-09-10.)
First Peoples' Cultural Council. 2018b. *Report on the status of B.C. First Nations languages*, 3rd edn. Brentwood Bay: First Peoples' Cultural Council. (https://fpcc.ca/wp-content/uploads/2020/07/FPCC-LanguageReport-180716-WEB.pdf) (Accessed 2021-09-10.)
First Peoples' Cultural Council. n.d. *About us: Overview*. (https://fpcc.ca/about-us/overview/) (Accessed 2021-09-10.)
FirstVoices. n.d.a. *dii?diitidq alphabet*. (https://www.firstvoices.com/explore/FV/sections/Data/Wakashan/diidiitidq/dii%C9%81diitidq/learn/alphabet) (Accessed 2021-09-24.)
FirstVoices. n.d.b. *Explore languages*. (https://www.firstvoices.com/explore/FV/sections/Data) (Accessed 2021-06-10.)
FirstVoices. n.d.c. *Halq'eméylem alphabet*. (https://www.firstvoices.com/explore/FV/sections/Data/Salish/Halkomelem/Halq'em%C3%A9ylem/learn/alphabet) (Accessed 2021-06-10.)
FirstVoices. n.d.d. *Líl̓wat alphabet.* (https://www.firstvoices.com/explore/FV/sections/Data/St%CC%95%C3%A1t%CC%95imc/Lil'wat/L%C3%ADl%CC%93wat/learn/alphabet) (Accessed 2021-06-10.)
FirstVoices. n.d.e. *nsyilxcən alphabet*. (https://www.firstvoices.com/explore/FV/Workspaces/Data/nsyilxc%C9%99n/nsyilxc%C9%99n/nsyilxc%C9%99n/learn/alphabet) (Accessed 2021-06-10.)
FirstVoices. n.d.f. *nłeʔkepmxcin alphabet*. (https://www.firstvoices.com/explore/FV/sections/Data/scw%CC%93exmxc%C3%8Cn/n%C5%82e%CA%94kepmxcin/n%C5%82e%CA%94kepmxcin/learn/alphabet) (Accessed 2021-06-10.)

FirstVoices. n.d.g. *Secwepemc alphabet*. (https://www.firstvoices.com/explore/FV/sections/Data/Secwepemc/Secwepemctsin/Secwepemc/learn/alphabet) (Accessed 2021-06-10.)
Fleishman, Glenn. 2016. Has the Internet killed curly quotes? *The Atlantic*, December 28. (https://www.theatlantic.com/technology/archive/2016/12/quotation-mark-wars/511766/) (Accessed 2021-09-10.)
Gaultney, Victor. 2010. *Gentium Plus* [Typeface]. SIL International. (https://software.sil.org/gentium/) (Accessed 2022-05-28.)
Gaultney, Victor & Annie Olsen. n.d. *Andika* [Typeface]. SIL International. (https://software.sil.org/andika/) (Accessed 2022-05-28.)
Gillam, Richard. 2003. Architecture: Not just a pile of code charts. In *Unicode demystified: A practical programmer's guide to the encoding standard*, 61–110. Boston: Addison-Wesley Professional.
Google. 2013a. *Noto Sans* [Typeface]. Google Fonts. (https://fonts.google.com/noto/specimen/Noto+Sans) (Accessed 2022-05-28.)
Google. 2013b. *Noto Serif* [Typeface]. Google Fonts. (https://fonts.google.com/noto/specimen/Noto+Serif) (Accessed 2022-05-28.)
Haralambous, Yannis. 2007. *Fonts & encodings*. Sebastopol, CA: O'Reilly Media, Inc.
Hargus, Sharon. 2010. Athabaskan phonetics and phonology. *Language and Linguistics Compass* 4(10). 1019–1040. doi:10.1111/j.1749-818x.2010.00245.x
Harvey, Christopher. 2004. *Aboriginal Sans* [Typeface]. LanguageGeek. (http://www.languagegeek.com/font/fontdownload.html) (Accessed 2022-05-28.)
Harvey, Christopher. 2007. *Aboriginal Serif* [Typeface]. LanguageGeek. (http://www.languagegeek.com/font/fontdownload.html) (Accessed 2022-05-28.)
Heiltsuk Cultural Education Centre. 1986. *Alphabet chart*. Bella Bella: Heiltsuk Cultural Education Centre. (Text by John Rath. Illustrated by Shirl Hall.) (https://heiltsuk.arts.ubc.ca/alphabet-chart-rory-housty/) (Accessed 2021-06-10.)
Hermes, Mary, Phil Cash Cash, Keola Donaghy, Joseph Erb, & Susan Penfield. 2016. New domains for Indigenous language acquisition and use in the USA and Canada. In Coronel-Molina, Serafín M. & Teresa L. McCarty (eds.), *Indigenous language revitalization in the Americas*, 269–291. New York: Routledge.
Kuhn, Markus. 2007. ASCII and Unicode quotation marks. (https://www.cl.cam.ac.uk/~mgk25/ucs/quotes.html) (Accessed 2021-09-10.)
Matteson, Steve. 2013a. *Arimo* [Typeface]. Google Fonts. (https://fonts.google.com/specimen/Arimo) (Accessed 2022-05-28.)
Matteson, Steve. 2013b. *Tinos* [Typeface]. Google Fonts. (https://fonts.google.com/specimen/Tinos) (Accessed 2022-05-28.)
Microsoft. n.d. Smart quotes in Word. (https://support.microsoft.com/en-us/office/smart-quotes-in-word-702fc92e-b723-4e3d-b2cc-71dedaf2f343) (Accessed 2021-09-10.)
Morison, Stanley & Victor Lardent. 1932. *Times New Roman* [Typeface]. Philadelphia: Monotype Corporation. (https://docs.microsoft.com/en-us/typography/font-list/times-new-roman) (Accessed 2022-05-31.)

Nicholas, Robin & Patricia Saunders. 1982. *Arial* [Typeface]. Philadelphia: Monotype Corporation. (https://docs.microsoft.com/en-us/typography/font-list/arial) (Accessed 2022-05-31.)

Pine, Aidan & Mark Turin. 2018. Seeing the Heiltsuk orthography from font encoding through to Unicode: A case study using convertextract. In Soria, Claudia, Laurent Besacier, & Laurette Pretorius (eds.), *Proceedings of the LREC 2018 workshop "CCURL 2018 – Sustaining knowledge diversity in the digital age,"* Miyazaki, 12 May, 27–30.

Pullum, Geoffrey K. & William A. Ladusaw. 1996. *Phonetic symbol guide*. Chicago: University of Chicago Press.

Schillo, Julia & Mark Turin. 2020. Applications and innovations in typeface design for North American Indigenous languages. *Book 2.0* 10(1). 71–98.

Shen, Juliet. 2010. Aesthetic innovation in Indigenous typefaces: Designing a Lushootseed font. *Glimpse* (7). 20–29.

SIL International. 1992. *Doulos SIL* [Typeface]. SIL International. (https://software.sil.org/doulos/) (Accessed 2022-05-28.)

SIL International. 1997. *Charis SIL* [Typeface]. SIL International. (https://software.sil.org/charis/) (Accessed 2022-05-28.)

Tiro Typeworks (@TiroTypeworks). 2021. Why we need a Latin theta character. The Greek letter takes stroke modulation and shapes that are not appropriate for Latin script orthographies. *Twitter*, July 2. (https://twitter.com/TiroTypeworks/status/1411006460318740482) (Accessed 2021-09-26.)

Typing Baba. n.d. *Nepali keyboard for online Nepali typing*. (https://www.typingbaba.com/keyboard/online-nepali-keyboard.php) (Accessed 2021-09-27.)

Ulanovsky, Julieta, Sol Matas, Juan Pablo del Peral, & Jacques Le Bailly. 2017. *Montserrat* [Typeface]. Google Fonts. (https://fonts.google.com/specimen/Montserrat) (Accessed 2022-05-28.)

Unicode Consortium. 2021. *The Unicode® standard: Version 14.0 – Core specifications*. Mountain View: Unicode Consortium. (https://www.unicode.org/versions/Unicode14.0.0/UnicodeStandard-14.0.pdf) (Accessed 2022-02-28.)

Unicode, Inc. 2019. *The Unicode® standard: A technical introduction*. (https://unicode.org/standard/principles.html) (Accessed 2021-07-11.)

Unicode, Inc. 2020a. 0027. In *Unicode utilities: Character properties*. (Part of Unicode version 13.0.) (https://util.unicode.org/UnicodeJsps/character.jsp?a=0027) (Accessed 2021-06-11.)

Unicode, Inc. 2020b. 0313. In *Unicode utilities: Character properties*. (Part of Unicode version 13.0.) (https://util.unicode.org/UnicodeJsps/character.jsp?a=0313) (Accessed 2021-06-11.)

Unicode, Inc. 2020c. 0315. In *Unicode utilities: Character properties*. (Part of Unicode version 13.0.) (https://util.unicode.org/UnicodeJsps/character.jsp?a=0315) (Accessed 2021-06-11.)

Unicode, Inc. 2020d. 006A. In *Unicode utilities: Character properties*. (Part of Unicode version 13.0.) (https://util.unicode.org/UnicodeJsps/character.jsp?a=006A) (Accessed 2021-06-11.)

Unicode, Inc. 2020e. 00A8. In *Unicode utilities: Character properties*. (Part of Unicode version 13.0.) (https://util.unicode.org/UnicodeJsps/character.jsp?a=00A8) (Accessed 2021-09-25.)
Unicode, Inc. 2020f. 00E4. In *Unicode utilities: Character properties*. (Part of Unicode version 13.0.) (https://util.unicode.org/UnicodeJsps/character.jsp?a=00E4) (Accessed 2021-09-10.)
Unicode, Inc. 2020g. 00F6. In *Unicode utilities: Character properties*. (Part of Unicode version 13.0.) (https://util.unicode.org/UnicodeJsps/character.jsp?a=00F6) (Accessed 2021-09-10.)
Unicode, Inc. 2020h. 00FC. In *Unicode utilities: Character properties*. (Part of Unicode version 13.0.) (https://util.unicode.org/UnicodeJsps/character.jsp?a=00FC) (Accessed 2021-09-10.)
Unicode, Inc. 2020i. 02BC. In *Unicode utilities: Character properties*. (Part of Unicode version 13.0.) (https://util.unicode.org/UnicodeJsps/character.jsp?a=02BC) (Accessed 2021-06-11.)
Unicode, Inc. 2020j. 2019. In *Unicode utilities: Character properties*. (Part of Unicode version 13.0.) (https://util.unicode.org/UnicodeJsps/character.jsp?a=2019) (Accessed 2021-06-11.)
Unicode, Inc. 2021. *The Unicode standard, version 14.0: Archived code charts*. (https://www.unicode.org/Public/UCD/latest/charts/CodeCharts.pdf) (Accessed 2021-09-24.)
Unicode, Inc. n.d. *Confusable characters*. (https://util.unicode.org/UnicodeJsps/confusables.jsp) (Accessed 2021-11-29.)
W̱SÁNEĆ Leadership Council. 2019. *Finding our talk episode* - SENĆOŦEN [Video]. YouTube. (https://www.youtube.com/watch?v=Onh9lZwkvto) (Accessed 2021-10-05.)

Julia Schillo
schilloji@gmail.com
orcid.org/0000-0003-0560-2564

Mark Turin
mark.turin@ubc.ca
orcid.org/0000-0002-2262-0986