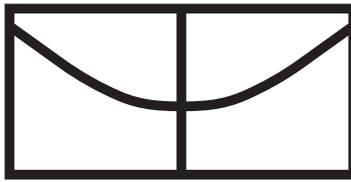


UNIVERSIDADE DE BRASÍLIA
FACULDADE DE COMUNICAÇÃO
PROGRAMA DE PÓS-GRADUAÇÃO
LINHA DE IMAGEM, SOM E ESCRITA

RAFAEL DIETZSCH

A tipografia das línguas indígenas brasileiras

Brasília, 2021



UNIVERSIDADE DE BRASÍLIA
FACULDADE DE COMUNICAÇÃO
PROGRAMA DE PÓS-GRADUAÇÃO
LINHA DE IMAGEM, SOM E ESCRITA

RAFAEL DIETZSCH

A tipografia das línguas indígenas brasileiras

Tese apresentada ao Programa de Pós-Graduação em
Comunicação da Universidade de Brasília como requisito
para obtenção do grau de Doutor em Comunicação
Social, pela linha de Imagem, Som e Escrita.

Orientador: Prof. Dr. Gustavo de Castro da Silva
Co-orientador: Prof. Dr. Rogério José Camara

Brasília, 2021

RAFAEL DIETZSCH

A TIPOGRAFIA DAS LÍNGUAS INDÍGENAS BRASILEIRAS

Tese apresentada ao Programa de Pós-Graduação
em Comunicação Social da Universidade de Brasília,
e defendida sob avaliação da Banca Examinadora constituída por:

Prof. Dr. Gustavo de Castro da Silva
Orientador
Universidade de Brasília

Profa. Dra. Gabriela Pereira de Freitas
Membro Interno
Universidade de Brasília

Prof. Dr. Jorge Domingues Lopes
Membro Externo
Universidade Federal do Pará

Prof. Dr. Hugo Cristo Sant'Anna
Membro de Escolha Livre
Universidade Federal do Espírito Santo

Profa. Dra. Priscila Monteiro Borges
Suplente
Universidade de Brasília

Este trabalho é dedicado
aos povos indígenas brasileiros

Agradecimentos

À minha amada, Aline, pelo amor, carinho, apoio e paciência.
Obrigado por estar ao meu lado nesta travessia. Sem você eu não teria conseguido.

Ao meu filho querido, Arthur. A criança mais parceira e querida deste mundo.

Aos meus pais, João e Zeni, pelo amor e apoio incondicional em todos os sentidos. Sempre. E por me ensinarem que a educação é a coisa mais importante que um homem pode carregar consigo.

À minha irmã Laurinha e ao Chico, ao João Luís e ao Mateus, pelo carinho.

Ao professor Gustavo de Castro, pela confiança e por acreditar nesta pesquisa.

Ao professor Rogério Camara, meu coorientador, por acreditar nesta pesquisa e por me mostrar caminhos que eu sozinho não teria achado.

À Tânia Montoro, pelas contribuições no momento da qualificação.

Ao meu querido amigo André Maya. Pela troca de ideias de alto nível, sempre. Mas, sobretudo, pela amizade e parceria.

À professora Selma Oliveira e ao professor Wagner Rizzo, por acreditarem nesta ideia, incondicionalmente, desde o primeiro dia.

Ao amigo Kalapi Gajjar-Bordawekar, pela amizade e pelo suporte no desenvolvimento. Sem você, estas ferramentas não existiriam.

Aos amigos Gustavo Soares e Eduardo Borges, pelas contribuições para a automação dos tediosos processos de levantamento de dados desta tese.

Às amigas Priscila Borges e Bruna Moreira, pelo olhar atento e pelas valiosas críticas ao longo do processo.

Ao amigo Marcus Vinícius Garcia, por acreditar no projeto e por abrir as portas para a colaboração com o Inventário Nacional da Diversidade Linguística.

A Pablo Cosgaya e Marcela Romero, pela parceria, pelo intercâmbio de ideias e por serem um ponto de referência na comunidade de designers de tipos latino-americana.

Aos membros da banca, Gabriela Freitas, Jorge Domingues e Hugo Cristo. Obrigado pela colaboração e participação.

Aos colegas (professores e técnicos) da Faculdade de Comunicação, do Departamento de Audiovisuais e Publicidade, e do Departamento de Design, pelo apoio.

A Ronan Nascimento, por me ajudar a conter a ansiedade e manter o foco.

A Debbie Anderson, Yannis Haralambous, Gerry Leonidas, Eric Fischer, Sebastian Drude, John Hudson, Thiago Chacon, Jorge Lopes, Rafael Saraiva, Henrique Beier, Daniela Franca, Fernanda Werneck e Giulia Santos, pelas contribuições das mais diversas formas.

Resumo

DIETZSCH, Rafael. **A tipografia das línguas indígenas brasileiras**. 2021.

Tese (Doutorado em Comunicação) – Programa de Pós-Graduação em Comunicação, Faculdade de Comunicação, Universidade de Brasília, 2021.

Orientador: Prof. Dr. Gustavo de Castro da Silva

Defesa: 30/08/2021.

A proposta deste trabalho é a estruturação de um estudo de referência sobre a tipografia das línguas indígenas brasileiras, tema sobre o qual há poucas pesquisas e escassa documentação científica. Sob a ótica da comunicação visual, este trabalho procura preencher essas lacunas, descrevendo o Estado da Arte e propondo um método para desenvolvimento de ferramentas de escrita para essas línguas. ¶ Esta tese parte de um problema decorrente da particularidade da ortografia de muitas línguas indígenas brasileiras, que utilizam caracteres acentuados pouco comuns ou, até mesmo, ausentes em grande parte das fontes tipográficas digitais disponíveis. Do ponto de vista tipográfico, é preciso atenção a essas particularidades, a fim de facilitar e viabilizar a comunicação textual (em meios digitais) no cotidiano das minorias indígenas, em seus idiomas originários. Nesse sentido, esta pesquisa propõe-se à construção de um modelo de investigação que possa abraçar a tipografia das línguas indígenas brasileiras e, para tal, propõe dois principais desenvolvimentos. ¶ Primeiro, uma descrição detalhada da metodologia desenvolvida para o levantamento de dados desta pesquisa, cujo propósito essencial é a construção de uma catalogação sistemática e atualizada das ortografias das línguas indígenas brasileiras. A catalogação ortográfica tem como objetivo principal a especificação de uma lista de requisitos para o suporte linguístico de fontes para línguas indígenas brasileiras. ¶ Em segundo lugar, o desenvolvimento de fontes tipográficas digitais e layouts de teclado, ferramentas necessárias para a composição e editoração de textos em línguas indígenas brasileiras. ¶ Finalmente, deu-se atenção especial a questões técnicas bastante ignoradas no campo das línguas indígenas, especialmente em relação à codificação dessas línguas. A compreensão dessas questões técnicas nos ajuda a entender o contexto de uso dessas codificações e da comunicação nessas línguas. ¶ Como resultado, um vasto corpo documental foi coletado, analisado e discutido, com o intuito de ajudar na compreensão das práticas adotadas na tipografia das línguas indígenas, não somente no processo de design de tipos e codificação, mas também nas relações com a comunicação.

Palavras-chave: Tipografia; Design de Tipos; Codificação; Línguas Indígenas Brasileiras; Ortografia; Inclusão Digital; Comunicação Visual; Comunicação Digital

Abstract

The purpose of this work is to provide a framework for the study of the typography of Brazilian indigenous languages, a topic on which there is little research and little scientific documentation. This work addresses this gap from the perspective of visual communication, describing the state-of-the-art in practice and proposing a method for developing writing technologies for these languages. ¶ This thesis is rooted in the challenging particularities of the orthographies of many Brazilian indigenous languages; these employ accented characters that are unusual, or even absent, in most of the digital typefaces available today. From a typographic point of view, attention must be paid to these particularities, in order to facilitate textual communication (in digital media) in the daily lives of indigenous minorities, in their original languages. In this regard, this research proposes the construction of a research framework that can embrace the typography of Brazilian indigenous languages and, for this purpose, it proposes two main developments. ¶ First, a detailed description of the methodology developed for the data collection of this research, whose essential purpose is the construction of a systematic and updated cataloging of the orthographies of Brazilian indigenous languages. ¶ Second, the development of digital typefaces and keyboard layouts, tools that are essential for the composition and edition of texts in Brazilian indigenous languages. ¶ Finally, special attention was given to technical issues that were largely neglected in the field of indigenous languages, especially in relation to the encoding of these languages. Engaging with these technical issues is critical for understanding the context of using these encodings and communicating in these languages. ¶ To support this work, a vast body of documents was collected, analysed and discussed, with the aim of helping to understand the practices adopted in typography of indigenous languages, not only in the type design and encoding process, but also in the relationships with the field of communication.

Keywords: Typography; Typeface Design; Encoding; Brazilian indigenous languages; Orthography; Digital Inclusion; Visual Communication; Digital Communication

Lista de figuras

- Figura 1** População indígena total em 2010
- Figura 2** Línguas indígenas brasileiras no censo de 2010
- Figura 3** ESCREVER LEGENDA
- Figura 4** *Arte de grammatica da lingoa mais vsada na costa do Brasil*
- Figura 5** Manuscrito maia
- Figura 6** *Breve compendio...*
- Figura 7** Página inicial da gramática da gramática do guarani
- Figura 8** Folha de rosto (emespanhol) de um livro de doutrinas em guarani, de autoria do Cacique Yapuguay
- Figura 9** Página do livro de doutrinas em guarani, de autoria do Cacique Yapuguay
- Figura 10** *Mapa Etno-histórico do Brasil e Regiões Adjacentes*
- Figura 11** Diagrama do sistema geral de comunicação, conforme Shannon & Weaver
- Figura 12** Interpretação de dados numéricos, de acordo com diferentes padrões de codificação
- Figura 13** Exemplo de página de código
- Figura 14** Desenvolvimento da escrita chinesa: da escrita arcaica à escrita moderna
- Figura 15** Desenvolvimento da escrita egípcia
- Figura 16** Desenvolvimento da escrita chinesa
- Figura 17** Exemplo de digrafia em de placa de sinalização de trânsito
- Figura 18** Amostra de caracteres do alfabeto fonético internacional

- Figura 19** Espécimes de sistemas de escrita, compostos em Noto Sans Regular
- Figura 20** Variações da letra M
- Figura 21** Glifos com formas idênticas na maiúscula latina, grega e cirílica
- Figura 22** Sequência de composição de caractere pré-composto, com seus respectivos códigos Unicode
- Figura 23** Sequência de composição de caractere acentuado, formado por letra base e marca combinada, e seus respectivos códigos Unicode
- Figura 24** Exemplo de composição demonstrando as potencialidades das marcas combinadas da fonte Brill Roman
- Figura 25** Ilustração que explica o conceito de modularidade do tipo móvel
- Figura 26** Gaveta de tipos móveis e componedor
- Figura 27** Linotype Machine
- Figura 28** Representação esquemática do mecanismo de fotocomposição
- Figura 29** Representação esquemática do mecanismo de fotocomposição
- Figura 30** Exemplo de desenho original da fonte Demos, de Gerard Unger, mostrando a grade de construção modular do CRT
- Figura 31** Mesa de trabalho de um designer de tipos do final dos anos 1980
- Figura 32** Desenhos de Erik Spiekermann, de junho de 1985, em papel transparente, para digitalização no Ikarus
- Figura 33** Diferença entre curvas de Bézier cúbicas (PostScript) e quadráticas (TrueType)
- Figura 34** Quadrado de Polybius
- Figura 35** Telégrafo de Chappe, desenho a nanquim
- Figura 36** Telégrafo de Chappe, gravura
- Figura 37** Mapas das primeiras redes de telégrafos óticos no século XIX
- Figura 38** Telégrafo de Hughes
- Figura 39** Layout do teclado de Hughes
- Figura 40** Teclado de Baudot
- Figura 41** Fita perfurada representando o código Baudot

- Figura 42** Código Baudot de 5 unidades
- Figura 43** Esquema de codificação de Murray, conforme registrado em patente de 1899
- Figura 44** Murray keyboard perforator model No. 5-PM. Fabricado por Reid Brothers, Engineers, Limited, Camden, Inglaterra, 1914–1916
- Figura 45** Layout de teclado de Murray, de 1911
- Figura 46** International Telegraph Alphabets
- Figura 47** Proposta de layout de teclado para codificações de seis bits
- Figura 48** International Telegraph Alphabet No. 2, arranjado em ordem binária
- Figura 49** codificação FIELDATA
- Figura 50** Codificação ASCII, 1967
- Figura 51** Estrutura da codificação ISO 2022
- Figura 52** Codificação ISO 8859-1
- Figura 53** Exemplo de codificação de oito bits: Windows Codepage 1252
- Figura 54** Espécime de caracteres acentuados gerados a partir de um sistema utilizando codificação ASCII
- Figura 55** Monotype, estojo de matrizes
- Figura 56** Planos do Unicode
- Figura 57** Blocos do Unicode
- Figura 58** Tabela de código Co Controls and Basic Latin
- Figura 59** Tabela de código CI Controls and Latin-1 Supplement
- Figura 60** Codificação de uma sequência de bits em caracteres
- Figura 61** Captura de tela do software Glyphs
- Figura 62** Processo simplificado de funcionamento da codificação de caracteres
- Figura 63** Ordenação bidirecional
- Figura 64** Tabela de código Spacing Modifier Letters
- Figura 65** Tabela de código Combining Diacritical Marks
- Figura 66** Métrica do acento agudo, U+00B4, do bloco Spacing Modifier Letters
- Figura 67** Métrica do acento agudo, U+0301, do bloco Combining Diacritical Marks

- Figura 68** Simulação da diminuição progressiva dos valores das proteções laterais, até o zero
- Figura 69** Skolar (type specimen)
- Figura 70** OpenType features: refinamentos tipográficos
- Figura 71** Substituições e ligaturas contextuais
- Figura 72** Reordenamento de caracteres
- Figura 73** Bidirecionalidade
- Figura 74** Uma fonte é feita de tabelas
- Figura 75** As tabelas GSUB e GPOS contém features
- Figura 76** Os features contém lookups
- Figura 77** Hierarquia de organização das tabelas na fonte OpenType
- Figura 78** Fluxo do processamento linguístico no OpenType
- Figura 79** Variações de Lookups de substituição
- Figura 80** GSUB Lookup. Tipo 1. Substituição simples
- Figura 81** GSUB Lookup. Tipo 2. Substituição múltipla
- Figura 82** GSUB Lookup. Tipo 3. Substituição de alternativos
- Figura 83** GSUB Lookup. Tipo 4. Substituição de ligatura
- Figura 84** Lookups de posicionamento
- Figura 85** GPOS Lookup. Tipo 4. Posicionamento de marca em relação à base
- Figura 86** GPOS Lookup. Tipo 6. Posicionamento de marca em relação a outra marca
- Figura 87** Figura 87: Teclado virtual no iOS 14.6. Fonte: do autor.
- Figura 88** Layout de teclado ABNT. Fonte: TECLADO, 2021.
- Figura 89** Layout de teclado ANSI. Fonte: KEYBOARD LAYOUT, 2021.
- Figura 90** Exemplos de medições dos livros de Manutius.
- Figura 91** Exemplos de medições dos livros de Manutius
- Figura 92** Molde de fundição de tipos
- Figura 93** Matrizes originais de Claude Garamond, séculoXVI

- Figura 94** A parte de cima da imagem mostra o molde conforme gravura da série feita para a Académie Royale des Sciences, na França, na última década do século XVI
- Figura 95** Exemplos de medições das frações do quadratim
- Figura 96** Na parte inferior da gravura é possível ver os tipos e espaços em perspectiva isométrica (Fig.4), o componedor (Fig.5) e um bloco justificado (Fig.6)
- Figura 97** Exemplos de medições dos livros de Manutius
- Figura 98** Modelo arquetípico (para o tipo romano)
- Figura 99** Critérios de legibilidade para ortografias
- Figura 100** Co Controls and Basic Latin. CI Controls and Latin-I Supplement
- Figura 101** Amostra de texto na língua baniwa-coripaco, utilizando somente caracteres do bloco Latin-I
- Figura 102** Amostra de texto na língua ikpeng, utilizando somente caracteres do bloco Latin-I
- Figura 103** Latin Extended-A
- Figura 104** Latin Extended-B
- Figura 105** Latin Extended Additional
- Figura 106** Amostra de texto na língua kaiabi, utilizando caracteres dos blocos Latin-I, Latin Extended-A e Latin Extended Additional
- Figura 107** IPA Extensions
- Figura 108** Amostra de texto na língua guarani kaiowá
- Figura 109** Combining Diacritical Marks
- Figura 110** Amostra de texto na língua tuyuka.
- Figura 111** Amostra de texto na língua mehinaku
- Figura 112** Amostra de texto na língua suiá
- Figura 113** Amostra de texto na língua mehinaku
- Figura 114** Amostra de texto na língua mehinaku
- Figura 115** Amostra de texto na língua tuyuka
- Figura 116** Amostra de texto na língua tuyuka

- Figura 117** Amostra de texto na língua waujá
- Figura 118** Amostra de texto na língua yanomami
- Figura 119** Amostra de texto na língua yudja
- Figura 120** Amostra de texto na língua yudja
- Figura 121** Amostra de texto na língua wajãpi
- Figura 122** Esboços para o tipo Kingfisher, de Jeremy Tankard
- Figura 123** Exemplo de planilha utilizada para levantamento de dados
- Figura 124** Interface do WRIT
- Figura 125** Amostra de caracteres da fonte Sylfaen, de John Hudson, Ross Mills e Geraldine Wade
- Figura 126** Amostra de caracteres da fonte Sylfaen
- Figura 127** Captura de tela do Hyperglot
- Figura 128** Captura de tela do Hyperglot
- Figura 129** Captura de tela do Charset Builder
- Figura 130** Captura de tela do Charset Builder
- Figura 131** Captura de tela do Charset Builder
- Figura 132** Captura de tela da janela de diálogo para adição de novos caracteres no software Glyphs
- Figura 133** Exemplo de documentação fotográfica realizada no acervo do Ministério da Educação
- Figura 134** Sistema de classificação das fontes primárias
- Figura 135** Tabela produzida no levantamento de dados preliminar
- Figura 136** Estrutura básica das tabelas do levantamento no aplicativo online Google Sheets
- Figura 137** Tabela de referência do Bloco Latin-1 Supplement.
- Figura 138** Arquivo produzido pelo script ContadorDiacriticos de Gustavo Soares
- Figura 139** Captura de tela do software Zotero
- Figura 140** Exemplo de relatório gerado pelo Zotero
- Figura 141** Exemplo de sistematização das ortografias na publicação do PRODOCLIN

- Figura 142** Imagem usada para divulgação do projeto de catalogação ortográfica em dezembro de 2020
- Figura 143** Compilação das respostas da pesquisa do Google Forms
- Figura 144** Basic Latin e Latin-1 Supplement(Latin-1).Amostra gerada no aplicativo online Charset Buider
- Figura 145** tabelaCombinada
- Figura 146** Captura de tela do software Glyphs
- Figura 147** Captura de tela do software Glyphs
- Figura 148** Grupos linguísticos 1, 2 e 3
- Figura 149** Conjuntos de caracteres dos grupos 1, 2 e 3
- Figura 150** Conjunto básico de diacríticos
- Figura 151** Captura de tela do software Glyphs
- Figura 152** Variação da forma dos diacríticos para maiúsculas e minúsculas
- Figura 153** Captura de tela do software Glyphs
- Figura 154** Três variações da forma do mesmo diacrítico
- Figura 155** Variações da forma dos diacríticos
- Figura 156** Captura de tela do software Glyphs
- Figura 157** Variações dos diacríticos comb.low
- Figura 158** Comparação entre as variações comb (esq.) e comb.low (dir.)
- Figura 159** Comparação entre as variações comb.low (topo) e comb (abaixo)
- Figura 160** Captura de tela do software Glyphs
- Figura 161** Comparação entre as variações comb.narrow. low (topo) e comb.narrow (abaixo)
- Figura 162** Captura de tela do software Glyphs
- Figura 163** Variações na largura do til
- Figura 164** Representação do centro ótico da letra <g>
- Figura 165** Caracteres criados com o uso de componentes
- Figura 166** Ilustração mostrando o posicionamento das âncoras nas vogais
- Figura 167** Ilustração mostrando o posicionamento das âncoras em alguns diacríticos

- Figura 168** Conjunto de caracteres gerado automaticamente com o uso de componentes
- Figura 169** Caracteres pré-compostos com diacríticos sobrepostos
- Figura 170** Variantes dos caracteres <g> e <G> em diferentes fontes
- Figura 171** Amostras dos caracteres <u> e <U>
- Figura 172** Amostras dos caracteres <i> e <I>
- Figura 173** Diferentes formas de entrada para o mesmo caractere
- Figura 174** Exemplo de marca combinada com duas âncoras
- Figura 175** Chivo (specimen)
- Figura 176** Faustina (specimen)
- Figura 177** Variações de pesos das fontes Chivo (specimen) e Faustina
- Figura 178** Amostra de caracteres acentuados da fonte Faustina
- Figura 179** Amostra de caracteres acentuados da fonte Chivo.
- Figura 180** Amostra de texto em línguas indígenas na fonte Faustina
- Figura 181** Amostra de texto em línguas indígenas na fonte Chivo
- Figura 182** Esqueama de mapeamento do layout de teclado ABNT modificado

Sumário

1. Introdução	20
2. Contextualização	26
3. Justificativa e relevância	28
3.1 Tecnologia e promoção da inclusão digital	28
3.2 Tipografia: preenchendo lacunas entre diferentes campos	31
3.3 Escrever, antes de descrever	34
4. Objetivos	36
4.1 Objetivo geral	36
4.2 Objetivos específicos	36
5. Escopo/Delimitação do Corpus	38
6. Sobre as populações e línguas indígenas no Brasil	40
7. Sobre o estudo de línguas indígenas no Brasil	45
7.1 Presença missionária no Brasil Colonial	45
7.2 Desenvolvimentos paralelos na América hispânica	48
7.3 Século XIX e início do século XX	55
7.4 Presença missionária no século XX	55
7.5 Estudos contemporâneos	57
8. Codificação e linguagem	60
8.1 Terminologia e definições	63
8.1.1 Língua	63
8.1.2 Código	65
8.1.3 Ortografia	66
8.1.4 Informação	67

8.1.5 Codificação	71
8.1.6 Escrita	75
8.1.7 Unicode	90
8.1.8 Caracteres e glifos	93
8.1.9 Diacríticos e marcas combinadas	96
8.1.10 Tipografia	98
8.2 Codificação de caracteres	93
8.2.1 Codificação de caracteres: histórico	111
8.2.2 Unicode	140
8.2.3 Como funciona a codificação de texto em computadores.	152
8.2.4 OpenType	166
8.2.5 Mecanismo de modelagem de texto (<i>text-shaping engine</i>)	185
8.2.6 Teclados e métodos de entrada	185
9. Padronização	189
9.1 Linguística e padronização	191
9.2 Padronização e planejamento linguístico	193
9.3 Motivações e implicações da padronização ortográfica	194
9.4 O papel da imprensa na padronização das línguas	195
9.5 Tipografia e padronização	199
10. Ortografias	212
10.1 Referências acerca do desenvolvimento de ortografias para línguas ágrafas	212
10.2 Critérios essenciais	214
10.3 Fatores científicos: o ideal fonêmico	216
10.4 Fatores políticos: regulação e restrições governamentais	216
10.5 Aceitação pelas partes envolvidas	217
10.6 Facilidade de aprendizagem	218
10.7 Transferibilidade	218
10.8 Profundidade ortográfica, sub-representação e super-representação	219
10.9 Legibilidade e leiturabilidade	221

10.10 Identidade: todas as ortografias são políticas	221
10.11 Dialetos	223
10.12 Tecnologia: fontes, teclados e dispositivos	224
11. Características tipográficas das línguas indígenas brasileiras	225
11.1 Escopo	225
11.2 Codificações e conjuntos de caracteres nas ortografias das línguas indígenas brasileiras	227
11.3 Problemas tipográficos e de codificação em línguas indígenas brasileiras	236
11.4 Substituições de caracteres e diacríticos	237
11.5 Uso dos recursos de tachado e sublinhado	240
11.6 Diacríticos empilhados	242
11.7 Problemas de harmonia de largura e espaçamento	242
12. Metodologia	245
12.1 Metodologia: referências iniciais	245
12.2 John Hudson e o uso de bancos de dados no design de tipos	248
12.3 Ferramentas atuais de bancos de dados para codificação e tipografia	255
12.4 Desenvolvimento do método de levantamento de dados	261
12.5 Critérios para classificação dos dados	275
13. Desenvolvimento	278
13.1 Catalogação ortográfica	278
13.2 Diretrizes de desenvolvimento	284
13.3 Desenvolvimento de fontes	305
13.4 Desenvolvimento de teclados	313
Conclusão	316
Referências Bibliográficas	321
Apêndices	345
Anexos	440

1. Introdução

Considerado um dos países de maior diversidade cultural e étnica do mundo, o Brasil também é um país com uma grande diversidade linguística. Além da língua portuguesa (que tem o status de língua oficial) e das línguas crioulas, afro-brasileiras e de imigração, o Censo 2010 (IBGE, 2010a) compila uma lista de 274 línguas indígenas, faladas por indivíduos pertencentes a 305 etnias diferentes¹.

Apesar de sua importância cultural, pouco se sabe a respeito da tipografia das línguas indígenas brasileiras. Pode-se dizer que não existem pesquisas sistemáticas sobre o tema, e quase não há documentação científica sobre o assunto. Sob a ótica da comunicação visual, este trabalho procura preencher essas lacunas, descrevendo o Estado da Arte e propondo um método para desenvolvimento de ferramentas de escrita para essas línguas.

O interesse por este tema tem suas origens no trabalho de dissertação de mestrado deste autor na Universidade de Reading, Inglaterra, entre 2011 e 2012. Em resumo, essa dissertação identifica e analisa as características tipográficas e os problemas de composição de texto em línguas indígenas brasileiras. Esses problemas, em sua grande maioria, estão relacionados a uma particularidade da ortografia de muitas línguas indígenas brasileiras, que é a utilização de caracteres acentuados pouco comuns (como ã ou ã) ou, até mesmo, ausentes em grande parte das fontes tipográficas digitais disponíveis. A ausência desses caracteres específicos em muitas das fontes digitais existentes

1. Segundo o IBGE (2010a) “[...] no que diz respeito aos números totais de língua e etnia, há ainda a necessidade de estudos linguísticos e antropológicos mais aprofundados, pois algumas línguas declaradas podem ser variações de uma mesma língua, assim como algumas etnias também se constituem em subgrupos ou segmentos de uma mesma etnia.”

ou a falta de conhecimento de como digitá-los ou compô-los tem como resultado uma série de inconsistências tipográficas, que poderiam ser evitadas com o uso de métodos e tecnologias apropriadas.

Em 2017, como forma de continuação do estudo desenvolvido no mestrado, iniciou-se a pesquisa de doutorado, que busca investigar o tema de maneira mais profunda, com um estudo mais abrangente acerca da tipografia das línguas indígenas brasileiras, a fim de facilitar a promoção da inclusão digital por meio de tecnologias de escrita. Em outras palavras, a ideia central deste projeto é construir conhecimento e ferramentas para viabilizar a comunicação textual em meios digitais no cotidiano dessas minorias indígenas, em seus idiomas originários. Nesse sentido, levantou-se algumas hipóteses.

Primeiramente, o levantamento de dados desta pesquisa indica uma prática de escrita extremamente limitada em muitas dessas línguas, em decorrência da carência de ferramentas que permitam uma atividade cotidiana de escrita e de leitura. O que dimensiona e faz com que um sistema de escrita se desenvolva é justamente a atividade coletiva e cotidiana de escrita. Nesse sentido, colocar em campo ferramentas padronizadas que potencializem a prática da escrita representa uma medida pragmática para facilitar a comunicação, podendo levar a um impacto muito grande para essas comunidades. Por isso, este trabalho enfatiza o caráter da tipografia como ferramenta.

Em segundo lugar, este trabalho procura contribuir com um tipo de abordagem teórico-metodológica inspirada no trabalho de John Hudson e ainda pouco utilizada na pesquisa em comunicação digital, tipografia e línguas indígenas; em particular, no tratamento e manipulação da informação utilizada para o desenvolvimento de ferramentas de escrita. A partir de um grande volume de dados sobre a ortografia das línguas indígenas, essa abordagem aponta para utilização dessas informações em vários contextos: e.g. documentação de línguas, especificação de conjuntos de caracteres tipográficos etc. No entanto, para que essas informações representem de fato um lastro para o desenvolvimento teórico-metodológico e tenham aplicações práticas no desenvolvimento de ferramentas, é preciso estabelecer uma relação entre essas informações e a codificação digital, com base em padrões existentes (e.g. Unicode, ISO 8859-1, MS Windows 1250 etc.), uma vez que o suporte linguístico de muitas ferramentas é pensado a partir das

codificações supracitadas. Em linhas gerais, não basta apenas realizar um enorme levantamento de dados; é preciso contextualizar esse levantamento.

Nesse sentido é importante levantar aqui algumas questões:

1. Enquanto o português é a língua oficial e hegemônica no Brasil, como efetivamente inserir as línguas indígenas brasileiras na comunicação textual cotidiana de cidadãos que não têm o português como língua de origem?
2. Como inserir e promover essas línguas como ferramentas efetivas em produtos de comunicação (notícias, publicações, redes sociais etc.)?
3. Como resolver os problemas de ordem tecnológica, para que se possa estruturar uma sistematização dessas ferramentas e métodos?

Recentemente, a tecnologia digital impulsionou o avanço de pesquisas no campo da comunicação e das ciências humanas. Os maiores avanços são relacionados ao grande volume de informações que pesquisadores podem acessar e manipular rapidamente. Apesar disso, para muitos trabalhos do campo da tipografia, a tecnologia contribuiu pouco, já que procedimentos de análise ainda consistem na checagem de amostras letra a letra, a olho nu ou com ajuda de lentes (OLOCCO, 2019, p. 25).

Com o intuito de contribuir para a melhoria dessas técnicas de análise, esta pesquisa propõe-se à construção de um modelo de investigação que possa abranger a tipografia das línguas indígenas brasileiras e, para tal, propõe dois principais desenvolvimentos.

Primeiro, o levantamento de dados, cujo propósito essencial é a construção de uma catalogação sistemática e atualizada das ortografias das línguas indígenas brasileiras. Só assim será possível detectar a grande maioria das ocorrências de diacríticos (e outras particularidades tipográficas) e elaborar uma documentação consistente para o futuro desenvolvimento ferramentas de escrita. Em segundo lugar, o desenvolvimento de fontes tipográficas digitais e layouts de teclado, ferramentas necessárias para a composição e editoração de textos em línguas indígenas brasileiras.

Finalmente, deu-se atenção especial a questões técnicas bastante ignoradas no campo das línguas indígenas, especialmente em relação à codificação dessas línguas.

A compreensão dessas questões técnicas nos ajuda a entender o contexto de uso dessas codificações e da comunicação nessas línguas.

Como resultado, um vasto corpo documental foi coletado, analisado e discutido, com o intuito de ajudar na compreensão das práticas adotadas na tipografia das línguas indígenas, não somente no processo de design de tipos e codificação, mas também nas relações com a comunicação. Estes são temas que não foram devidamente documentados ainda.

Os parágrafos a seguir tentam explicar como a tese foi estruturada. Grosso modo, os primeiros sete capítulos têm um caráter introdutório e de contextualização do campo. Depois disso, outros quatro capítulos trazem a reflexão teórica-metodológica do trabalho. E os últimos dois capítulos descrevem a metodologia e sua aplicação prática.

Os capítulos que sucedem a introdução (*Contextualização, Justificativa e Relevância, Objetivos e Delimitação do corpus*) procuram delimitar o escopo e os objetivos, além contextualizar e justificar a importância desta pesquisa para as línguas indígenas brasileiras e para a comunicação escrita por meio dessas línguas, explicando o caráter pioneiro da pesquisa e sua relevância para a documentação linguística. Procuram também explicar a relevância da tipografia nos dias atuais, especialmente da tipografia como “ponte” entre diferentes campos. Por último, destaca-se que, além de a escrita ser um fator que auxilia na manutenção das línguas, o desenvolvimento de ferramentas que permitam a comunicação escrita digital em línguas indígenas é uma forma de promover a inclusão digital dos povos indígenas, sem privá-los de sua própria língua, preservando, assim, sua identidade cultural.

Nos capítulos seguintes (*Sobre as populações e línguas indígenas no Brasil; Sobre o estudo de línguas indígenas no Brasil*), reúnem-se alguns dados estatísticos sobre as populações indígenas brasileiras (em especial, sobre a situação linguística desses povos, de forma geral) e traça-se uma linha do tempo sobre o estudo das línguas indígenas brasileiras, que cobre, de forma bastante resumida, alguns marcos históricos principais.

O capítulo *Codificação e linguagem*, o mais longo deste trabalho, discute questões relacionadas à codificação, uma vez que os problemas relacionados a ela representam

uma das maiores dificuldades na composição de texto em línguas indígenas brasileiras e, por sua vez, na comunicação nessas línguas. Assim, para uma melhor compreensão das relações entre linguagem e codificação, esse capítulo revisa alguns conceitos elementares sobre codificação de caracteres e seu funcionamento em computadores, sistemas operacionais, aplicativos e teclados. Uma vez compreendidas as questões relativas à codificação, torna-se possível ampliar a discussão, para que se tenha melhor compreensão do problema na tese. Por essa razão, esse capítulo está dividido em duas grandes seções. A primeira delas faz um levantamento da terminologia e das definições acerca dos processos de codificação usados ao longo da tese. A segunda parte procura explicar em detalhes o funcionamento da codificação de texto em computadores. Nesse sentido, é importante listar algumas das principais referências desse capítulo.

O capítulo seguinte, *Padronização*, se concentra em questões relacionadas à padronização da escrita e procura reunir pontos importantes acerca do tema no contexto desta tese. Apesar de representar um recorte temático suficientemente extenso e complexo, a ideia de padronização é algo inerente à escrita e recorrente na literatura de várias áreas, como comunicação visual, tipografia, tecnologia da informação e linguística. Por essa razão, o capítulo busca demonstrar como a padronização permeia os campos supracitados e está organizado conforme descrição a seguir. Primeiramente, são abordados alguns tópicos relacionados à padronização no campo da linguística. Em segundo lugar, são levantadas algumas referências históricas sobre padronização. Por último, o tema padronização é relacionado ao campo da tipografia. Com isso, pretende-se levantar exemplos e evidências sobre padronização da literatura de distintas áreas (tipografia, linguística, história), demonstrando a importância do tema para a comunicação escrita nas línguas indígenas brasileiras.

Em seguida, o capítulo *Ortografias* procura explicar os fatores envolvidos no desenvolvimento de ortografias para línguas ágrafas, a fim de ilustrar algumas questões envolvidas na comunicação escrita nas línguas indígenas brasileiras. Reúne-se também uma breve revisão de literatura sobre o assunto.

O capítulo *Características tipográficas das línguas indígenas brasileiras* tem como objetivo examinar, a partir de uma perspectiva tipográfica, as condições técnicas e os

problemas relativos à composição de texto, codificação das ortografias e uso de diacríticos nas línguas indígenas brasileiras, e é dividido em três seções. A primeira delas delimita o escopo da análise. A segunda explica como as ortografias das línguas indígenas brasileiras utilizam o padrão Unicode. E a terceira tenta ilustrar alguns problemas de composição tipográfica em línguas indígenas brasileiras.

Finalmente, os dois mais importantes capítulos deste trabalho. O primeiro deles, *Metodologia*, tem como propósito a descrição detalhada da metodologia utilizada para o levantamento de dados desta pesquisa. Esta metodologia foi aplicada, em especial, no trabalho de catalogação ortográfica, que tem como objetivo principal a especificação de uma lista de requisitos para o suporte linguístico de fontes para línguas indígenas brasileiras. Esse método pode ser replicado e utilizado em outros sistemas e bancos de dados, e também em outros contextos (e.g. outros países da América Latina). Esse capítulo tem início com uma breve revisão de literatura sobre a metodologia de desenvolvimento de design de tipos, com algumas indicações para alguns métodos mais relevantes para esta tese. Em seguida, dá-se a descrição do desenvolvimento dessa metodologia, das diretrizes iniciais e de suas finalidades. E, por fim, descreve os critérios para classificação dos dados coletados, fundamentais para análise e classificação das línguas pesquisadas.

Por último, o capítulo *Desenvolvimento* demonstra como aplicar a metodologia na customização de fontes e teclados para línguas indígenas brasileiras e descreve o desenvolvimento de projeto dessas ferramentas. Além disso, explica como funciona a catalogação ortográfica – que reúne, ao todo, 239 tabelas com informações sobre III línguas – e como empregá-la.

Nos apêndices, ao final deste trabalho, foram incorporadas todas as tabelas com os dados da catalogação ortográfica e também alguns outros documentos relativos ao desenvolvimento deste trabalho (correspondência por e-mail com outros pesquisadores, glossário etc.).

2. Contextualização

Considerado um dos países de maior diversidade cultural e étnica do mundo, o Brasil é o maior país da América Latina, território no qual é a única nação lusófona. E o português, apesar de oficial, não é o único idioma falado pela população brasileira (RODRIGUES, 2001, n.p.). Além das populações descendentes de imigrantes que, de alguma maneira, em pequenas comunidades, ainda preservam seus idiomas de origem, existem milhares de indígenas autóctones que utilizam suas línguas originárias no dia a dia. Segundo Rodrigues (2001, n.p.), “a singularidade lingüística do Brasil está em que uma dessas línguas, o português, é hoje extremamente majoritária e as demais são todas extremamente minoritárias”. Devido ao baixo número de falantes e à baixa taxa de transmissão, a grande maioria dessas línguas está seriamente ameaçada.

Nesse sentido, além das línguas crioulas, afro-brasileiras e de imigração, o Censo 2010 (IBGE, 2010a) compila uma lista de 274 línguas indígenas, faladas por indivíduos pertencentes a 305 etnias diferentes. Tendo em vista esse cenário, é importante contextualizar as línguas indígenas brasileiras, especialmente do ponto de vista das tecnologias de escrita.

Primeiramente, é preciso levar em conta que as línguas indígenas brasileiras são originalmente ágrafas, baseadas na tradição oral, sem registro escrito. Em muitos casos, a transcrição dessas línguas para o alfabeto latino é um fenômeno recente, e a vasta maioria das fontes primárias analisadas nesta pesquisa foi publicada nos últimos 30 anos – coincidentemente, no mesmo período em que o computador pessoal e as fontes tipográficas digitais começam a se tornar acessíveis ao usuário comum.

Desde 2004, a partir da instalação de políticas públicas de incentivo à educação escolar de populações indígenas em suas línguas originárias, surgem também programas para treinamento de nível superior de professores indígenas e de fomento à

produção de materiais didáticos em suas línguas nativas (CUNHA, 2008, p. 153). Como resultado desse investimento, dezenas de publicações em línguas indígenas começam a surgir, dando forma a idiomas nunca impressos anteriormente. E essas publicações são fontes primárias¹ primordiais para o levantamento de dados desta pesquisa. Apesar dos recentes (e escassos) desenvolvimentos em prol da educação dos povos indígenas em suas línguas originárias, grande parte dos pesquisadores envolvidos no estudo de línguas indígenas ainda se concentra na descrição linguística, enquanto ainda há pouca informação sobre como, de fato, escrever essas línguas (por escrever entende-se composição tipográfica, composição de texto em plataforma digital). A ideia por trás deste projeto não é *descrever* essas línguas, mas criar meios para efetivamente *escrever* essas línguas. Nesse sentido, esta pesquisa de doutorado tem uma orientação voltada para a elaboração de métodos e instrumentos para promover essas línguas nativas a ferramentas de comunicação efetivas do cotidiano.

Nesse sentido, é fundamental contextualizar este projeto no campo da comunicação e de seu caráter interdisciplinar. Por se tratar de uma área que envolve a editoração de documentos em plataformas digitais, assim como a sua reprodução gráfica, o estudo parte do campo da comunicação visual, dialogando fortemente com as áreas de linguística, ortografia, design, história e informática. Sob essa perspectiva, é imprescindível situar este objeto e deixar claro que este estudo se concentra na configuração visual de documentos em línguas indígenas, e não no conteúdo dos textos nessas línguas.

1. Essas fontes primárias são, em grande parte, material didático e literatura em línguas indígenas. Muitas dessas edições são bilíngues, contendo também conteúdo em português. A maioria desse material foi editado e publicado por ONGs e centros de pesquisa ligados ao tema, e, em muitos casos, financiado pelo governo brasileiro.

3. Justificativa e relevância

O objetivo desta seção é justificar a importância desta pesquisa para as línguas indígenas brasileiras e para a comunicação escrita por meio dessas línguas. Nesse sentido, esta seção procura explicar, inicialmente, o caráter pioneiro desta pesquisa e sua relevância para a documentação linguística. Também levanta questionamentos sobre a relevância da tipografia nos dias atuais, especialmente da tipografia como “ponte” entre diferentes campos. Por último, procura explicar por que o desenvolvimento de ferramentas de escrita são tão importantes para a comunicação desses povos. Além de a escrita ser um fator que auxilia na manutenção das línguas, o desenvolvimento de ferramentas que permitam a comunicação escrita digital em línguas indígenas é uma forma de promover a inclusão digital dos povos indígenas, sem privá-los de sua própria língua, preservando, assim, sua identidade cultural.

3.1 Tecnologia e promoção da inclusão digital

Primeiramente, deve-se ressaltar que um dos principais propósitos deste projeto é utilizar a tecnologia para a promoção da inclusão e preservação da identidade cultural dos povos indígenas. A maioria das línguas indígenas brasileiras está fortemente ameaçada de extinção; i.e., tem menos de mil falantes (RODRIGUES, 2001, n.p.; IBGE, 2010a). Nesse sentido, esta pesquisa tem papel importante na preservação desse patrimônio cultural, em alinhamento com outras políticas públicas de inclusão e promoção dos direitos dessas minorias, e.g. Inventário Nacional da Diversidade Linguística (GARCIA et al., 2016).

Em segundo lugar, é importante ressaltar a relevância e originalidade desta pesquisa para a área da comunicação visual e design, envolvendo conhecimentos do campo da linguística e da informática. Enquanto existe documentação linguística relevante sobre

os idiomas indígenas brasileiros, há poucos estudos sobre suas características tipográficas e tecnológicas, e seus aspectos visuais, sem os quais a comunicação por meio da escrita fica prejudicada. Nesse sentido, além da pesquisa para a dissertação de mestrado deste autor (2012), um dos poucos trabalhos conhecidos nesse campo é um artigo de Kollontai Diniz (2007), pioneiro na descrição de características tipográficas das línguas indígenas brasileiras. Outros estudos conhecidos acerca do tema estão relacionados a grupos indígenas da Argentina, Paraguai, Colômbia, Venezuela e México, em especial por parte de projetos de pesquisa do Programa de Pós-graduação em Tipografia da Universidade de Buenos Aires. Nesse programa, os alunos de mestrado têm desenvolvido trabalhos relevantes de desenvolvimento de fontes com suporte linguístico para línguas indígenas (como o guarani), em paralelo às suas pesquisas de dissertação. A seção *Bibliografia Consultada* lista muitos desses desenvolvimentos.

Em linhas gerais (e de forma resumida), percebeu-se nos trabalhos supracitados que, em grande parte das fontes digitais disponíveis à época, a particularidade mais notável na ortografia de muitas línguas indígenas brasileiras é a utilização de caracteres acentuados pouco comuns (como ã ou ã) ou, até mesmo, inexistentes (g com til). A ausência desses caracteres específicos em muitas das fontes digitais ou a falta de conhecimento técnico sobre composição tipográfica e editoração têm como resultado uma série de inconsistências gráficas, que poderiam ser evitadas com o uso de métodos e tecnologias apropriados.

Embora tais inconsistências possam parecer para alguns apenas um problema técnico, como um erro de digitação em português, tais inconsistências gráficas em línguas ameaçadas de extinção, com poucos falantes configuram um problema bastante grave para a comunicação escrita entre os indígenas. Sem uma escrita consolidada e sem a possibilidade de se estabelecer uma escrita consistente, dada pela falta ou dificuldade em se encontrar caracteres da escrita em fontes digitais, a comunicação escrita em língua indígena fica muito prejudicada, o que pode contribuir para o não uso da língua escrita entre os povos falantes da própria língua.

A presente pesquisa justifica-se também por algumas outras questões. Uma delas é preencher lacunas tecnológicas na área de comunicação visual, mais especificamente

sobre os aspectos tipográficos de línguas indígenas brasileiras. Conforme explicitado nos objetivos da pesquisa, o trabalho em andamento propõe um levantamento abrangente e uma catalogação sistemática e atualizada das ortografias dessas línguas e, conseqüentemente, das suas características tipográficas. Entende-se que somente por esses meios é possível trabalhar em aplicações práticas consistentes (e.g. projeto e customização de fontes tipográficas, desenvolvimento de publicações), trazendo inúmeros benefícios para as comunidades indígenas.

Ainda, é importante relatar os esforços para a criação de um Acordo de Cooperação Técnica¹ entre a Universidade de Brasília e o Departamento do Patrimônio Imaterial do Instituto do Patrimônio Histórico e Artístico Nacional (IPHAN). Este último abriga uma divisão técnica que cuida do Inventário Nacional da Diversidade Linguística (INDL)², que desenvolve ações e projetos de promoção e valorização da diversidade linguística brasileira. Quando uma língua é incluída no INDL, ela é automaticamente declarada *Referência Cultural Brasileira*. Após a conclusão desta tese, as ferramentas (e.g. mapas de teclados, fontes) e banco de dados desenvolvidos durante a pesquisa poderão ser incorporados ao INDL.

Além do INDL, há também interesse neste projeto por parte de pesquisadores oficialmente associados ao Unicode, como a Professora Deborah Anderson, do

1. Em 2019, iniciou-se uma discussão com a equipe responsável pelo INDL para que fosse firmado um Acordo de Cooperação Técnica entre o IPHAN e a UnB. Essas ideias começaram a tomar forma e tramitar, em ambas as instâncias, em janeiro de 2020. No entanto, por conta da pandemia do Covid-19, a tramitação dos processos acabou demorando mais que o esperado e ainda estão pendentes.

2. Instituído pelo decreto 7.387 de 2010, o *Inventário Nacional da Diversidade Linguística* é uma política voltada para o reconhecimento da diversidade linguística como patrimônio cultural, por meio da identificação, documentação e ações de apoio e fomento. Por ser um instrumento com a dupla finalidade de pesquisar as línguas e reconhecê-las como patrimônio cultural, o INDL visa ao mapeamento, à caracterização e ao diagnóstico das diferentes situações relacionadas à pluralidade linguística brasileira. Ou seja, para que uma língua seja incluída no Inventário, é necessário, antes de tudo, produzir conhecimento sobre ela, documentar seus usos e realizar um diagnóstico sobre as suas condições de vitalidade. A produção de conhecimento sobre as línguas, incluindo a documentação audiovisual e os diagnósticos sobre vitalidade linguística, é elemento estruturante dessa política, pois parte considerável das línguas existentes ainda é pouco conhecida.

Departamento de Linguística da UC Berkeley³. Os resultados deste trabalho podem contribuir, ainda, com iniciativas de relevância e alcance internacional, como o *Script Encoding Initiative* (SEI)⁴, liderado pela professora Anderson. Além disso, como representante da UC Berkeley no Unicode Consortium, Anderson acredita que a presente pesquisa também é relevante para o *Unicode Common Locale Data Repository* (CLDR)⁵. Ademais, deve-se ressaltar que, no Departamento de Linguística na UC Berkeley, há especialistas em línguas amazônicas interessados em contribuir com a pesquisa. A possibilidade de colaboração entre essas instituições não somente é mutuamente benéfica, como também fortalece laços entre diferentes grupos de pesquisadores, produz impacto positivo e expande a rede de intercâmbio de informações sobre o tema.

3.2 Tipografia: preenchendo lacunas entre diferentes campos

Apesar da existência dos padrões internacionais de codificação de texto em computadores, de gramáticas e regras ortográficas criadas pelas academias de línguas, e de extensa documentação de caráter técnico e conceitual no campo tipográfico, há ainda lacunas não documentadas na literatura sobre escrita, tipografia e informática. Essas lacunas encontram-se, principalmente, nas áreas de transição entre esses campos ou como se relacionam: questões de como as línguas se relacionam com os padrões de codificação de texto, e como a escrita (i.e. a tipografia) atua como “ponte” entre os campos da linguística e da tecnologia da informação.

3. O contato com a professora Anderson se deu entre 2018/2019, por conta da tentativa de realização do doutorado sanduíche na UC Berkeley, impossibilitado por falta de financiamento (bolsa de estudos) e pela pandemia em 2020. Nos anexos consta a carta de aceite para realização do estágio doutoral em Berkeley.

4. O SEI é um projeto que se dedica à preparação de propostas formais para a codificação de escritas ainda não codificadas pelo Unicode. Ver: <https://linguistics.berkeley.edu/sei/>

5. O CLDR é um banco de dados mantido pelo Unicode, com o maior e mais extenso repositório padrão de dados de localidade disponível. Em linhas gerais, é utilizado na localização de softwares, i.e. na adaptação à localidade, adaptação às necessidades de uma região. O CLDR também é utilizado por desenvolvedores de teclados no Vale do Silício, que trabalham na criação de mapas de teclados para línguas minoritárias. Ver: <http://cldr.unicode.org>

Assim, além da existência de lacunas entre os campos da comunicação visual, linguística⁶ e tecnologia da informação, é preciso um campo que possa preencher essas lacunas, entre a comunicação e a linguística, assim como entre a linguística e a informática, e entre a comunicação e a informática.

Grosso modo, pode-se dizer que a tipografia, considerada uma subárea da comunicação visual, vem exercendo este papel de “ponte” entre diferentes campos do conhecimento há pelo menos 500 anos. Enquanto campo, pode-se dizer que a tipografia consiste, principalmente, em um ofício, uma área do conhecimento (e não uma ciência exata), que remonta ao século XV, com manuais e documentação técnica, histórica e comercial que datam do séc. XVII⁷. Nestes quase seis séculos de tradição, há abundante documentação sobre o tema, sobretudo do ponto de vista histórico. No trecho a seguir, John Hudson (2002, p. 21, tradução nossa) tenta explicar a relevância da tipografia e, principalmente, do design de tipos nos dias atuais:

Está na natureza do design de tipos seguir no vagão de bagagens. É uma indústria secundária, e por mais de 550 anos desempenhou seu papel na logística do meio editorial, da religião, propaganda, política e negócios e, recentemente, da tecnologia de informação. Se, como acredito, o design de tipos está passando um verdadeiro renascimento internacional no começo do século XIX, isto é em grande parte possibilitado pelo desejo da indústria de computadores de vender hardware e software para pessoas que falam outras línguas que não o inglês. A indústria de computadores não substituiu antigos mercados para o design de tipos (embora os tenha revolucionado, junto com a maioria das outras coisas), mas ao tornar cada usuário de computador no mundo um cliente para novos tipos de fontes – apesar de indiretamente, por meio do licenciamento de fontes para companhias de hardware e software, como parte do “pacote”, junto aos seus produtos –, desenvolvedores de software expandiram enormemente a nova demanda por novas e melhores fontes digitais internacionalmente. O tipo e a

6. Jefferson Maia (2018), desenvolveu sua tese de doutorado a partir do artigo *Toward a typographical linguistics de David Crystal* (1998). Uma das ideias levantadas por Maia é que a tipografia representa uma lacuna tangível no estado da arte da descrição linguística.

7. Os primeiros manuais de artes gráficas datam do século XVII, como a obra de Joseph Moxon, *Mechanic Exercises: or the Doctrine of Handy Works, Applied to the Art of Printing* (1683). Há ainda outras obras sobre artes gráficas anteriores à Revolução Industrial, como os manuais de Pierre-Simon Fournier (*Manuel Typographique*, 1764) e Giambattista Bodoni (*Manuale Tipografico*, 1818).

tipografia estão agora intimamente envolvidos – estão implícitos – na internacionalização da computação e, portanto, nos desenvolvimentos culturais e econômicos que esse processo apoia.

Apesar da longa tradição, e por tratar-se principalmente de uma área projetual (v. definições na p. 99), um ofício, só recentemente começaram a surgir, no campo da tipografia, teorias específicas. Nesse sentido, Gerard Unger é um dos pioneiros na tentativa de criar uma teoria específica para a área de design de tipos, com a publicação do livro *Theory of Type Design*, em 2018. Unger já havia esboçado muitos dos conceitos da obra acima mencionada em seu livro *Enquanto você lê*, publicado originalmente em holandês em 1997 e traduzido para o português em 2016.

Em *Theory of Type Design*, Unger (2018, p. 11, tradução nossa) observa algo muito particular à atividade do design de tipos:

De todos os objetos projetados, as letras são os mais disseminados, muito familiares, ainda que surpreendentemente diversificadas em sua aparência. As formas das letras vão desde as quase anônimas e utilitárias, em grandes quantidades, até umas poucas formas individuais e excepcionais. Entre esses extremos, designers desenvolveram quase todas as variações possíveis; parece que não há limite para a inventividade humana quando se trata de diferentes formas de letra.

A partir dessa premissa, explica que a questão principal de sua teoria não está centrada na forma, que não existe uma teoria para uma “forma perfeita”, e que isso depende do contexto (cultural, tecnológico etc.). Sua abordagem propõe uma visão ampla, interdisciplinar, e afirma que sua teoria não deve ser meramente descritiva, certamente não deve ser prescritiva, mas investigativa e reflexiva. E que tal teoria deve ser crítica e questionar conceitos, noções e ideias que usamos para dar sentido às formas das letras (UNGER, 2018, p. 11).

Reforçando as ideias de Unger, no prefácio dessa obra (UNGER, 2018, p. 7–8), Gerry Leonidas levanta algumas perguntas sobre como podemos avaliar a maturidade de um campo de conhecimento. Primeiro, os limites/escopo do campo e, conseqüentemente, seu posicionamento e relações com outros campos estão bem claros e bem definidos? Em segundo lugar, há um corpo de conhecimento aceito como fundacional para toda a

atividade por uma comunidade ativa no meio? Terceiro, há rotas estabelecidas para o aprendizado e a prática no campo? E, por último, há caminhos estabelecidos para pesquisa, reflexão, geração de conhecimento e correção dentro do campo? Leonidas afirma que, em diferentes graus, o design de tipos é um campo novo, em crescimento em todos os quatro aspectos, e prossegue destacando seu caráter interdisciplinar. Uma ideia notável desse prefácio, com a qual Unger corrobora, é afirmação de que os interessados em expandir seus conhecimentos nesse campo frequentemente têm que traçar seu próprio caminho e trabalhar sem o suporte e recursos que um campo maduro oferece.

3.3 Escrever, antes de descrever

Uma das referências importantes, nesse ponto, é o artigo *O que quer a linguística e o que se quer da linguística – a delicada questão da assessoria linguística no movimento indígena* (1999), do pesquisador Gilvan Müller de Oliveira. Em linhas gerais, o artigo discute a atuação de linguistas em projetos de educação escolar indígena. Oliveira (1999, p. 26) aponta para a importância da escrita nesse contexto, e destaca que o compromisso do linguista não deve ser exclusivamente o de elaboração de descrições das línguas indígenas, já que as descrições não são essenciais ao processo de criação de tradições escritas para línguas antes ágrafas.

Um dos argumentos centrais do artigo (OLIVEIRA, 1999, p. 27) consiste na crítica de que os linguistas envolvidos em projetos de educação escolar indígena, muitas vezes, “têm compromisso muito maior com a tradição da própria disciplina do que com as necessidades dos professores indígenas e dos projetos que os contratam.”

Oliveira (1999, p. 29) defende ainda a ideia de que descrições linguísticas não são condição *sine qua non* para a constituição das línguas indígenas como línguas escritas. Nessa linha, explica (1999, p. 30–31) que por mais de trezentos anos o português funcionou como língua escrita, antes de surgirem as primeiras gramáticas da língua: a de Fernão de Oliveira, em 1536; e João de Barros, em 1540; e a normatização ortográfica é ainda posterior a isso.

Nesse sentido, as ideias centrais desta tese giram em torno do desenvolvimento e documentação de métodos e ferramentas para efetivamente *escrever* essas línguas indígenas brasileiras em meios digitais, com o intuito de promover essas línguas nativas

a ferramentas de comunicação efetivas do cotidiano. Acredita-se que o uso da escrita em língua indígena, em situações além do ambiente escolar ou familiar (e.g. nos meios de comunicação, no comércio e administração), pode contribuir para a manutenção da língua viva.

4. Objetivos

4.1 Objetivo geral

A partir de demandas de recursos tecnológicos para composição de texto em línguas indígenas brasileiras, detectadas na publicação *Inventário Nacional da Diversidade Linguística* (GARCIA et al., 2016), e com base em pesquisas prévias sobre as características tipográficas das línguas indígenas brasileiras, criar mecanismos para que a **tipografia** se torne uma **ferramenta de comunicação indispensável para a política da diversidade linguística brasileira** e para os falantes de línguas indígenas brasileiras em meios digitais. E, a partir destas demandas, desenvolver métodos e técnicas que demonstrem **como a tipografia pode beneficiar a manutenção e representação gráfica correta dessas línguas indígenas brasileiras**, contribuindo para a promoção e **fortalecimento dos usos sociais dessas línguas nos meios de comunicação** e para a **inclusão digital dos povos indígenas brasileiros**

4.2 Objetivos específicos

1. Realizar levantamento de dados abrangente, a partir de fontes primárias (i.e. publicações), para identificar o maior número possível de ocorrências de diacríticos, resultando numa catalogação ortográfica atualizada e consistente.
2. Organizar as línguas catalogadas em três principais grupos, com base nos conjuntos de caracteres utilizados para composição de textos. O primeiro é formado por línguas que podem ser compostas com o layout de teclado ABNT, o mesmo usado para o português. O segundo é formado por línguas que utilizam caracteres que dependem de outros layouts de teclado, além do ABNT. E o terceiro, e mais complexo, é formado por línguas que utilizam caracteres não codificados pelo Unicode.

3. Documentar problemas (como ruídos) de composição tipográfica, em particular nos casos de caracteres não codificados pelo Unicode, e de casos de composição complexa (e.g., diacríticos empilhados), com a finalidade de solucioná-los.
4. Desenvolver projeto de customização de algumas famílias tipográficas *open-source* para auxílio na composição de texto em línguas indígenas, tornando-as disponíveis para a comunidade de usuários em plataforma online aberta.
5. Desenvolver projeto de customização de teclados, para garantir o bom funcionamento das fontes digitais específicas para línguas brasileiras.

5. Escopo/Delimitação do Corpus

Os povos indígenas brasileiros constituem, originalmente, culturas ágrafas, ou seja, de tradição oral, que tradicionalmente não se expressam pela escrita, não têm registro escrito. Nesse sentido, em comparação à língua portuguesa, os esforços de documentação de caráter científico dessas línguas são extremamente recentes.

O trabalho da linguística indígena no Brasil consiste, em grande medida, em descrições linguísticas de caráter científico. Uma consequência natural do desenvolvimento desses estudos é o surgimento de livros de caráter didático ou literário, voltados para os falantes de línguas indígenas. Esses livros utilizam a própria convenção ortográfica da língua e empregam somente o alfabeto latino, a exemplo da frase a seguir, em língua guarani: *Heta oĩ oikóva yvýrupinte* (A maioria andava a pé) (“Dicionário Português-Guarani Glosbe”, [s.d.]).

O escopo deste estudo abrange a análise de línguas indígenas para as quais já existe uma ortografia estabelecida (como é o caso da língua guarani), em publicações que apresentem amostras de texto significativas nessas línguas. Por isso, determinou-se, como objeto principal de estudo, as fontes primárias impressas (livros, cartilhas), nas diversas línguas indígenas brasileiras que já têm uma ortografia estabelecida. Essas fontes primárias formam um corpus de aproximadamente 150 publicações.

Há, no entanto, duas exceções a essa regra. A primeira diz respeito às análises de algumas publicações de caráter científico (e.g. dicionários, gramáticas) em que há uma descrição clara da ortografia de uma determinada língua, com explicação detalhada sobre o alfabeto utilizado, i.e., quais letras e quais diacríticos. A segunda exceção diz respeito aos resultados coletados em uma pesquisa online, na qual cada colaborador faz

uma listagem dos caracteres acentuados utilizados na língua (ou línguas) com a qual tem familiaridade, contato, envolvimento. É possível fazer múltiplas contribuições por meio desse formulário.

O capítulo *Metodologia* apresenta uma descrição mais detalhada das fontes primárias e do levantamento de dados para esta tese.

6. Sobre as populações e línguas indígenas no Brasil

De acordo com o documento *O Brasil Indígena* (IBGE, 2010a), do Instituto Brasileiro de Geografia e Estatística, “das 896 mil pessoas que se declaravam ou se consideravam indígenas, 572 mil, ou 63,8%, viviam na área rural e 517 mil, ou 57,7%, moravam em Terras Indígenas oficialmente reconhecidas”. No entanto, a população indígena era bem maior antes da chegada dos colonizadores europeus. Antes da colonização, estima-se que mais de 1000 grupos distintos existiram no Brasil, numa população de 2 a 4 milhões de pessoas (ISA, 2018). Sobre o extermínio desses grupos, Darcy Ribeiro relata: “Quanto ao contingente indígena, contamos com estudos dos fatores responsáveis por seu extermínio, entre os quais sobrepõem, no nível biótico, os efeitos das moléstias introduzidas pelo europeu e, no nível social, as condições de opressão a que foi submetido” (RIBEIRO, 1956 apud RIBEIRO, 1995, p. 230). O documento do IBGE (2010a) fornece também algumas estatísticas sobre as línguas indígenas brasileiras.

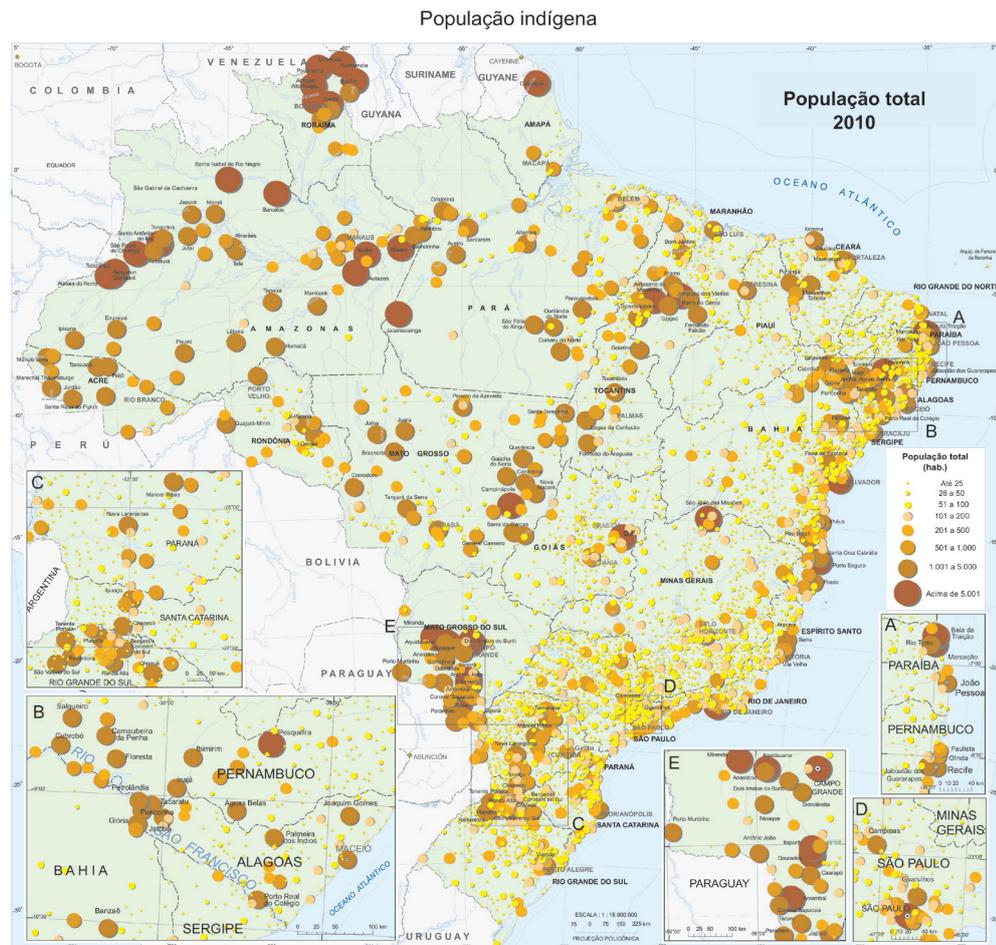


Figura 1: População indígena total em 2010. Fonte: IBGE, 2010b.

Os resultados do Censo 2010 apontam para 274 línguas indígenas faladas por indivíduos pertencentes a 305 etnias diferentes. Essas declarações ultrapassaram as estimativas iniciais feitas pela Funai. Entretanto, no que diz respeito aos números totais de língua e etnia, há ainda a necessidade de estudos linguísticos e antropológicos mais aprofundados, pois algumas línguas declaradas podem ser variações de uma mesma língua, assim como algumas etnias também se constituem em subgrupos ou segmentos de uma mesma etnia (IBGE, 2010a).

O documento *O Brasil Indígena* (IBGE, 2010a) fornece uma listagem dessas línguas (Fig.2) em agrupamentos de troncos e famílias, onde, de acordo com o documento, somente 40 línguas (14,6%) das 274 são faladas por mais de 1000 falantes; 12 línguas (4,4%) são faladas por mais de 4.000 falantes; somente 5 línguas (1,8%) são faladas por mais de 10.000 falantes. Na listagem, 51 línguas (18,6%) são faladas por menos de 10 indivíduos.

Segundo Aryon Rodrigues, antes da colonização, estima-se que existiram por volta de 1200 línguas diferentes. No mesmo trabalho, o autor descreve 43 famílias de línguas indígenas no Brasil. Algumas delas formam grandes agrupamentos e outras são compostas de uma única língua, caracterizando o que se conhece por “línguas isoladas”. Esse isolamento é decorrente do processo de colonização que exterminou as outras línguas da mesma família. Apesar de existir documentação dos séculos passados sobre algumas famílias específicas, que permite determiná-las ao menos aproximadamente, todas as línguas pertencentes a essas famílias estão extintas e, portanto, a família também está extinta (RODRIGUES, 2005, p. 35).

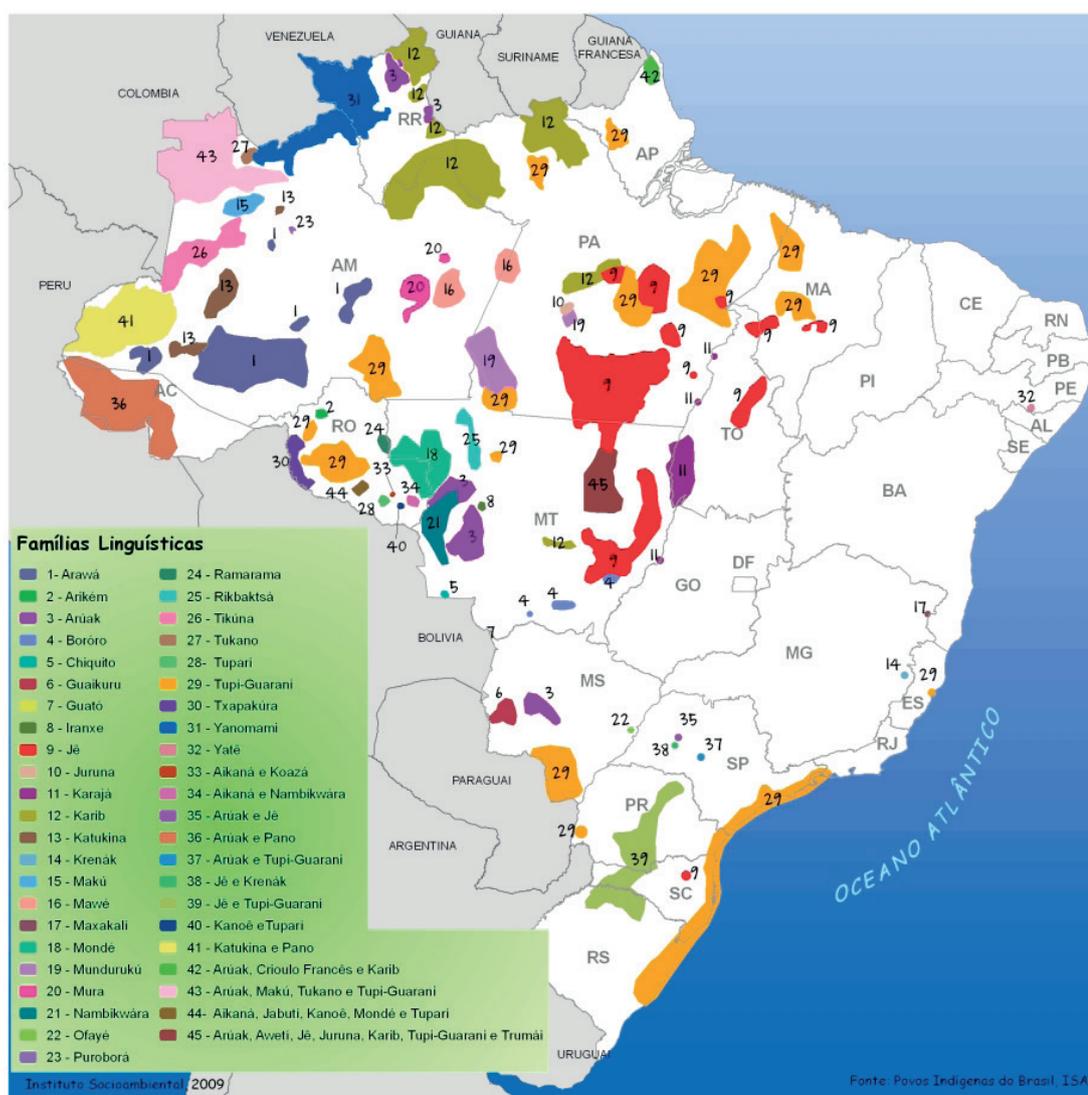


Figura 3: Mapa das línguas indígenas do Brasil. Fonte: ISA, 2018.

Por outro lado, algumas famílias linguísticas compartilham propriedades estruturais muito similares, que só podem ser explicadas por uma origem comum mais remota e que justifica uma classe mais abrangente, o tronco linguístico. O tronco tupi, bem estabelecido, compreende dez famílias. O tronco macro-jê, de caráter ainda bastante hipotético, abrange outras 12 famílias (RODRIGUES, 2005, p. 35–36).

Além disso, é importante destacar a ocorrência de mais de uma língua franca. Também chamadas de “línguas gerais”, foram utilizadas em diferentes períodos e regiões do Brasil. É o caso do tupinambá (ou tupi antigo), utilizado ao longo da costa nos séculos XVI e XVII como a língua predominante do contato entre os portugueses e os nativos, e a primeira língua indígena brasileira a ser escrita (RODRIGUES, 1986, p. 21). Outro exemplo é a língua geral amazônica (ou nheengatu), que em 2002, junto às línguas baniwa e tukano, foi decretada¹ como língua oficial no município de São Gabriel da Cachoeira, Amazonas.

Por último, nota-se que o Brasil é caracterizado por uma grande pluralidade linguística: um grande número de línguas para uma população de baixa densidade (RODRIGUES, 1970, p. 4034). Tikuna, por exemplo, é a língua com o maior número de falantes (cerca de 30.000). Do ponto de vista fonológico, fonético e gramatical, existem características muito particulares sobre as línguas indígenas brasileiras. Essas características compreendem: línguas com um número abundante de fonemas, assim como a língua com o menor número de fonemas já registrada no mundo; sons não observados em nenhum outro lugar do mundo; estruturas gramaticais incomuns e singulares (RODRIGUES, 2001).

1. Lei Municipal 145, de 2002.

7. Sobre o estudo de línguas indígenas no Brasil

O objetivo deste capítulo é traçar uma linha do tempo sobre o estudo das línguas indígenas brasileiras e cobrir, de forma bastante resumida, alguns marcos históricos principais. Para um panorama mais detalhado, indica-se, como complementação, os seguintes artigos: *A linguística indígena no Brasil* (1999), de Lucy Seki; *Sobre as línguas indígenas e sua pesquisa no Brasil* (2005), de Aryon Dall’Igna Rodrigues. Pode-se dizer que, a partir desses trabalhos, é possível ter uma visão geral do estado da arte do estudo da linguística indígena no Brasil.

7.1 Presença missionária no Brasil Colonial

Desde a chegada dos europeus às Américas, existe interesse em documentar as línguas indígenas deste continente. No Brasil, a documentação das línguas indígenas foi inicialmente desempenhada pelos jesuítas, que desembarcam no Brasil em 1549. De acordo com Ronaldo de Oliveira Batista (2005, p. 121), o objetivo destas obras de referência (gramáticas, dicionários) era “facilitar, por meio da aprendizagem das línguas, o contato entre jesuítas e indígenas, tendo em vista a colonização e catequização.”

Durante todo o Período Colonial era proibido imprimir no Brasil. Assim, todas as obras dos missionários eram impressas na Europa. De acordo com Kury (2008), esse tipo de proibição era um recurso para evitar que informações sobre a colônia, de importante valor comercial, se espalhassem e caíssem nas mãos de nações rivais.

Um aspecto da ocupação no período colonial é que os colonizadores portugueses se estabeleceram quase que exclusivamente ao longo da costa do Atlântico. Logo, tiveram maior exposição ao idioma dos Tupinambás que, conforme o título da gramática de Anchieta, era a língua mais usada na costa do Brasil. Como consequência, somente

três línguas nativas foram documentadas em gramáticas e dicionários pelos missionários no período colonial (RODRIGUES, 2005, p. 35).

As mais antigas documentações conhecidas de uma língua indígena brasileira estão em publicações francesas da segunda metade do século XVI. *La Cosmographie Universelle* (1575), de André Thevet, e *Histoire d'un voyage fait en la terre du Brésil, autrement dite Amérique* (1578), de Jean de Léry, apresentam fragmentos de texto impressos em Tupinambá, ou Tupi Antigo. Thevet, um frade franciscano, e Léry, um missionário calvinista, acompanharam a incursão francesa à Baía da Guanabara, conhecida como França Antártica. Léry foi o primeiro a fazer anotações sobre a gramática do Tupinambá (RODRIGUES, 1986, p. 21).

Em 1595, José de Anchieta publicou a *Arte de grammatica da lingua mais usada na costa do Brasil*, a primeira gramática do Tupinambá (Fig.4). Em 1621, outro jesuíta, Luis Figueira, publicou uma nova gramática do Tupinambá, *Arte da lingua brasilica*. Os jesuítas também trabalharam em dicionários, como o *Vocabulário da Lingua Brasilica*, em livros de orações, e em gramáticas de outras línguas, como *Arte de grammatica da lingua brasilica da naçam Kiriri* (1699), do padre Luis Vicencio Mamiani, além de documentos sobre algumas outras línguas, das quais os manuscritos foram perdidos (RODRIGUES, 2005, p. 35; 1970, p. 4035).

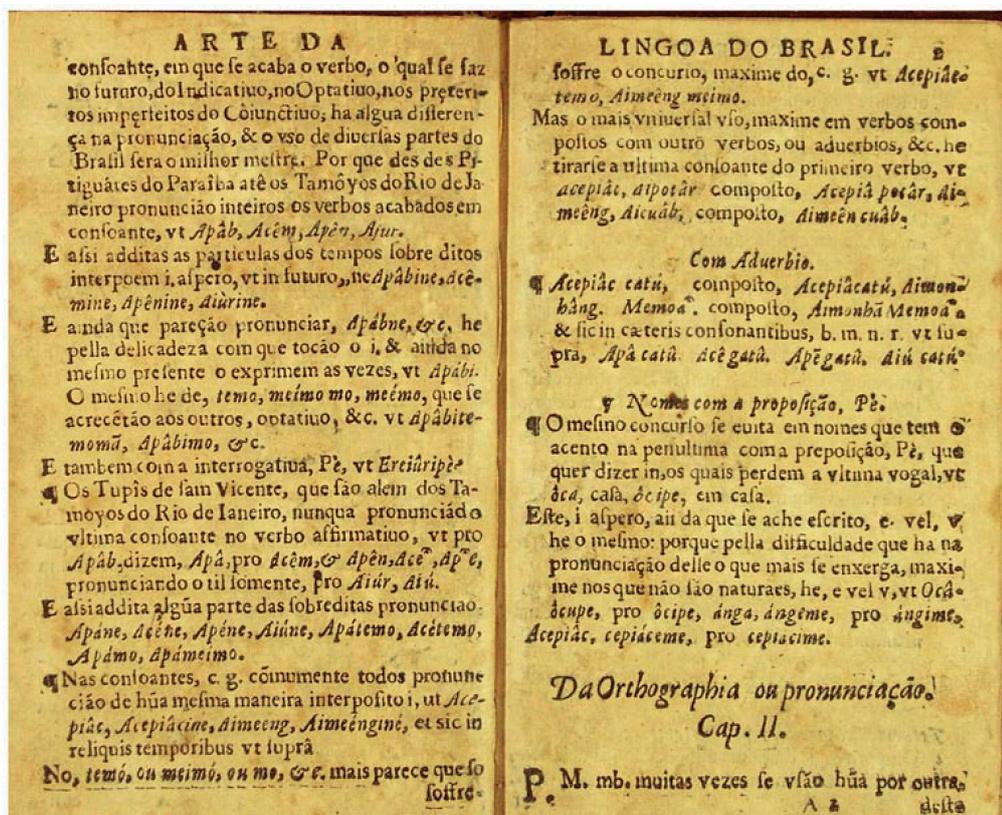
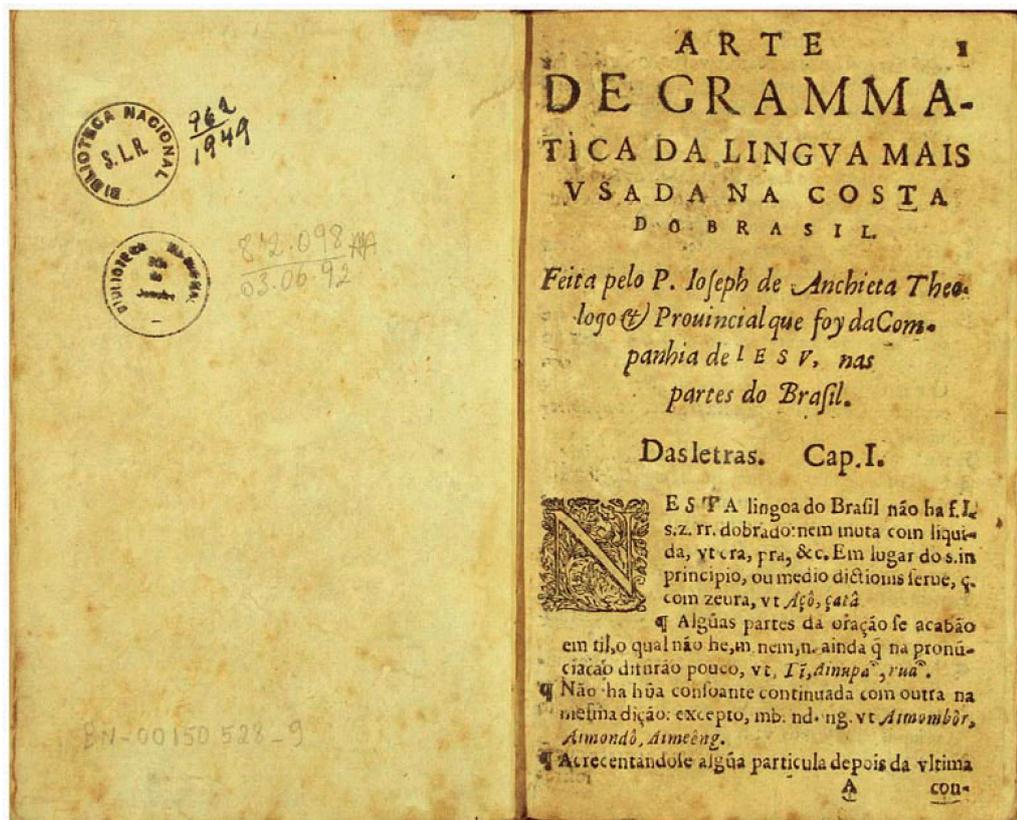


Figura 4: Arte de grammatica da lingua mais vsada na costa do Brasil. Fonte: ANCHIETA, 1595.

7.2 Desenvolvimentos paralelos na América hispânica

Enquanto nas colônias portuguesas a proibição para imprimir durou até o século XIX, na América hispânica atividades dessa natureza se iniciaram no século XVI. No México, o primeiro livro foi impresso em 1539 e a primeira universidade, fundada nas décadas seguintes (UNNA, 2004, p. 28). A presença de uma universidade resultou, obviamente, em demanda por livros e outros impressos.

É importante mencionar também que, antes da chegada dos espanhóis à região do México e Guatemala, a civilização maia já utilizava um sistema de escrita hieroglífica, o único exemplo conhecido de sistema de escrita nativa das Américas, documentado não somente em inscrições, mas também em livros conhecidos como “Códices Maias”. O livro maia, ou *codex*, era feito de uma única grande folha de papel, que era, então, dobrada como uma sanfona, com papel feito da casca de figueiras nativas (ROSS, 2010, p. 19). Além disso, o fato de que os maias já produziam o papel amate (ou amatl) desde aproximadamente o século IV ilustra o avançado estágio de desenvolvimento desse povo (BINNQÜIST, 2003, p. 3).¹

Infelizmente, a maioria desses livros foi destruída por colonizadores espanhóis, em particular pelo bispo Diego de Landa que, em 1562, sob alegação de que o conteúdo tratava de superstições e mentiras do demônio, queimou grande parte da cultura escrita dos maias em seu *Auto de Fé* (ROSS, 2010, p. 39).

1. Para mais informações sobre o papel amate, consultar a tese de doutorado de Rosaura Citlalli López Binnqüist (2003).

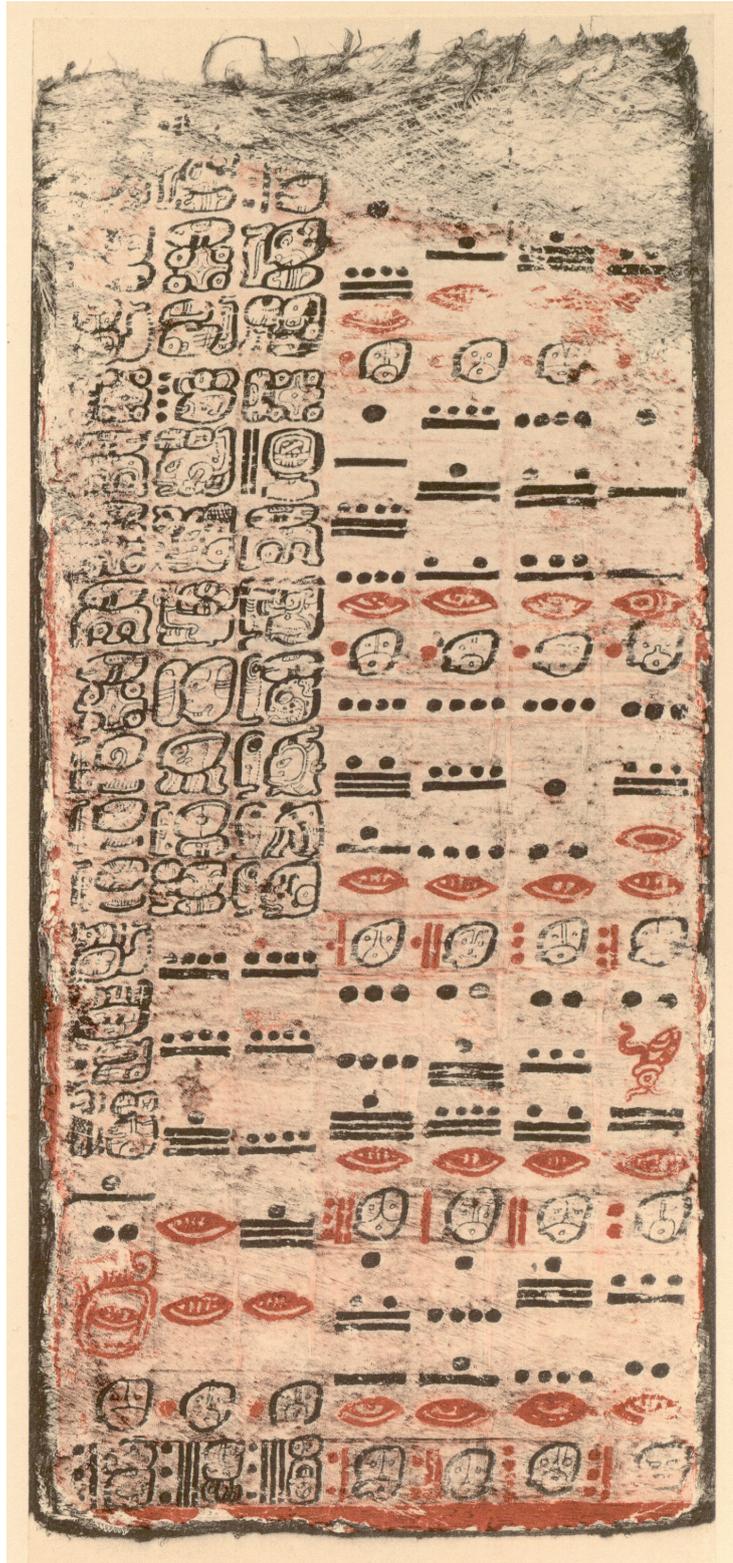


Figura 5: Manuscrito maia. Imagem de facsimile do Códice de Dresden

Entretanto, existem registros de documentação de línguas indígenas na América Central anteriores ao incidente supracitado. Como exemplo digno de nota, o primeiro livro impresso nas Américas, *Breve y más compendiosa doctrina Cristiana en lengua mexicana e castellana* (Fig.5), um livro de preces bilíngue em espanhol e náhuatl², impresso com tipos góticos (UNNA, 2004, p. 26-31). Além desse exemplo, Marina Garone Gravier (2004, p. 44) lista uma extensa bibliografia³ produzida por missionários no território mexicano entre os sécs. XVI e XVIII, que compreende obras religiosas, gramáticas e dicionários em línguas nativas e também em edições bilíngues.

As dificuldades de transcrição de línguas ágrafas com o alfabeto latino, em razão da ausência de caracteres especiais para representar a complexidade dos sons de muitas destas línguas, é uma questão discutida desde o período colonial. O tratamento a alguns idiomas como o otomí, por exemplo, foi negligenciado dada a sua complexidade, enquanto outros, como o náhuatl, tiveram maior atenção, porque (além de outras razões) eram mais fáceis de se reproduzir com o material tipográfico disponível (GRAVIER, 2004, p. 42).

Levando em conta essa complexidade, Gravier (2013b) observa um desenvolvimento único do ponto de vista tipográfico: o livro *Breve compendio...* (Fig.6), publicado em 1785, em espanhol e otomí. Na obra, é notável a inventividade das formas desses caracteres especiais e, ainda mais impressionante, o empenho técnico necessário para a sua fabricação, que provavelmente foi realizada no México. No entanto, pode-se perceber a dificuldade de harmonização entre os tipos originais, utilizados para o texto em espanhol, e aqueles criados para os sons do otomí.

2. Náhuatl, por ser língua bastante difundida no território mexicano, era chamada pelos colonizadores de *lengua mexicana*, assim como aconteceu no Brasil com o tupinambá, ou tupi antigo, que era chamado de *língua brasilica*; ambas tiveram um papel de língua franca no período.

3. Nos títulos da lista aparecem outros nomes de línguas como otomí, maya, huasteca e zapoteca.

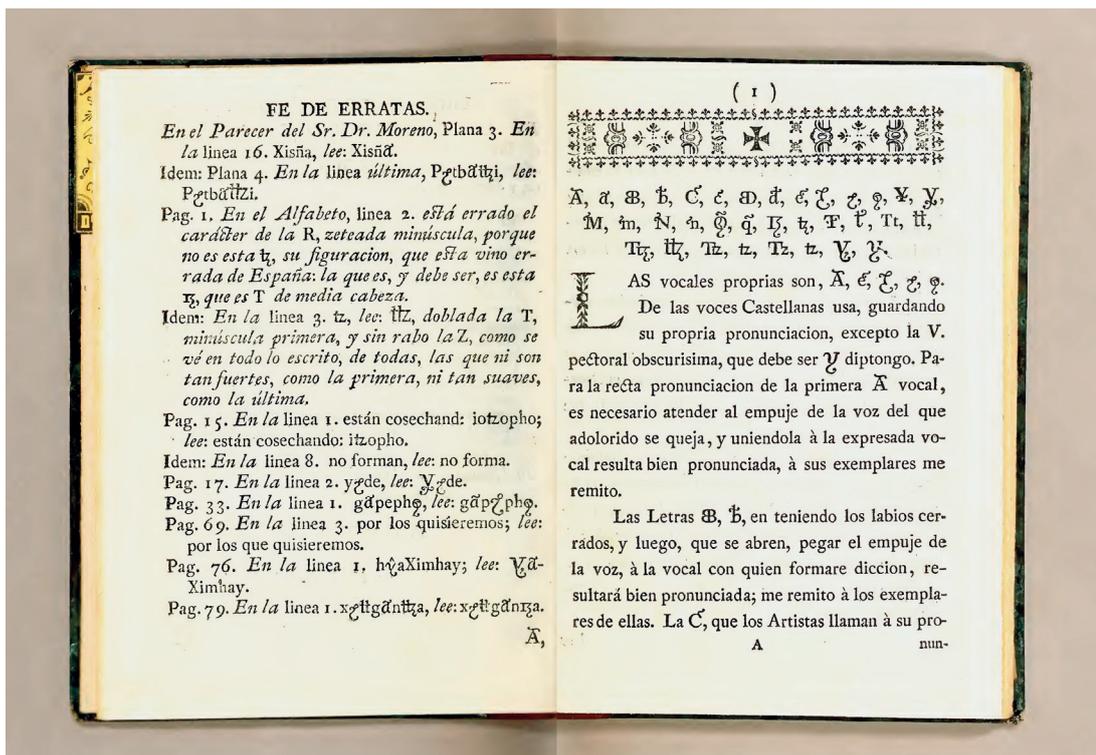


Figura 6: *Breve compendio...*No topo, à direita, pode-se observar amostras de caracteres especiais, produzidos para a obra. Ligaturas para representação de dígrafos, formas espelhadas, caracteres acentuados e formas pouco comuns e/ou de origem desconhecida. Fonte: RAMÍREZ; LEÓN, 1785.

Os missionários jesuítas chegaram à região mais ao sul da América do Sul no começo do século XVII (após a União Ibérica, em 1580), e se espalharam por territórios que hoje são parte do Paraguai, Uruguai, norte da Argentina e região sul do Brasil (principalmente Paraná, Santa Catarina e Rio Grande do Sul). Assim como no litoral brasileiro e América Central, “o aprendizado da língua local era uma condição básica para a evangelização, assim como requisito indispensável para ser missionário da ordem” (MORALES, 2002, p. 16 apud COSGAYA, 2003, p. 22). A língua em questão era o guarani, que por ser a mais utilizada pelos indígenas da região, se coloca em condição similar ao náhuatl e ao tupinambá, idiomas dominantes em suas respectivas regiões.



A R T E
DE LA LENGVA
G V A R A N I.
P R E L V D I O.



QVATRO pronunciaciones tiene esta lengua muy necessarias, para hablar propriamente, cuyas notas se ponen aqui, y seruiràn para entèder el Bocabulario, y tesoro desta légua.

La primera pronunciacion es narigal, que se forma en la nariz, cuya nota es esta, ^ˆ puesta sobre la vocal q̄ se ha de pronunciar con la nariz, como tātā fuerte, aduirriendo, que muy frequentemente la silaba narigal haze narigales la antetecedente, y conseqente, vt nūī, aguja. ahâ-ângatú, por, catú, y varia el accento en breue, y largo.

La segunda es, vna pronunciacion gutural, que se forma in gutture contrayendo la lengua ázia dentro, su nota es esta, [˘] sobre la (y) en q̄ siempre cae, vt ta ĩra, hijo, y siempre es largo su accento.

La tercera, incluye las dos dichas, su nota es esta [˜] sobre la y, en que siempre cae, y se ha de pronunciar con nariz, y ingutture juntamente, como aroÿrõ, yo despre precio, y siempre tiene acento largo.

La quarta pronunciacion es gutural, contracta que se
A haze

Figura 7: Página inicial da gramática da gramática do guarani, de Antonio Ruiz de Montoya.

A página mostra o uso de diacríticos na ortografia da língua. Facsímile da edição original impressa em 1640. Fonte: RUIZ DE MONTOYA; PLATZMANN, 1876.

Ao contrário do que aconteceu na colônia portuguesa, no território das missões uma prensa foi instalada e os primeiros livros foram impressos nos primeiros anos do século XVIII. Como os jesuítas tinham permissão para construir e instalar somente uma prensa, eles tinham que carregá-la de uma missão para outra. Os tipos utilizados nesse contexto, dos quais as matrizes são de origem desconhecida, utilizam diacríticos próprios para o guarani, além de apresentarem espaçamento e alinhamento muito irregular (Fig.7). Essa imprensa operou até 1727 e ainda não se sabe que fim levaram os tipos e a prensa. Sabe-se que a dificuldade de conseguir papel foi um dos motivos para o fim da iniciativa (COSGAYA, 2003, p. 22).

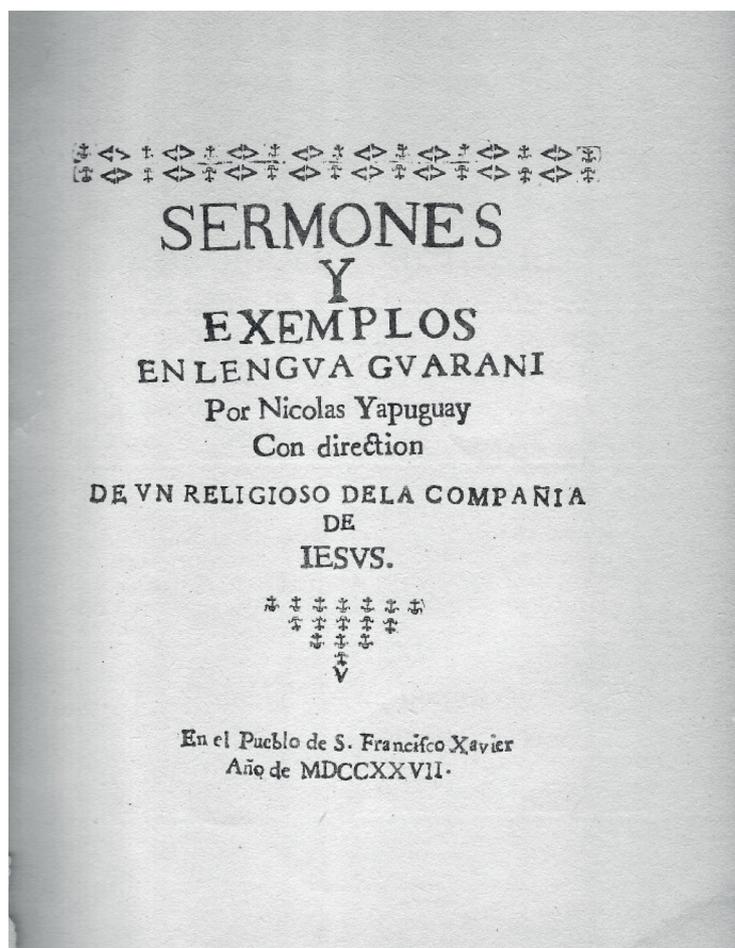


Figura 8: Folha de rosto (em espanhol) de um livro de doutrinas em guarani, de autoria do Cacique Yapuguay. Impresso na missão de São Francisco Xavier; um dos poucos exemplos de livros impressos nas missões. Facsimile da primeira edição de 1727, com uma introdução de Guillermo Furlong. Fonte: YAPUGUAY, 1953.

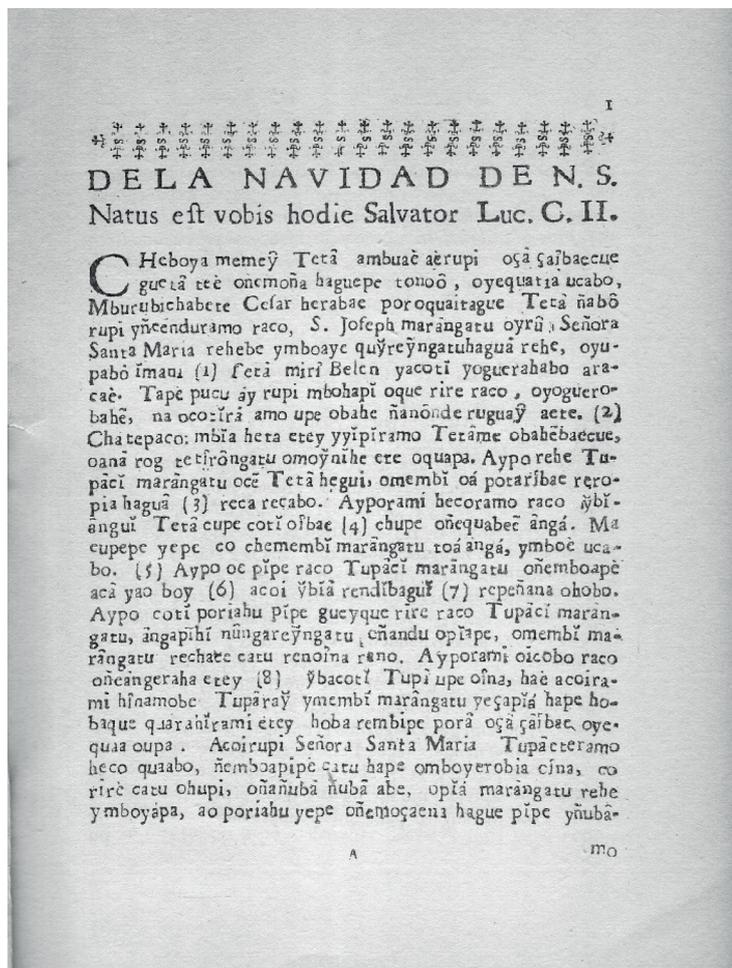


Figura 9: Página do livro de doutrinas em guarani, de autoria do Cacique Yapuguay. O trabalho segue ortografia similar a de Montoya. Fonte: YAPUGUAY, 1953.

Além da gramática do peruano Antonio Ruiz de Montoya⁴ (Fig.9), um exemplo singular da aptidão para as letras entre os guaranis é o trabalho do cacique Nicolás Yapuguay, que escreveu e publicou dois livros de doutrinas: *Explicación de el catecismo en lengua guarani*, de 1724, e *Sermones y exemplos*, de 1727 (NEWMAN, 2009, p. 185).

Em 1759, por decreto do Marquês de Pombal, os jesuítas são expulsos das colônias, resultando num longo período sem a publicação de estudos significativos sobre línguas nativas brasileiras.

4. A primeira edição foi impressa em Madrid, em 1640, antes da chegada da prensa às missões na região Sul.

7.3 Século XIX e início do século XX

Somente a partir da transferência da corte de D. João VI para o Rio de Janeiro, em 1808, é que novos estudos sobre línguas nativas começam a surgir. A introdução da imprensa, assim como a fundação de instituições de ensino e educação superior, coincide com o interesse pelo Brasil por parte dos naturalistas europeus, que começam a aportar no território brasileiro para missões e expedições com o apoio do governo local. Muitos naturalistas (e.g. Karl Friedrich Philipp von Martius, Johann Baptist von Spix) contribuíram para o estudo de línguas nativas, através do registro e coleta de fragmentos de vocabulário, da mesma maneira em que coletavam espécimes de flora e fauna (RODRIGUES, 1970, p. 4035).

Ainda segundo Rodrigues (1970, p. 4034), o estudo de línguas ameríndias como disciplina científica tem início efetivamente na América do Norte na segunda metade do século XIX. No começo do século XIX, o trabalho de Franz Boas (*Handbook of American Indians Languages*, de 1911) “teve influência decisiva na orientação puramente científica que se passou a dar ao estudo das línguas ameríndias nos Estados Unidos [...] No México e na América Central só por volta de 1935 começou a ser superada a fase predominantemente amadoresca dos estudos de línguas ameríndias, e na América do Sul só em 1950.”

Rodrigues (1970, p. 4034) observa ainda:

No Brasil, por exemplo, o estudo científico das línguas ameríndias começou, na realidade, com as expedições do antropólogo alemão Karl von den Steinen à região das cabeceiras do rio Xingu, em 1884 e 1887. Ao extraordinário êxito científico dessas viagens deve-se a realização de uma série de outras expedições alemãs ao Brasil, entre as quais as de Paul Ehrenreich, Max Schmidt e Theodor Koch-Grünberg foram de particular importância para o acréscimo dos conhecimentos linguísticos. A essa série de antropólogos visitantes deve somar-se o nome de Curt (Unkel) Nimuendajú, que, tendo emigrado para o Brasil, de 1905 a 1945 contribuiu de maneira particularmente extensiva para o conhecimento das línguas ameríndias deste país.

7.4 Presença missionária no século XX

A partir da segunda metade do século XX, a documentação científica de línguas indígenas se intensifica e passa acontecer de forma mais sistemática na América Latina. Além

do aumento do interesse de pesquisadores do campo da linguística, em universidades e centros de pesquisa, é importante mencionar a atuação do SIL International⁵, instituição missionária que agrega um grande número de pesquisadores com treinamento específico em linguística descritiva (RODRIGUES, 1970, p. 4034).

De acordo com David Stoll, essa organização cresceu na América Latina por meio de alianças com políticos e intelectuais (como Lázaro Cardenas, Vargas Llosa e Darcy Ribeiro), e não com o apoio das igrejas evangélicas locais (STOLL, 1985 apud BARROS, 2004, p. 45). Uma estratégia utilizada por essa instituição para ganhar acesso aos países latino-americanos é que seus membros se apresentavam como linguistas, e não como religiosos, tendo em vista que muitos governos não olhavam a atividade missionária com bons olhos (BARROS, 2004, p. 47-50).

Assim, por conta de suas credenciais acadêmicas, Kenneth Pike ganha acesso ao país e participa da fundação da área de linguística no Museu Nacional no Rio de Janeiro em 1956. Algum tempo depois, Pike consegue assinar convênio com o governo e trazer outros missionários da instituição, além de obter permissão para operar rádio e aviões no território nacional (BARROS, 2004, p. 75-77). No entanto, por um período nos anos 1970, em razão da omissão da identidade missionária, a organização não teve seus convênios renovados com muitos países, inclusive o Brasil (STOLL, 1985 apud BARROS, 2004, p. 78-79).

Aqui é importante mencionar que, no campo da pesquisa em linguística indígena, existem muitas divergências em relação às ações da SIL; isto em razão de seu objetivo principal, que é a evangelização. Nesse sentido, é possível dizer que existe, no campo da linguística indígena, uma polarização: colocando como extremos, de um lado, os pesquisadores afiliados à linguística antropológica e, de outro, aqueles que tem alguma afiliação com entidades missionárias (BARROS, 2004; FRANCHETTO, 2008).

5. SIL International, instituição anteriormente nomeada *Summer Institute of Linguistics* e, por vezes, identificada em português como *Sociedade Internacional de Linguística*, é uma organização cujo objetivo primário é a tradução Bíblia. O artigo *A missão Summer Institute of Linguistics e o indigenismo latino-americano: história de uma aliança (décadas de 1930 a 1970)*, de Maria Cândida Drumond Mendes Barros (2004), aborda a história da aliança entre essa missão evangélica e intelectuais e políticos latino-americanos, e procura reconstruir o contexto político e intelectual dessas alianças.

No entanto, é preciso reconhecer o mérito dessa instituição. Não somente em termos de documentação linguística, mas especialmente pela contribuição para o desenvolvimento técnico das áreas da tecnologia de informação e tecnologia de escrita. O SIL é responsável pela criação e desenvolvimento de várias tecnologias, e.g. famílias de fontes *open-source*, teclados etc.

7.5 Estudos contemporâneos

Nos últimos 50 anos, a linguística indígena cresceu consideravelmente no Brasil, com um aumento substancial no número de publicações científicas, instituições de pesquisa e programas de pós-graduação na área. De acordo com a pesquisadora Lucy Seki (1999, p. 258):

Até bem recentemente a expressão “linguística indígena” não seria apropriada em nosso contexto, não só em decorrência da situação incipiente da Linguística no País, mas também porque nossas línguas indígenas praticamente não tinham lugar nesta ciência. O fato de hoje a expressão soar natural indica uma mudança promissora de perspectiva: pressupõe a delimitação, dentro da Linguística Brasileira, de um campo de estudos voltado para as nossas línguas e constitui um indicativo de que elas começaram a ocupar o espaço que lhes era devido nessa ciência.

Nesse sentido, é importante citar a atuação do professor Aryon Dall’Igna Rodrigues (falecido em 2014), maior referência brasileira no assunto. Com mais de 70 anos de atuação, foi responsável pela formação de gerações de pesquisadores e pela fundação do Laboratório de Línguas Indígenas da Universidade de Brasília, um dos mais importantes centros de referência da área, abrigando um dos primeiros cursos de pós-graduação em linguística indígena no país. Além da Universidade de Brasília, Rodrigues atuou também em outras universidades; para mencionar a produção de outros centros de pesquisa, talvez o mais notável seja a Universidade de Campinas.

Outros importantes centros de pesquisa são o Museu Paraense Emílio Goeldi, em Belém, e o Museu Nacional da UFRJ, no Rio de Janeiro. No Museu Goeldi, a linha de pesquisa Linguística Indígena se concentra na realidade linguística da região amazônica. O Museu Nacional da UFRJ, abriga o programa de Mestrado Profissional em Linguística

e Línguas Indígenas (PROFLIND). Recentemente, em virtude do incêndio que destruiu grande parte de suas coleções, teve suas exposições fechadas por tempo indeterminado. É triste notar como a história se repete: em 1911 um incêndio consumiu as instalações da Imprensa Nacional, destruindo grande parte do acervo de materiais tipográficos acumulado desde a fundação da instituição em 1808.

Na primeira década do século XXI, a partir da instalação de políticas públicas de incentivo à educação escolar de populações indígenas em suas línguas originárias, surgem também programas para treinamento de nível superior de professores indígenas e de fomento à produção de materiais didáticos em suas línguas nativas (CUNHA, 2008, p. 149).

Por último, é importante mencionar o IPHAN como um importante centro de documentação e políticas públicas. Entre as suas iniciativas, destaca-se o Inventário Nacional da Diversidade Linguística (INDL), uma política voltada para o reconhecimento da diversidade linguística como patrimônio cultural, por meio da identificação, da documentação e de ações de apoio e fomento. A produção de conhecimento sobre as línguas é elemento estruturante dessa política, pois parte considerável das línguas existentes ainda é pouco conhecida (IPHAN, 2014).

Nessa linha, é importante mencionar duas publicações recentes do IPHAN, produzidas nesse contexto: os dois volumes do *Guia de pesquisa e documentação do INDL* (GARCIA et al., 2016) e a reedição do *Mapa Etno-histórico do Brasil e Regiões Adjacentes* de Curt Nimuendajú (2017), uma importante documentação da linguística indígena no Brasil, trabalho desenvolvido pela equipe técnica do INDL, com parceria da Universidade Federal do Pará, do Museu Paraense Emílio Goeldi, do Museu Nacional/UFRJ e do Instituto Brasileiro de Geografia e Estatística (IBGE).

8. Codificação e linguagem

Como um dos reflexos das faculdades da linguagem, de articular e comunicar pensamentos, a espécie humana desenvolveu a escrita. E, entre as diversas ferramentas de escrita utilizadas pelo homem, a mais recente (e mais complexa) é o computador¹, que consiste numa máquina não somente para a escrita, mas também para leitura e armazenamento de textos e para comunicação e intercâmbio de dados; tudo isso embrulhado em um único pacote (HARALAMBOUS, 2007, p. 1).

Nesse espaço digital para escrita, questões relacionadas à codificação² têm relevância, em maior ou menor grau, para todos que lidam com texto em um computador. Para o presente trabalho, a questão da codificação tem grande importância, uma vez que os problemas relacionados a ela representam uma das maiores dificuldades na composição de texto em línguas indígenas brasileiras e, por sua vez, na comunicação nessas línguas. Assim, para uma melhor compreensão das relações entre linguagem e codificação, é imprescindível revisar alguns conceitos elementares sobre codificação de caracteres e seu funcionamento em computadores, sistemas operacionais, aplicativos e teclados.

Uma vez compreendidas as questões relativas à codificação, torna-se possível ampliar a discussão, para que se tenha melhor compreensão do problema na tese. Por essa razão, este capítulo está dividido em duas grandes seções. A primeira delas faz um levantamento da terminologia e das definições acerca dos processos de codificação usados

1. O termo computador (exceto quando especificado) refere-se não somente a desktops e laptops, mas também a outros dispositivos eletrônicos controlados por um sistema operacional, como *smartphones* e *tablets*.

2. Ao longo deste trabalho, utiliza-se o termo codificação como uma forma abreviada para codificação de caracteres (em ambientes computacionais), termo que será descrito com maiores detalhes neste capítulo.

ao longo da tese. A segunda parte procura explicar em detalhes o funcionamento da codificação de texto em computadores. Nesse sentido, é importante listar algumas das principais referências deste capítulo.

O livro *Language culture type: international type design in the age of Unicode*, editado pela Association Typographique Internationale (ATypI) em 2002, foi um importante ponto de partida para este capítulo. Esse trabalho reúne textos de renomados especialistas em tipografia, escrita e tecnologia. Dois textos dessa obra têm maior relevância para este trabalho: o capítulo introdutório *Voices, languages and scripts around the world*, de Robert Bringhurst, e o capítulo *Unicode, from text to type*, de John Hudson. Esses dois textos fornecem uma boa introdução sobre sistemas de escrita e sobre relações entre codificação e tipografia.

No campo da escrita, o linguista alemão Florian Coulmas é considerado uma das maiores autoridades. Da bibliografia pesquisada sobre escrita sua obra (1991; 1999; 2003) apresenta com bastante precisão as definições e terminologias relativas à área. Ainda no campo da escrita, o trabalho *O traço: teoria da escrita* (2013), do designer e educador holandês Gerrit Noordzij, analisa a escrita de uma “perspectiva grafológica”, demonstrando as qualidades caligráficas subjacentes em todas as letras, independentemente da tecnologia que as produziu.

Sobre o conceito de informação, é importante mencionar algumas referências que têm um sentido especial neste trabalho; em particular, a Teoria Matemática da Comunicação, de Claude Shannon e Warren Weaver. No campo da comunicação, é uma das poucas teorias que trata de problemas de codificação.

Como um dos principais objetivos desta pesquisa é a produção de documentação, de ferramentas e de metodologia para o trabalho com tipografia e línguas indígenas em meio digital, é importante definir o *The Unicode Standard* como principal referência sobre codificação de texto em meio digital. Neste trabalho, o termo Unicode é utilizado de algumas maneiras. Primeiramente, é importante mencionar a publicação *The Unicode Standard: version 13.0 – Core Specification* (2020a), editada pelo Unicode Consortium. Essa publicação define o padrão Unicode (Unicode Standard), que é o padrão universal de codificação de caracteres para processamento de texto em computadores, que

possibilita o intercâmbio de informação textual internacionalmente. Os termos Unicode e padrão Unicode referem-se à própria especificação, i.e., ao conteúdo do livro *The Unicode Standard*, portanto, do padrão Unicode (Unicode Standard). Quando for necessário fazer referência à publicação em si, esta será feita com destaque em itálico, como na frase anterior. O Unicode Consortium é a organização responsável pela publicação e manutenção do Unicode.

Também é importante mencionar a especificação OpenType, *OpenType Specification* (MICROSOFT CORPORATION; ADOBE INC., 2020, n.p.). O OpenType é um formato de fonte digital compatível com o Unicode, com amplo suporte linguístico e compatível com os principais sistemas operacionais atuais. Esta tecnologia é uma importante ponte de ligação entre os campos da codificação, linguagem, tipografia.

Nesse contexto, é importante mencionar a obra *Fonts & Encodings: From Advanced Typography to Unicode and Everything in Between* (2007), do matemático grego Yannis Haralambous. Pode-se dizer que esse trabalho é uma espécie de “versão comentada” do Unicode, e faz uma ponte entre a documentação Unicode e o campo tipográfico. Haralambous explica como se deu a evolução da codificação de caracteres, desde a década de 1960, explicando o funcionamento do Unicode de forma mais clara e fazendo a ponte com o campo da tipografia. Nessa linha, autores como Jukka K. Korpela (2006) e Peter Constable (2001) também trouxeram contribuições relevantes.

No campo da tipografia, mais especificamente no campo do design de tipos, o Prof. Gerard Unger é uma referência importante. Em 2018, publicou *Theory of type design*, seu último trabalho. Na tentativa de criar uma teoria específica para a área de design de tipos, Unger nos ajuda a compreender e a definir a tipografia como campo do conhecimento. Alguns outros autores, como Paul Luna (2018), Robin Kinross (2010) e Gerry Leonidas (2018) trazem alguns apontamentos breves, porém bastante relevantes, sobre a tipografia como campo projetual.

Ao longo desta seção, menciona-se alguns outros autores que também discutem as relações entre tipografia e tecnologia, como John Hudson (2002), José Scaglione (2006; 2019), Richard Southall (2005), Alice Savoie (2014), Gerard Unger (2018) e Ferdinand Ulrich (2018).

8.1 Terminologia e definições

Por se tratar de tema que envolve muitos termos, de diferentes áreas (informática, comunicação, linguística, ortografia, tipografia, design, artes gráficas), considera-se importante uma introdução à terminologia usada neste capítulo e ao longo da tese. Não se pretende aqui esgotar as definições de termos como código, língua, codificação, escrita etc., num sentido amplo, mas aproximá-lo dos problemas desta pesquisa.

Nesta seção, o formato adotado é o de uma espécie de terminologia comentada. O texto apresenta-se de forma um pouco fragmentada, como é da natureza de obras de referência. Entende-se como necessário o acréscimo de comentários acerca de termos essenciais a este trabalho.

Nem todos os termos estão listados, como tradicionalmente em um glossário. Foram criadas algumas categorias principais, com subdivisões quando necessário. Muitos termos (e.g. grafema, fonema) são explicados dentro de um contexto de uso específico. Vários termos aqui explicados são essenciais para a compreensão de conceitos em seções posteriores a esta.

Ao longo da tese, adotou-se, para a representação de grafemas, os sinais de menor e maior, e.g. <a>, <é>, <Õ>, com base na notação utilizada na literatura da linguística: /a/ = fonêmico, [a] = fonético, <a> = grafêmico. Por extensão, utilizou-se a essa notação para representação não somente de grafemas, mas de caracteres, de uma forma geral.

8.1.1 Língua

O dicionário *Michaelis* define língua como “conjunto de palavras ou signos vocais e regras combinatórias estabelecidas, de que fazem uso os membros de uma comunidade para se comunicar e interagir; idioma” (LÍNGUA, 2015).

No mesmo verbete é listada uma definição atribuída a Ferdinand de Saussure, que descreve a língua como um “sistema abstrato de signos, subjacente à fala e à escrita, usado por uma comunidade e que se opõe à sua realização individual; *langue*” (LÍNGUA, 2015).

A partir dessas definições, o termo língua, neste trabalho, é utilizado no sentido de idioma; como um código verbal utilizado por uma comunidade para a comunicação.

Dada a temática desta tese, é importante lembrar que as línguas indígenas brasileiras eram línguas originalmente ágrafas. Ou seja, são línguas que não tinham registro

escrito, de tradição oral. Por isso, a título de complementação, é importante mencionar a questão da relação dialética entre língua e fala. Nesse sentido, Coelho Netto, em *Semiótica, informação e comunicação* (1980, p. 18–19) observa:

Na teoria da linguagem de Saussure surge a oportunidade e a necessidade de distinguir-se entre o social e o individual e entre o essencial e o acessório ou acidental. O social e o essencial recaem no domínio da língua, cabendo à fala o recorte do que é individual e acidental. Em outras palavras, a língua (no dizer de R. Barthes, a linguagem menos a fala) apresenta-se ao indivíduo como um sistema preexistente, uma instituição social que acumulou historicamente uma série de valores e sobre a qual, em princípio, o indivíduo não tem nenhuma ascendência enquanto indivíduo.

A fala, inversamente, é um ato individual de utilização da língua, um modo de combinar os elementos da língua no ato de comunicação. Se a língua é um sistema (conjunto de elementos com relação determinada entre si), a fala é um processo (sequência de atos) que atualiza, que dá existência concreta a essa língua, tornando a comunicação um fenômeno e não mais uma simples potencialidade.

Língua e fala mantém uma relação dialética entre si, de tal modo que se pode afirmar não existir língua sem fala ou fala sem língua. [...]

Usando como exemplo uma linguagem natural qualquer, seria possível dizer que sua língua seria o conjunto de todos os termos constantes de um dicionário ideal mais as normas de combinação possível entre eles. A fala seria uma mensagem qualquer formada a partir desse conjunto [...]

É importante mencionar que a visão de Saussure, na qual a fala/discurso exerce papel central nos estudos sobre linguagem, ainda tem forte influência em vários campos; especialmente no campo da linguística. Florian Coulmas (2003, p. 10) observa que, no campo das ciências da linguagem, ainda predomina um pensamento ortodoxo que corrobora o posicionamento de Saussure. Um dos argumentos de Coulmas é sobre o fato de que os linguistas (por influência de uma lógica aristotélica, à qual Saussure adere) tendem a negligenciar a problemática da escrita:

Língua e escrita são dois sistemas distintos de signos; a única razão de ser do segundo é representar o primeiro; o objeto lingüístico não se define pela combinação da palavra escrita e da palavra falada; esta última, por si só, constitui tal objeto. (SAUSSURE, 2006, p. 34)

Coulmas (2003, p. 13) critica a questão da primazia da fala/discurso sobre a escrita na linguística. Nesse sentido, observa que Jacques Derrida “utilizou o termo gramatologia para designar uma teoria da escrita que ele entende como uma crítica ao logocentrismo da tradição intelectual ocidental desde Aristóteles, que considera o signo (escrita) como mero suplemento ao invés de uma força epistêmica por direito próprio” (COULMAS, 1999, p. 173, tradução nossa).

8.1.2 Código

Vilém Flusser (2007, p. 130) explica que: [...] um código é um sistema de símbolos. Seu objetivo é possibilitar a comunicação entre os homens. Essa passagem, mesmo quando tirada de seu contexto original, nos ajuda a entender, de forma sucinta, a importância do código para o campo da comunicação. Nesse sentido, o texto a seguir reúne uma série de definições para complementar essas ideias.

Florian Coulmas, em *The Blackwell Encyclopedia of Writing Systems* (1999, p. 89, tradução nossa), apresenta a seguinte definição de código:

Qualquer sistema de símbolos usado para comunicar mensagens. Na teoria da informação, o termo é usado para sistemas que atribuem os elementos de um sistema de signos aos de outro, podendo assim representar o mesmo conteúdo de informação.

No dicionário *Michaelis* (CÓDIGO, 2015), encontram-se as seguintes definições de código, específicas às áreas da informática, linguística e comunicação, particularmente úteis a este trabalho:

INFORM Regras usadas para converter instruções ou dados de uma forma para outra; COMUN, LING Sistema de signos devidamente organizados e convencionados, a fim de permitir a elaboração e transmissão de mensagens.

Ainda nesse verbete, é apresentada a expressão *código linguístico*:

Conjunto de signos linguísticos de uma língua e das regras gramaticais que os regem, possibilitando a comunicação entre as pessoas que falam essa língua (CÓDIGO, 2015).

Por último, Coelho Netto (1980, p.19) observa ainda que “[...] o código [é] o esquema que permite a elaboração da mensagem e o ponto de referência a partir do qual é possível determinar o significado desta mensagem [...]”.

Voltando à definição de que o código consiste essencialmente em um sistema de regras, um esquema, a partir do qual é possível estruturar e transmitir informação, neste trabalho é necessário observar o termo por dois ângulos.

Primeiramente, sob o ponto de vista que a língua escrita é um código, um sistema de sinais gráficos que a representam, e que esses sinais, devidamente organizados e convencionados, permitem a elaboração e transmissão de mensagens. A construção da informação textual em meios digitais depende desse código, i.e. desse conjunto de sinais gráficos que constituem a língua escrita, ferramenta primordial para a comunicação. O artifício da escrita como código exerce papel central neste trabalho.

Em segundo lugar, também é importante mencionar que, sob a ótica da informática, a definição de código (“regras usadas para converter instruções ou dados de uma forma para outra”) implica a conversão da informação textual em dados binários, a única forma de informação que um computador pode processar e/ou armazenar. Essa noção de código é fundamental para o entendimento da codificação de caracteres em meios digitais.

8.1.3 Ortografia

Segundo Michael Cahill (2001, p. 1, tradução nossa), “ortografia é um sistema para representar um idioma na forma escrita; inclui não apenas caracteres individuais (grafemas), mas também quebras de palavras, pontuação, diacríticos, capitalização (i.e. uso de maiúsculas), como dividir e hifenizar palavras no final das linhas”.

Complementando, Florian Coulmas (1999, p. 379, tradução nossa) define ortografia como “grafia correta e a parte da gramática que trata das regras de grafia correta. Uma ortografia é uma seleção normativa das possibilidades de uma escrita para escrever um determinado idioma de maneira uniforme e padronizada. [...] Em línguas escritas em ordem alfabética, os aspectos da escrita mais comumente codificados por meio de regras ortográficas são correspondência grafema-fonema, divisão de palavras, hifenização, letras maiúsculas e a grafia de palavras emprestadas [...]”.

Acerca das metodologias de desenvolvimento de ortografias para línguas ágrafas, é bastante recorrente a referência a Kenneth Pike, em especial à obra *Phonemics: A Technique for Reducing Languages to Writing*. Segundo o próprio autor (1947, p. 248, tradução nossa), a redução de uma língua à escrita (*reducing a language to writing*) consiste na “análise dos sons de uma língua, especialmente em termos de seus fonemas, e a atribuição de símbolos ortográficos para esses sons, para a preparação de uma literatura prática na língua; isso envolve uma análise parcial da gramática como pré-requisito para um tratamento adequado de alguns dos problemas fonêmicos, como aqueles que envolvem a divisão de palavras”.

Nesse sentido, Florian Coulmas (1999, p. 435, tradução nossa) ressalta ainda a forte associação da redução à escrita ao uso do alfabeto: “este termo é limitado [...] a sistemas deliberadamente criados, que tomam um alfabeto com correspondência grafema-fonema próxima como o padrão de excelência”. Observa também que a redução à escrita é um pré-requisito necessário para a submissão de uma língua à análise gramatical.

Logo, entende-se o termo *redução para a escrita* como sinônimo para desenvolvimento de ortografia, utilizado principalmente no contexto de línguas ágrafas.

8.1.4 Informação

Esta seção tem como objetivo conceituar o termo informação a partir da obra *Contribuições recentes à teoria matemática de comunicações*, de Warren Weaver. Trata-se do primeiro de dois textos do livro *Teoria matemática da comunicação*, de Shannon e Weaver, publicado em 1949, que “[...] consiste, parcialmente, de uma introdução expositiva à teoria geral e deve ser lida primeiro por aqueles que desejam panorâmica do amplo escopo da teoria [...]” (SHANNON; WEAVER, 1975, p. 1). A segunda parte do livro traz o texto de Claude Shannon na íntegra, que trata, principalmente, das formulações matemáticas de sua teoria. O texto de Shannon foi originalmente publicado em 1948 no *Bell System Technical Journal* e o de Weaver, em 1949 na *Scientific American*.

Reconhecida no campo da comunicação, a teoria de Shannon e Weaver é uma das poucas, nesse campo, que trata de problemas de codificação. Weaver explica que, na tese, o termo comunicação é pensado em um sentido amplo, e classifica os problemas

de comunicação em três níveis: técnico (relacionado à *exatidão* da transmissão dos símbolos de comunicação), semântico (relacionado à *precisão* na transferência) e de eficiência (condicionados ao êxito com que a mensagem transmitida ao receptor levará este à conduta desejada e prevista). A discussão reflete, predominantemente, sobre o problema técnico que, de acordo com Weaver, tem uma influência predominante sobre os problemas de semântica e eficiência (SHANNON; WEAVER, 1975, p. 4–6).

Weaver (SHANNON; WEAVER, 1975, p. 9) descreve que “...a palavra *informação* é utilizada com um sentido especial, que de maneira alguma deverá ser confundido com o de seu uso generalizado. Especificamente, *informação* não deve, por equívoco, ser compreendida como significado”. Deve-se entender, a partir dessa premissa, que Shannon e Weaver se referem à informação no sentido de informação binária, medida em *bits*. Edgar Morin (1977, p. 276) explica que “Shannon determina a informação como grandeza observável e mensurável e esta torna-se a pedra angular da teoria da comunicação [...]”.

Segundo Mauro Wolf (1987, p. 98–99), “a origem do modelo [da teoria matemática da comunicação] vai buscar-se aos trabalhos de engenharia das telecomunicações. [...] Todos estes estudos [Nyquist, Hartley] têm por objectivo melhorar a velocidade de transmissão de mensagens, diminuir as suas distorções e aumentar o rendimento global do processo de transmissão de informação. [...] A teoria matemática da comunicação é, essencialmente, uma teoria sobre a *transmissão* óptima das mensagens e o esquema do ‘sistema geral de comunicação’, proposto por Shannon, é o seguinte”:

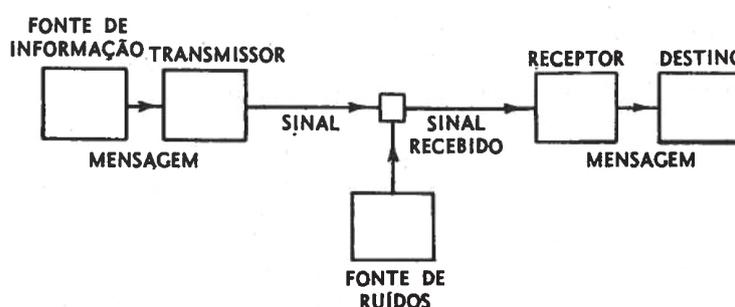


Figura II: Diagrama do sistema geral de comunicação, conforme Shannon & Weaver.

Fonte: SHANNON; WEAVER, 1975, p.7.

Sobre o diagrama da Fig.10, Edgar Morin (1977, p. 276) explica:

Esta teoria nasceu de preocupações práticas. A sociedade Bell procurava o modo mais económico e mais seguro para transmitir as mensagens. Assim o quadro original da teoria é o dum sistema de comunicações onde um emissor transmite uma mensagem a um receptor através dum canal dado. Por hipótese, emissor e receptor têm um repertório comum (código que contém as categorias de sinais utilizáveis); assim, a mensagem em código é transmitida, do emissor ao receptor, através do canal, sob a forma de signos ou sinais que podemos decompor em unidades de informação chamadas *bits* (*binary digits*).

O *bit* pode definir-se como um acontecimento que desfaz a incerteza dum receptor colocado diante duma alternativa, na qual as duas saídas são, para ele, equiprováveis. Quanto mais numerosas forem as eventualidades que este receptor pode encarar, mais acontecimentos informativos comporta a mensagem, mais aumenta a quantidade de *bits* transmitidos.

O termo bit^3 é um acrónimo de *binary digit*, dígito binário. Em termos matemáticos, 1 *bit* significa 2 elevado à primeira potência. Logo, $2^1 = 2$, ou seja, duas alternativas possíveis⁴, somente dois valores: 0 e 1. Assim, $2^2 = 4$, ou dois bits, quatro alternativas possíveis ou quatro bits de informação. Outra unidade, o byte, consiste em um conjunto de bits adjacentes, constituído por oito bits: um byte = oito bits⁵ (HARALAMBOUS, 2007, p. 27–29; HUDSON, 2002, p. 26; SHANNON; WEAVER, 1975, p. 10).

Morin (1977, p. 278) aponta, inclusive, que “... o *bit* não é uma unidade de sentido. A informação shannoniana chega a ser inteiramente muda ou cega quanto à significação, à qualidade, ao valor e ao alcance da informação para o receptor”. Para ilustrar

3. Bit: termo cunhado pelo estatístico norte-americano John Tukey, em 1946.

4. Em matemática, denomina-se lógica booleana (de George Boole) ao ramo da álgebra centrada nos operadores booleanos; i.e. conjunto de valores limitado a *falso* e *verdadeiro*, ou 0 e 1. No Unicode, as propriedades booleanas também são comumente chamadas de propriedades binárias (UNICODE CONSORTIUM, 2020a, p. 96).

5. Byte: termo cunhado por Werner Buchholz em 1956, durante o desenvolvimento do computador IBM Stretch. Um byte era originalmente definido como uma sequência de um a seis bits. No final daquele ano, A IBM decidiu estabelecer como padrão o byte de 8 bits, amplamente adotado pela indústria posteriormente (HUDSON, 2002, p. 42).

a propriedade da equiprobabilidade da informação, o autor (MORIN, 1977, p. 278) utiliza os seguintes exemplos:

Isolda espera o regresso de Tristão; sabe que uma vela branca anunciará o regresso do seu amante, uma vela negra, sua morte; para ela, os dois ramos desta alternativa são equiprováveis. Ora, que se passará na informação shannoniana com as alternâncias de esperança e de desesperança de Isolda, com as suas e com a sua espera, com o amor desmedido, com o espectro da morte? Quando, no mar infinito, surgir uma vela, branca ou negra, o observador shannoniano factura: um *bit*! Consideremos o poema *La Rivière de Cassis*. É um agrupamento original e complexo, e, portanto, improvável na sucessão de letras e de palavras, e pode ser decomposto num total n de *bits*, equivalente ao número de decisões que o receptor deveria tomar para identificar as letras ou as palavras que constituem o poema. Todavia, tal lista não nos diz nada acerca do sentido do poema: este comportaria a mesma quantidade de informações se as letras estivessem dispostas ao acaso, ou seja, se se tornassem puro ruído. A quantidade de informação nem sequer nos dá uma indicação sobre a originalidade ou beleza do poema: o aumento dos *bits* só nos fala do aumento aritmético da improbabilidade, o que não está directamente ligado à qualidade poética. Um poema que, com um número de letras igual, comportasse um número menor ou maior de *bits* não seria por isso menos ou mais poético.

Ainda sobre o conceito de informação, Umberto Eco esclarece: “Aparentemente, o problema do teórico da informação parece ser o de ‘pôr em código’ uma mensagem, [...] mas, na realidade, o teórico da informação não está imediatamente interessado na correlação entre sinais binários e o seu possível conteúdo alfabético. Está interessado na forma informação, como medida estatística [...] mais económica de transmitir os sinais, sem gerar ambiguidade e neutralizando ruídos no canal ou erros de transmissão [...] O problema da teoria da informação é a sintaxe interna do sistema binário, não o facto de as sequências expressas pelo sistema binário poderem exprimir, como seu conteúdo, letras alfabéticas ou qualquer outra sequência de valor” (ECO, 1972, p. 8 apud WOLF, 1987, p. 101).

Em relação a isso, o texto de Wolf (1987, p. 102) traz um exemplo de Robert Escarpit sobre a codificação de mensagens, fazendo uma analogia com o trabalho de um telegrafista. A perspectiva do telegrafista é “pôr em código” as mensagens, sendo

seu significado pouco relevante para o mesmo. Enquanto o remetente (fonte de informação) e o destinatário (destino) estão de fato interessados no conteúdo e significado transmitidos, para o empregado dos correios (transmissor) estes são irrelevantes.

O objetivo desses processos é, principalmente, evitar a deformação da fidelidade do sinal codificado na fonte da informação (emissor) e na transmissão. Weaver (1975, p. 14, grifo nosso) reforça essa ideia ao declarar que “**quando visto do ângulo da engenharia, um sistema de comunicação tem que encarar o problema de manipulação de qualquer mensagem que a fonte [de informação] produzir**”.

8.1.5 Codificação

Em termos gerais, pode-se definir codificação como processo, ato ou efeito de codificar, de reduzir a código, de criar os sinais de um código (CODIFICAR, 2015; CODIFICAÇÃO, 2015).

O trecho a seguir define codificação de uma outra perspectiva:

Codificação refere-se ao processo de representar informações de alguma forma. A linguagem humana é um sistema de codificação pelo qual representamos informações em termos de sequências de unidades lexicais e em termos de sequências de sons ou gestos. A linguagem escrita é um sistema derivado de codificação pelo qual aquelas sequências de unidades lexicais, sons ou gestos são representadas em termos dos símbolos gráficos que compõem algum sistema de escrita (CONSTABLE, 2001, tradução nossa).

Nesta tese, o termo codificação tem uma definição principal e é empregado, primeiramente, como sinônimo de *codificação de caracteres* (em inglês, *character encoding*; ou somente *encoding*); ou seja, à conversão da informação textual em dados binários, processo que será mais bem descrito na seção *Codificação de caracteres* deste trabalho. A citação a seguir ilustra brevemente esse conceito:

Um computador é uma máquina que armazena e manipula números; [isto] significa que texto, para ser inserido, armazenado, editado e digitado em um computador, deve ser representado como uma sequência de números. Estes números devem ser, por sua vez, mapeados a métodos de entrada de caracteres, e.g. teclas de um teclado, e à representação visual dos caracteres em uma fonte. Um

padrão de codificação de caracteres, como o Unicode, define a relação entre as menores unidades/componentes da escrita necessárias para entrada e armazenamento de texto – i.e. caracteres – e os números que os representam em texto codificado em computador (HUDSON, 2002, p. 26, tradução nossa).

Complementando a citação acima, Peter Constable (2001) atenta para o fato de que uma codificação de caracteres compreende, essencialmente, no mínimo dois componentes: um conjunto de caracteres e algum sistema para representá-los em termos das unidades de processamento usadas no computador, i.e. números binários. Há várias maneiras de fazer isso, várias formas de associar caracteres a números binários. Dependendo da convenção que se assume, os números podem significar coisas diferentes, como demonstra a Fig. 12 a seguir. Nesse sentido, Constable observa ainda que padrões de codificação são importantes por, pelo menos, dois motivos: por fornecerem uma base para os desenvolvedores (para a criação de software com comportamentos de texto apropriados) e por permitirem intercâmbio de dados entre usuários.

Dados armazenados

66	105	108	108	32	115	108	111	119	108	121	32	16	117	114	110	101	100
----	-----	-----	-----	----	-----	-----	-----	-----	-----	-----	----	----	-----	-----	-----	-----	-----

Como os dados são interpretados de acordo com a codificação ASCII

B i l l s l o w l y t u r n e d

Como os dados podem ser interpretados de acordo com outra codificação

ב י ל ל ש ל א מ ת ר ב ע ה

Figura 12: Interpretação de dados numéricos, de acordo com diferentes padrões de codificação. Imagem criada a partir do original de Peter Constable. Composto em Brill Roman e Lucida Grande (hebraico). Fonte: CONSTABLE, 2001.

Ainda, é importante entender que o termo codificação é utilizado como sinônimo para página de código (em inglês, *codepage*). Ou seja, a palavra codificação é utilizada para descrever a própria página de código, uma tabela que associa caracteres a números (KORPELA, 2006, p. 174), numa ordem específica, como no exemplo a seguir, na Fig. 13. Haralambous (2007, p. 45) observa que o termo página de código (em inglês,

codepage) foi cunhado pela Microsoft. Em resumo, o termo codificação é usado para o produto, o resultado final de uma codificação, que é a página de código ou a tabela que descreve um conjunto de caracteres.

Ä ₈₀	Å ₈₁	Ç ₈₂	É ₈₃	Ñ ₈₄	Ö ₈₅	Ü ₈₆	á ₈₇	à ₈₈	â ₈₉	ä _{8A}	ã _{8B}	å _{8C}	ç _{8D}	é _{8E}	è _{8F}
ê ₉₀	ë ₉₁	í ₉₂	ì ₉₃	î ₉₄	ï ₉₅	ñ ₉₆	ó ₉₇	ò ₉₈	ô ₉₉	ö _{9A}	õ _{9B}	ú _{9C}	ù _{9D}	û _{9E}	ü _{9F}
† _{A0}	° _{A1}	¢ _{A2}	£ _{A3}	§ _{A4}	• _{A5}	© _{A6}	ß _{A7}	® _{A8}	© _{A9}	™ _{AA}	´ _{AB}	¨ _{AC}	≠ _{AD}	Æ _{AE}	Ø _{AF}
∞ _{B0}	± _{B1}	≤ _{B2}	≥ _{B3}	¥ _{B4}	μ _{B5}	∂ _{B6}	Σ _{B7}	Π _{B8}	π _{B9}	∫ _{BA}	^a _{BB}	^o _{BC}	Ω _{BD}	æ _{BE}	ø _{BF}
¿ _{C0}	¡ _{C1}	¬ _{C2}	√ _{C3}	ƒ _{C4}	≈ _{C5}	Δ _{C6}	« _{C7}	» _{C8}	· _{C9}	NBSP _{CA}	À _{CB}	Ã _{CC}	Õ _{CD}	Œ _{CE}	œ _{CF}
– _{D0}	— _{D1}	“ _{D2}	” _{D3}	‘ _{D4}	, _{D5}	÷ _{D6}	◇ _{D7}	ÿ _{D8}	Ÿ _{D9}	/ _{DA}	⌘ _{DB}	< _{DC}	> _{DD}	fi _{DE}	fl _{DF}
‡ _{E0}	· _{E1}	, _{E2}	” _{E3}	% _{E4}	Â _{E5}	Ê _{E6}	Á _{E7}	Ë _{E8}	È _{E9}	Í _{EA}	Î _{EB}	Ï _{EC}	Ì _{ED}	Ó _{EE}	Ô _{EF}
⌘ _{F0}	Ò _{F1}	Ú _{F2}	Û _{F3}	Ü _{F4}	ı _{F5}	^ _{F6}	~ _{F7}	- _{F8}	˘ _{F9}	· _{FA}	° _{FB}	˘ _{FC}	¨ _{FD}	‘ _{FE}	˘ _{FF}

Figura 13: Exemplo de página de código. Apple Standard Roman, parte da tabela complementar ao ASCII. Fonte: HARALAMBOUS, 2007, p.47.

Em segundo lugar, pela relação deste trabalho com o campo da linguística, considera-se também importante listar uma segunda definição para o termo codificação, conforme o verbete *codification*, da publicação *The Blackwell Encyclopedia of Writing Systems*:

A regulação de uma língua por regras prescritivas baseadas em uma descrição analítica de sua estrutura. [...] Por meio da codificação, as regras recebem status normativo, ou seja, são modelos designados para orientar o uso correto. As línguas são codificadas em vários graus e em diferentes níveis linguísticos: lexical, fonológico, ortográfico e gramatical. A codificação normalmente pressupõe a escrita, mesmo porque seria difícil memorizar uma descrição abrangente de todos os níveis de um sistema de linguagem sem um registro escrito. [...] **De qualquer maneira, a codificação, no sentido moderno, depende da escrita, mesmo no que diz respeito à pronúncia.** Dicionários e gramáticas são as principais ferramentas de codificação. A tarefa de codificação é frequentemente confiada a academias de línguas, como a Academie Française, que publicam os resultados de suas deliberações, fornecendo à língua em questão um arcabouço normativo formulado por escrito (COULMAS, 1999, p. 89–90, tradução nossa, grifo nosso).

Conforme indicado por Florian Coulmas, o desenvolvimento ou especificação de uma ortografia faz parte de um processo, de um nível de codificação linguístico. Trazendo a definição de Coulmas para o contexto desta tese, pode-se ilustrar o processo de desenvolvimento de uma ortografia para uma língua indígena brasileira. Em linhas gerais, pode-se dizer que a criação de uma convenção escrita para os sons de uma língua indígena consiste em um ato de codificar a escrita dessa determinada língua. O desenvolvimento da ortografia para uma língua ágrafa consiste num processo de codificação que transforma o intangível e efêmero (a fala) em algo tangível, permanente e documentado (escrita).

É importante demonstrar a diferença entre os dois conceitos de codificação aqui descritos, equivalentes aos termos (em inglês) *codification* e *encoding*, conforme parágrafo a seguir.

As academias de línguas, responsáveis pela especificação de ortografias, trabalham com a codificação (*codification*) de maneira descritiva, documental. Isso significa, por exemplo, que o texto do *Vocabulário Ortográfico da Língua Portuguesa* (publicação organizada pela Academia Brasileira de Letras), especifica quais letras do alfabeto latino⁶ são utilizadas na notação escrita da língua portuguesa, quais são os dígrafos utilizados nessa língua, como se dá a acentuação gráfica das palavras oxítonas, como se dá emprego do acento grave etc. No entanto, o *Vocabulário Ortográfico...* não faz nenhuma referência às relações da língua portuguesa com as codificações de caracteres (*encodings*) e mapas de teclados utilizados no Brasil. Questões como essas, de caráter mais técnico, são padronizadas de acordo com normas da Associação Brasileira de Normas Técnicas (ABNT), e também de acordo com as especificações do Unicode e da International Organization for Standardisation (ISO).

Se na língua portuguesa o elo de ligação entre a ortografia (*codification*) e a codificação de caracteres (*encoding*) não é bem documentado, o que se pode dizer a respeito

6. Nota-se que o *Vocabulário Ortográfico...* indica o alfabeto utilizado no português, mas somente as letras de A a Z, e o C cedilha <ç>. Não lista, junto ao alfabeto, os diacríticos necessários para a escrita do português: grave <`>, agudo <´>, circunflexo <^> e til <~>. E também não indica quais letras recebem esses diacríticos. Além disso, ao listar o C cedilha <ç> junto a dígrafos, coloca este em uma outra categoria, como se a cedilha <, > não fosse também um diacrítico.

das línguas indígenas brasileiras? Os capítulos *Metodologia e Desenvolvimento* apontam para possíveis respostas a essas perguntas, uma vez que procuram descrever soluções técnicas para esses problemas.

Em computadores, no entanto, o conceito de codificação tem especificidades que serão descritas nas seções a seguir; particularmente, na seção *Codificação de caracteres*.

8.1.6 Escrita

Da mesma forma que código e língua, o termo escrita tem uma gama de conceitos particulares a algumas áreas, como literatura, filosofia, linguística, semiótica, entre outros. Esta seção procura explorar a concepção de escrita como tecnologia de comunicação.

Por esse ângulo, o termo escrita está relacionado a duas noções principais nesta tese. Primeiro, a uma noção de código, de sistema de representação gráfica da linguagem, por meio sinais (e.g. letras, ideogramas etc.); em inglês, *writing*. Em segundo lugar, o termo é utilizado para descrever conjunto de sinais gráficos próprios de um sistema de escrita, por exemplo, a escrita latina ou a escrita chinesa (ESCRITA, 2015); em inglês, *script*. Observa-se que, em inglês, existem duas palavras diferentes (*writing* e *script*) para dois conceitos diferentes, enquanto em português o termo escrita é utilizado para as duas coisas.

Além destas duas definições de escrita (*writing* e *script*), enumerou-se também algumas diferenças entre escrita e alfabeto, além de se listar definições de sistema de escrita, caracteres e glifos, e alguns outros termos relacionados.

8.1.6.1 Escrita: sistema de representação gráfica da linguagem

Esta seção tem como objetivo conceituar escrita enquanto sistema de representação gráfica da linguagem, por meio de sinais (e.g. letras, ideogramas etc.); em inglês, *writing*. A partir dessa ótica, também apresenta a escrita, do ponto de vista gráfico, como campo de estudo.

Florian Coulmas define escrita (*writing*) nos seguintes termos (ênfase para a primeira definição):

Pode-se distinguir pelo menos seis significados para escrita: (1) um sistema de registro da linguagem por meio de marcas visíveis ou táteis; (2) a atividade de colocar tal sistema em uso; (3) o resultado de tal atividade, um texto; (4) a forma particular de tal resultado, um estilo de escrita, como letras blocadas; (5) composição artística; (6) uma ocupação profissional. Enquanto neste livro minha principal preocupação é com (1), as relações com os outros significados não são acidentais ou sem importância. Os vários usos da “escrita” revelam os muitos aspectos da sociedade e cultura tocados pelo que o antropólogo cultural Jack Goody apropriadamente chamou de tecnologia da mente (COULMAS, 2003, p. 1–2, tradução nossa).

Em *Letterletter*, Gerrit Noordzij, dá exemplos de como é possível examinar o termo escrita (*writing*), quando observado sob a ótica de diferentes campos:

Meu estudo da escrita não tem nada a ver com a linguística, e a minha grafologia [...] rejeita a adivinhação quiromântica. Grafologia deve ser entendida como o estudo da escrita por si mesma [por si só]. Inclui tipografia, que deve ser entendida como escrita com caracteres pré-fabricados.

Grafologia tem uma posição em outros campos de interesse. Esses estudos podem se satisfazer pela observação de um único aspecto da escrita, mas esse isolamento não deve permitir visões distorcidas da escrita.

Paleografia isola a escrita antiga em livros. Epigrafia isola a escrita antiga em paredes [monumentos, edificações]. Diplomacia isola escritos antigos em cartas. Pedagogia isola a escrita infantil. Psicologia isola a percepção da escrita ou funções motoras da escrita. Estética isola a apreciação da escrita. História da imprensa [das artes gráficas] isola a escrita tipográfica. Matemática isola a topologia da escrita. Antropologia cultural isola convenções da escrita.

Não há objeções contra o isolamento como tal, mas eu insisto na contínua reformulação das concepções isoladas em confronto com a grafologia geral.

O que se segue é dedicado, principalmente, a uma tentativa de generalizar fenômenos tipográficos: tipografia de uma perspectiva grafológica (NOORDZIJ, 2000, p. 3, tradução nossa).

A obra supracitada contribui para (e antecede) as bases de *The stroke: theory of writing* (NOORDZIJ, 2005, 2013), na qual descreve estruturas elementares dos traços da escrita e analisa as qualidades e propriedades das letras. Um de seus méritos é ter elaborado um modelo que consegue descrever as propriedades das formas manuscritas

com precisão paramétrica (a partir de parâmetros como contraste⁷, espessura de traço etc.), demonstrando as qualidades caligráficas implícitas em todas as letras, independentemente da tecnologia utilizada para criar essas letras.⁸

Apesar de utilizar algumas concepções um pouco imprecisas (e.g. matemática, antropologia cultural), o autor as utiliza para chegar à delimitação de seu campo de interesse: tipografia de uma perspectiva grafológica. Entende-se isto como uma delimitação de escopo, uma tentativa de entender o termo escrita como campo, com aproximação à área projetual da comunicação visual e, por isso, relevante para este trabalho.

Outra contribuição de Noordzij, de caráter extremamente sintético e preciso, é a sua definição de tipografia: escrita com letras pré-fabricadas (NOORDZIJ, 2000, p. 3). Essa definição reforça o entendimento da escrita como uma tecnologia de comunicação.

Nessa perspectiva, o livro *Language culture type: international type design in the age of Unicode* (2002) constitui importante referência para a tipografia em diferentes sistemas de escrita (árabe, grego, japonês etc.), com textos assinados por renomados especialistas em cada uma dessas escritas. Robert Bringhurst assina o texto introdutório, *Voices, languages and scripts around the world*, com observações relevantes sobre linguagem e escrita. Por meio de comparações e metáforas, o autor (BRINGHURST, 2002, p. 3) faz uma analogia da diversidade linguística com a natureza e com a biodiversidade, como no trecho a seguir (mantido no idioma original):

Drop a word in the ocean of meaning and concentric ripples form. To define a single word means to try to catch those ripples. No one's hands are fast enough. Now drop two or three words in at once. Interference patterns form, reinforcing one another here and canceling each other there. To catch the meaning of the words is not to catch the ripples that they cause; it is to

7. Em caligrafia e tipografia, o termo contraste se refere à diferença de espessura entre traços finos e grossos. Na escrita latina, há uma tendência de traços verticais grossos e traços horizontais mais finos que os verticais, gerando essa relação de contraste.

8. O trabalho de Noordzij toma como referência duas principais ferramentas de escrita: a pena chata (ou pena quadrada) e a pena flexível (bico de pena). Em tese, seu modelo de análise pode ser aplicado a qualquer alfabeto e sistema de escrita, e também a outras ferramentas (e.g. esferográfica, pincel, lápis). Entretanto, apesar de estabelecer princípios muito esclarecedores, Noordzij se concentra somente na escrita latina, e seus exemplos utilizam modelos clássicos da caligrafia ocidental.

catch their interaction. This is what it means to listen; this is what it means to read. It is incredibly complex, yet humans do it every day, and very often laugh and weep at the same time. Writing, by comparison, seems altogether simple, at least until you try.

Writing is the solid form of language, the precipitate. Speech comes out of our mouths, our hands, our eyes in something like a liquid form and then evaporates at once. It appears to me that this is part of a natural cycle: one of the ways the weather forms on the ocean of meaning. What else are the words we drop like pebbles in that ocean if not condensing droplets of evaporated speech, recycled bits of the ocean of meaning itself? Yet language can also solidify – into iridescent, sharp, symmetrical crystals, or into structures more like hailstones or shale beds or mud. In solid as in liquid form, the intersecting meanings may reinforce each other or rub each other out.

É importante destacar (na citação acima) uma conhecida frase deste autor, “a escrita é a forma sólida da linguagem”, um dos pontos de partida para sua argumentação. Entre as diversas questões tratadas pelo autor, com reflexões pertinentes sobre a linguagem e a escrita, considera-se significativas as ideias descritas a seguir.

Primeiramente, Bringham observa que *uma escrita não é uma língua, mas um sistema de representação*, capaz de captar grande parte (mas nunca a totalidade) de uma língua em sua rede. Explica que, durante 95% do tempo de seu tempo na terra, a espécie humana não sentiu necessidade do controle que sistemas de escrita permitem. No entanto, a linguagem pode se adaptar e se adapta à escrita, da mesma forma que animais e plantas se adaptam à agricultura e à pecuária. Existem variedades linguísticas que nunca existiriam ou sobreviveriam sem a proteção oferecida pela escrita (BRINGHURST, 2002, p. 4). Esta observação, de que *uma escrita não é uma língua, mas um sistema de representação*, pode ter outras interpretações, especialmente se examinada pelo viés da semiótica ou do pós-estruturalismo. Em primeiro lugar, é imprescindível deixar claro que, para efeito de processamento de dados, não pode haver subjetividade nessas descrições. Escrita e língua são conceitos objetivos, que definem coisas diferentes. No processamento de textos em computadores, há duas categorias a serem preenchidas, cada qual com seu valor, e.g. escrita = latina, língua = português.

Em segundo lugar, o autor informa que assim como as espécies na natureza, as línguas se dividem e se subdividem, formando famílias, troncos, agrupamentos. Essa

mesma divisão (em famílias e agrupamentos) ocorre com as escritas; mas enquanto *as escritas são essencialmente inventadas*, as línguas não são. E enquanto o mundo dos manuscritos e impressos precisa de artifícios para sua sustentação (escolas, treinamento), a oralidade floresce naturalmente. Essa é uma das razões pelas quais as árvores taxonômicas das línguas e escritas nem sempre coincidem (BRINGHURST, 2002, p. 4-5).

A respeito disso, é importante resgatar uma observação de Flusser (2007, p. 89-90):

A comunicação humana é um processo artificial. Baseia-se em artifícios, descobertas, ferramentas e instrumentos, a saber, em símbolos organizados em códigos. Os homens, comunicam-se uns com os outros de uma maneira não “natural”: na fala, não são produzidos sons naturais, como por exemplo, no canto dos pássaros, e a escrita não é um gesto natural como a dança das abelhas.

Nesse sentido, enquanto Bringhurst diferencia a natureza das línguas (naturais) e das escritas (artificiais), para Flusser, tanto a linguagem oral como a escrita são artificiais. Apesar do contraponto em relação à fala, ambos concordam sobre a natureza artificial da escrita. Não é o objetivo deste trabalho discutir essas questões, mas simplesmente reforçar o caráter artificial da escrita.

Bringhurst complementa ainda que as línguas e escritas evoluem e estão sujeitas a mudanças, assim como as espécies na natureza. Seus territórios se expandem e se contraem, se fundem e se dividem. Nenhum deles é imortal, e todos estão sujeitos à extinção. Destaca também que a transição da linguagem falada para a escrita e sua cristalização num sistema que capture e preserve graficamente as sutilezas da fala dependem de uma cultura preparada para sustentar tais artifícios. Partindo do zero, sem referências de modelos importados, há registros de transição de culturas orais para escritas, como no caso da Mesopotâmia, há cerca de 5000 anos; do norte da China, há cerca de 4500 anos; e da Guatemala e sul do México, há cerca de 2000 anos. É importante observar que, nesses contextos, a escrita se inicia a partir de desenhos e, conforme esses desenhos evoluem, passam a representar palavras, e depois sílabas, e tornam-se cada vez mais abstratos (BRINGHURST, 2002, p. 5), conforme demonstram as figuras 14 a 16.

Nesse sentido, o autor destaca (BRINGHURST, 2002, p. 6) uma observação de Leonard Bloomfield, de que a linguagem escrita veio do desenho, é uma ramificação do desenho, mas que, nascendo do desenho, se torna outra coisa. Mesmo sabendo que há estágios intermediários entre as duas coisas, quando a escrita se distingue totalmente do desenho, ela assume algumas características, expostas a seguir (em quatro pontos) de forma resumida (BRINGHURST, 2002, p. 6).

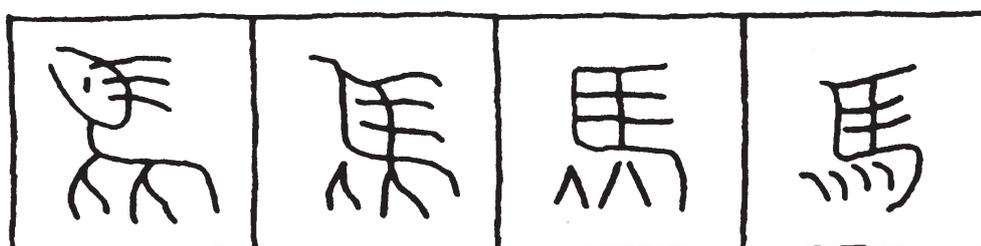


Figura 14: Desenvolvimento da escrita chinesa: da escrita arcaica à escrita moderna.

Fonte: FRUTIGER, 1999, p. 88.

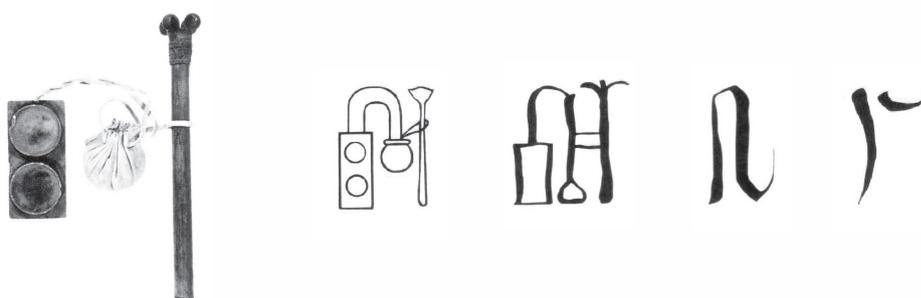


Figura 15: Desenvolvimento da escrita egípcia. O hieróglifo para escriba representava a paleta do Antigo Império, a bolsa com cordão para as barras de tinta seca e um porta pincéis de junco. As mudanças neste glifo demonstram o processo evolutivo (da esquerda para a direita): hieróglifo (2700 a.C.); escrita hieroglífica manuscrita (c.1500 a.C.); escrita hierática (c.1300 a.C.); escrita demótica (c.400 a.C.). Fonte: MEGGS; PURVIS, 2009, p. 30.



Figura 16: Desenvolvimento da escrita chinesa. Li (vasilha cerâmica de três pernas), período Baixo Neolítico. A evolução do caractere Li (da esq. para a dir.) originou-se desse recipiente: da pictografia de osso de oráculo, grafia de bronze (1000a.C.) e estilo regular (200a.C.). Fonte: MEGGS; PURVIS, 2009, p. 53.

Em primeiro lugar, *a escrita é abstrata*. Isso quer dizer que é possível desenhar utilizando letras, mas na escrita propriamente dita, o conteúdo pictórico desaparece. Na famosa frase de Eric Gill, “letras são coisas, e não imagens de coisas”. A leitura consiste, principalmente, num processo de abstração e automatismo⁹, e os leitores simplesmente não associam imagens ou figuras aos sinais da escrita no momento da leitura. Flusser (2007, p. 89-90) também reforça essa ideia:

O caráter artificial da comunicação humana (o fato de que o homem se comunica com outros homens por meio de artifícios) nem sempre é totalmente consciente. Após aprendermos um código, tendemos a esquecer a sua artificialidade [...]

Em segundo lugar, *os sistemas de escrita são codificados* e, geralmente, consistem em um conjunto de símbolos suficientes para reproduzir a língua a qual servem. Enquanto doze letras do alfabeto latino são suficientes para se escrever no idioma havaiano, milhares de caracteres são necessários para a grafia de um texto em chinês. Independentemente do sistema empregado, é possível escrever algo que nunca foi escrito antes sem a necessidade de invenção de novos símbolos (BRINGHURST, 2002, p. 6).

9. Gerard Unger tem textos relevantes sobre o tema da leitura (e.g. automatismos). Ver *Enquanto você lê* (2016) e *Theory of type design* (2018).

Esses símbolos são definidos em termos de outra coisa. A outra coisa é usualmente a fala, mas não necessariamente a fala. O que essa outra coisa precisa ser é a linguagem¹⁰ (BRINGHURST, 2002, p. 6).

O sistema é autocontido tanto estilisticamente quanto simbolicamente. Isso quer dizer que os símbolos da escrita, assim como elementos de outras formas de linguagem (música, dança), conforme evoluem, assumem uma lógica própria. Por exemplo, a escrita latina não nasceu tal qual a conhecemos e utilizamos hoje. Escritas se desenvolvem e adquirem uma gramática interna¹¹ na medida em que amadurecem, sendo escritas¹² (BRINGHURST, 2002, p. 6).

8.1.6.2 Escrita: conjunto de sinais de um sistema de escrita

Esta seção tem como objetivo descrever a segunda noção do termo escrita (*script*). Conforme anunciado no início das definições sobre escrita, o termo é utilizado para descrever conjunto de sinais gráficos próprios de um sistema de escrita, por exemplo, escrita latina (*latin script*) ou escrita chinesa (*chinese script*) (ESCRITA, 2015).

A esse respeito, o glossário¹³ do Unicode define escrita (*script*) como “uma coleção de letras e outros sinais escritos usados para representar a informação textual em um ou mais sistemas de escrita. Por exemplo, o russo é escrito com um subconjunto da escrita cirílica; o ucraniano é escrito com um subconjunto diferente da escrita cirílica. O sistema de escrita japonês usa várias escritas” (SCRIPT, 2020, tradução nossa).

10. Na frase original, *What it has to be is language*, não fica claro se o autor está se referindo à língua ou linguagem. O contexto sugere que ele está tratando de linguagem num sentido mais amplo, de um sistema que carrega, transmite conteúdo.

11. O termo é inerente à escrita e às funções para seu uso, i.e. sua sintaxe gráfica.

12. Sobre a gramática interna das escritas, ver *grammar of legibility*, no capítulo *Type and Language*, de *Theory of type design* (UNGER, 2018, p. 29–34).

13. Referência ao *Glossary of Unicode Terms*, documento em formato de página web. A publicação *The Unicode Standard: version 13.0 – Core Specification* (UNICODE CONSORTIUM, 2020a) não traz esse glossário incorporado, mas indica um link de uma página web com os termos do glossário: <https://www.unicode.org/glossary/>. Os termos do glossário do Unicode serão tratados como uma entrada/verbete de dicionário.

8.1.6.3 *Escritas vs. alfabetos*

Florian Coulmas (1999, p. 9–11, tradução nossa) explica que alfabeto é um:

“[...] sistema de escrita caracterizado pela relação de mapeamento sistemático entre seus signos (grafemas) e as unidades mínimas [distintivas] da fala (fonemas). [...] O termo deriva do latim, do qual se estendeu para o uso geral, para qualquer conjunto de signos elementares de uma escrita fonética. A escrita alfabética pode ser definida como um sistema de registro da linguagem que faz uso desse inventário de letras ou de qualquer variante historicamente relacionada dele”.

Esse mesmo autor observa que, muitas vezes, a definição de alfabeto é confundida com a definição de escrita (e vice-versa): “em contextos não técnicos, o termo [escrita] é freqüentemente usado de forma intercambiável com ‘sistema de escrita’ ou ‘alfabeto’, do qual deve ser diferenciado, entretanto [...]” (COULMAS, 1999, p. 454, tradução nossa).

Assim, em termos de conjuntos e agrupamentos, pode-se afirmar que o termo escrita é mais abrangente do que o termo alfabeto, i.e. pode-se dizer que um alfabeto está contido em uma escrita, mas o contrário não é verdadeiro. O alfabeto havaiano é um exemplo disso: emprega 17 letras (12 letras + 5 vogais acentuadas + apóstrofo), constituindo um subconjunto da escrita latina: A a E e H h I i K k L l M m N n O o P p U u W w Ā ā Ē ē Ī ī Ō ō Ū ū ‘

A partir dessas definições, pode-se ilustrar essa questão com alguns casos nos quais o termo alfabeto é empregado como sinônimo para escrita. O primeiro é no uso do termo “alfabeto cirílico” em contextos nos quais o correto seria “escrita cirílica”. Nesse caso, da escrita cirílica, com a qual se escreve em russo, búlgaro, servo-croata e outros idiomas eslavos, os termos alfabeto e escrita têm um significado muito próximo, já que a escrita cirílica é alfabética. No entanto, nos casos de digrafia (o uso de duas ou mais escritas diferentes para a mesma língua), como é o caso do servo-croata, o mesmo idioma pode ser escrito tanto com a escrita latina quanto com a escrita cirílica, conforme exemplo da Fig.17. Isso quer dizer que alfabeto sérvio pode significar tanto um conjunto de letras da escrita cirílica, como da escrita latina, tornando-se necessária a especificação, ou do alfabeto cirílico sérvio, ou do alfabeto latino.



Figura 17: Exemplo de digrafia em de placa de sinalização de trânsito. A placa contém o nome de uma mesma cidade em duas escritas: Жељуша, no alfabeto cirílico sérvio, e Željuša, no alfabeto latino. Fonte: ORTH, 2016.

Outro exemplo comum é o uso dos termos alfabeto chinês e alfabeto coreano. A escrita chinesa não é alfabética, e sim logossilábica. E a escrita coreana é baseada num silabário, denominado Hangul. Seguem exemplos, encontrados nos jornais *The New York Times* e *The Guardian*:

[...] The South Korean president, Kim Dae-jung, decided this week that Chinese would be used alongside the Korean **alphabet**, Hangul, on street signs, at tourist sites and in official documents [...] (GITTINGS, 1999).

[...] When he lands in Seoul, Crapser, 41, will be able to write his name in Korean and recite the Korean **alphabet** [...] (PERRY, 2016).

[...] Now, one South Korean woman, Lee Ki-nam, is determined to wring more recognition from the world with an unusual export: the Korean **alphabet**. Ms. Lee is using a fortune she made in real estate to try to take the **alphabet** to places where native peoples lack indigenous written systems to record their languages. [...] Her project had its first success and generated headlines in July, when children from an Indonesian tribe began learning the Korean **alphabet**, called Hangul [...] (SANG-HUN, 2009).

A exemplo do uso genérico do termo alfabeto, o termo letra¹⁴ é utilizado até mesmo em sistemas de escrita nos quais as menores unidades de valor semântico (naquela determinada escrita) são sílabas ou ideogramas, conforme sugerido no trecho a seguir:

O conceito de letra é usado em muitos contextos. Os padrões de linguagem de computador [...] não especificam exatamente o que é uma “letra”, “sílabas”, “ideógrafo” ou “dígito”, deixando as definições implicitamente para um padrão de codificação de caracteres [...] Em particular, enquanto o Unicode Standard inclui vários “alfabetos” e “silabários”, também inclui sistemas de escrita que se situam em algum lugar no meio. Como resultado, não é feita nenhuma tentativa de estabelecer uma distinção acentuada de propriedades entre letras e sílabas (UNICODE CONSORTIUM, 2020a, p. 187, tradução nossa).

Por último, é importante mencionar o IPA, *International Phonetic Alphabet*, ou alfabeto fonético internacional, um sistema alfabético de notação fonética, desenvolvido em 1888 pela Association Phonétique Internationale, que tinha como objetivo criar um sistema padronizado de representação dos sons da fala, independente do idioma, aplicável a todos os idiomas. O IPA foi pensado para representar transcrição da linguagem oral: fonemas, entonação e separação de palavras e sílabas (COULMAS, 1999, p. 235; International Phonetic Association, 1999).

Apesar da descrição linguística de caráter científico de línguas indígenas brasileiras estar fora do escopo da análise deste projeto, considera-se importante mencionar o alfabeto IPA e reforçar que esta tese se concentra na análise do registro escrito dessas línguas e, por isso, se concentra no alfabeto latino.

14. Florian Coulmas define letra como “uma de uma classe de formas que são reconhecidas como instâncias de conceitos gráficos abstratos que representam as unidades básicas de um sistema de escrita alfabético. Cada uma dessas unidades tem um nome, por exemplo *ei*, *si*, *kei*. Assim, a letra *eme* é a classe de todos os *Ms*, não importa a sua forma [cursiva, itálica, gótica, geométrica, pixelada etc]; mais precisamente, é a classe de todas as formas de letras concebíveis às quais o nome *eme* se aplica [...]” (Coulmas 1999, p. 291–292, tradução nossa).

De acordo com a forma de representação dos sons ou palavras da linguagem humana, pode-se classificar os sistemas de escrita em algumas categorias¹⁵, conforme proposta do Unicode, listada abaixo (UNICODE CONSORTIUM, 2020a, p. 254–258). Para esta tese, sistemas de escrita alfabéticos têm maior relevância, pois todas as línguas indígenas brasileiras analisadas nesta tese empregam a escrita latina, de caráter alfabético. Por esse motivo, as descrições de cada categoria se resumirão à listagem abaixo, cada uma seguida de um exemplo, a título de ilustração.

1. Alfabetos: latino, grego, cirílico, armênio.
2. Abjads (consonantários): árabe, hebraico.
3. Abugidas (alfassilabários): devanagari, gujarati, thai, tâmil.
4. Logossilabários: escrita ideográfica chinesa (Han).
5. Silabários simples: silabário indígena canadense unificado (UCAS), Cherokee, ge'ez (etíope), hiragana, katakana.
6. Silabários (*Featural Syllabaries*): hangul.

15. Com base em propostas taxonômicas de Peter T. Daniels e de Ignace J. Gelb, Robert Bringhurst (2002, p. 16–23) propõe um sistema de classificação dos sistemas de escrita com as seguintes categorias: semográfico, silábico, alfabético, prosódico. Apesar de constituir um sistema válido (e provavelmente há outros), optou-se pela adesão aos termos do Unicode. É interessante notar que, assim como Bringhurst, o Unicode também cita como referência a obra de Peter T. Daniels (UNICODE CONSORTIUM, 2020a, p. 255).

No que concerne aos sistemas alfabéticos, destaquemos alguns pontos importantes. Coulmas (1999, p. 9–11, tradução nossa), define alfabeto como um “sistema de escrita caracterizado pela relação de mapeamento sistemático entre seus signos (grafemas) e as unidades mínimas da fala (fonemas) [...]”. Em sistemas de escrita alfabéticos, vogais e consoantes têm o mesmo status enquanto letras dentro do sistema. A correspondência entre letras e sons pode ser mais ou menos exata, apesar de muitos alfabetos não exibirem uma correspondência homóloga (idêntica, pareada) entre letras ou grupos de letras usadas para a representação de distintos sons de uma língua. É muito comum o uso de mais de uma letra para representar um único som, e.g. o dígrafo <ch>, em português, ou as vogais <o> ou <e> duplas, em inglês. Ainda sobre correspondência entre letras e sons, há exemplos de línguas transparentes, nas quais o que se escreve é exatamente (ou muito próximo) da pronúncia, como o alemão e o italiano, e línguas opacas, como o inglês, cuja pronúncia é complexa e arbitrária (UNICODE CONSORTIUM, 2020a, p. 254; COULMAS, 1999, p. 11).

Reforçando a ideia de que “escritas se desenvolvem e adquirem uma gramática interna na medida em que amadurecem, sendo escritas” (BRINGHURST, 2002, p. 6, tradução nossa), pode-se observar também que, quando exportado de um idioma para outro, um sistema de escrita quase sempre sofre algumas alterações. É característica de línguas modernas e cosmopolitas criar algum espaço para palavras estrangeiras, assim como é marcante em escritas modernas e cosmopolitas criar algum espaço para letras estrangeiras. O alfabeto latino era originalmente composto de 20 letras: **A B C D E F G H I L M N O P Q R S T V X**, conjunto ao qual foram incorporados o **K** etrusco e o **Y** e **Z** gregos, e posteriormente as letras **J U W**. Com o desenvolvimento da escrita latina, surgiram as minúsculas, e, ao longo desse processo evolutivo, também foram adicionados ao repertório os sinais de pontuação e numerais indo-arábicos. Posteriormente, a escrita latina continuou a se adaptar às diferentes línguas. No caso do francês, por exemplo, se somaram ao conjunto vários caracteres acentuados: **à â ç é è ê ë î ï ô ù û ü**, além da ligatura **œ**. Ao se adaptar ao norueguês, letras como **c q x** caíram em desuso, enquanto caracteres como **å ø æ** passaram a fazer parte do repertório. Em línguas como o taitiano, não são utilizadas as letras **b c d g l q s x** (BRINGHURST, 2002, p. 11-12).

Nessa lógica, pode-se dizer que os sistemas de escritas do inglês e francês, por exemplo, consistem essencialmente em “dialetos gráficos” da escrita latina. E isso vale para o restante dos sistemas de escrita, que, em sua grande maioria, são como o inglês e o francês – variantes da escrita. Sistemas de escrita raramente são criados do zero se há algo existente que possa ser copiado ou adaptado. De maneira geral, **não há uma razão linguística que justifique a criação de uma escrita que seja exclusiva para uma determinada língua** (BRINGHURST, 2002, p. 12).

Da mesma forma que o inglês e o francês, as línguas indígenas brasileiras também têm seus “dialetos gráficos”. Da listagem de 274 línguas do censo de 2010 (IBGE, 2010), encontrou-se evidências de documentação escrita (i.e. publicações) de menos da metade dessas línguas. As características gráficas dessas línguas são assunto de grande interesse nesta tese (ver detalhes no capítulo *Características tipográficas das línguas indígenas brasileiras*).

Por último, a título de referência, será listado a seguir o alfabeto completo utilizado para a escrita da língua portuguesa no Brasil (ACADEMIA BRASILEIRA DE LETRAS, 2010):

A B C D E F G H I J K L M N O P Q R S T U V W X Y Z

a b c d e f g h i j k l m n o p q r s t u v w x y z

À Á Â Ã Ç È É Ê Í Ò Ó Ô Õ Ú

à á â ã ç é ê í ò ó ô õ ú

8.1.7 Unicode

O Unicode fornece um número único para cada caractere, não importa a plataforma, não importa o programa, não importa a língua (UNICODE CONSORTIUM, 2008).

Neste trabalho, utiliza-se o *The Unicode Standard* como principal referência sobre codificação de texto em meio digital. A publicação organizada pelo Unicode Consortium define o padrão universal de codificação de caracteres para processamento de texto em computadores, que possibilita o intercâmbio de informação textual internacionalmente. Com o surgimento do Unicode, a indústria de tecnologia da informação se beneficiou enormemente com maiores possibilidades de interoperabilidade global, estabilidade

e intercâmbio de dados, simplificação dos softwares e custos de desenvolvimento reduzidos. Hoje em dia, a esmagadora maioria dos sistemas operacionais, navegadores de internet, dispositivos móveis, protocolos de internet e linguagens de programação (como HTML, Java e XML) adotam o Unicode como padrão de codificação de caracteres (UNICODE CONSORTIUM, 2020a, p. 1).

Cunhado por Joe Becker, da Xerox, o termo Unicode representa o conceito de *codificação de caracteres única, universal e uniforme*. A primeira versão do *Unicode Standard* foi publicada em 1990¹⁶, e está em atualização constante desde então (HUDSON, 2002, p. 29).

Haralambous (2007, p. 54) observa que, na documentação do Unicode, é onde pela primeira vez se define de fato o que é um caractere, conceito-chave para o entendimento do Unicode e que será mais bem descrito na seção a seguir. Maiores informações sobre os princípios de funcionamento do Unicode encontram-se na seção *Codificação de caracteres* deste capítulo.

8.1.8 Caracteres e glifos

Esta seção tem como objetivo definir glifo (*glyph*) e caractere (*character*) e, especialmente, ressaltar as diferenças entre esses termos. Ambos estão associados à composição de texto digital. Na seção *Codificação de caracteres* serão discutidas questões específicas da codificação Unicode acerca desses dois termos.

Em relação ao termo caractere, partiremos da seguinte definição: “[...] segundo o Unicode Consortium, um caractere é a menor componente da linguagem escrita com valor semântico; refere-se ao significado abstrato e/ou forma, e não à somente uma forma específica. Em linhas gerais, um caractere é uma unidade de informação que corresponde a uma letra ou grafema” (HENESTROSA; MESEGUER; SCAGLIONE, 2019, p. 147).

Diferentemente do termo letra, que define, de forma genérica, um elemento de um alfabeto, caractere é um termo mais abrangente, pois representa principalmente, mas não exclusivamente, as letras (ou elementos de outros sistemas de escrita, como

16. Documentação sobre a história do Unicode encontra-se disponível nos apêndices C e D, do *The Unicode Standard* (UNICODE CONSORTIUM, 2020a, p. 949–966).

ideogramas). O termo também se aplica a algarismos, símbolos e pontuação, entre outros sinais¹⁷ que constituem texto em linguagem natural e notação técnica (CARACTERE, 2015; UNICODE CONSORTIUM, 2020a, p. 15).

Grosso modo, para o Unicode, um caractere é um código, um número, que descreve uma lista de propriedades. Nesse sentido, é importante entender um atributo determinante dos caracteres: a representação de uma unidade de informação para um computador. Para o Unicode, cada caractere é representado por um ponto de código (*code point*); ou seja, uma sequência de quatro¹⁸ números e letras, utilizando a notação hexadecimal¹⁹ (e.g. 004D), que define “o menor componente da linguagem escrita que tem valor semântico; refere-se ao significado abstrato e/ou forma, ao invés de uma forma específica” (CHARACTER, 2020).

A cada ponto de código são associadas uma série de propriedades de cada caractere, como nome, categoria (letra, numeral, pontuação), escrita (latina, árabe), direcionalidade (da esquerda para a direita ou vice-versa, de cima para baixo) etc. (UNICODE CONSORTIUM, 2020a, p. 157–192). O nome do caractere (e.g. *Latin Capital Letter M*, que corresponde ao código U+004D) talvez seja das propriedades mais importantes, pois é também uma descrição de algo que não existe concretamente, e sim como uma noção, uma ideia.

Uma dessas propriedades é a noção, ideia ou conceito chamado de caractere abstrato (*abstract character*), termo sobre o qual é importante observar alguns pontos.

17. Além de marcas e sinais visíveis (e.g. letras, pontuação), há caracteres que definem espaços em branco, i.e. não são visíveis, não definem nenhuma marca ou sinal gráfico. Os códigos de controle (*control codes*) são caracteres comumente usados na formatação de texto. Desempenham, por exemplo, funções na formatação de parágrafos, como quebra de linha, retorno, tabulação etc. (UNICODE CONSORTIUM, 2020a, p. 882–884). Outro exemplo de espaço em branco é o espaço utilizado para separar palavras, normalmente acessível aos usuários por meio da barra de espaço de qualquer teclado.

18. O Unicode também possui esquemas de codificação de 32bits, cujos pontos de código são sequências de oito dígitos, como 0000004D. Esse esquema de codificação é usado, por exemplo, para extensões do sistema de escrita ideográfico chinês e, portanto, de menor relevância para a discussão deste trabalho (UNICODE CONSORTIUM, 2020a, p. 33–41).

19. Sistema de numeração posicional que representa os números em base 16, portanto empregando 16 símbolos: dez numerais, de 0 a 9, e seis letras, de A a F.

Primeiro, caractere abstrato não é uma coisa que existe diretamente em um computador, e também não constitui um objeto ou artefato tangível, mas nocional, especulativo; um caractere abstrato não tem forma concreta, e não deve ser confundido com glifo (CONSTABLE, 2001; UNICODE CONSORTIUM, 2020a, p. 88). Nessa perspectiva, pode-se dizer que o nome do caractere (e.g. *Latin Capital Letter M*, que corresponde ao código U+004D) é uma descrição desse caractere abstrato.



Figura 20: Variações da letra M. A propriedade de caractere abstrato do M engloba todas as formas de letra M concebíveis: cursivas, itálicas, góticas, geométricas, pixeladas etc.

Fonte: COULMAS, 1999, p. 291–292.

Para o Unicode, um caractere codificado (*encoded character, coded character*) é constituído pela associação de um caractere abstrato (*abstract character*) e um ponto de código (*code point*) e é essa unidade de informação à qual nos referimos como atributo determinante para o processamento de texto em computadores.

É importante, neste ponto, fazer um parêntese com a definição de grafema, segundo o Unicode:

(1) A menor unidade da escrita que se possa distinguir, no contexto de um determinado sistema de escrita. Por exemplo, e <d> são grafemas distintos em sistemas de escrita do inglês, porque existem palavras distintas como big e

dig. Contrariamente, uma letra a minúscula itálica <a> e uma letra a minúscula romana <a> não são grafemas distintos, porque não se pode distinguir nenhuma palavra com base nessas duas formas diferentes. (2) O que um usuário entende como um caractere (GRAPHEME, 2020, tradução nossa).

Como o objetivo deste trabalho é pensar a escrita do ponto de vista tipográfico, em que o termo grafema é pouco usado, propõe-se a aderência ao termo grafema sob a seguinte condição: como um conceito simplificado e análogo/paralelo a fonema (menor unidade sonora de uma língua, unidade mínima da fala). E do ponto de vista etimológico, no campo da linguística, a definição do parágrafo anterior pode ser considerada superficial; para definições mais fundamentadas, consultar *The Blackwell Encyclopedia of Writing Systems* (COULMAS, 1999, p. 174–175).

Uma vez compreendidos os conceitos de caractere e grafema, torna-se importante entender o conceito de glifo e a diferença entre glifo e caractere. De acordo com Haralambous (2007, p. 55, tradução nossa):

Um glifo é a imagem de um símbolo usado em um sistema de escrita (em um alfabeto, um silabário, um conjunto de ideogramas, etc.) ou em um sistema de notação (como música, matemática, cartografia, etc.).

Em linhas gerais, um glifo representa “a aparência gráfica de um caractere, enquanto este último é uma unidade conceitual. Por exemplo, o [mesmo] glifo que representa o caractere A maiúsculo <A> do alfabeto latino também pode representar o caractere Alfa maiúsculo <Α> do alfabeto grego” (UNICODE CONSORTIUM 2020a, p. 148, tradução nossa).

Assim, o mesmo glifo , pode ser representado por, pelo menos, três códigos diferentes, conforme ilustra a Fig.21. Pode-se observar que, enquanto a forma maiúscula é idêntica nos três sistemas de escrita (latino, grego e cirílico), o mesmo não ocorre com as minúsculas.

B	U+0042	LATIN CAPITAL LETTER B	b	U+0062	LATIN SMALL LETTER B
Β	U+0392	GREEK CAPITAL LETTER BETA	β	U+03B2	GREEK SMALL LETTER BETA
В	U+0412	CYRILLIC CAPITAL LETTER VE	В	U+0432	CYRILLIC SMALL LETTER LETTER VE

Figura 21: Glifos com formas idênticas na maiúscula latina, grega e cirílica. Observa-se que enquanto as maiúsculas são idênticas, as minúsculas variam. Composto em Brill Roman, de John Hudson. Fonte: do autor.

Ainda, destacando a diferença entre glifos e caracteres:

A diferença entre identificar um caractere e renderizá-lo na tela ou no papel é crucial para entender o papel do Unicode Standard no processamento de texto. **O caractere identificado por um ponto de código Unicode é uma entidade abstrata**, como “Latin Capital Letter A” ou “Bengali Digit Five”. **A marca feita na tela ou no papel, chamada glifo, é uma representação visual do caractere. O padrão Unicode não define imagens de glifo.** Ou seja, o padrão define como os caracteres são interpretados, não como os glifos são renderizados. Por fim, o mecanismo de renderização de software ou hardware de um computador é responsável pela aparência dos caracteres na tela. O Padrão Unicode não especifica a forma, tamanho ou orientação precisa dos caracteres na tela (UNICODE CONSORTIUM, 2020a, p. 6, tradução nossa, grifo nosso).

Em relação ao *uso* do termo caractere, é importante também ressaltar que há uma definição técnica do Unicode, e há um uso corrente do termo caractere por profissionais de comunicação visual (e também pela maioria das pessoas).

Observemos, por exemplo, esses conceitos do ponto de vista de um designer de tipos:

Os designers de tipos criam novos glifos, ou seja, imagens que representam caracteres. Os designers de tipos geralmente não criam novos caracteres, ou seja, o significado dessas imagens, ou, mais tecnicamente, ‘o menor componente da linguagem escrita que tem valor semântico.’ Não inventamos alfabetos, apenas reinterpretemos os existentes. [...]

Em síntese: caracteres são o que você digita, glifos são o que você vê (SCHEICHELBAUER, 2013, tradução nossa, grifo nosso).

Nesse sentido, o trecho supracitado de Scheichelbauer alerta para como o uso do termo caractere por profissionais de comunicação visual tem um certo distanciamento da definição técnica do Unicode. A maioria dos designers de tipos, tipógrafos e designers gráficos entende por caractere um elemento de uma fonte tipográfica, e.g., o <W> da fonte Times New Roman Bold, ou o <Z> da fonte Georgia Italic. Para designers de tipos, um caractere é algo que tem forma, que tem uma representação visual; ou seja, uma associação entre caractere e glifo. Logo, nesse contexto, o termo *caractere* descreve uma unidade de informação dentro de uma fonte que, para efeitos de processamento de dados, é formado por três componentes: o contorno (*outline*), que é o desenho da letra em si; a métrica (*metrics*), que é a quantidade de espaço em branco ao redor de cada letra; e o ponto de código (*code point*), da codificação Unicode²⁰.

Essa questão é relevante, pois, afinal de contas, toda a lógica de funcionamento da tipografia em meio digital considera como caractere não só um código que descreve propriedades abstratas, mas algo que tem uma representação visual. Assim, pode-se dizer que, no uso corrente, um caractere descreve um código e uma aparência gráfica.

8.1.9 Diacríticos e marcas combinadas

Diacríticos (do grego, “que distingue”), também chamados de acentos, marcas diacríticas (*diacritical marks*), ou simplesmente marcas (*marks*), são sinais adicionados a letras-base que alteram sua pronúncia e/ou significado. Amplamente usadas para ampliar o alfabeto latino, essas marcas podem ser feitas acima, abaixo, sobrepostas ou em qualquer lugar ao redor da letra. Diacríticos posicionados no topo das letras <â ê í ã ü> são a forma mais comum, embora também existam formas de diacríticos sobrepostos <Ł ł>, conectados à base <ą ę> ou sob a base <ș ș> (COULMAS, 1999, p. 26; GAULTNEY, 2002, p. 2).

O Unicode documenta duas principais formas de associação de diacríticos a letras da escrita latina: com caracteres compostos ou pré-compostos (*composites, precomposed*

20. David Březina dedica, em sua tese de doutorado, *Coherence in typeface design: visual similarity of characters in Cyrillic, Devanagari, and Latin* (2018), um capítulo inteiro para uma conceituação mais completa em relação ao termo caractere sob a ótica do design de tipos.

characters) e com marcas combinadas (*combining marks, combining diacritical marks*). A primeira delas é considerada uma forma de composição estática, e a segunda, dinâmica.

O caractere pré-composto é a forma mais comum, e consiste em único caractere, construído a partir da associação de um caractere base e um ou mais diacríticos. Por se tratar de um único caractere, formado pela união de dois glifos (letra base + diacrítico), tem somente um ponto de código. Por exemplo: quando o usuário digita um acento agudo, seguido pela letra A, o sistema operacional substitui essa combinação por um caractere pré-composto: <Á> (U+00C1), conforme a Fig.22.

$$\begin{array}{ccccccc} \acute{\prime} & + & A & = & \acute{A} & & \\ U+00B4 & & U+0041 & & U+00C1 & & \end{array}$$

Figura 22: Sequência de composição de caractere pré-composto, com seus respectivos códigos Unicode. Composto em Brill Roman, de John Hudson. Fonte: do autor.

Marcas combinadas levam em conta o advento da composição dinâmica (*dynamic composition*), um dos dez princípios básicos de design do Unicode (UNICODE CONSORTIUM, 2020a, p. 14), que diz que letras acentuadas podem ser dinamicamente compostas. Por exemplo: em condições ideais, quando o usuário digita uma letra, seguida por uma marca combinada (e.g. acento agudo combinado), o diacrítico automaticamente cai em cima da letra. Diferentemente dos caracteres pré-compostos, essa combinação de letra base e marca combinada tem dois pontos de código (Fig.23). Uma característica das marcas combinadas é que elas podem se combinar entre si e formar glifos com diacríticos múltiplos, sem limite de número ou posição dos sinais diacríticos (UNICODE CONSORTIUM, 2020a, p. 54; HARALAMBOUS, 2007, p. 60–61), conforme exemplificado por John Hudson²¹ na Fig.24 (dynamic diacritics).

21. Em apresentação na conferência Typo Berlin (HUDSON, 2014). PDF da apresentação disponibilizado pelo autor.

$$\begin{array}{ccc}
 A & + & \text{◌} & = & \acute{A} \\
 \text{U+0041} & & \text{U+0301} & & \text{U+00410301}
 \end{array}$$

Figura 23: Sequência de composição de caractere acentuado, formado por letra base e marca combinada, e seus respectivos códigos Unicode. O círculo pontilhado (DottedCircle, U+25CC) é uma convenção utilizada para uma marca solta, que não está posicionada sobre uma letra base. Composto em Brill Roman. Fonte: do autor.

dynamic diacritics

ḋŷṅḁṁīç ḍiãçrītīçş

ḋŷṅḁṁīč ḍiãçrītīčş

ḋŷṅḁṁīç ḍiãçrītīçş

Figura 24: Exemplo de composição demonstrando as potencialidades das marcas combinadas da fonte Brill Roman. Fonte: HUDSON, 2014, p. 19–22.

8.1.10 Tipografia

Esta seção traz algumas definições do campo tipográfico, seguindo a linha de pensamento deste capítulo sobre terminologia. Para que essas definições não fiquem muito extensas,

ao final desta tese foi incluído um glossário de termos tipográficos. Além disso, incluiu-se também uma seleção de exemplos ilustrados de anatomia e terminologia tipográfica.

8.1.10.1 *Tipografia: definições gerais*

Nesta tese, o termo tipografia pode ser lido de três maneiras: como campo projetual, como resultado dessa atividade projetual e como ferramenta e tecnologia de escrita.

Primeiramente, é importante definir o campo projetual. Paul Luna, autor de *Typography: A Very Short Introduction*, abre essa obra com uma definição que é, ao mesmo tempo, ampla e precisa: “Tipografia é design para leitura. É um conjunto de escolhas visuais projetadas para tornar a mensagem mais acessível, mais facilmente transmissível, mais significativa, ou mais atraente” (LUNA, 2018, p. 1, tradução nossa). Essa declaração reforça a ideia de que especificações tipográficas são decisões de design fundamentais, que influem na legibilidade, nos fatores tecnológicos e nas associações emocionais. A ideia de tipografia como projeto, planejamento, especificação é algo que permeia esta tese.

Nesse sentido, em *Modern Typography*, Robin Kinross (2010, p. 19) observa que esse posicionamento faz parte de uma cultura, um pensamento moderno, que move o foco de atenção dos produtos para as ideias que informam a produção. Essa ênfase em ideias é algo que se relaciona com sua tese sobre o que é que constitui a modernidade: a discussão, descrição e ordenamento da prática, e não a mera prática e meros produtos.

Em *Notas para uma normatização da nomenclatura tipográfica*, Priscila Farias (2004) também corrobora a ideia de tipografia como uma atividade projetual:

Começaremos por definir tipografia como o conjunto de práticas e processos envolvidos na criação e utilização de símbolos visíveis relacionados aos caracteres ortográficos (letras) e para ortográficos (números, sinais de pontuação, etc.) para fins de reprodução. Isso inclui tanto o design de tipos quanto o design com tipos.

A partir dessas premissas, configuram-se duas principais áreas de projeto: a do design de tipos (*type design, typeface design*) e a do design com tipos, design tipográfico (*typographic design*). A primeira, como o próprio nome diz, consiste no design de formas de letras (*letterforms*) e outros sinais. A segunda, o design tipográfico (*typography*,

typographic design), refere-se ao arranjo e detalhamento do texto (combinado a imagens e espaços) de acordo com um planejamento, projeto (UNGER, 2018, p. 229). As expressões “design tipográfico” e “design com tipos” são pouco usuais e, por isso, é mais comum associar o trabalho do design com tipos com atividades como design de livros, design editorial e até áreas mais abrangentes, como design gráfico, direção de arte, webdesign etc.

O design de tipos é o campo de trabalho do designer de tipos (*type designer, typeface designer*). Já o design com tipos, com uma aplicação mais ampla, denomina não somente o domínio do tipógrafo (*typographer*), profissional que, antes do advento do computador, era responsável pela composição²² com tipos de chumbo manualmente na impressão tipográfica tradicional, i.e. compositor. Por extensão, pode-se denominar tipógrafo aquele que utiliza tipos para compor um texto, num contexto projetual. Além disso, utiliza-se, como sinônimo para tipógrafo, o termo designer tipográfico (*typographic designer*): ao profissional que cria e executa especificações de projetos gráficos, e.g. jornais, livros etc. (UNGER, 2018, p. 229; HENESTROSA; MESEGUER; SCAGLIONE, 2019, p. 149).

Em segundo lugar, complementado as definições acima, é importante definir o resultado dessas atividades projetuais, ou seja, das atividades do design de tipos: fontes digitais e processos técnicos relacionados, i.e. engenharia de fontes; e do trabalho do design com tipos: livros, jornais, revistas, identidades visuais, websites etc.

Por último, deve-se observar tipografia como uma tecnologia de escrita, sua forma mecanizada; segundo Gerrit Noordzij (2000, p. 3), “tipografia é escrita com letras pré-fabricadas” conforme explica a seção a seguir.

8.1.10.2 Fundamentos tipográficos: escrita com letras pré-fabricadas e o conceito de espaço na tipografia

Esta seção é importante para a definição, principalmente, dos conceitos de métrica (*metrics*) e espaçamento (*spacing, fitting*). Na tipografia, o espaço em branco ao redor das letras, ou espaço negativo, é tão importante quanto o próprio desenho da letra e, nesse

22. Composição; *typesetting*

ponto, pode-se considerar que o espaço ao redor das letras é um importante componente que viabiliza a tipografia como algo padronizado, estável, modular. Em linhas gerais, o espaço ao redor das letras é o que torna possível a tipografia como escrita com letras pré-fabricadas. Na obra *Entre parágrafos: fundamentos tipográficos* (2017, p. 13), Cyrus Highsmith ilustra essa ideia de forma muito simples e clara:

Gutenberg não inventou livros, e sua invenção não mudou imediatamente a aparência do livro aos olhos do leitor. Ele inventou uma nova abordagem para se criar um parágrafo de texto. Em vez de escrever cada linha à mão, desenhando cada palavra e cada letra, Gutenberg inventou uma forma de simular um parágrafo escrito à mão.

Uma peça de tipo de metal ilustra a essência da ideia de Gutenberg: colocar cada letra dentro de sua própria caixa. Ao se colocarem coisas com formas estranhas em caixas de tamanho padrão, elas ficam mais fáceis de manusear. Pense na comparação entre carregar quatro pares de sapato e carregar quatro pares de sapato dentro de caixas de sapato.

Foi basicamente isso que Gutenberg fez com o tipo móvel. Lidar com um parágrafo de texto é bem mais complicado do que carregar uma pilha de sapatos, mas o conceito é parecido. Gutenberg colocou as letras dentro de caixas para que elas ficassem mais fáceis de pegar e de arranjar.



Figura 25: Ilustração que explica o conceito de modularidade do tipo móvel.

Fonte: HIGHSMITH, 2017, p. 12–15

A partir dessa ideia, pode-se observar que, em manuscritos, o espaço em branco não é algo padronizado. E com o advento da tipografia, Gutenberg cria uma forma de padronizar os espaços. Ao colocar as letras dentro de “caixas”, cria o conceito-chave

dos tipos móveis, que é a padronização dos espaços em branco ao redor das letras. Sobre os espaços, Highsmith (2017, p. 15) observa:

Um parágrafo tem dois elementos visuais básicos: as partes pretas e as partes brancas. A parte preta são as letras, e a parte branca é o fundo. Mas em vez de pensar nelas como camadas separadas, como tinta e papel, imagine as partes pretas e brancas juntas, como duas partes de um quebra-cabeça. As letras pretas são arranjadas para formar palavras, que são arranjadas para formar linhas. Os elementos brancos – as outras partes do quebra-cabeça – são os espaços em branco dentro de uma letra (contraforma), entre as letras (entreletra), entre as palavras (entrepalavra) e entre as linhas (entrelinha).

Highsmith (2017, p. 23) introduz ainda um conceito relevante para o entendimento da padronização dos espaços em branco, que é a ideia de *espaço do glifo*, descrito por ele como algo equivalente à “caixa” onde as letras tipográficas se inserem. O autor explica que a altura do espaço do glifo define o corpo do tipo. Observa ainda:

Pensar em termos de espaço em branco (ou negativo) não é algo especificamente relacionado à tipografia ou mesmo à caligrafia. Trata-se de um elemento básico do design bidimensional. O espaço do glifo, no entanto, é um conceito exclusivamente tipográfico. É o mecanismo que torna o tipo móvel possível. O espaço do glifo forma o espaçamento padrão (spacing) de um tipo, e esse espaçamento é o ponto de partida para a estrutura do parágrafo.”

8.1.10.3 Tecnologia tipográfica

O objetivo desta seção é traçar um breve panorama tecnológico do campo. Ferramentas e técnicas sempre tiveram um papel importante para o desenvolvimento do design de tipos, métodos de composição tipográfica, técnicas e sistemas de impressão. Dentre os principais desenvolvimentos tecnológicos de uma história de quase seis séculos da letra impressa, observa-se quatro principais momentos, dando-se maior ênfase para o período da tecnologia digital.

Da década de 1450, do primeiro livro impresso com tipos móveis por Gutenberg, até o século XIX, utilizou-se o sistema denominado composição manual. Esse sistema se baseava na composição manual com tipos de chumbo que, por sua vez, também

eram produzidos de maneira artesanal²³, por meio de fundição com matrizes de cobre e moldes ajustáveis.

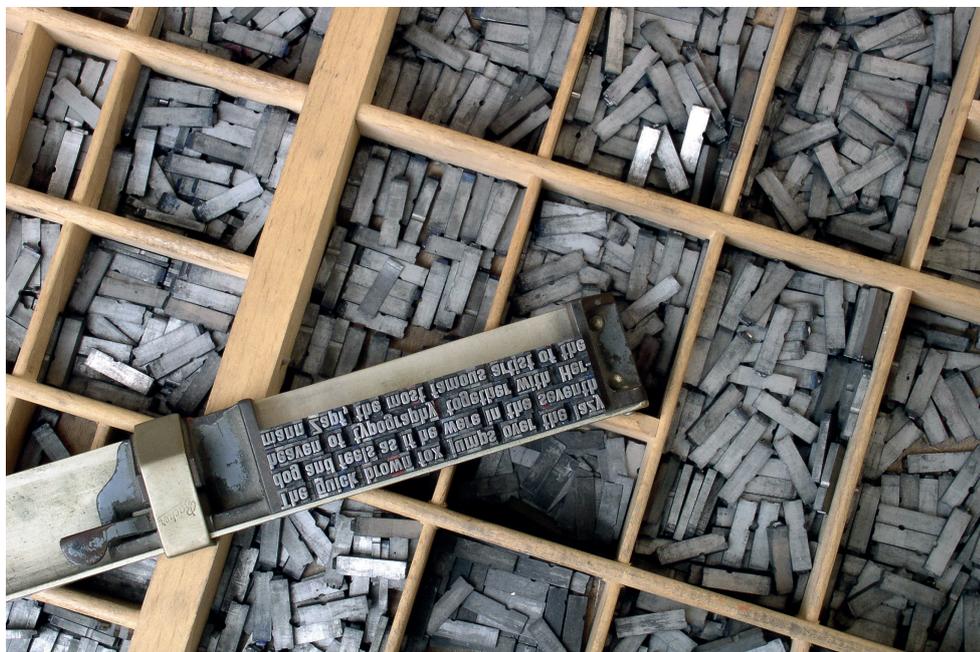


Figura 26: Gaveta de tipos móveis e componedor. Fonte: HEIDELBACH, 2004

Esses processos de fabricação e composição manuais se preservaram por mais de quatrocentos anos. No século XIX, no contexto da revolução industrial, “uma série de avanços tecnológicos mudaram radicalmente o modo de produzir e imprimir tipos. Os mais importantes são, provavelmente, a impressora a vapor, o pantógrafo tipográfico e as máquinas de linotipia e monotipia”²⁴ (SCAGLIONE, 2019, p. 113).

23. Para mais informações sobre o processo de fabricação dos tipos móveis, em português, ver *Contrapunção*, de Fred Smeijers (2015).

24. Duas invenções possibilitaram a composição mecânica (i.e. composição a quente, *hot metal typesetting*): o linotipo (Linotype), inventado por Ottmar Mergenthaler em 1885, e o monotipo (Monotype), desenvolvido por Tolbert Lanston e J.S. Bancroft a partir de 1890. Esses dois mecanismos propiciaram não somente a composição mecanizada, mas também a mecanização da fundição de tipos (SOUTHALL, 2005, p. 35–46).

PLATE LXV.

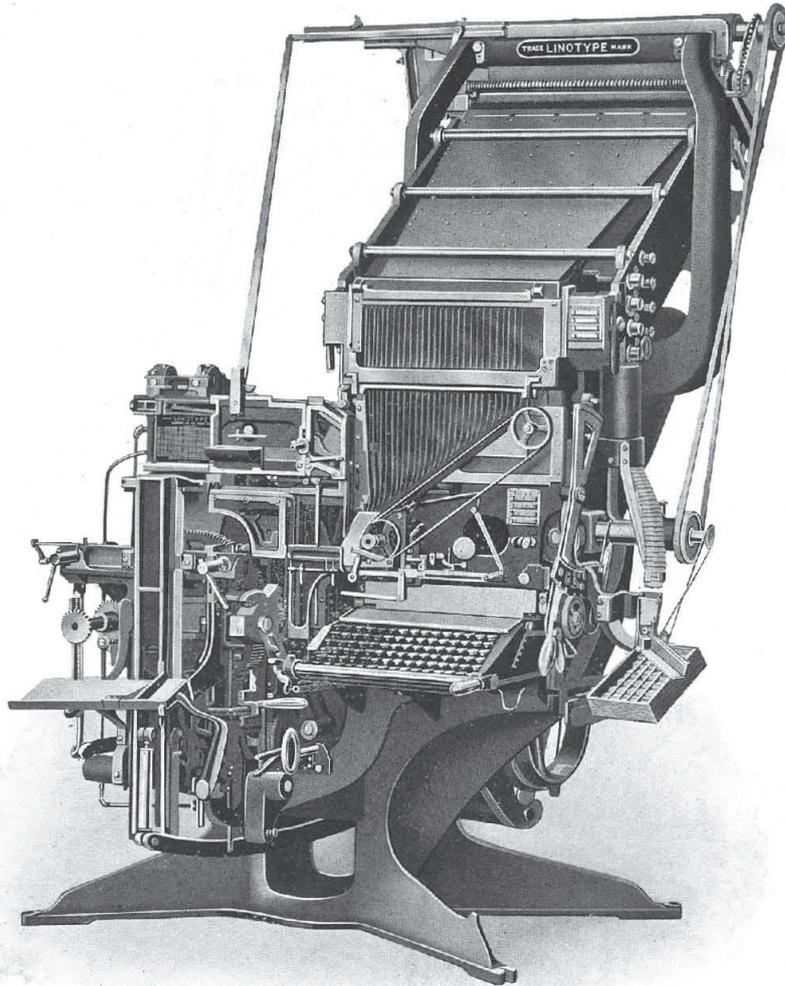


FIG. 410.—*Linotype; quadruple magazine machine; model 9; English and American.*
To face plate LXVI.]

Figura 27: Linotype Machine. Fonte: LEGROS; GRANT, 1916, Plate LXV, n.p.

Na década de 1950²⁵, surge a fotocomposição, tecnologia que utiliza matrizes óticas para a produção de linhas de texto em meio fotográfico, em substituição às matrizes de metal e aos tipos de chumbo. Pode-se dizer que a fotocomposição situa-se como elo de transição entre os sistemas mecânicos e os sistemas digitais, pois a partir da década de 1960, começam a surgir unidades de fotocomposição controladas por computador. É também nesse período que as matrizes óticas desses equipamentos começam a ser substituídas por fontes digitais; i.e. matrizes físicas passam a ser substituídas por informações binárias (SOUTHALL, 2005, p. 79–92; SAVOIE, 2014, p. 75–112).

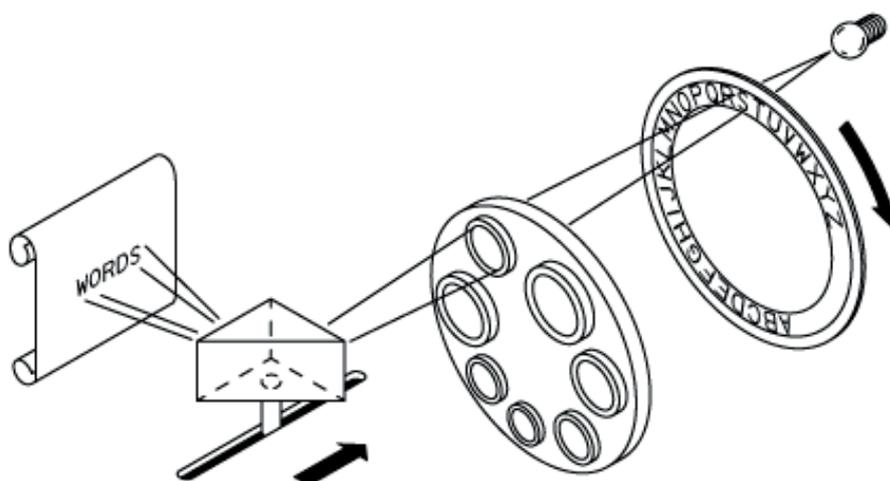


Figura 28: Representação esquemática do mecanismo de fotocomposição. Fonte: LEHNI, 2011

25. Em 1946 já existiam equipamentos (Intertype Fotosetter) em uso no US Government Printing Office; somente em 1950 se dá a primeira demonstração pública do Fotosetter. A primeira geração de equipamentos (e.g. Monotype Filmsetter, Intertype Fotosetter) era baseada em tecnologias anteriores; i.e. modificações em equipamentos de composição a quente, adaptando-as ao processo fotográfico, enquanto a segunda geração foi concebida especificamente para o processo fotográfico, não derivando de nenhum outro equipamento anterior. É o caso do Lumitype-Photon, criado por Rene Higonnet e Louis Moyroud, caracterizado pela tecnologia inovadora do disco de matrizes, capaz de acomodar até 1440 caracteres (o equivalente a 16 fontes de 90 caracteres). Com o dispositivo, era possível dar saída em 12 tamanhos diferentes (de 6pt a 72pt), e estes poderiam ser misturados numa mesma linha, sem interferência na justificação ou alinhamento. Além disso, o equipamento trabalhava com cálculos binários para determinar larguras de caracteres e justificação. A abordagem versátil e inovadora fez da Lumitype-Photon o primeiro equipamento de fotocomposição bem-sucedido comercialmente (SAVOIE, 2014, p. 93–97).

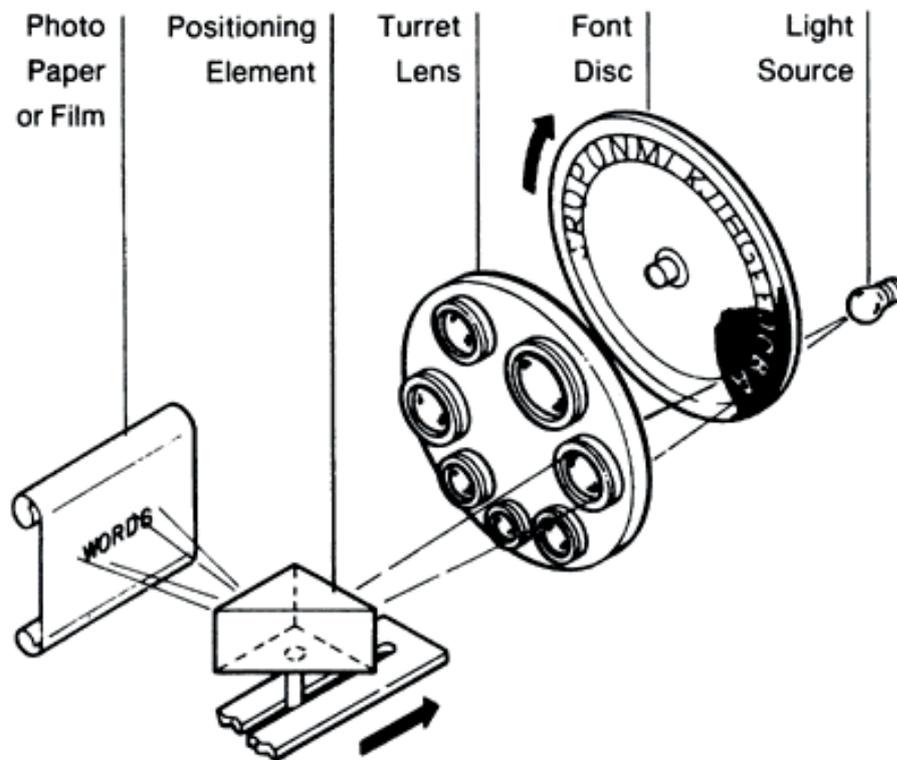


Figura 29: Representação esquemática do mecanismo de fotocomposição.

Fonte: INTERNATIONAL PAPER COMPANY, 1983

As primeiras fontes digitais²⁶ eram fontes *bitmap*. Gerard Unger (1979, p. 135) explica que nos primeiros equipamentos de composição digital, as imagens dos caracteres eram formadas a partir de informação binária: o equipamento projetava feixes de luz, emitidos por um tubo de raios catódicos (CRT), sobre filme ou papel fotossensível, gravando a imagem da letra na superfície do meio fotográfico.

26. Southall (2005, p. 143-145), Unger (1979, passim), Savoie (2014, p. 101) e Ulrich (2018, passim) apontam para o Digiset como o primeiro equipamento de composição digital. O equipamento só passa a ser comercializado mais amplamente a partir da década de 1970: Unger (1979, p. 137) e Ulrich (2018) indicam que a fonte Marconi, de Hermann Zapf (1976), foi a primeira fonte desenhada especificamente para a Digiset, seguida pelo desenvolvimento das fontes Edison (também de Zapf), Demos e Praxis (de Unger).

*Designing for the
Cathode Ray Tube.*

Demos was conceived specifically for the cathode ray tube (CRT) grid, a recent innovation in typeface design. (Hermann Zapf's Marconi, which appeared in 1976, was the first type design made especially for CRT typesetting.) I made sketches and original drawings on the CRT grid for type sizes 8- to 16-point Didot (no metric measurements yet), with 100 horizontal positions per em horizontally and 120 positions vertically. Coordinates, which together form the characters, are read into the computer and are recalled without the aid of a visible matrix. The Digiset equipment produced since 1976 can set from 6- to 16-point from this master (Figure 6).

Figure 6.

An example of an original drawing showing the CRT grid.

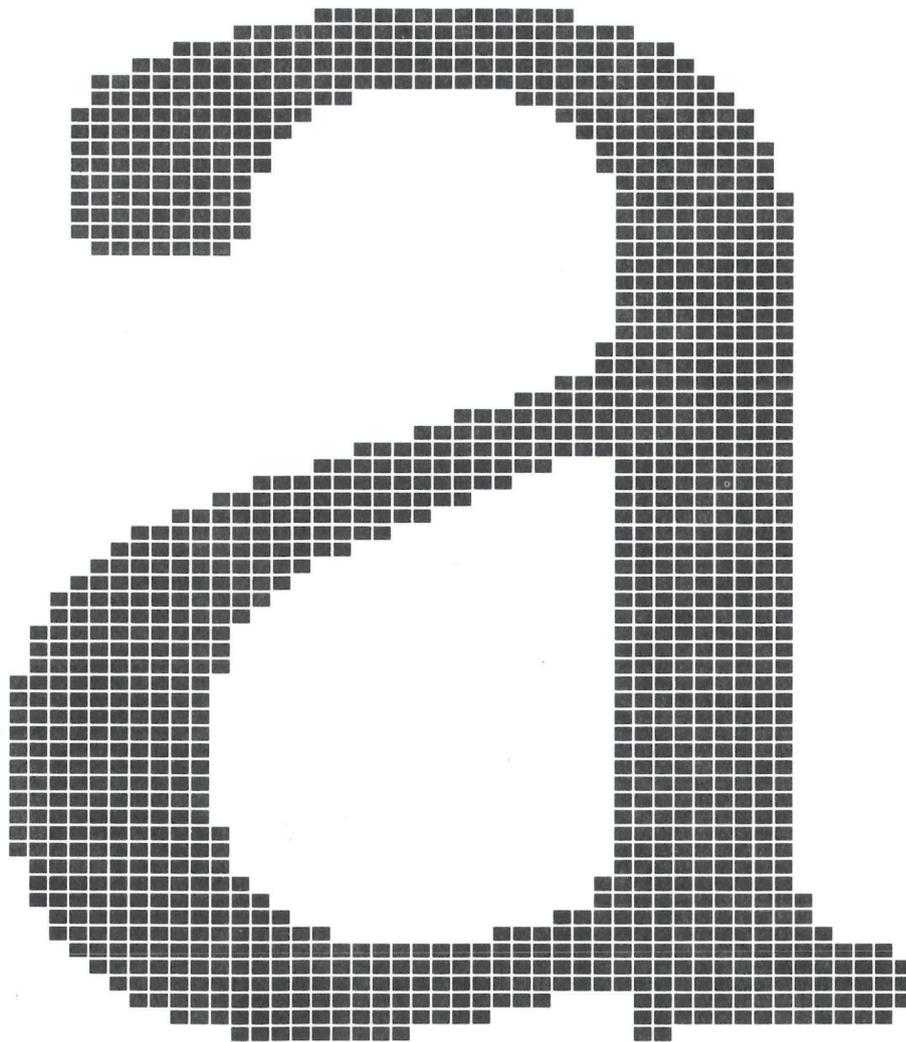


Figura 30: Exemplo de desenho original da fonte Demos, de Gerard Unger, mostrando a grade de construção modular do CRT. Fonte: UNGER, 1979, p.137.

Essas primeiras fontes bitmaps tinham algumas limitações. Primeiramente, eram tecnologias proprietárias, pensadas para equipamentos específicos e embutidas dentro das máquinas. Em segundo lugar (e talvez sua maior desvantagem), o problema de que fontes bitmaps não eram escalonáveis: quando se aumentava o corpo da fonte, os pixels começavam a se tornar visíveis. As curvas de Bézier²⁷ foram a solução encontrada para tornar os tipos escalonáveis, sem perda de resolução. O desenvolvimento mais notável nesse sentido é o Ikarus²⁸, sistema que consistia em uma mesa para digitalização de desenhos analógicos, que podiam ser convertidos em curvas vetoriais, para a criação de contornos de caracteres digitais em alta-resolução. Pode-se dizer que o Ikarus lançou as bases das curvas vetoriais dos formatos de fontes atuais (KAROW, 2013; ULRICH, 2018).



Figura 31: Mesa de trabalho de um designer de tipos do final dos anos 1980. Computador Apple Macintosh, software Ikarus M, teclado, mouse e desenhos de letras para digitalização na mesa digitalizadora Aristo. Materiais da coleção de Erik Spiekermann, fotografados por Norman Posselt, Berlin. Fonte: POSSELT, 2018.

27. Curvas de Bézier, um tipo de curva paramétrica, base do desenho vetorial digital, foram desenvolvidas pelos engenheiros Paul de Casteljaou e Pierre Bézier, por volta de 1960 (ULRICH, 2018).

28. Desenvolvido por Dr. Peter Karow, a partir de 1972, na firma URW, em Hamburgo, e apresentado na conferência da ATypI em 1975, em Varsóvia. O sistema utilizava uma mesa digitalizadora Aristo, originalmente empregada na digitalização de desenhos para a indústria naval e automotiva (KAROW, 2013; ULRICH, 2018).

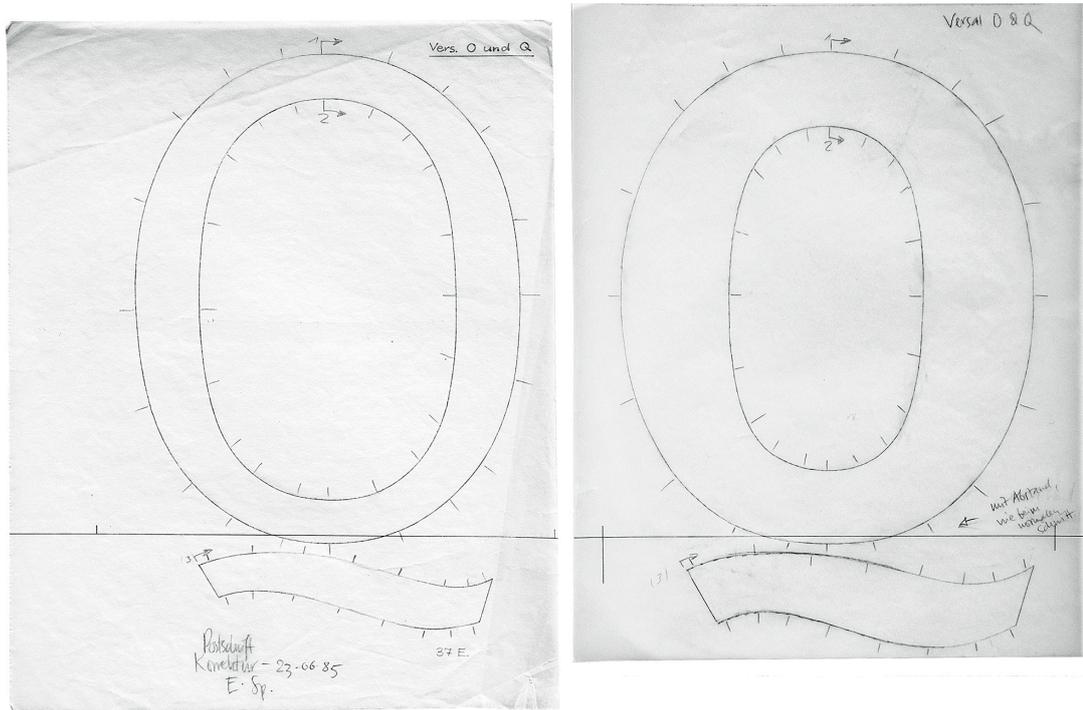


Figura 32: Desenhos de Erik Spiekermann, de junho de 1985, em papel transparente, para digitalização no Ikarus. Fonte: ULRICH, 2018.

Entre as iniciativas de desenvolvimento de formatos vetoriais do período, destaca-se o PostScript Type 1 (1983), da Adobe, criado a partir da linguagem PostScript, uma linguagem de descrição de página²⁹, também utilizada para descrever formatos de fontes vetoriais. Em 1991, a Apple lança o formato TrueType³⁰, como uma alternativa ao formato PostScript (KAROW, 2013; ULRICH, 2018). O PostScript é a origem do formato de fonte *device-independent*: termo que se aplica ao software capaz de funcionar em uma ampla variedade de dispositivos, independentemente do hardware local no qual o software é usado.

29. Page Description Language, PDL. O PostScript, desenvolvido por John Warnock e Charles Geschke, foi a linguagem que possibilitou a comunicação entre diversos dispositivos (como computador, OS, monitor, impressora), ganhando amplo reconhecimento a partir do momento em que foi implementada às impressoras LaserWriter da Apple, em 1985 (ULRICH, 2018).

30. A Apple iniciou o desenvolvimento do formato TrueType para evitar os custos de licenciamento da tecnologia PostScript (SCAGLIONE, 2019, p. 117)

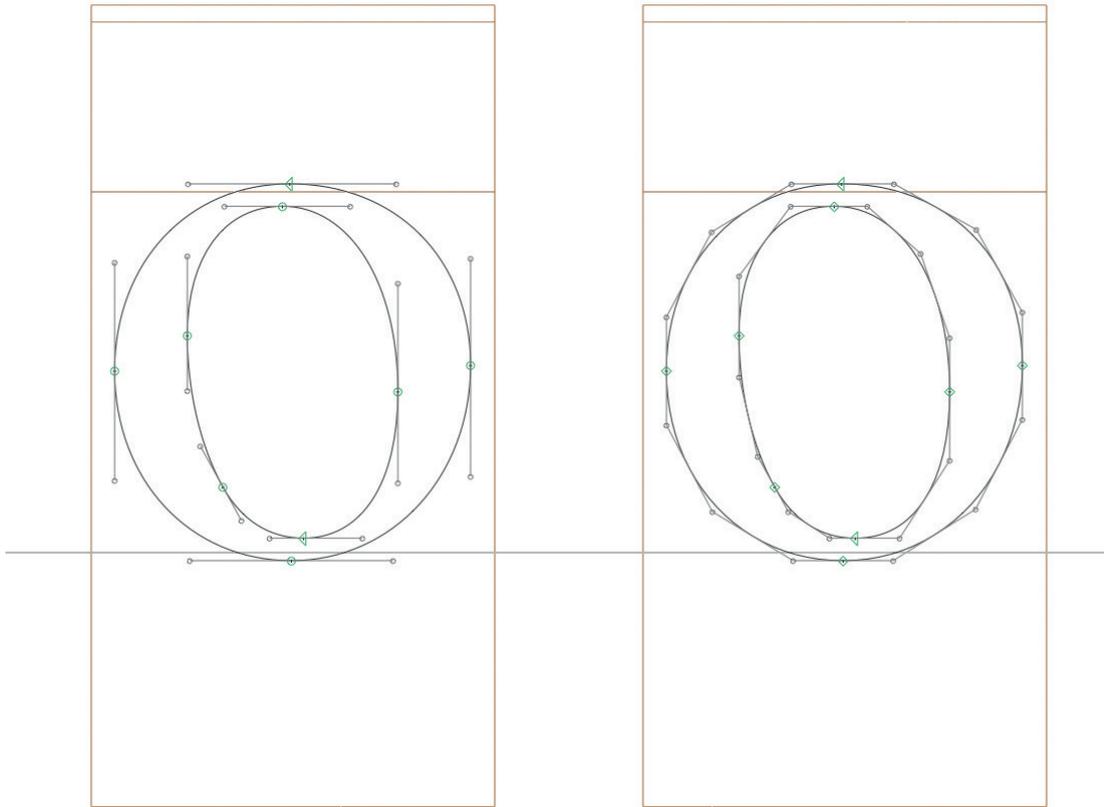


Figura 33: Diferença entre curvas de Bézier cúbicas (PostScript) e quadráticas (TrueType). Observa-se que as curvas quadráticas (ou splines) utilizam um número maior de pontos para resolver o mesmo desenho; no caso, a letra o da fonte Times New Roman. Ilustrações geradas a partir de captura de tela do software Glyphs. Fonte: do autor.

Como resposta à existência de duas tecnologias proprietárias concorrentes, incompatíveis, a Adobe e a Microsoft desenvolvem o formato OpenType. A especificação OpenType foi desenvolvida como uma extensão do TrueType³¹ e publicada originalmente em abril de 1997 (MICROSOFT CORPORATION, 2020a; MICROSOFT CORPORATION; ADOBE INC., 2020).

O OpenType resolve vários problemas dos formatos anteriores. Primeiramente, é multiplataforma, compatível com os principais sistemas operacionais (MacOS, Windows, Linux). Em segundo lugar, resolve problemas de compatibilidade entre diferentes

31. Extensão do formato TrueType 'sfnf', que pode incorporar tanto curvas PostScript Type 1 (cúbicas) como curvas TrueType (quadráticas). Fontes OpenType com curvas PostScript apresentam um nome de arquivo com extensão/sufixo .otf, enquanto fontes com curvas TrueType podem ter o sufixo otf, .ttf ou .ttc (ADOBE INC., 2008, p. 2).

codificações, já que incorpora o padrão Unicode que, por sua vez, permite um melhor suporte multilíngue dentro de um mesmo arquivo de fonte³². Por último, apresenta novos recursos tipográficos avançados, como ligaturas, substituições contextuais, reordenamento, bidirecionalidade, entre outros (ADOBE INC., 2008; HUDSON, 2002; MICROSOFT CORPORATION; ADOBE INC., 2020; SCAGLIONE, 2006; SCAGLIONE; 2019). Alguns desses recursos serão descritos em detalhes na seção *Como funciona a codificação de texto em computadores* deste capítulo.

8.2 Codificação de caracteres

A história das comunicações elétricas ou eletrônicas realmente significa a história das comunicações seriais. Serial significa um símbolo por vez, um após o outro, em uma sequência combinada. O conceito não é arbitrário, parece ser inerente à linguagem humana. As palavras são ditas uma de cada vez, as palavras têm um começo e um fim. Embora a visão seja “paralela”, de amplo espectro, tanto as linguagens baseadas no alfabeto quanto as icônicas olham para um símbolo ou ideograma por vez.

A comunicação baseada em caracteres é fundamentalmente diferente de coisas como a telefonia. Os caracteres reduzem as comunicações a símbolos distintos (incorretamente chamados de “digitais”), enquanto coisas como telefones e fac-símile (“fax”) são continuamente variáveis (razoavelmente chamadas de “analógicas”, pois as vibrações em um fone de ouvido são uma analogia, uma cópia falha, das vibrações que sua voz faz em um microfone) (JENNINGS, 2004, tradução nossa).

8.2.1 Codificação de caracteres: histórico

Esta seção tem como objetivo traçar um panorama histórico acerca do desenvolvimento das codificações no campo da comunicação. Nesse sentido, serão levantados alguns pontos.

Primeiramente, é importante observar que, apesar de haver alguns divisores de águas, muitos desenvolvimentos tecnológicos no campo das codificações de texto foram

32. Fontes OpenType são compatíveis com o padrão multi-byte do Unicode: uma mesma fonte pode conter até 65,536 caracteres (16 bits) e acomodar mais de uma escrita, e.g. latina, grega e cirílica.

desenvolvidos ao longo de períodos, até que de fato se estabelecessem como padrão. Em outras palavras, um padrão não se torna um padrão do dia para a noite.

Em segundo lugar, observa-se que é uma tendência em muitos campos do desenvolvimento tecnológico, que o padrão se estabeleça em razão de seu uso amplo. Isso se deve a vários motivos, como a dominação do mercado por certas empresas, a difusão nos principais meios, a disponibilidade de certos equipamentos ou a preferência cultural por uma parcela da população.

8.2.1.1 Origens

Desde a antiguidade, utiliza-se sistemas de codificação para transmissão de informação a longas distâncias. Haralambous (2007, p. 28) relata que, por volta de 350 A.C., na Grécia, conforme descrição do historiador Polybius, o general Aeneas utilizou um sistema que consistia em duas fileiras de cinco tochas cada, presas a uma parede, para codificar mensagens de texto. As tochas acesas e apagadas formavam combinações com as quais era possível codificar as 24 letras do alfabeto grego, com base em uma tabela, conhecida como *Quadrado de Polybius* (Fig.34).

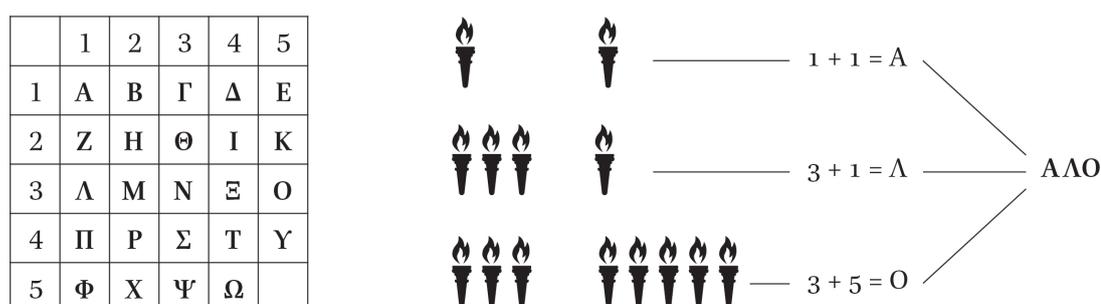


Figura 34: *Quadrado de Polybius*. Fonte: do autor; “Polybiussquare”, 2021.

Na década de 1790, o engenheiro francês Claude Chappe estabeleceu a primeira ligação³³ telegráfica entre Paris e Lille, usando semáforos visíveis a distâncias de 10 a 15

33. Chappe foi sucedido, ainda na mesma década, por Abraham Niclas Edelerantz, que criou sistema similar na Suécia, conforme descrito por Holzmann e Pehrson (1994, p. 124–129).

quilômetros. A codificação do texto se dava por meio dos posicionamento dos espelhos dos semáforos (Figura 35). As linhas de transmissão do telégrafo óptico eram formadas por um sistema de semáforos em cadeia (Figura 36), que permitiam a transmissão de mensagens a centenas de quilômetros em poucos minutos, constituindo as primeiras redes sem fio de transmissão de informação a longas distâncias (HARALAMBOUS, 2007, p. 28; HOLZMANN; PEHRSON, 1994, p. 124–129).

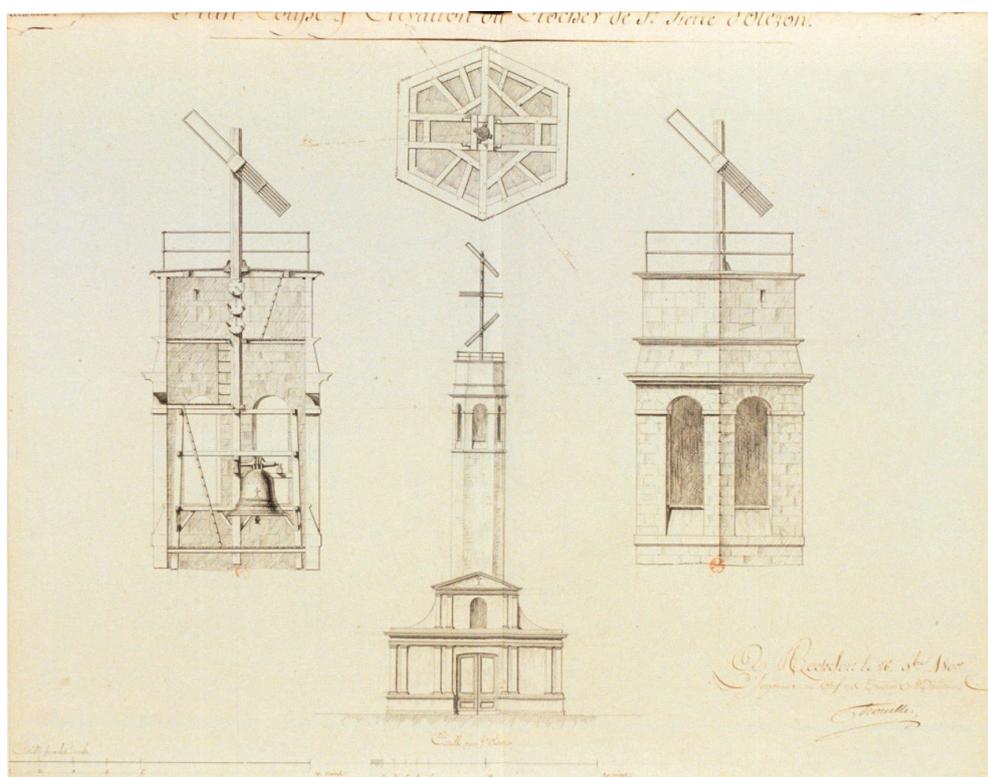


Figura 35: Telégrafo de Chappe, desenho a nanquim. Fonte: TROUILLE, 1807.

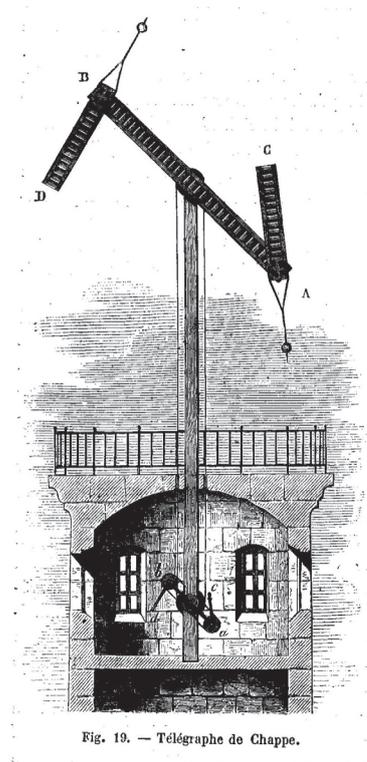


Figura 36: Telégrafo de Chappe, gravura. Fonte: FIGUIER, 1867–1891.

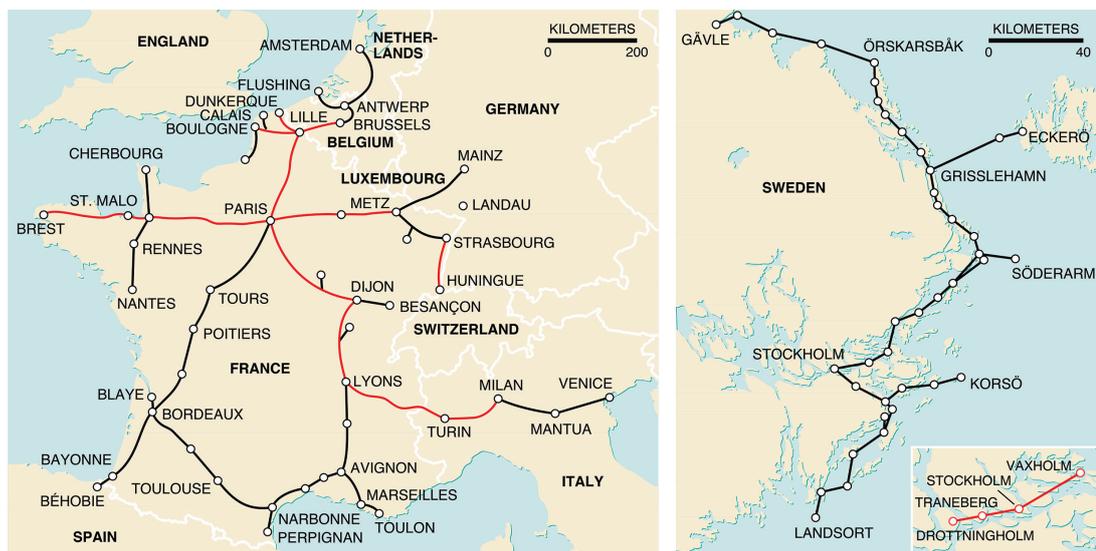


Figura 37: Mapas das primeiras redes de telégrafos óticos no século XIX. França(à esquerda) e Suécia(à direita). As linhas vermelhas na imagem à esquerda indicam redes construídas antes de 1810. Fonte: HOLZMANN; PEHRSON, 1994, p. 126.

Em 1837, Samuel Morse patenteou o código Morse³⁴, sistema que utiliza um número variável de pulsos eletromagnéticos longos e curtos (*dahs* e *dits*) para codificação de cada letra. Assim, tem-se somente duas unidades básicas: o traço (*dah*) e o ponto (*dit*). É considerado o primeiro sistema de codificação reconhecido internacionalmente (COULMAS, 1999, p. 348; HARALAMBOUS, 2007, p. 28; JENNINGS, 2004). Ainda no século XIX, com base em uma invenção de David Edward Hughes, de 1855 (Figura 38), Émile Baudot criou, em 1874, um dispositivo de codificação para transmissão telegráfica com um teclado de cinco teclas, similares às de um piano (Figura 39). Utilizando duas teclas com a mão esquerda e três teclas com a mão direita, era possível produzir até cinco perfurações por vez na fita de papel do telégrafo (Figura 40). Cada conjunto de cinco perfurações representa um caractere³⁵. Esse esquema de codificação, perfuração e leitura da fita de papel em sistemas telegráficos é conhecido como Código Baudot, Código Baudot de 5 unidades ou Código Baudot de 5 canais (*Baudot 5-unit code*) (Figura 38) (FISCHER, 2000, p. 1–2; HARALAMBOUS, 2007, p. 28; JENNINGS, 2004).

34. A invenção do código Morse é creditada a Alfred Vail, assistente de Samuel Morse.

35. No código Morse, cada caractere tem um comprimento diferente (e.g., a letra E é formada por um ponto, a letra S é formada por três traços), configurando um sistema de codificação de comprimento variável (*variable-width encoding*). O Código Baudot representa uma mudança significativa para um sistema de codificação de comprimento fixo (*fixed-width encoding*) de cinco unidades, de mais fácil decodificação: todos os caracteres, sem exceção, são formados por conjuntos de cinco perfurações na fita de papel. A grande maioria dos esquemas de codificação atuais são de comprimento fixo, como o próprio Unicode.

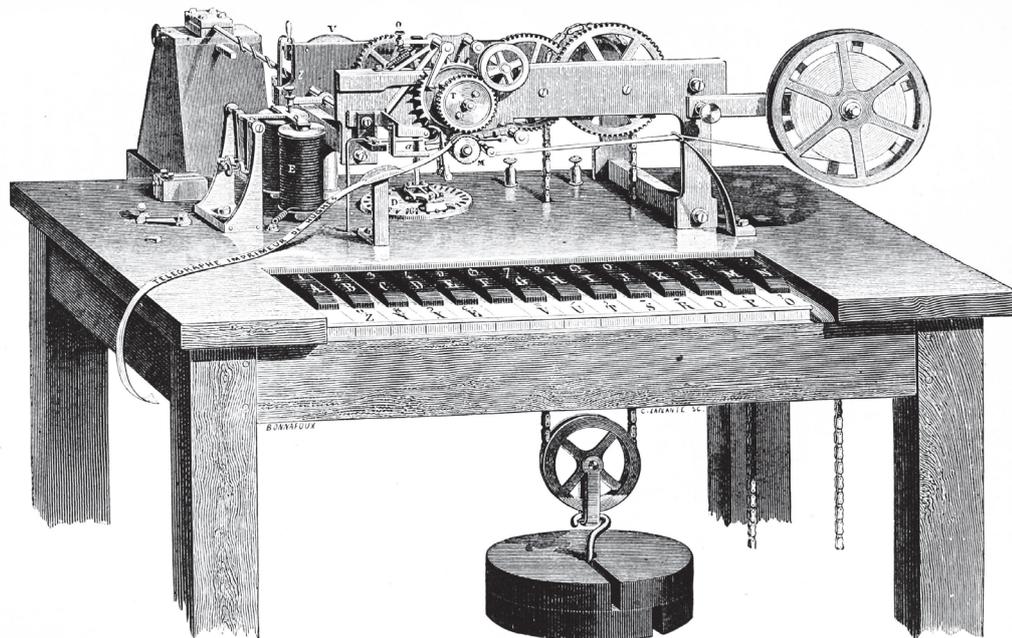


Figura 38: Telégrafo de Hughes. Fonte: DESCHANEL, 1873, p. 413.

1	2	3	4	5	6	7	8	9	0	.	,	;	:
A	B	C	D	E	F	G	H	I	J	K	L	M	N
	"	&)	(=	/	§	-	+	'	!	?	
	Z	Y	X	É	V	U	T	S	R	Q	P	O	

Figura 39: Layout do teclado de Hughes. Fonte: FISCHER, 2000, p. 3.

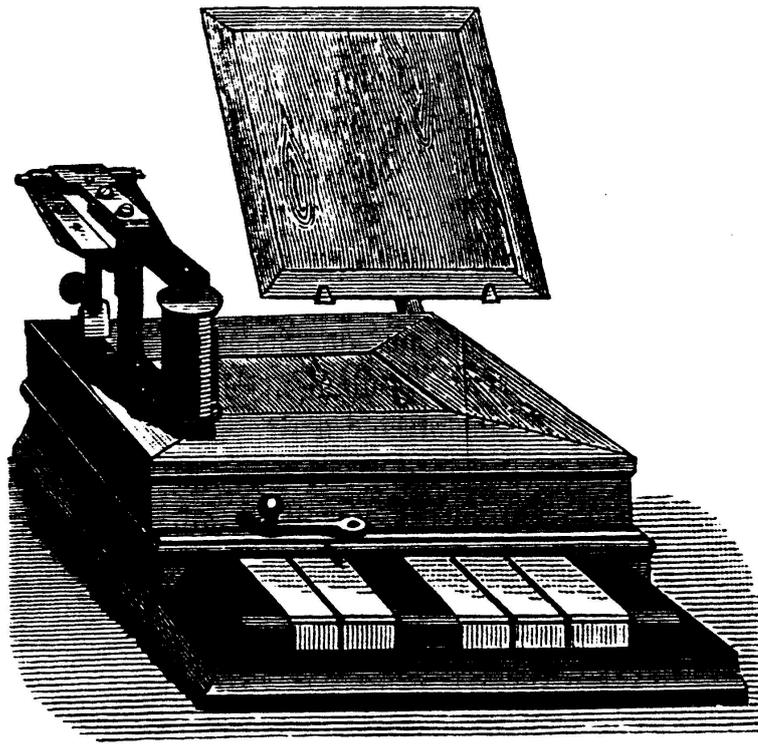


Figura 40: Teclado de Baudot. Fonte: FISCHER, 2000, p. 3.

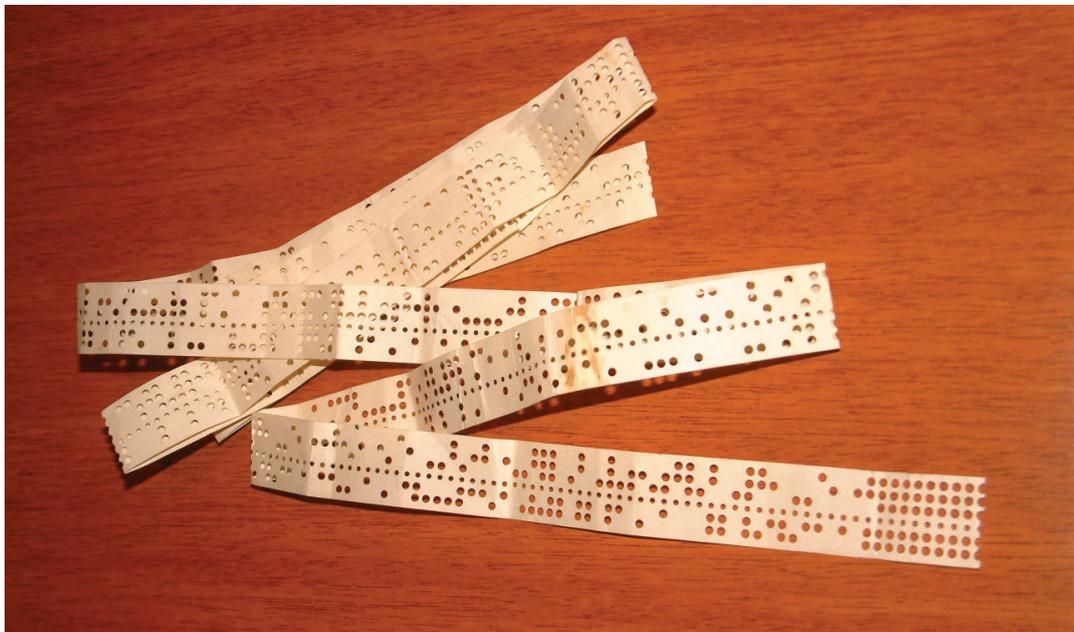


Figura 41: Fita perfurada representando o código Baudot. No centro da fita pode-se perceber uma linha com furos menores e constantes, separando a mão esquerda (dois furos) da mão direita (três furos). Fonte: OLIVEIRA, 2010.

	let	fig		let	fig
· · · · ·	unused		· · · ○○	*	*
○ · · · ·	A	1	○ · · ○○	K	(
○○ · · ·	É	&	○○ · ○○	L	=
· ○ · · ·	E	2	· ○ · ○○	M)
· ○○ · ·	I	°	· ○○ ○○	N	N ^o
○○○ · ·	O	5	○○○ ○○	P	%
○ · ○ · ·	U	4	○ · ○ ○○	Q	/
· · ○ · ·	Y	3	· · ○ ○○	R	-
· · ○ ○ ·	B	8	· · ○ · ○	S	;
○ · ○ ○ ·	C	9	○ · ○ · ○	T	!
○○○ ○ ·	D	0	○○○ · ○	V	'
· ○○ ○ ·	F	f	· ○○ · ○	W	?
· ○ · ○ ·	G	7	· ○ · · ○	X	,
○○ · ○ ·	H	h	○○ · · ○	Z	:
○ · · ○ ·	J	6	○ · · · ○	t	.
· · · ○ ·	figure space		· · · · ○	letter space	

Figura 42: Código Baudot de 5 unidades. Fonte: FISCHER, 2000, p. 3.

Com cinco teclas, é possível codificar até 32 caracteres ($2^5 = 32$). No entanto, Baudot utilizou um recurso similar ao da tecla *shift* dos teclados modernos, dobrando esse número ($2^6 = 64$) e, assim, abrindo espaço suficiente para as 26 letras do alfabeto, numerais e sinais de pontuação (FISCHER, 2000, p. 1–2; HARALAMBOUS, 2007, p. 28; JENNINGS, 2004). Pode-se dizer que o Código Baudot de 5 canais é a primeira codificação binária utilizada nas comunicações e, grosso modo, consiste em uma codificação de cinco bits (2^5).

Baudot utilizou os códigos [letter space] e [figure space] para alternar entre duas classes³⁶: de letras para numerais e vice-versa, conforme as duas colunas da Fig.42

36. Em inglês, utiliza-se também o termo *case* para representar diferentes classes de caracteres: maiúsculas, caixa alta (*uppercase*), e minúsculas, caixa baixa (*lowercase*). Por derivação, na literatura há também ocorrência do termo *figure case* (*figure*, i.e. numeral), em oposição a *letter case*. O termo *case*, em inglês, significa caixa, estojo, e faz referência à caixa tipográfica, uma gaveta com divisórias onde eram guardados os tipos de metal. Os termos *uppercase* e *lowercase* fazem referência à posição na caixa tipográfica: na parte superior (*upper*), mais distantes do tipógrafo, ao fundo da gaveta, ficavam as maiúsculas; na parte inferior (*lower*), mais próximas ao tipógrafo, ficavam as minúsculas. Os tipógrafos faziam as composições sentando-se em frente à gaveta, compondo peça por peça. E é por isso que as minúsculas ficavam mais próximas à abertura da gaveta, mais próximas ao tipógrafo, já que elas ocorrem com maior frequência na composição de textos.

(Baudot 5-unit code). Ao digitar o comando [letter space], o equipamento produz um espaço em branco no texto e começa a operar somente na classe das letras; e ao digitar o comando [figure space], o equipamento produz um espaço em branco e passa a operar somente na classe dos numerais³⁷ (FISCHER, 2000, p. 2; JENNINGS, 2004). A sequência para digitação do texto COST 700 DOLLARS 1929, por exemplo, ficaria assim: C O S T [figure space] 7 0 0 [letter space] D O L L A R S [figure space] 1 9 2 9.

Ao final do século XIX, era inevitável que o teclado de Baudot fosse substituído por algo similar às máquinas de escrever. Assim, em 1899, o neozelandês Donald Murray patenteou o primeiro teletipo; grosso modo, uma máquina de escrever telegráfica (Figuras 43, teletipo Murray). O dispositivo (Fig.44) utilizava uma codificação baseada no Código Baudot. Os caracteres no teclado eram distribuídos conforme layout demonstrado na Fig.45 (FISCHER, 2000, p. 2–5; JENNINGS, 2004).

37. Na literatura (e.g. Fischer, Jennings) há também ocorrência dos termos *letter shift* e *figure shift*, como uma tecla ou função para alternar entre letras e números.

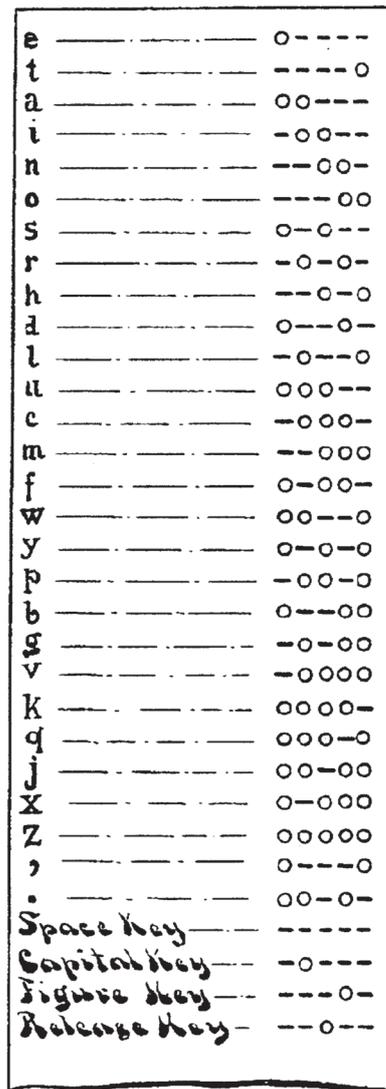


Figura 43: Esquema de codificação de Murray, conforme registrado em patente de 1899.

Fonte: FISCHER, 2000, p. 4.



Figura 44: Murray keyboard perforator model No. 5-PM. Fabricado por Reid Brothers, Engineers, Limited, Camden, Inglaterra, 1914–1916. FONTE: SCIENCE MUSEUM GROUP.

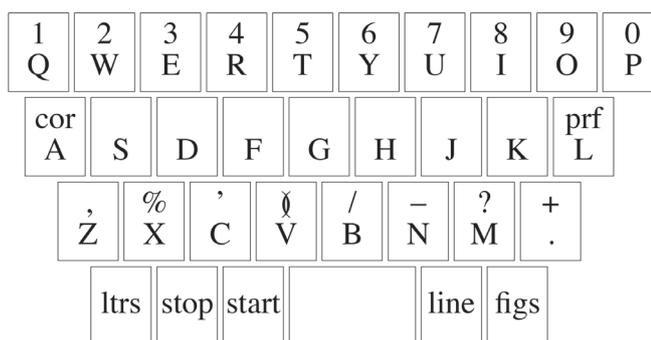


Figura 45: Layout de teclado de Murray, de 1911. Fonte: FISCHER, 2000, p. 5.

A partir da evolução dos sistemas Baudot e Murray (entre outros), e das comunicações telegráficas, surge a necessidade de padronização internacional das codificações para esse fim. As discussões a esse respeito se iniciaram em 1926, em Berlim, com as primeiras reuniões do Comité Consultatif International des Communications Télégraphiques (CCIT), e levaram cinco anos até que essa organização definisse, em 1931, dois padrões: International Telegraph Alphabets No. 1 e No. 2, respectivamente (Fig.46). Pouco tempo depois, o alfabeto No. 1 caiu em desuso, enquanto o International Telegraph Alphabet No. 2 (ITA2) firmou-se como padrão internacional (FISCHER, 2000, p. 6–9; JENNINGS, 2004).

let	fig	let	fig	let	fig	let	fig				
00000	A	1	00000	Q	/	00000	A	-	00000	Q	1
00000	B	8	00000	R	-	00000	B	?	00000	R	4
00000	C	9	00000	S	.	00000	C	:	00000	S	'
00000	D	0	00000	T	nat	00000	D	wru	00000	T	5
00000	E	2	00000	U	4	00000	E	3	00000	U	7
00000	F	nat	00000	V	'	00000	F	nat	00000	V	=
00000	G	7	00000	W	?	00000	G	nat	00000	W	2
00000	H	+	00000	X	,	00000	H	nat	00000	X	/
00000	I	nat	00000	Y	3	00000	I	8	00000	Y	6
00000	J	6	00000	Z	:	00000	J	bell	00000	Z	+
00000	K	(00000	carriage return		00000	K	(00000	carriage return	
00000	L	=	00000	line feed		00000	L)	00000	line feed	
00000	M)	00000	letter space		00000	M	.	00000	letter shift	
00000	N	nat	00000	figure space		00000	N	,	00000	figure shift	
00000	O	5	00000	* *		00000	O	9	00000	space	
00000	P	%	00000	idle		00000	P	0	00000	not used	

Figura 46: International Telegraph Alphabets No. 1(esq.)e No. 2(dir.). Fonte: FISCHER, 2000, p. 9.

Jennings (2004) observa ainda que, com o ITA2, surgem os caracteres de controle: códigos que não têm como resultado a impressão ou gravação de um símbolo, i.e. não

são visíveis, mas que desempenham funções na formatação, como quebra de linha (*line feed*), retorno do carro (*carriage return*) etc., conforme ilustrado na figuras 46 (ITA 1 e 2).

Codificações de cinco bits, como o ITA2 (e suas variações), que se baseiam no código Baudot, foram utilizadas por décadas. No entanto, com o crescente aumento do uso de computadores, na década de 1950, surge também a necessidade de padrões de codificação compatíveis com os novos equipamentos. Período coincidente com o pós-guerra, teve como reflexo o surgimento e aperfeiçoamento de novas tecnologias. Em linhas gerais, nesse período se define alguns dos padrões de codificação de texto em computadores nos quais o Unicode se baseou.

Nesse sentido, é importante observar que, nesse estágio inicial do desenvolvimento dos computadores modernos, entre 1930 e 1960, apenas alguns visionários como Alan Turing e Vannevar Bush imaginaram que o computador poderia ser usado para processamento de texto; antes disso, era pensado somente como uma máquina de calcular (HARALAMBOUS, 2007, p. 28; JENNINGS, 2004). O trabalho de Turing, *On computable numbers, with an application to the Entscheidungsproblem*, de 1936, é considerado pioneiro, fundacional para a computação moderna. Grosso modo, Turing criou ideia de programa de computador e também a ideia de uma máquina programável de uso geral, que executa tarefas de acordo com as especificações do programa (JACKSON, 2012).

8.2.1.2 Processo de transição das codificações em equipamentos de telecomunicações para o processamento de dados em computadores

Desenvolvimentos relacionados à codificação, subsequentes ao ITA2, e ocorridos logo após a Segunda Guerra Mundial estão, em linhas gerais, ligados a necessidades de compatibilidade entre diferentes sistemas de equipamentos; notavelmente, entre sistemas de telecomunicações e computadores. Como outros fatores que impulsionaram o desenvolvimento de novas codificações, pode-se apontar necessidades linguísticas regionais e necessidade de aumento do número de caracteres. Esta seção descreve algumas questões relevantes desse período de transição.

Na segunda metade da década de 1950, começam a surgir, no âmbito do CCIT, discussões acerca de necessidades regionais específicas como, por exemplo, de línguas que

utilizam caracteres acentuados. Uma dessas discussões diz respeito à possibilidade de se alocar marcas diacríticas nos espaços reservados para uso nacional na codificação ITA2 (v. Fig.46: ITA 2, código [nat], na segunda coluna). Esses espaços reservados poderiam ser personalizados: para caracteres utilizados somente em alguns países, mas não utilizados internacionalmente (e.g. sinais monetários como £, ou caracteres como ø, å, œ). No entanto, chegou-se à conclusão de que esses espaços reservados para uso nacional não seriam uma boa solução para a codificação de letras acentuadas (FISCHER, 2000, p. 8–9).

Nesse sentido, John Hudson (2002, p. 26) observa que a natureza básica de uma codificação é determinada por seu tamanho, e este é determinado pelo número de bits ou bytes que essa codificação aloca para cada caractere.

Por essa razão, entre outras, o CCIT inicia, em 1956, estudo de viabilidade da adoção de um código de seis unidades, em substituição ao código de cinco unidades. Isso possibilitaria, além do aumento do número de pontos de código (i.e. aumento do número de caracteres), a inclusão de marcas diacríticas e caracteres adicionais necessários para algumas línguas. Em reunião em 1958, o CCITT³⁸ ainda considerava prematura a ideia da adoção de um novo padrão mas, mesmo assim, preparou uma lista de marcas diacríticas que deveriam ser adotadas futuramente, em novas codificações: agudo <´ >, grave <` >, circunflexo <^ >, trema <¨ > e til <~ > (FISCHER, 2000, p. 9).

Uma das vantagens de codificações de seis bits ($2^6 = 64$) é a possibilidade de se trabalhar com um layout de teclado muito parecido com o de máquinas de escrever (Fig.47) (FISCHER, 2000, p. 9-10). Além das quatro fileiras de teclas (com numerais e letras em teclas separadas, independentes), pode-se observar também na Fig.47 que os códigos de controle, e.g. campainha [bell], apagar [rub out], retorno do carro [cr], nova linha [lf], neste teclado de 1960, já estão consolidados e bem organizados como funções essenciais³⁹.

38. Em janeiro de 1957, o Comité Consultatif International des Communications Télégraphiques (CCIT) e o Comité Consultatif International Téléphonique (CCIF) foram fundidos em uma única instituição, o Consultative Committee for International Telephony and Telegraphy (CCITT) (FISCHER, 2000, p. 9).

39. Codificações como essas foram adotadas, mesmo que por curto período, em alguns sistemas de empresas como AT&T e Western Union (FISCHER, 2000, p. 9).

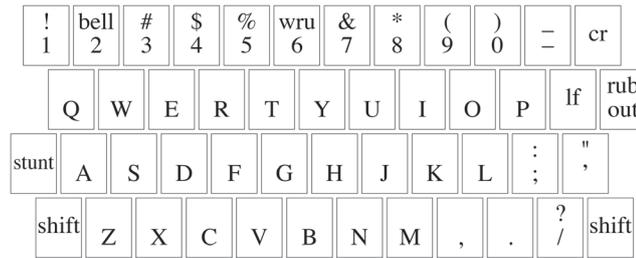


Figura 47: Proposta de layout de teclado para codificações de seis bits. Dezembro de 1960.

Fonte: FISCHER, 2000, p. 10.

8.2.1.3 Compatibilidade: origens do padrão ASCII

A partir da necessidade de um padrão de codificação compatível entre telecomunicações e processamento de dados, o exército dos Estados Unidos iniciou, em 1957, o desenvolvimento da codificação FIELDATA, parte de um sistema integrado de equipamentos de transmissão e processamento de dados, de mesmo nome. Nesse período, a maior parte dos equipamentos de telecomunicações utilizava alguma versão do ITA2, que causava incompatibilidades com computadores, pois quando arranjada em ordem binária, ficava completamente desordenada, conforme ilustra a Fig.48 (FISCHER, p. 11). Além disso, padrões como o ITA2 foram utilizados em computadores simplesmente porque era o que havia disponível (FISCHER, 2000, p. 9-10).

00000	· · · · ·	null	10000	· · · · 0	T	4
00001	0 · · · ·	E 3	10001	0 · · · 0	Z	+
00010	· 0 · · ·	line fd	10010	· 0 · · 0	L)
00011	00 · · ·	A -	10011	00 · · 0	W	2
00100	· · 0 · ·	space	10100	· · 0 · 0	H	nat
00101	0 · 0 · ·	S ' ,	10101	0 · 0 · 0	Y	6
00110	· 00 · ·	I 8	10110	· 00 · 0	P	0
00111	000 · ·	U 7	10111	000 · 0	Q	1
01000	· · · 0 ·	car ret	11000	· · · 00	O	9
01001	0 · · 0 ·	D nat	11001	0 · · 00	B	?
01010	· 0 · 0 ·	R 4	11010	· 0 · 00	G	nat
01011	00 · 0 ·	J bell	11011	00 · 00	figures	
01100	· · 00 ·	N ,	11100	· · 000	M	.
01101	0 · 00 ·	F nat	11101	0 · 000	X	/
01110	· 000 ·	C :	11110	· 0000	V	=
01111	0000 ·	K (11111	00000	letters	

Figura 48: International Telegraph Alphabet No. 2, arranjado em ordem binária.

Fonte: FISCHER, 2000, p. 11.

Por essa razão, surge o FIELDATA: em linhas gerais, uma codificação de 7 bits, de uso exclusivo militar, que nunca teve aplicação comercial, embora posteriormente tivesse muita influência no desenvolvimento da codificação ASCII. Pode-se mencionar, como inovações do FIELDATA, os agrupamentos e ordenação em ordem binária, de classes (letras, numerais) e funções (caracteres de controle ou imprimíveis) (HARALAMBOUS, 2007, p. 29; JENNINGS, 2004).

MS ₄₀	UC ₄₁	LC ₄₂	LF ₄₃	CR ₄₄	SP ₄₅	A ₄₆	B ₄₇	C ₄₈	D ₄₉	E _{4A}	F _{4B}	G _{4C}	H _{4D}	I _{4E}	J _{4F}
K ₅₀	L ₅₁	M ₅₂	N ₅₃	O ₅₄	P ₅₅	Q ₅₆	R ₅₇	S ₅₈	T ₅₉	U _{5A}	V _{5B}	W _{5C}	X _{5D}	Y _{5E}	Z _{5F}
) ₆₀	- ₆₁	+ ₆₂	< ₆₃	= ₆₄	> ₆₅	- ₆₆	\$ ₆₇	* ₆₈	(₆₉	" _{6A}	: _{6B}	? _{6C}	! _{6D}	, _{6E}	STOP _{6F}
0 ₇₀	1 ₇₁	2 ₇₂	3 ₇₃	4 ₇₄	5 ₇₅	6 ₇₆	7 ₇₇	8 ₇₈	9 ₇₉	' _{7A}	; _{7B}	/ _{7C}	. _{7D}	SPEC _{7E}	IDLE _{7F}

Figura 49: codificação FIELDATA. Fonte: HARALAMBOUS, 2007, p. 29.

No mesmo período, em paralelo ao FIELDATA, ocorreram outras tentativas de normalização dos padrões de codificação na indústria de computadores. Por exemplo, o EBCDIC⁴⁰ da IBM, uma das primeiras codificações de oito bits, que se baseava no sistema de codificação dos cartões perfurados de Hollerith. Foi utilizada, sobretudo, em computadores de grande porte da IBM, como o IBM System/360⁴¹, de 1964 (HARALAMBOUS, 2007, p. 31–32).

Por esse motivo, a indústria de telecomunicações empenhou esforços para desenvolver uma codificação padrão internacional. A própria IBM, além da AT&T, Bell e Teletype Corporation (entre outras), tinha necessidade imediata de adoção de um novo padrão. Além do fato de que essas empresas praticamente monopolizavam o setor, tinham também grande aporte financeiro para promover essas mudanças. Tudo isso levou a American

40. O EBCDIC (Extended Binary Coded Decimal Interchange Code) é uma codificação de oito bits, baseada na codificação do sistema dos cartões perfurados inventado por Herman Hollerith em 1890, por sua vez, inspirado no sistema de cartões perfurados para máquinas de tear do francês Joseph-Marie Jacquard, de 1801. No final do século XIX, Hollerith fundou a Tabulating Machine Company que, mais tarde, em 1924, se tornaria a International Business Machines, ou IBM (HARALAMBOUS, 2007, p. 31–32).

41. Segundo Haralambous (2007, p. 31–32), o segundo projeto industrial mais caro da década de 1960 nos Estados Unidos, ficando atrás somente do projeto Apollo, da NASA.

Standards Association (ASA)⁴² a definir o ASCII (acrônimo de American Standard Code for Information Interchange), uma nova codificação de sete bits. Assim, o ASCII-1963, uma versão preliminar do padrão, sem letras minúsculas, nasceu em junho de 1963 e foi atualizado em 1967, com a inclusão de letras minúsculas. Com base no ASCII, a International Organization for Standardisation publicou, em seguida, a codificação ISO/IEC 646, da qual a International Reference Version (ISO 646:IRV) é praticamente idêntica, e é equivalente ao ASCII⁴³. A norma ISO/IEC 646 também foi ratificada pela European Computer Manufacturers Association (ECMA), tendo a norma ECMA-6 como equivalente ao ISO 646:IRV (FISCHER, 2000, p. 28; HARALAMBOUS, 2007, p. 29–31, JENNINGS, 2004).

NUL ₀₀	SOH ₀₁	STX ₀₂	ETX ₀₃	EOT ₀₄	ENQ ₀₅	ACK ₀₆	BEL ₀₇	BS ₀₈	HT ₀₉	LF _{0A}	VT _{0B}	FF _{0C}	CR _{0D}	SO _{0E}	SI _{0F}
DLE ₁₀	DC1 ₁₁	DC2 ₁₂	DC3 ₁₃	DC4 ₁₄	NAK ₁₅	SYN ₁₆	ETB ₁₇	CAN ₁₈	EM ₁₉	SUB _{1A}	ESC _{1B}	FS _{1C}	GS _{1D}	RS _{1E}	US _{1F}
SP ₂₀	! ₂₁	" ₂₂	# ₂₃	\$ ₂₄	% ₂₅	& ₂₆	' ₂₇	(₂₈) ₂₉	* _{2A}	+ _{2B}	, _{2C}	- _{2D}	. _{2E}	/ _{2F}
0 ₃₀	1 ₃₁	2 ₃₂	3 ₃₃	4 ₃₄	5 ₃₅	6 ₃₆	7 ₃₇	8 ₃₈	9 ₃₉	: _{3A}	; _{3B}	< _{3C}	= _{3D}	> _{3E}	? _{3F}
@ ₄₀	A ₄₁	B ₄₂	C ₄₃	D ₄₄	E ₄₅	F ₄₆	G ₄₇	H ₄₈	I ₄₉	J _{4A}	K _{4B}	L _{4C}	M _{4D}	N _{4E}	O _{4F}
P ₅₀	Q ₅₁	R ₅₂	S ₅₃	T ₅₄	U ₅₅	V ₅₆	W ₅₇	X ₅₈	Y ₅₉	Z _{5A}	[_{5B}	\ _{5C}] _{5D}	^ _{5E}	_ _{5F}
` ₆₀	a ₆₁	b ₆₂	c ₆₃	d ₆₄	e ₆₅	f ₆₆	g ₆₇	h ₆₈	i ₆₉	j _{6A}	k _{6B}	l _{6C}	m _{6D}	n _{6E}	o _{6F}
p ₇₀	q ₇₁	r ₇₂	s ₇₃	t ₇₄	u ₇₅	v ₇₆	w ₇₇	x ₇₈	y ₇₉	z _{7A}	{ _{7B}	_{7C}	} _{7D}	~ _{7E}	DEL _{7F}

Figura 50: Codificação ASCII, 1967. As duas primeiras linhas são ocupadas exclusivamente por códigos de controle. Fonte: HARALAMBOUS, 2007, p. 29–30.

O surgimento da codificação ASCII foi um divisor de águas. Todos os padrões posteriores a utilizaram como referência e ponto de partida. No Unicode, por exemplo, os primeiros 128 pontos de código são idênticos ao ASCII. O aumento considerável no número de caracteres/pontos de código ($2^7 = 128$) possibilitou uma série de novos recursos como, por exemplo, a inclusão de minúsculas e o rompimento definitivo com o passado das codificações de cinco bits, nas quais era preciso fazer uso do recurso dos

42. Em 1969, a American Standards Association (ASA) se tornaria a American National Standards Institute (ANSI).

43. A única diferença entre os dois padrões é que a ISO 646 IRV substituiu o cifrão <\$> do ASCII pelo símbolo monetário genérico <¤>.

comandos [letter space] e [figure space] para alternar entre as classes de letras e numerais e vice-versa. Além disso, o agrupamento e a ordem dos caracteres foram resolvidos de uma maneira mais inteligente: nas primeiras duas colunas, os códigos de controle, seguidas pelas colunas dos sinais de pontuação e símbolos não alfabéticos, numerais, maiúsculas e, por último, as duas colunas das minúsculas. Esse arranjo não somente incorpora as codificações anteriores (e.g. ITA2, FIELDATA), como também elimina suas ambiguidades e inconsistências (JENNINGS, 2004).

Por último, é importante observar que, até os dias de hoje, o ASCII é o padrão utilizado na sintaxe básica de muitas linguagens de programação, como linguagens de comando e linguagens de marcação, entre outras notações utilizadas em programação (KORPELA, 2006, p. 175).

8.2.1.4 O surgimento do padrão ISO 8859: transição para o padrão de oito bits

Apesar de ter alavancado processos de padronização e internacionalização na década de 1960, o ASCII ainda apresentava um conjunto de caracteres limitado; não somente por ter sido pensado inicialmente para o mercado estadunidense e para a variação do inglês norte-americano, mas também pela limitação dos 128 pontos de código da codificação de sete bits.

Por essa razão, não demorou muito para que começassem a surgir variantes do ASCII mundo afora; versões nacionais, com alterações no conjunto de caracteres, com base na norma ISO/IEC 646⁴⁴. Essas alterações consistiam, basicamente, na substituição de caracteres pouco utilizados por outros que eram indispensáveis em certas localidades, como diacríticos, sinais monetários, entre outros. Na Inglaterra, por exemplo, o cifrão <\$> foi substituído pelo sinal de libra esterlina <£> (CONSTABLE, 2001; HARALAMBOUS, 2007, p. 31).

Haralambous (2007, p. 33-35) relata que no início da década de 1970, o maior problema em relação às diversas versões localizadas do ASCII era causado pela utilização

44. A norma ISO/IEC 646 incorporou as variações do ASCII propostas por órgãos de normalização nacionais, e.g. Deutsches Institut für Normung (Alemanha) e British Standards Institution (Reino Unido), e também de variações do ASCII propostas por empresas (e.g. IBM) para mercados específicos.

simultânea de diferentes codificações num mesmo documento ou transmissão; mudar de uma codificação para outra no meio de um documento ou transmissão ainda era uma dificuldade. Foi por essa razão que, em 1973, surgiu a norma ISO 2022 que, grosso modo, não é uma codificação, mas um esquema que possibilita o uso de codificações distintas dentro do mesmo conjunto de dados.

	x0	x1	x2	x3	x4	x5	x6	x7	x8	x9	xA	xB	xC	xD	xE	xF
0x																
1x																
2x																
3x																
4x																
5x																
6x																
7x	C0			GL				C1					GR			
8x																
9x																
Ax																
Bx																
Cx																
Dx																
Ex																
Fx																

Figura 51: Estrutura da codificação ISO 2022. As oito colunas à esquerda (C0 e GL) são ocupadas pelo ASCII. C0 e C1 indicam posições de caracteres de controle, e GL e GR indicam posições de caracteres imprimíveis. As letras C e G remetem a *Control* e *Graphic*, respectivamente.

Fonte: HARALAMBOUS, 2007, p. 34.

Em linhas gerais, esse método consiste em somar os 128 caracteres do ASCII a outros 128 caracteres adicionais. Enquanto o conjunto do ASCII se conservava intacto, as outras 128 posições na tabela podiam ser utilizadas para outros caracteres acentuados, símbolos, sinais matemáticos, pontuação etc. Nesse sentido, pode-se dizer que o ISO 2022 é o elo entre as codificações de sete e oito bits.

A partir do padrão ISO 2022, com o qual era possível combinar várias codificações em um único fluxo de dados, a International Organization for Standardisation passou a definir codificações que são, em linhas gerais, versões expandidas do ASCII. Essa família de codificações de oito bits é conhecida como ISO 8859. O padrão ISO 8859-1 (Latin-1), o primeiro da família, foi publicado em 1987, e tinha como objetivo cobrir as línguas da Europa Ocidental (HARALAMBOUS, 2007, p. 35–37).

NBSP A0	í A1	ç A2	£ A3	¤ A4	¥ A5	¡ A6	§ A7	¨ A8	© A9	ª AA	« AB	¬ AC	SHY AD	® AE	¯ AF
° B0	± B1	² B2	³ B3	´ B4	µ B5	¶ B6	· B7	¸ B8	¹ B9	º BA	» BB	¼ BC	½ BD	¾ BE	¿ BF
À C0	Á C1	Â C2	Ã C3	Ä C4	Å C5	Æ C6	Ç C7	È C8	É C9	Ê CA	Ë CB	Ì CC	Í CD	Î CE	Ï CF
Ð D0	Ñ D1	Ò D2	Ó D3	Ô D4	Õ D5	Ö D6	× D7	Ø D8	Ù D9	Ú DA	Û DB	Ü DC	Ý DD	Þ DE	ß DF
à E0	á E1	â E2	ã E3	ä E4	å E5	æ E6	ç E7	è E8	é E9	ê EA	ë EB	ì EC	í ED	î EE	ï EF
ð F0	ñ F1	ò F2	ó F3	ô F4	õ F5	ö F6	÷ F7	ø F8	ù F9	ú FA	û FB	ü FC	ý FD	þ FE	ÿ FF

Figura 52: Codificação ISO 8859-1. A imagem mostra somente os caracteres imprimíveis da codificação ISO 8859-1, complementares ao ASCII. Fonte: HARALAMBOUS, 2007, p. 36.

8.2.1.5 Localização: multiplicação dos padrões de oito bits

A partir da década de 1980, com a popularização dos computadores pessoais, houve um aumento considerável na demanda em novos mercados e, assim, por equipamentos preparados com codificações compatíveis com os padrões linguísticos desses novos mercados. Nesse período, se intensificaram os processos de localização (*localisation*), termo que, em tecnologia da informação, significa adaptar à localidade, adaptar às necessidades de uma região.

A proliferação desses novos padrões de codificação teve origem em duas fontes: órgãos de normalização técnica e fabricantes independentes de software (CONSTABLE, 2001).

Por parte dos fabricantes, novas codificações surgiram por dois motivos principais. Primeiro, para suprir necessidades de um produto particular, em relação a um mercado particular. Por exemplo, a IBM codepage 437 para o sistema operacional DOS⁴⁵, é uma versão estendida do ASCII, à qual foram adicionados caracteres necessários para o grupo de línguas da Europa Ocidental⁴⁶, além de elementos gráficos para elementos de interface (i.e. linhas e formas geométricas). Além da codificação 437, a IBM criou também padrões para outros mercados, como a IBM codepage 852, para línguas do leste

45. MS-DOS, sistema operacional desenvolvido pela Microsoft para o IBM-PC no início da década de 1980.

46. No jargão, Europa Ocidental, ou Western European, se refere, grosso modo, a dois grandes grupos de línguas da porção ocidental da Europa: românicas (e.g. português, francês, espanhol etc.) e germânicas (alemão, inglês, dinamarquês, holandês etc.).

européu⁴⁷, e a IBM codepage 855, para cirílico. Em segundo lugar, outros fabricantes criaram padrões de codificação para suprir as necessidades de seus produtos e clientes. Como exemplo, podemos citar o mercado de publicações, que tinha necessidades específicas de composição tipográfica de alta qualidade, para além das capacidades do ASCII. Essas codificações eram usadas internamente e não foram amplamente difundidas. Nessa linha, os computadores de grande porte são outro exemplo, e estes empregavam seus próprios padrões de codificação, diferentes daqueles embutidos nos modelos de varejo (CONSTABLE, 2001).

Em produtos de varejo, empresas como Apple, IBM e Microsoft estão entre os principais responsáveis pela proliferação de diferentes padrões de codificação. A Apple criou vários padrões (diferentes da Microsoft e da IBM), adaptados à natureza distinta das interfaces gráficas da linha Macintosh. Da mesma forma, quando a Microsoft iniciou o desenvolvimento do Windows, surgiu a necessidade de novas codificações; é o caso da Windows Codepage 1252, comumente chamada de “Western”, “Latin 1” or “ANSI”⁴⁸ (CONSTABLE, 2001).

Além dos fabricantes de software e hardware, órgãos de normalização técnica também foram responsáveis pela difusão de padrões de codificação. É o caso do ASCII, que começou como um padrão nos Estados Unidos, mas depois foi normatizado pela International Organization for Standardisation. Há ainda exemplos de padrões que tiveram origem em nível internacional, como a família ISO 8859 (CONSTABLE, 2001).

Diante desse cenário, pode-se listar uma série de problemas relativos às codificações de oito bits. Primeiramente, Peter Constable (2001) observa que, diante da profusão de diferentes padrões de codificação, não é de se surpreender que, em muitos casos, um determinado padrão oferecesse suporte para certos caracteres que outros não

47. Europa Oriental, ou Eastern European, se refere, grosso modo, ao grupo de línguas eslavas da Europa Central, como tcheco, polonês, croata, húngaro etc. Todas essas línguas eslavas escritas com a escrita latina.

48. Em referência ao American National Standards Institute (ANSI), que adotou esta codificação como norma.

ofereciam. No entanto, nenhuma dessas codificações herdadas⁴⁹ era suficientemente abrangente, completa. Além disso, em muitos casos, diferentes padrões de codificação davam suporte para os mesmos conjuntos de caracteres, mas de maneiras incompatíveis. Por essa ótica, John Hudson (2002, p. 26–27) explica que apesar de padrões como Apple Mac Roman⁵⁰, ISO 8859-1 e Windows Codepage 1252⁵¹ terem sido pensados para dar suporte ao mesmo grupo de línguas (Western European), todos eles diferem na correspondência entre pontos de código e caracteres.

Nesse sentido, John Hudson (2002, p. 26–27) descreve o problema da inevitável ambiguidade entre diferentes codificações de oito bits, já que estas utilizam os mesmos números (i.e. pontos de código) para representar diferentes caracteres (Fig.53). Por exemplo, ao se comparar somente codificações da Microsoft, pode-se constatar que o mesmo ponto de código (F0) representa diferentes caracteres em quase todas as codificações de oito bits, conforme a lista a seguir:

Codepage (description) / Postscript name / character

Windows Codepage 1250 (Central European) / dcroat / đ

Windows Codepage 1251 (Cyrillic) / unio440 / p

Windows Codepage 1252 (Western European) / eth / ð

Windows Codepage 1253 (Greek) / pi / π

Windows Codepage 1254 (Turkish) / gbreve / ğ

Windows Codepage 1257 (Baltic) / scaron / š

Windows Codepage 1250 (Vietnamese) / dcroat / đ

49. A expressão *legacy encoding* é empregada para codificações anteriores ao Unicode. Neste trabalho será traduzida como codificação herdada ou codificação legada. De modo geral, a expressão é utilizada para designar codificações de oito bits.

50. De acordo com Haralambous (2007, p. 47), as codificações da Apple eram versões estendidas do ASCII. No entanto, a parte que excede o ASCII, era totalmente diferente do ISO 8859-1.

51. De acordo com Haralambous (2007, p. 45–46), a codificação MS Windows Codepage 1252 é baseada na codificação ISO 8859-1; a diferença é que a Microsoft incluiu novos caracteres nas linhas 8 e 9 (as duas primeiras após o bloco do ASCII, Fig.52), posições que na codificação ISO 8859-1 eram reservadas a caracteres de controle, conforme se pode observar na Fig.53.

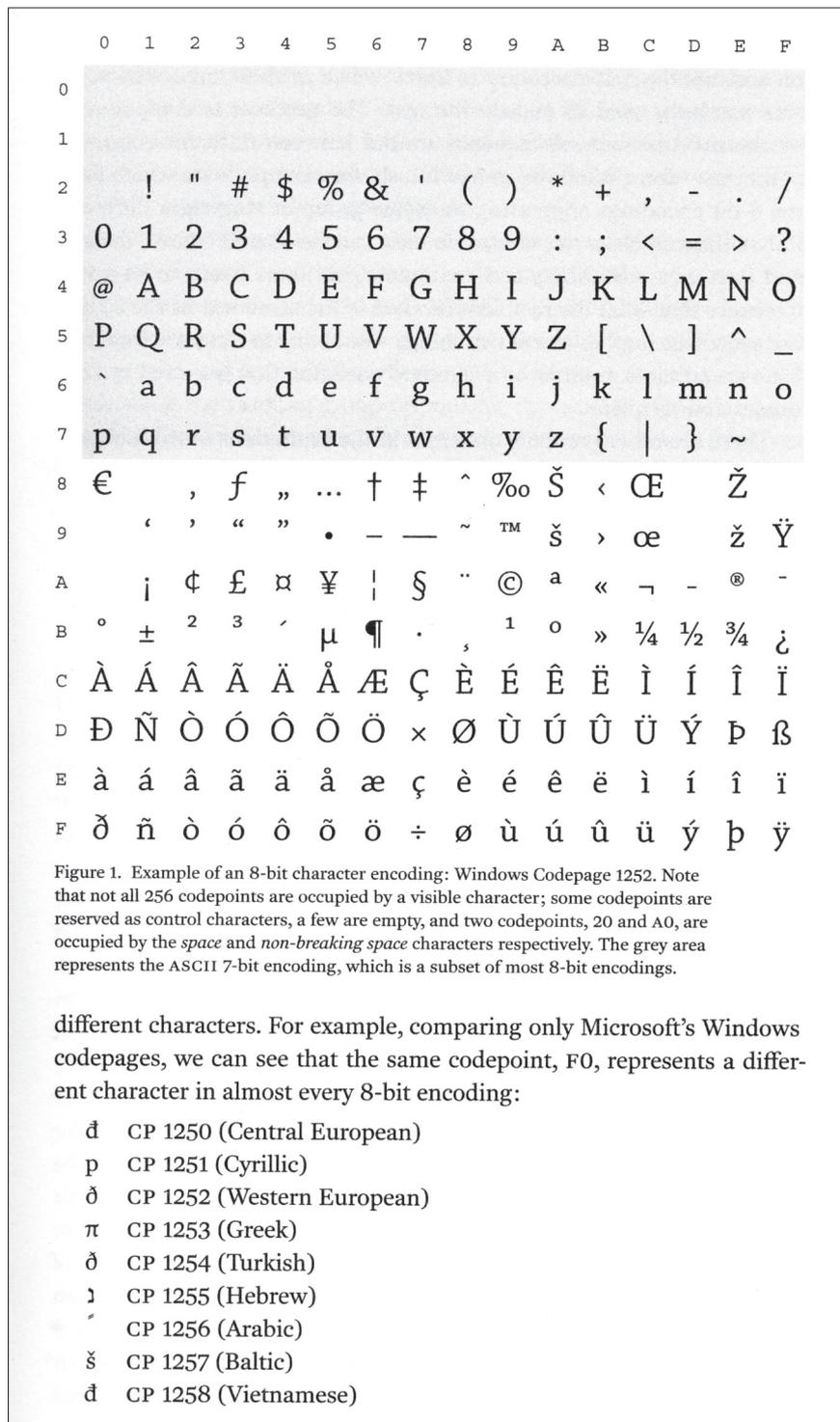


Figura 53: Exemplo de codificação de oito bits: Windows Codepage 1252. Pode-se observar que nem todos os pontos de código são ocupados por um caractere visível; alguns pontos de código são reservados aos caracteres de controle, alguns estão vazios, e dois pontos de código, 20 e A0, são ocupados pelo espaço e pelo espaço não-quebrável/espaço duro (*nonbreakingspace*), respectivamente. A área cinza representa a codificação ASCII, que é um subconjunto da maioria das codificações de oito bits. Fonte: HUDSON, 2002, p.27.

Logo, para possibilitar a visualização de um documento em um computador, é preciso saber qual padrão (entre os muitos existentes) foi utilizado na codificação de um texto. O problema se agrava quando há necessidade de intercâmbio de documentos entre diferentes plataformas (e.g. de Windows para Macintosh), nas quais há diferenças até mesmo nas codificações de oito bits que dão suporte para o mesmo grupo de línguas (e.g. Apple Mac Roman e Windows Codepage 1252). É preciso haver um jeito de assegurar que o que o destinatário recebe é o mesmo texto que o remetente enviou. Isso implica duas coisas: ou a habilidade de conversão de uma codificação de uma plataforma para outra, ou uma codificação padrão que seja utilizada por ambas (HUDSON, 2002, p. 28).

Em segundo lugar, há problemas que são inerentes às limitações dos 256 pontos de código das codificações de oito bits, especialmente se olharmos para além dos requisitos relativamente simples da grande maioria das línguas europeias. Muitos dos sistemas de escrita orientais fazem uso de mais de 256 caracteres. O silabário etíope, por exemplo, utiliza cerca de 345 caracteres. O conjunto de caracteres Han (denominado *hanza* na China, *kanji no* Japão e *hanja* na Coréia) pode chegar a 80.000. Obviamente, 256 caracteres não resolvem o problema desses sistemas de escrita (HUDSON, 2002, p. 28).

Há duas formas de se aumentar o número de caracteres em uma codificação: criando uma coleção de codificações de oito bits, que podem ser usadas em conjunto, ou aumentando o número de bits. A segunda, de mais fácil implementação, reflete em um aumento do número de bytes a serem processados. Isso significa um pulo de oito bits ($2^8 = 256$) para dezesseis bits ($2^{16} = 65.536$) (HUDSON, 2002, p. 28); este último, o método adotado para a constituição do padrão Unicode.

Antes de concluir esta seção, é importante observar que, apesar da atual ubiquidade e onipresença do Unicode, as codificações legadas, por algumas razões, não podem ser ignoradas. Talvez a principal delas seja a compatibilidade reversa, i.e. os programas atuais têm que ser capazes de interagir com dados em termos de codificações legadas (CONSTABLE, 2001).

8.2.1.6 Unicode: surgimento de um novo padrão

Devido aos problemas descritos na seção anterior, como incompatibilidade entre plataformas, ambiguidades e inconsistências entre codificações similares e limitações do número de caracteres, diversas empresas e entidades uniram esforços na tentativa de solucioná-los, conforme John Hudson (2002, p. 28–29, tradução nossa) descreve a seguir:

Em 1986, engenheiros da Xerox e da Apple começaram a discutir uma solução comum para os desafios de codificação do chinês [e japonês] [...]. Essa noção, expressa como Han unification – de tratamento dos sistemas de escrita chines e japonês não como escritas separadas, mas como expressões diferentes da mesma escrita – tornou-se um dos princípios chave nas propostas que resultariam no Padrão Unicode. Ao longo do ano seguinte, essas discussões se expandiram para considerar a ideia de um padrão universal de codificação de caracteres de comprimento fixo para todas as escritas e idiomas do mundo.

No final de 1987, Joe Becker, da Xerox, cunhou o termo Unicode, a partir de “codificação de caracteres única, universal e uniforme”. A primeira versão do Unicode Standard foi publicada em 1990, quando os comitês ad hoc originais evoluíram para o Unicode Consortium, com representação de muitas das maiores empresas de software do mundo com interesse na internacionalização.

Anteriormente, em 1984, a International Organisation for Standardisation (ISO) formou um grupo de trabalho para desenvolver um padrão de codificação de caracteres universal similar [ao Unicode]: ISO/IEC 10646. A principal diferença entre os padrões do Unicode e do ISO é que o Unicode Consortium é uma organização industrial, com representantes diretos de organismos nacionais de normalização. Em 1991, [...] percebendo a futilidade de manter duas codificações de caracteres universais, porém incompatíveis, o WG2 e o corpo editorial do Unicode – o Unicode Technical Committee, ou UTC – concordaram em fundir os dois padrões, mas seus repertórios de caracteres são mantidos e estendidos cooperativamente, e o Unicode é reconhecido pela ISO como a forma oficial de implementar a ISO/IEC 10646.

Essa história, bastante abreviada, deixa de fora muitos detalhes e falha em reconhecer os muitos homens e mulheres que fizeram do padrão Unicode o seu trabalho por mais de quinze anos. O que mesmo a história mais condensada do Unicode e ISO/IEC 10646 deixa claro é a percepção comum, em meados da década de 1980, de que a internacionalização de softwares exigiria uma codificação universal de caracteres para fornecer um identificador único para cada caractere nos muitos sistemas de escrita do mundo.

Além disso, é importante mencionar que nenhuma tecnologia se torna padrão e acessível do dia para a noite. Levou algum tempo até que o Unicode fosse implementado pelas grandes empresas em diferentes programas, plataformas e sistemas. Levou quase uma década para que o padrão fosse implementado em sistemas operacionais como Windows e MacOS, e em programas da Adobe. Hoje, o Unicode é o padrão não somente nesses programas e sistemas, mas também para o padrão XML adotado pelo consórcio WWW (KORPELA, 2006, p. 17; HUDSON, 2002, p. 29).

Questões técnicas relativas ao Unicode de maior relevância para esta tese serão mais bem descritas na seção 8.2.2 *Unicode*, a seguir.

8.2.1.7 Transição tecnológica: relações entre codificações de caracteres e a técnica tipográfica tradicional

Antes de concluir esta seção sobre a história das codificações, é importante levantar algumas questões, relacionando codificações de caracteres e tecnologias tipográficas.

Em primeiro lugar, pode-se observar que somente com as codificações de 7 bits o uso de maiúsculas e minúsculas em computadores torna-se corrente. Maiúsculas e minúsculas são variações da escrita padronizadas na técnica tipográfica desde o século XV, mas somente a partir de codificações como ASCII e EBCDIC é que passaram a fazer parte do repertório de codificações em computadores. Maiúsculas e minúsculas também foram, desde cedo, incorporadas aos teclados de máquinas de escrever. Em linotipos e monotipos (do final do século XIX), já estavam incorporadas aos equipamentos desde os primeiros modelos comercializados. Isso demonstra que existiu, no estágio inicial do desenvolvimento dos computadores modernos, um distanciamento entre a codificação de caracteres e a notação/convenção tipográfica padrão, em uso então há cinco séculos. Isso se deve, obviamente, a limitações técnicas: nos primórdios dos computadores, uma das virtudes de se projetar uma codificação era conseguir obter o máximo de caracteres com o mínimo possível de pontos de código. Assim, nota-se o atraso considerável para que os computadores conseguissem acompanhar as convenções tipográficas tradicionais, algo que só começa a acontecer de fato com o surgimento dos microcomputadores e das linguagens PostScript e TrueType, na década de 1980.

Além das limitações do número de caracteres, um dos inconvenientes do padrão ASCII está relacionado ao método empregado para acentuar letras. Segundo descrições de Fischer (2000, p. 22), Haralambous (2007, p. 31) e Korpela (2006, p. 175), para compor um diacrítico sobre uma vogal, era necessário digitar a letra base, seguida do comando *backspace* e, por último, o diacrítico. Ou seja, um método pouco ortodoxo, diferente do que já se fazia, à época, num teclado de máquina de escrever, com teclas mortas. Além disso, não havia uma relação de correspondência direta entre os diacríticos que seriam reproduzidos (i.e. impressos, saída) e o que se encontrava na codificação e nos teclados (i.e. entrada). Por exemplo: para utilizar o acento agudo, era necessário digitar o apóstrofo <'>; para utilizar o trema, era necessário digitar as aspas duplas <">; já o acento grave, acento circunflexo e til tinham pontos de código definidos, com uma relação de correspondência direta de entrada e saída. A partir desses fatos, pode-se observar algumas questões.

Primeiro, que a concepção de tecla morta (*dead key*) como método de entrada para letras acentuadas ainda não estava incorporada aos métodos de entrada do ASCII, enquanto hoje tal recurso é amplamente difundido para a composição de letras acentuadas. São denominadas teclas mortas porque nada acontece no momento em que são tecladas: o efeito modificador da tecla não pode ser visualizado até que a próxima tecla (uma “viva”) seja digitada (HARALAMBOUS, 2007, p. 167). Por exemplo, em um teclado no qual todos os diacríticos têm a função de teclas mortas, ao se digitar a tecla correspondente ao acento agudo, só se visualizará o resultado desse diacrítico depois de se digitar alguma vogal.

Em segundo lugar, conforme apontado⁵² por Eric Fischer e Yannis Haralambous, as chances desse sistema funcionar, com resultados visualmente aceitáveis (proporções e alinhamentos corretos entre diacríticos e letras) eram pequenas. Os mesmos autores afirmam também que esse método de entrada dos diacríticos só era empregado em

52. As questões levantadas neste parágrafo foram apontadas por Eric Fischer e Yannis Haralambous, em correspondência pessoal por e-mail. Ver anexos.

composição de texto com fontes monoespaciaadas (i.e. fontes nas quais todos os caracteres têm a mesma largura, como nas máquinas de escrever mecânicas). Os exemplos de composição com diacríticos, gerados a partir de um sistema utilizando codificação ASCII são bastante raros. A Fig.54 mostra um espécime desse tipo de composição, em publicação do National Bureau of Standards.

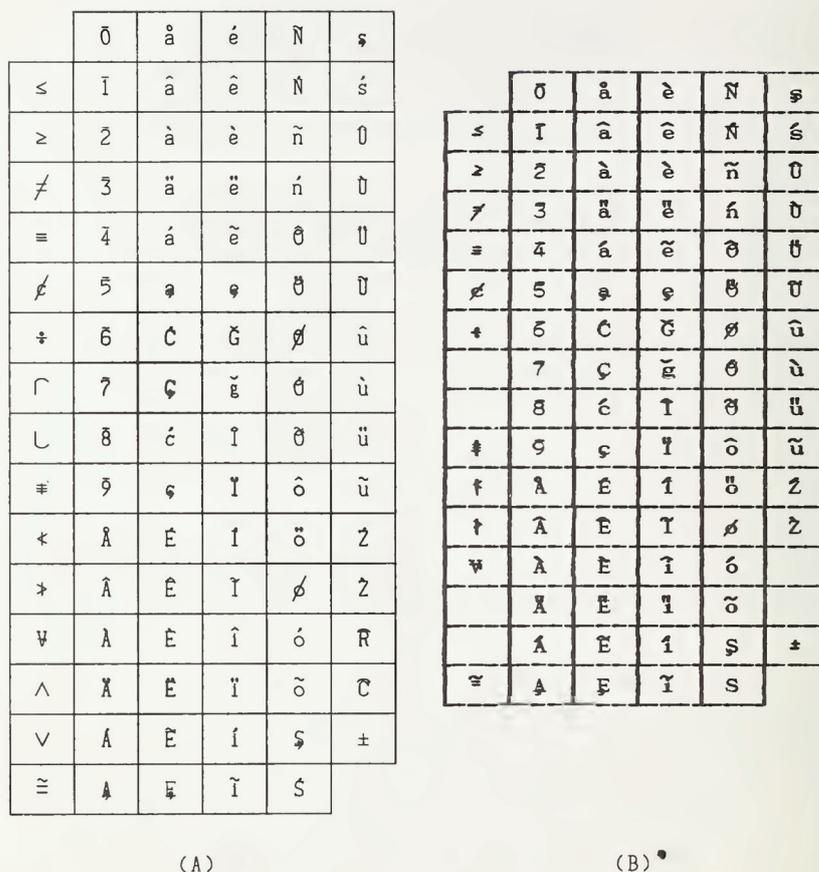


FIGURE 6

Composite symbols. Examples of new characters formed by overstriking pairs of characters from Figures 4 and 5.
a. Composites drawn by a computer driven incremental plotter (see Section 4.8) using the preferred 2:1 print window, Figure 4 characters.
b. Composites as produced on the NBS line printer showing the distortion caused by adoption of a 1.25:1 print window, Figure 5 characters.

Figura 54: Espécime de caracteres acentuados gerados a partir de um sistema utilizando codificação ASCII. Fonte: DUNCAN; GARVIN, 1974, p. 22.

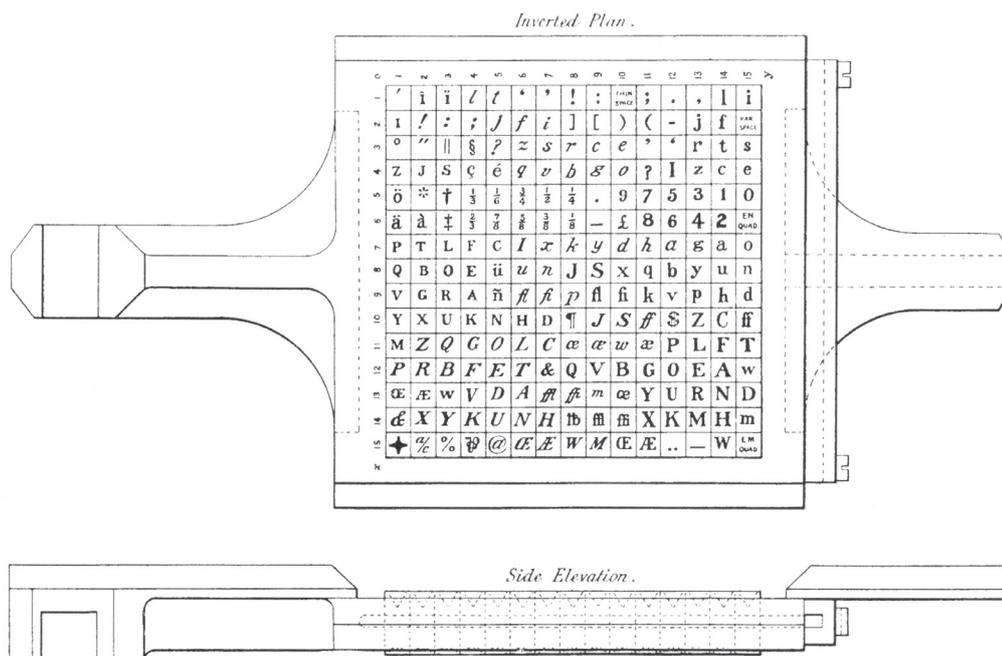
Haralambous observa⁵³ ainda que, nesse período, entre 1960 e 1980, as resoluções eram extremamente baixas: 144dpi (para impressos) era considerado um luxo. Destaca que ninguém jamais usaria caracteres acentuados por codificação ASCII em um sistema no qual se pudesse perceber a diferença entre um bom posicionamento dos diacríticos e um posicionamento ruim. Para esse tipo de coisa, não se usava ASCII, e sim equipamentos caros, com suas próprias codificações proprietárias. Nessa época, muito antes do Postscript, os dispositivos de saída tinham baixíssima resolução, e a renderização de letras acentuadas era tão ruim quanto o resto dos outros caracteres.

Outro ponto a se observar é a relação, no campo da tipografia, do que se entende como codificação (*encoding*) e como conjunto de caracteres (*character set*).

Em primeiro lugar, pode-se dizer que a noção de codificação é relativamente recente no meio tipográfico e, também, pouco utilizada. Designers de tipo referem-se muito mais ao jargão *conjunto de caracteres* do que ao termo codificação. Grosso modo, pode-se dizer que antes do advento da tipografia digital não há muita evidência de padrões e normas no campo da tipografia. Conjuntos de caracteres sempre foram personalizados e adaptados a clientes ou mercados específicos, para empresas etc.

Nesse sentido, Alice Savoie (2014, p. 61) observa que, enquanto a Linotype se popularizou no mercado de jornais, a Monotype encontrou sua clientela em mercados que exigiam maior refinamento tipográfico, como, por exemplo, o mercado editorial livreiro, o de periódicos acadêmicos de alta qualidade e o de composição complexa (e.g. tabelas, fórmulas matemáticas). Um dos fatores que proporcionavam ao sistema Monotype um maior refinamento está relacionado à possibilidade de personalização do estojo de matrizes do equipamento, permitindo aos clientes montarem conjuntos de 225 caracteres (15 × 15) altamente personalizáveis, de acordo com as suas necessidades.

53. Em correspondência pessoal por e-mail. Ver anexos.



17 An early Monotype matrix-case. It has fifteen rows of matrices, with fifteen matrices in each row.

Figura 55: Monotype, estojo de matrizes. Fonte: SOUTHALL, 2005, p. 37.

Richard Southall (2005, p. 36–37) explica como eram configurados os estojos de matrizes dos equipamentos da Monotype. As posições dos caracteres dentro do estojo de matrizes do monotipo eram baseadas nas larguras dos caracteres. Conforme se pode observar na Fig.55, as linhas de caracteres no estojo de matrizes são organizadas de acordo com a sua largura, em unidades correspondentes a múltiplos de 18. Assim, a primeira linha, ao topo, comporta os caracteres mais estreitos, medindo 5 unidades, ou 5/18. A última linha, abaixo, comporta os caracteres mais largos, todos medindo 18 unidades. Ou seja, não só é impossível traçar uma relação entre as configurações do estojo de matrizes com esquemas de codificação digital, como também pode-se dizer que essa forma de organizar as matrizes dentro do estojo teve pouca ou nenhuma influência nos esquemas de codificação em computadores.

É importante observar que o pós-guerra, período de transição em tecnologias de telecomunicações, é também um período de transição da tecnologia tipográfica. Na década de 1950 surgem os primeiros sistemas de fotocomposição. Utilizando matrizes óticas e

filme fotográfico para composição de texto, a fotocomposição foi pioneira na substituição de sistemas de composição mecânicos (i.e. linotipo, monotipo) por sistemas eletrônicos. Apesar de utilizar componentes analógicos (filme, matrizes óticas), estes eram controlados já por sistemas eletrônicos. Trata-se de um divisor de águas na história das artes gráficas, pois situa-se como elo de transição entre os sistemas mecânicos e os sistemas digitais (i.e. tipografia digital, editoração eletrônica) (SOUTHALL, 2005, p. 79–92).

Richard Southall (2005, p. 85) observa que, ainda na década de 1950, as matrizes de equipamentos de fotocomposição como Photon 200 e 500 comportavam até oito fileiras de 180 caracteres, totalizando 1440 caracteres por matriz. Essas fileiras eram organizadas em linhas concêntricas, num disco de vidro que agrupava todas as matrizes. Cada fileira agrupava dois estilos de uma fonte, cada um ocupando 90 caracteres. Era comum que os discos da Photon tivessem 16 fontes, ou 16 estilos de uma fonte, num mesmo disco.

De acordo com Southall (2005, p. 87) e Savoie (2014, p. 73), as matrizes dos equipamentos da Photon também eram personalizáveis, fabricadas de acordo com as demandas dos clientes. Isso significa que, além da possibilidade de personalizar o conjunto de caracteres, era possível personalizar o conjunto de fontes que cada disco continha. Era normal que na mesma fileira se acomodasse, por exemplo, as versões romana e itálica de uma determinada fonte.

8.2.2 Unicode

O objetivo desta seção é reunir algumas descrições sobre o modelo de codificação do Unicode.

Primeiramente, é importante mencionar que o Unicode possui uma documentação muito extensa. Por esta razão, esta seção levanta alguns pontos relevantes para esta tese. Naturalmente, há outros trechos neste trabalho com discussões diretamente relacionadas ao Unicode, que complementam esta seção.

Em segundo lugar, nota-se a carência de referências bibliográficas sobre o Unicode em português. Em inglês, além da publicação original (*The Unicode Standard*), se conhece algumas obras comentadas (e.g. Haralambous, Korpela), artigos, documentação, notas técnicas etc. Em português não foram encontradas referências, além de breves descrições (e.g. Wikipedia) e alguns textos na Internet. A busca pelo termo Unicode nas

bases de dados científicas (e.g. Scielo) também resultaram em poucos resultados, estes com pouca relação com o tema desta tese; comunicação, línguas indígenas, tipografia.

O projeto do Unicode começou com um objetivo principal de unificar as diferentes e conflitantes maneiras de se codificar caracteres, substituindo-as por um único padrão universal. Antes disso, havia uma proliferação de diferentes padrões de codificação, inconsistentes e incompletos, que atendiam a apenas uma pequena fração das línguas faladas no mundo, encarecendo o desenvolvimento de versões internacionais de softwares. Isso afetou, em particular, países em desenvolvimento (e suas minorias linguísticas), tendo em vista a inviabilidade econômica de se adaptar versões específicas de programas para mercados menores. Assim, o Unicode foi pensado como um método uniforme de codificação, mais eficiente e flexível do que os anteriores, satisfazendo as necessidades de computação técnica e multilíngue, com base num padrão uniforme (repertório grande o suficiente), eficiente (baseado em texto puro⁵⁴; simples de analisar) e inequívoco (um ponto de código representa sempre, e somente, o mesmo caractere) (UNICODE CONSORTIUM, 2020a, p. 4).

A descrição e designação de caracteres é apenas uma parte das especificações do Unicode. Além disso, há extensa documentação sobre processamento de texto, descrevendo, por exemplo, processos de quebra de linhas e divisão de palavras, formatação de números, datas e horários (adaptados à diferentes localidades), além de registros de métodos de entrada em sistemas de escritas complexos (UNICODE CONSORTIUM, 2020a, p. 6).

É interessante notar que, em comparação às codificações legadas, o Unicode é extremamente rico e detalhado. E por se tratar de uma documentação muito extensa, constarão nesta seção alguns dos aspectos mais relevantes do Unicode de forma resumida. Nesse sentido, destacam-se dois aspectos importantes: o modelo de codificação de caracteres e os princípios de design do Unicode. Além disso, serão detalhadas mais algumas propriedades/características, como a divisão em planos e blocos, como é constituído cada bloco, propriedades, categorias etc.

54. Tradução de *plain text*, i.e. texto sem formatação, texto simples. Itálico e negrito são exemplos de recursos utilizados na formatação de textos.

8.2.2.1 Modelo de codificação de caracteres do Unicode

Anteriormente (na seção terminologia e definições/codificação), foi dito que uma codificação de caracteres compreende, no mínimo, dois componentes: um conjunto de caracteres e algum sistema para representá-los em termos das unidades de processamento usadas no computador, i.e. números binários (CONSTABLE, 2001).

O modelo de codificação de caracteres (*character encoding model*) do Unicode é um pouco mais complexo e envolve quatro níveis de representação, conforme o *Unicode Technical Report #17: Unicode Character Encoding Model* (WHISTLER; DAVIS; FREYTAG, 2008). O relatório descreve um modelo para a estruturação de codificações de caracteres em quatro níveis⁵⁵:

- Repertório de caracteres abstratos (*Abstract Character Repertoire* – ACR): um conjunto não ordenado de caracteres a serem codificados, e.g. algum alfabeto ou conjunto de símbolos. Ou seja, um conjunto de descrições de caracteres, sem associação a pontos de código ou indicações de posição desses caracteres na tabela do Unicode.
- Conjunto de caracteres codificados (*Coded Character Set* – CCS): um repertório de caracteres abstratos, ao qual foram designados códigos, posições na tabela. Trata-se do mapeamento de um repertório de caracteres abstratos a designadores numéricos únicos (tipicamente números inteiros, na notação hexadecimal). Esses designadores numéricos são chamados de *ponto de código*, e a combinação de um *caractere abstrato* com seu ponto de código é um *caractere codificado*. Uma coleção de caracteres codificados em um determinado padrão é denominada página de código (*codepage*), i.e. codificação. Tipicamente, um padrão especificará nomes únicos (e.g. LATIN LETTER A), assim como pontos de código únicos.
- Formato⁵⁶ de codificação de caracteres (*Character Encoding Form* – CEF): É nesse nível que se começa a levar em consideração a representação real em um

55. Os quatro níveis desta lista são um resumo, com base nas descrições de Haralambous (2007, p. 62), Constable (2001) e Korpela (2006), além do próprio *Unicode Technical Report #17*.

56. Utilizou-se *formato* para a tradução do termo *form* neste contexto.

computador. Refere-se a um mapeamento dos pontos de código (em um conjunto de caracteres codificados) a sequências de comprimento fixo. Esses valores são conhecidos como *unidades de código*. Os valores mais comumente encontrados para unidades de código são 8, 16 e 32 bits.

- Esquema de codificação de caracteres (*Character Encoding Scheme – CES*): é uma representação de caracteres em bytes, sequências de oito bits; uma transformação reversível de um conjunto de sequências de unidades de código (de um ou mais CEFS para uma sequência serializada de bytes). Muitos sistemas para armazenamento e transmissão de dados operam em termos de bytes de 8 bits. Quando unidades de dados de 16 ou 32 bits são colocadas em um contexto de bytes de 8 bits, as unidades de dados podem ser facilmente divididas em blocos de 8 bits, pois seu tamanho é um múltiplo inteiro de 8 bits.

O quadro a seguir, conforme Korpela (2006, tradução nossa), fornece um resumo do modelo de codificação de caracteres:

NÍVEL DE CODIFICAÇÃO	EXPLICAÇÃO	EXEMPLO
Repertório de caracteres abstratos	Caracteres listados e descritos	“A” (Latin letter A)
Conjunto de caracteres codificados	Caracteres têm códigos	U+0041
Formato de codificação de caracteres	Sequências de unidade de código	00000041
Esquema de codificação de caracteres	Unidades de código mapeadas a sequências de octetos	00 00 00 41

8.2.2.2 Princípios de design do Unicode

O projeto (design) do Unicode se baseia em 10 princípios fundamentais, que nem sempre podem ser contemplados simultaneamente. O padrão procura encontrar um equilíbrio entre manter a consistência em prol da simplicidade e eficiência e manter a

compatibilidade para intercâmbio com os padrões existentes (UNICODE CONSORTIUM, 2020a, p. 14). Conforme colocado por Haralambous (2007, p. 58), num projeto da magnitude do Unicode, é importante definir princípios aos quais possamos, posteriormente, nos referir. Os princípios serão apresentados de maneira resumida, em um quadro com sua síntese (UNICODE CONSORTIUM, 2020a, p. 14, tradução nossa).

PRINCÍPIO	PROPOSIÇÃO
universalidade	O Unicode Standard provê um repertório único e universal.
eficiência	O texto [codificado] em Unicode é fácil de analisar e processar.
caracteres, não glifos	O Unicode Standard codifica caracteres, e não glifos.
semântica	Caracteres têm semântica bem definida
texto puro	Caracteres do Unicode representam texto puro.
ordem lógica	O padrão para representação de memória é a ordem lógica.
unificação	O Unicode Standard unifica caracteres duplicados em uma mesma escrita, em diferentes idiomas.
composição dinâmica	Formas acentuadas podem ser dinamicamente compostas.
estabilidade	Caracteres, uma vez definidos, não podem ser alterados, e as suas propriedades-chave são imutáveis.
convertibilidade	A conversibilidade precisa é garantida entre o padrão Unicode e outros padrões amplamente aceitos.

Mais à frente, alguns dos princípios relevantes para esta tese serão tratados com mais detalhe: texto puro, ordem lógica e composição dinâmica.

8.2.2.3 Espaço de codificação, blocos e planos

Colocando de forma simples, pode-se dizer que a estrutura elementar do Unicode é o caractere; e que os caracteres do Unicode estão organizados em um espaço de codificação. Dentro desse espaço, estão os planos. Dentro dos planos, os blocos. E dentro dos blocos, os caracteres.

Korpela (2006, p. 242) explica que o espaço de codificação (*codespace*) consiste em um intervalo de números inteiros que podem ser usados como designadores numéricos para caracteres. Numa codificação de oito bits, esse intervalo é de 0 a 255; numa codificação de dezesseis bits, o intervalo é de 0 a 65.535. O número total de designadores numéricos que um espaço de codificação comporta normalmente é contabilizado na notação decimal (e.g. 256 ou 65.536). No entanto, para numerar cada um desses caracteres, o Unicode usa a notação hexadecimal, i.e. os pontos de código usam a notação hexadecimal.

Nesse sentido, Korpela (2006, p. 242, tradução nossa) explica a organização do espaço de codificação do Unicode de uma maneira muito clara e simples:

Por razões práticas, o espaço de codificação foi dividido em partes chamadas de planos. Você pode visualizar um plano como uma enorme folha de papel com 65.536 (256 vezes 256) quadrados, cada um dos quais pode conter um caractere. Então imagine uma pilha de 17 dessas folhas. Lá você tem o espaço de codificação Unicode.

Grosso modo, cada um desses planos consiste em uma tabela gigantesca, equivalente a 16 bits ($2^{16} = 65.536$). Haralambous (2007, p. 70) explica que desses 17 planos, somente 6 são “povoados”, conforme se pode observar na Fig.56. O primeiro desses planos, correspondente aos primeiros 16 bits do Unicode, é denominado BMP (Basic Multilingual Plane) e abrange a maioria dos sistemas de escrita modernos, conforme ilustrado na Fig.57. Na prática, o BMP é o que é relevante para o contexto desta tese⁵⁷.

57. Por muitos anos, o Unicode consistiu apenas no BMP. Posteriormente, na ocasião da fusão com o padrão ISO 10646, foram criados os outros planos, que abrangem coisas muito específicas, como notação musical bizantina, símbolos matemáticos etc. O SIP (Supplementary Ideographic Plane) codifica e compila todos os novos caracteres ideográficos da escrita chinesa, adicionados anualmente (HARALAMBOUS, 2007, p. 70).

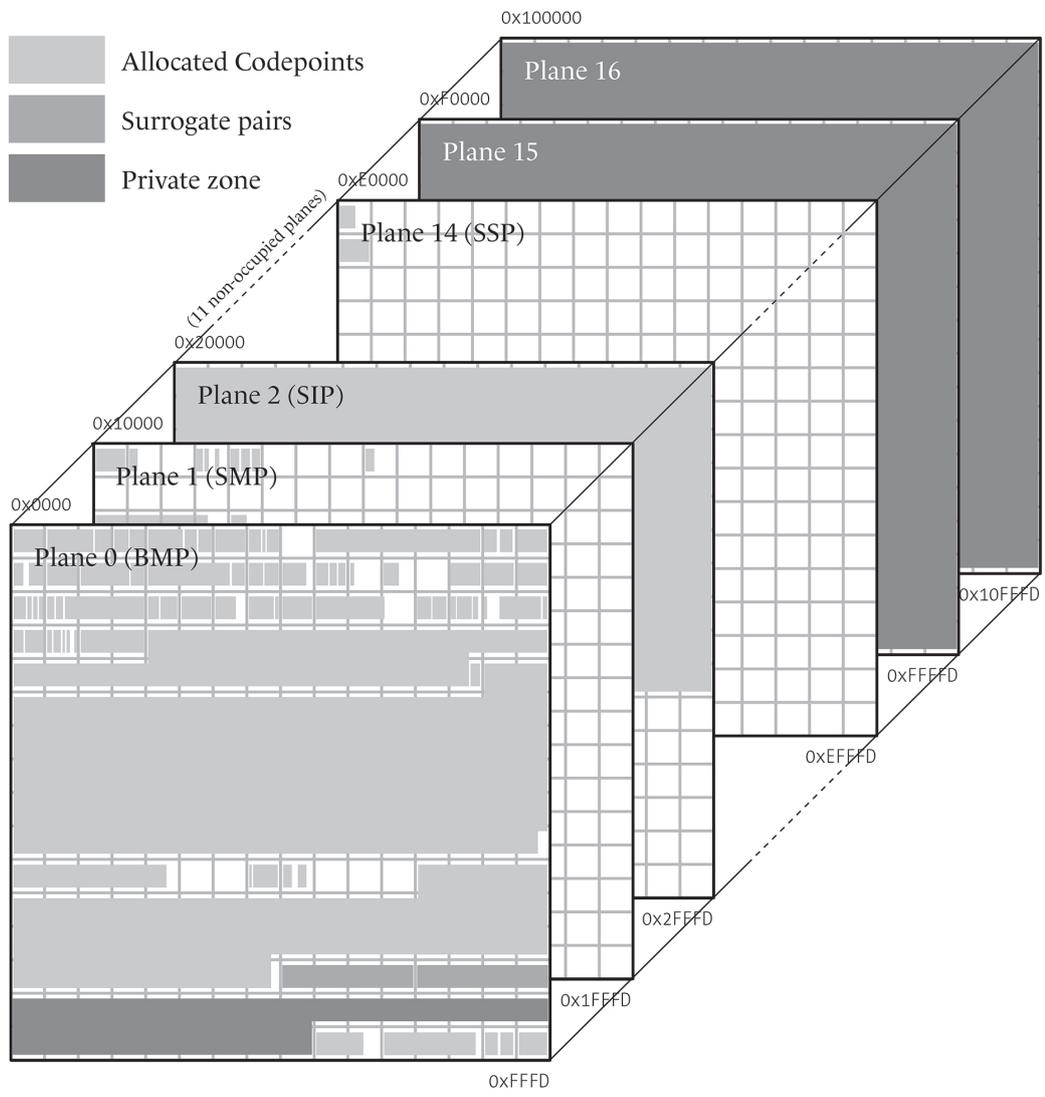


Figura 56: Planos do Unicode. Fonte: HARALAMBOUS, 2007, p.69.

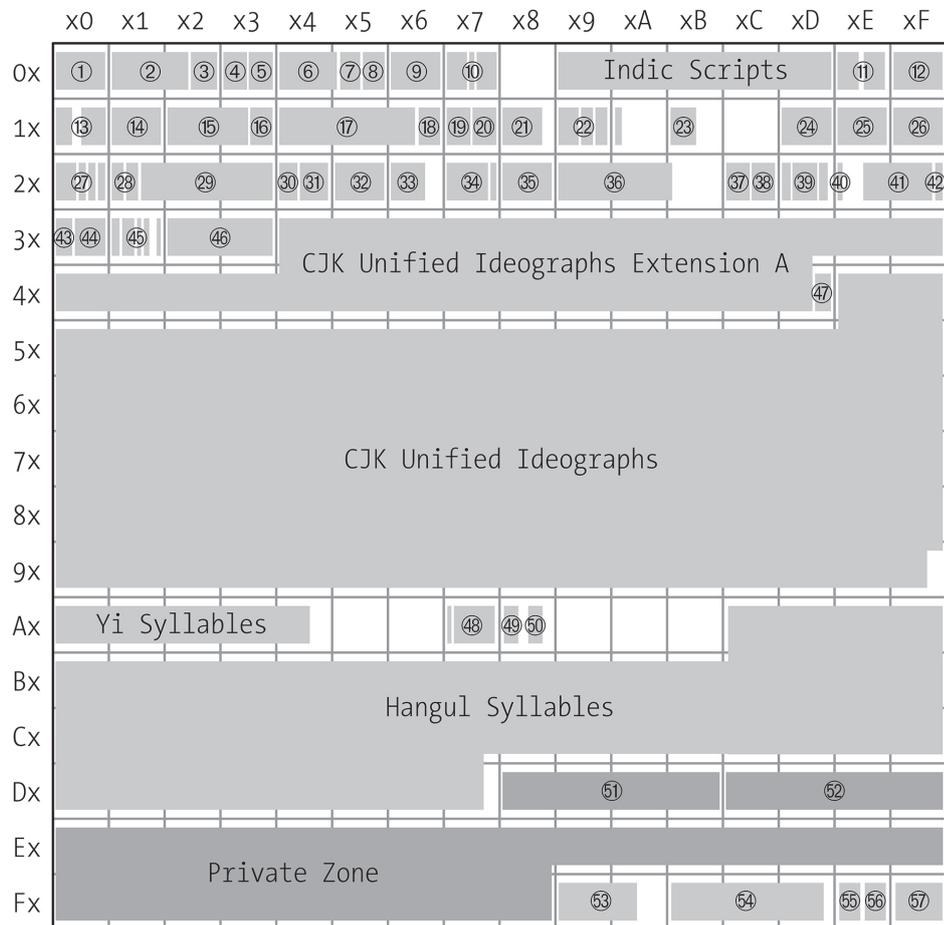


Figure 2-3: The roadmap of Unicode's Basic Multilingual Plane (BMP): ① ASCII and Latin 1, ② Latin Extended-A and -B, ③ phonetic alphabet and modifiers, ④ diacritical marks, ⑤ Greek (crippled by the monotonic reform) and Coptic, ⑥ Cyrillic, ⑦ Armenian, ⑧ Hebrew, ⑨ Arabic, ⑩ Syriac, Thaana and N'ko, ⑪ Thai and Lao, ⑫ Tibetan, ⑬ Myanmar and Georgian, ⑭ elements for forming hangul syllables, ⑮ Amharic, ⑯ Cherokee, ⑰ Canadian aboriginal scripts, ⑱ runes, ⑲ scripts of the Philippines, ⑳ Khmer; ㉑ Mongolian, ㉒ Limbu, Tai Le, etc., ㉓ Balinese, ㉔ phonetic extensions, ㉕ Latin Extended Additional, ㉖ Greek with accents and breathings (as it should be), ㉗ general punctuation, superscripts and subscripts, currency symbols, diacritical marks for symbols, ㉘ letterlike symbols, Roman numerals ㉙ arrows, mathematical and technical symbols, ㉚ graphic pictures for control codes, OCR, ㉛ enclosed alphanumerics, ㉜ geometric shapes, ㉝ miscellaneous symbols, ㉞ "Zapf dingbats", ㉟ braille, ㊱ supplemental arrows and mathematical symbols, ㊲ Glagolitic and Latin Extended-C, ㊳ Coptic disunified from Greek, ㊴ Khutsuri, Tifinagh and Ethiopic Extended, ㊵ Supplemental Punctuation, ㊶ ideographic radicals, ㊷ ideographic description characters, ㊸ ideographic punctuation, ㊹ kana, ㊺ bopomofo, hangul supplement, kanbun and CJK strokes, ㊻ enclosed ideographs and abbreviations used in ideographic writing systems, ㊼ Yijing hexagrams, ㊽ modified tone letters and Latin Extended-D, ㊾ Syloti Nagri, ㊿ Phags-pa, ㉀ high-half zone for surrogates, ㉁ low zone, ㉂ compatibility ideographs, ㉃ presentation forms A (Latin, Armenian, Hebrew, Arabic), ㉄ variation selectors and other oddities, ㉅ presentation forms B (Arabic), ㉆ full-width Latin letters and half-width katakana, specials.

Figura 57: Blocos do Unicode. Fonte: HARALAMBOUS, 2007, p.69.

Por uma questão de conveniência, os caracteres do Unicode estão agrupados por categorias linguísticas (e.g. grego, cirílico) e funcionais (e.g. pontuação, símbolos monetários). Mas por razões práticas, há desvios desse princípio geral, como os sinais de pontuação encontrados dentro do ASCII e mantidos lá, em vez de terem sido agrupados com outros sinais de pontuação no bloco correspondente. O agrupamento de caracteres (i.e. blocos) por escrita ou categoria funcional possibilita economia de espaço em implementações reais, como, por exemplo, em fontes tipográficas digitais (UNICODE CONSORTIUM, 2020a, p. 43).

Entender o espaço de codificação do Unicode é importante para saber como se dá a localização e organização dos caracteres. Os blocos consistem em grandes tabelas, com o número de caracteres sempre equivalente a um múltiplo de 16; logo, cada coluna da tabela tem sempre 16 linhas. Cada bloco é um intervalo de pontos de código de nome único. A numeração de bloco é contínua sempre começando por um múltiplo de 16, e os pontos de código não podem se sobrepor (UNICODE CONSORTIUM, 2020a, p. 88-89).

As figuras a seguir foram extraídas das tabelas de código (*code charts*) do Unicode⁵⁸. Nelas pode-se observar os dois primeiros blocos do BMP: Basic Latin (idêntico ao ASCII) e Latin-1 Supplement, ambos com 8 colunas, cada coluna com 16 linhas. Pode-se observar também os espaços reservados para caracteres de controle, nas duas primeiras colunas (Fig. 58 e 59).

58. Para maiores informações sobre como ler e interpretar a informação destas tabelas, consultar *Reading a Unicode block* (HARALAMBOUS, 2007, p. 120).

	000	001	002	003	004	005	006	007
0	NUL 0000	DLE 0010	SP 0020	0 0030	@ 0040	P 0050	` 0060	p 0070
1	SOH 0001	DC1 0011	! 0021	1 0031	A 0041	Q 0051	a 0061	q 0071
2	STX 0002	DC2 0012	" 0022	2 0032	B 0042	R 0052	b 0062	r 0072
3	ETX 0003	DC3 0013	# 0023	3 0033	C 0043	S 0053	c 0063	s 0073
4	EOT 0004	DC4 0014	\$ 0024	4 0034	D 0044	T 0054	d 0064	t 0074
5	ENQ 0005	NAK 0015	% 0025	5 0035	E 0045	U 0055	e 0065	u 0075
6	ACK 0006	SYN 0016	& 0026	6 0036	F 0046	V 0056	f 0066	v 0076
7	BEL 0007	ETB 0017	' 0027	7 0037	G 0047	W 0057	g 0067	w 0077
8	BS 0008	CAN 0018	(0028	8 0038	H 0048	X 0058	h 0068	x 0078
9	HT 0009	EM 0019) 0029	9 0039	I 0049	Y 0059	i 0069	y 0079
A	LF 000A	SUB 001A	* 002A	: 003A	J 004A	Z 005A	j 006A	z 007A
B	VT 000B	ESC 001B	+ 002B	; 003B	K 004B	[005B	k 006B	{ 007B
C	FF 000C	FS 001C	, 002C	< 003C	L 004C	\ 005C	l 006C	 007C
D	CR 000D	GS 001D	- 002D	= 003D	M 004D] 005D	m 006D	} 007D
E	SO 000E	RS 001E	. 002E	> 003E	N 004E	^ 005E	n 006E	~ 007E
F	SI 000F	US 001F	/ 002F	? 003F	O 004F	_ 005F	o 006F	DEL 007F

The Unicode Standard 13.0, Copyright © 1991-2020 Unicode, Inc. All rights reserved.

Figura 58: Tabela de código Co Controls and Basic Latin.

Fonte: UNICODE CONSORTIUM, 2020b, p.2.

	008	009	00A	00B	00C	00D	00E	00F
0	XXX 0080	DCS 0090	NB SP 00A0	◊ 00B0	À 00C0	Ð 00D0	à 00E0	ǎ 00F0
1	XXX 0081	PU1 0091	¡ 00A1	± 00B1	Á 00C1	Ñ 00D1	á 00E1	ñ 00F1
2	BPH 0082	PU2 0092	¢ 00A2	² 00B2	Â 00C2	Ò 00D2	â 00E2	ò 00F2
3	NBH 0083	STS 0093	£ 00A3	³ 00B3	Ã 00C3	Ó 00D3	ã 00E3	ó 00F3
4	IND 0084	CCH 0094	◊ 00A4	´ 00B4	Ä 00C4	Ô 00D4	ä 00E4	ô 00F4
5	NEL 0085	MW 0095	¥ 00A5	µ 00B5	Å 00C5	Õ 00D5	å 00E5	õ 00F5
6	SSA 0086	SPA 0096	¡ 00A6	¶ 00B6	Æ 00C6	Ö 00D6	æ 00E6	ö 00F6
7	ESA 0087	EPA 0097	§ 00A7	· 00B7	Ç 00C7	× 00D7	ç 00E7	÷ 00F7
8	HTS 0088	SOS 0098	¨ 00A8	¸ 00B8	È 00C8	Ø 00D8	è 00E8	ø 00F8
9	HTJ 0089	XXX 0099	© 00A9	¹ 00B9	É 00C9	Ù 00D9	é 00E9	ù 00F9
A	VTS 008A	SCI 009A	ª 00AA	º 00BA	Ê 00CA	Ú 00DA	ê 00EA	ú 00FA
B	PLD 008B	CSI 009B	« 00AB	» 00BB	Ë 00CB	Û 00DB	ë 00EB	û 00FB
C	PLU 008C	ST 009C	¬ 00AC	¼ 00BC	Ì 00CC	Ü 00DC	ì 00EC	ü 00FC
D	RI 008D	OSC 009D	® 00AD	½ 00BD	Í 00CD	Ý 00DD	í 00ED	ý 00FD
E	SS2 008E	PM 009E	® 00AE	¾ 00BE	Î 00CE	Þ 00DE	î 00EE	þ 00FE
F	SS3 008F	APC 009F	— 00AF	¿ 00BF	Ï 00CF	ß 00DF	ï 00EF	ÿ 00FF

The Unicode Standard 13.0, Copyright © 1991-2020 Unicode, Inc. All rights reserved.

Figura 59: Tabela de código C1 Controls and Latin-1 Supplement.

Fonte: UNICODE CONSORTIUM, 2020c, p.2.

8.2.2.4 Propriedades dos caracteres do Unicode

Esta seção consiste, basicamente, na listagem de algumas propriedades básicas do Unicode de forma resumida, conforme *Properties of Unicode characters*, de Haralambous (2007, p. 95–120).

Propriedades básicas

- Nome
- Bloco e escrita
- Idade

Categorias gerais

- Letras
- Marcas diacríticas
- Números
- Pontuação
- Símbolos
- Separadores
- Controle

Haralambous lista, ainda, outras propriedades gerais (não oficialmente listadas pelo Unicode), que possibilitam uma forma diferente de classificação e categorização:

- Espaços
- Caracteres alfabéticos
- Não-caracteres (*Noncharacters*)
- Caracteres ignoráveis (*Ignorable characters*)
- Caracteres desaprovaos (*Deprecated characters*)
- Exceções da ordem lógica
- *Soft-dotted letters*
- Caracteres matemáticos
- Aspas (*Quotation marks*)

- Hífens
- Pontuação terminal
- Diacríticos
- Extensores
- *Join control*

8.2.3 Como funciona a codificação de texto em computadores.

Em linhas gerais, esta seção tem como objetivo explicar o fluxo de informação no processamento de textos, desde processos básicos de codificação binária até a renderização. Incluiu-se também, nesta seção, explicações sobre alguns princípios do Unicode, como texto puro, ordem lógica e composição dinâmica.

8.2.3.1 Fluxo de informação no processamento de textos

Partiremos do princípio de que codificação se refere ao processo de representação de informação em alguma forma e do princípio de que um computador não armazena e *letras, numerais, imagens* e outras coisas; *bits* são a única coisa que computadores podem armazenar e processar. Um bit pode ter somente dois valores, como *sim* ou *não*, *verdadeiro* ou *falso*, *1* ou *0*, ou quaisquer outros termos que se queira usar para esses valores. Como computadores usam eletricidade, um bit é um pulso elétrico, que pode estar *ligado* ou *desligado*. Esses valores são usualmente representados como *1* e *0*. Para usar bits para representar alguma coisa, é preciso regras. Logo, para converter uma sequência de bits em algo como letras, números ou imagens, é preciso um esquema de codificação, como no exemplo da Fig.60, em que *01100010* significa **, *01101001* significa letra *<i>*, e assim por diante (ZENTGRAF, 2015, n.p.). Nesse caso, como cada sequência de oito números representa uma letra, configura-se uma codificação de oito bits.

01100010 01101001 01110100 01110011
b i t s

Figura 60: Codificação de uma sequência de bits em caracteres. Composto em Adelle Mono, de Veronika Burian e José Scaglione. Fonte: do autor.

Conforme explicado anteriormente, da associação entre um caractere abstrato (como a letra no exemplo acima) e um designador numérico único (um ponto de código), forma-se um caractere codificado. Uma coleção de caracteres codificados é chamada de página de código (*codepage*) ou codificação (*encoding*). Enquanto as codificações são uma parte “invisível” do processo de escrita em meios digitais, elas formam uma estrutura básica, um esqueleto conceitual para a construção de fontes tipográficas digitais, a parte visível. Nessa linha, Haralambous (2007, p. 1, tradução nossa) explica:

[...] fontes e codificações atendem a necessidades fundamentalmente diferentes. [...] elas formam uma dupla inseparável, como yin e yang, Céu e Terra, teoria e prática. [...] uma codificação emerge da tendência de conceituar informações; é o resultado de uma abstração, uma construção da mente. Uma fonte é um meio de representar visualmente a escrita, o resultado de uma expressão concreta, uma construção gráfica.

Uma codificação é uma tabela de caracteres – um caractere sendo uma entidade abstrata e intangível. Uma fonte é um recipiente de glifos, que são imagens, desenhos, marcas físicas de tinta preta em um fundo branco.

Nessa lógica, é importante entender que uma fonte não é somente uma coleção de glifos, organizados em uma tabela. Para que uma fonte funcione, é preciso que haja uma correspondência entre glifos e pontos de código. Em outras palavras, um computador não lê glifos, apenas caracteres codificados. A Fig.61 mostra uma captura de tela do editor de fontes Glyphs, mostrando como é visualizada a organização de um conjunto de caracteres em um arquivo de fonte, com cada caractere em seu compartimento. Pode-se observar, no extremo do canto inferior esquerdo, os detalhes do caractere

selecionado <a>, que tem o ponto de código U+0061 (marcado em azul claro). É dessa maneira que os desenhos de letra são codificados dentro de um arquivo de fonte, para que se tornem acessíveis nos programas e sistemas operacionais.

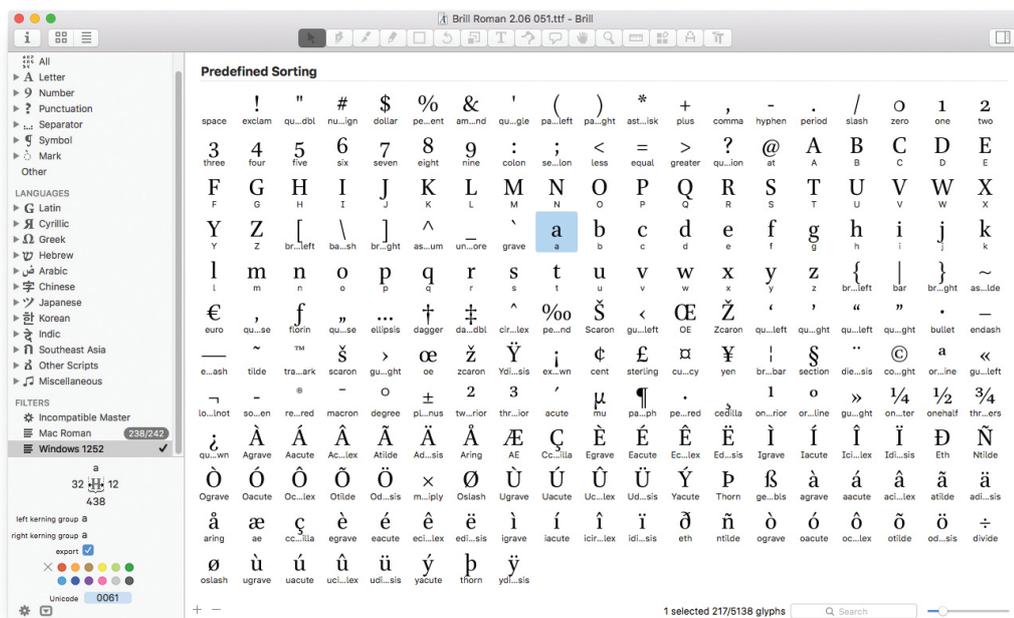


Figura 61: Captura de tela do software Glyphs. A seleção mostra o conjunto de caracteres que compõem a codificação Windows 1252. Fonte: do autor.

No livro *Como criar tipos: do esboço à tela* (2019, p. 114–115), José Scaglione explica:

[...] os computadores não sabem o que é uma letra A, ou B, ou C. Eles são dispositivos que armazenam e manipulam números, mas não conceitos linguísticos tão complexos quanto um fonema. Como funcionam, então? Sem entender o conceito de fonema e seu contexto linguístico, computadores armazenam informações sobre todos os caracteres de cada fonte na forma de códigos numéricos. Quando um usuário pressiona uma tecla, um código é enviado para o aplicativo de editoração eletrônica que está sendo utilizado. A informação é reencaminhada ao arquivo tipográfico (a fonte digital), solicitando que sejam retornados os conteúdos da caixa identificada pelo código. O aplicativo recebe de volta essa informação em formato digital, e, – às vezes com ajuda do sistema operacional – a interpreta e exhibe na tela.

Nesse sentido, a Fig.62 ilustra a citação do parágrafo acima, demonstrando o que acontece desde o momento em que o usuário digita uma letra (entrada) até o momento em que ele a visualiza na tela (saída).

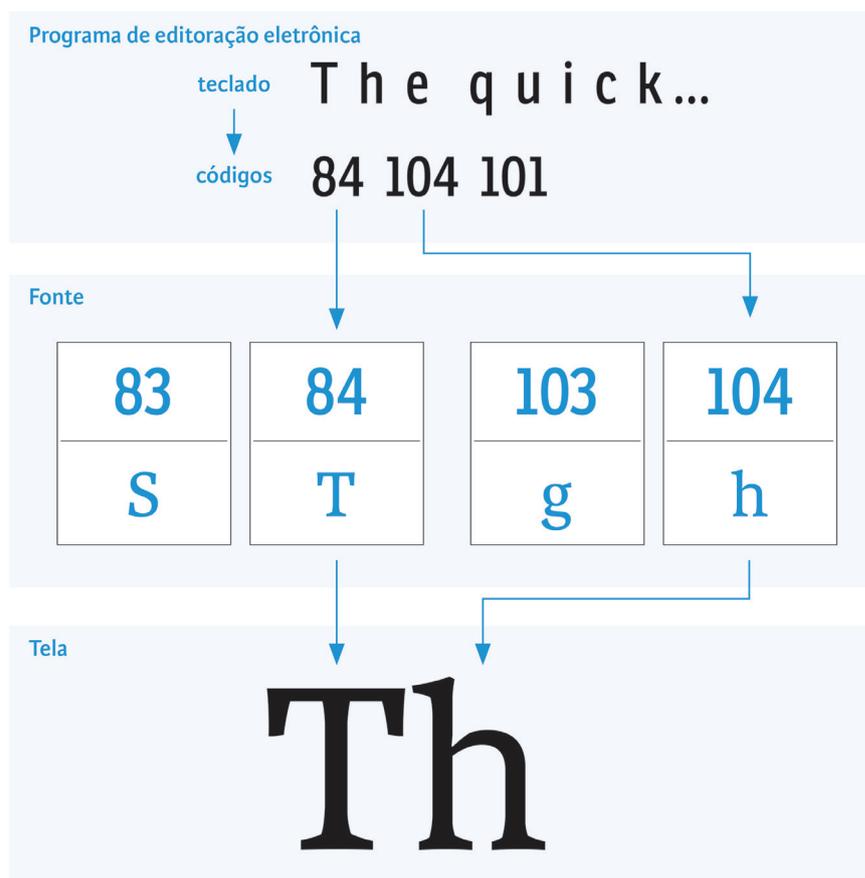


Figura 62: Processo simplificado de funcionamento da codificação de caracteres.

FONTE: SCAGLIONE, 2019, p. 115

A partir desse panorama geral do fluxo de informação no processamento de textos, as próximas seções têm como objetivo detalhar alguns desses processos relacionados aos métodos de entrada e renderização de caracteres.

8.2.3.2 Sequências de caracteres: texto puro e ordem lógica

Conforme sugerido por Jennings (2004, tradução nossa), na epígrafe da seção 8.2 *Codificação de caracteres*, a estrutura da composição de um texto configura um esquema

de comunicação serial, i.e. “um símbolo por vez, um após o outro, em uma seqüência combinada”. Essa descrição tem relação direta um dos princípios básicos do Unicode, a ordem lógica, que, por sua vez, se relaciona com o princípio do texto puro.

Em primeiro lugar, é importante entender o conceito de texto puro (*plain text*): uma seqüência de caracteres codificados, sem formatação. Texto formatado (*rich text*), por outro lado, é uma forma de representação textual baseada no texto puro, ao qual são adicionados atributos de formatação, como corpo (e.g. 12pt), cor (e.g. azul, preto), estilos (e.g. itálico, bold) etc.

Texto puro representa apenas o conteúdo dos caracteres, não sua aparência. Os caracteres, para se tornarem visíveis, terem uma aparência, dependem de um processo de renderização. Processos de renderização não são especificados pelo Unicode, que codifica apenas texto puro (UNICODE CONSORTIUM, 2020a, p. 19).

Nessa linha, Haralambous (2207, p. 59) explica que o Unicode só codifica texto (e não formatação). E no texto codificado em Unicode, a informação é construída somente com os próprios caracteres.

Na prática, trabalha-se no dia a dia com texto formatado. A formatação, no entanto, acontece em uma outra camada na estrutura de dados (atributos de formatação são aplicados em cima do texto puro) e é tratada como uma camada separada de dados, acima do nível de caractere e fora do escopo do padrão Unicode (KORPELA, 2006, p. 224).

Em segundo lugar, é importante entender o princípio da ordem lógica do Unicode, i.e. a ordem na qual o texto Unicode é armazenado na memória. Grosso modo, essa ordem corresponde à ordem em que o texto é digitado por meio do teclado. Em linhas gerais, também corresponde à ordem fonética, i.e. fonemas em série, um após o outro (UNICODE CONSORTIUM, 2020a, p. 19).

Na escrita latina, essa ordem lógica frequentemente corresponde a uma simples progressão linear de caracteres em uma única direção, da esquerda para a direita. Em outras escritas, como árabe e hebraico, o texto pode ser visualizado e/ou renderizado em uma ordem que difere de uma simples progressão linear. A Fig.63 mostra um exemplo de bidirecionalidade, ao se misturar caracteres das escritas hebraica e latina em uma mesma linha de texto, em que o texto em inglês se lê da esquerda para a direita,

e em hebraico, da direita para a esquerda. A direção de leitura predominante, dessa linha como um todo, é definida pelo primeiro caractere da sequência que, no caso, é da escrita latina (UNICODE CONSORTIUM, 2020a, p. 20).

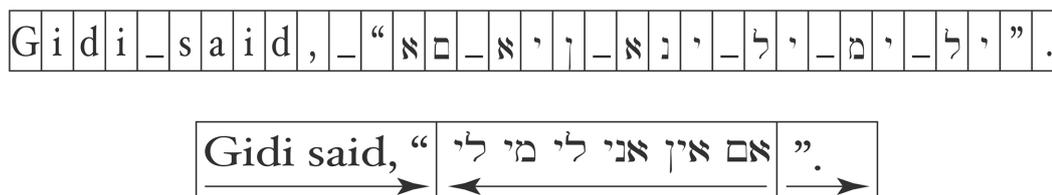


Figura 63: Ordenação bidirecional. A linha acima mostra a sequência de entrada na ordem lógica. Abaixo, a visualização do texto bidirecional. Fonte: UNICODE CONSORTIUM, 2020a, p. 20

Por último, é importante relacionar os princípios do texto puro e ordem lógica com a composição tipográfica, conforme observa John Hudson (2002, p. 31–32, tradução nossa):

O Unicode Standard não aborda diretamente os desafios da sofisticação tipográfica, mas ao prover uma robusta e inequívoca codificação de texto puro, fornece uma firme fundação até mesmo para a mais complexa tipografia. E o faz por meio da divisão entre codificação de texto e visualização (*display*) de texto. Em vez de tentar codificar todas as variantes tipográficas de caracteres – ligaduras, versaletes, *swash letters* [...] – que possam ser necessárias para composição de qualquer documento em qualquer escrita e língua, o Unicode codifica somente aqueles caracteres que são necessários para entrada e armazenamento de texto puro que possa subsequentemente ser exibido/visualizado utilizando-se muita ou pouca sofisticação tipográfica, conforme necessário. Para alguns sistemas de escrita, como no uso da escrita latina para a língua inglesa, essa codificação de texto é tipificada por uma relação de um para um (one-to-one) entre os caracteres armazenados e os mínimos critérios para a legibilidade humana. A língua inglesa não requer ligaduras ou versaletes para ser legível: precisa de tais recursos para ter uma boa aparência e articular seu significado de forma mais completa. Outros sistemas de escrita, como árabe e muitas escritas do sul da Ásia classificadas como índicas, requerem um processamento de alto nível para renderizar o texto armazenado com formas tipográficas apropriadas, para que sejam legíveis. Tais sistemas de escrita são comumente referidos, entre os especialistas em processamento de texto e internacionalização, como escritas complexas. [...]

Além das questões acima discutidas, o Unicode documenta ainda outras propriedades, menos relevantes para esta tese. A primeira delas diz respeito ao reordenamento de escritas complexas, que tem relação direta com a ordem lógica. Descreve também mais de uma dezena de propriedades de direcionalidade, também pouco relevantes para este contexto. Assim, a discussão a seguir procura abordar o funcionamento da ordem lógica na composição de caracteres acentuados, relevante, por sua vez, para a composição de textos em línguas indígenas.

8.2.3.3 *Diacríticos: métodos de entrada e características estruturais*

Além das questões acima mencionadas, é preciso observar também aspectos importantes acerca da composição de diacríticos, retomando conceitos da seção *Terminologia*, sobre composição estática e dinâmica. Esses dois esquemas têm formas diferentes de entrada, nos termos da ordem lógica. Ambos consistem numa progressão sequencial de caracteres, com algumas diferenças.

A ordem lógica para caracteres pré-compostos (composição estática), na prática corrente, consiste na digitação do diacrítico *antes* da letra base, prática esta utilizada desde as máquinas de escrever mecânicas. Quando o usuário digita a tecla morta do diacrítico <´ >, seguido pela tecla da letra base <a >, o sistema operacional ou aplicativo substitui essa combinação de dois caracteres por um único caractere pré-composto <á >. Por se tratar de um único caractere, formado pela união de dois glifos (diacrítico + letra base), tem-se somente um ponto de código. Por exemplo, o caractere <á > tem o código U+00E1 (ver as ilustrações da seção 8.1.9 *Diacríticos e marcas combinadas*).

A ordem lógica na composição dinâmica, segundo princípio normativo do Unicode (2020a, p. 109), indica a entrada da marca combinada **após** a letra base. Com base nesse princípio, sequências como <a >, <¨ >, <u > podem ser inequivocamente interpretadas como <äu >, e não <aü >. Outro exemplo: para a composição do g com til <ã >, a ordem lógica de entrada consiste na digitação da letra base <g >, seguida do til combinado <~ >. Diferentemente dos caracteres pré-compostos, essa combinação de letra base e marca combinada tem dois pontos de código. Do ponto de vista do Unicode, pode-se dizer que uma combinação como <ã >, formada por dois pontos de código, só

pode ser composta por meio da composição dinâmica. Por último, é importante mencionar que marcas combinadas não usam o recurso de teclas mortas (ver as ilustrações da seção 8.1.9 *Diacríticos e marcas combinadas*).

Outra característica dos métodos de entrada e codificação dos caracteres acentuados tem relação com a própria constituição dos caracteres e a forma como são classificados e agrupados no Unicode. Nesse sentido, há algumas propriedades relevantes a serem observadas, conforme descrições a seguir.

Composição estática e dinâmica fazem uso de caracteres de dois blocos distintos do Unicode: Letras Modificadoras Espaçadas (*Spacing Modifier Letters*) e Marcas Diacríticas Combinadas (*Combining Diacritical Marks*). Nesta tese, serão utilizados como sinônimos, respectivamente, os termos⁵⁹ marcas espaçadas (*spacing marks*), no contexto da composição estática, e marcas combinadas (*combining marks* ou *non-spacing marks*), para composição dinâmica. Como o próprio nome indica, uma das principais diferenças entre esses dois blocos está na especificação das métricas, i.e. quantidade de espaço em branco ao redor dos contornos dos glifos. Os parágrafos a seguir ilustram as particularidades de cada um desses blocos.

O Unicode Standard (2020a, p. 321) indica que o termo Letras Modificadoras (*Modifier Letters*) é empregado para designar modificadores: letras ou símbolos adjacentes a outros caracteres, que os modificam de alguma forma. Nos termos da ordem lógica, a modificação pode se aplicar ao caractere que precede o modificador ou ao caractere seguinte a ele. É comum que modificadores tenham função de diacrítico, indicando, por exemplo, uma mudança de pronúncia. Nesse contexto, a palavra *Letras* (do termo Letras Modificadoras) tem origem nas letras sobrescritas e subscritas utilizadas como modificadores em transcrições fonéticas (e.g. p^h). Apesar disso, a designação aplica-se não somente a letras, mas também (e principalmente) a diacríticos e outros símbolos, usados como modificadores no bloco *Spacing Modifier Letters* do Unicode.

59. *Spacing marks*, *combining marks* e *non-spacing marks* são termos correntes, utilizados internacionalmente na prática do desenho de tipos e produção de fontes.

	02B	02C	02D	02E	02F
0	h 02B0	?	˙	Y	^ 02F0
1	ĥ	ƒ	˘	l	<
2	j	<	˘	S	>
3	r	>	˘	X	o
4	ı	^	˘	ƒ	˘
5	ı	v	˘	┘	˘
6	Ɔ	^	+	┘	˘
7	w	v	-	┘	˘
8	y	'	˘	┘	:
9	/	-	•	┘	r
A	//	/	o	L	ı
B	‘	˘	˘	ı	L
C	’	ı	˘	v	J
D	˘	-	//	=	┘
E	˘	˘	˘	”	┘
F	˘	˘	X	v	←

The Unicode Standard 13.0, Copyright © 1991-2020 Unicode, Inc. All rights reserved.

Figura 64: Tabela de código Spacing Modifier Letters.

Fonte: UNICODE CONSORTIUM, 2020e, p.2.

	030	031	032	033	034	035	036
0	0300	0310	0320	0330	0340	0350	0360
1	0301	0311	0321	0331	0341	0351	0361
2	0302	0312	0322	0332	0342	0352	0362
3	0303	0313	0323	0333	0343	0353	0363
4	0304	0314	0324	0334	0344	0354	0364
5	0305	0315	0325	0335	0345	0355	0365
6	0306	0316	0326	0336	0346	0356	0366
7	0307	0317	0327	0337	0347	0357	0367
8	0308	0318	0328	0338	0348	0358	0368
9	0309	0319	0329	0339	0349	0359	0369
A	030A	031A	032A	033A	034A	035A	036A
B	030B	031B	032B	033B	034B	035B	036B
C	030C	031C	032C	033C	034C	035C	036C
D	030D	031D	032D	033D	034D	035D	036D
E	030E	031E	032E	033E	034E	035E	036E
F	030F	031F	032F	033F	034F	035F	036F

The Unicode Standard 13.0, Copyright © 1991-2020 Unicode, Inc. All rights reserved.

Figura 65: Tabela de código *Combining Diacritical Marks*.

Fonte: UNICODE CONSORTIUM, 2020d, p.2.

Em linhas gerais, numa sequência de caracteres em texto puro na escrita latina, diacríticos não aparecem isolados num texto. Eles são tipicamente associados ao caractere que os precede ou ao caractere que os sucede. Por exemplo, é difícil encontrar, num texto real, uma cedilha sozinha no meio de um texto ou um acento circunflexo isolado no meio de um parágrafo. São situações pouco prováveis, visto que a função desses caracteres é se associar às letras base.

É importante ressaltar uma característica estrutural dos caracteres do bloco *Spacing Modifier Letters*. Grosso modo, marcas espaçadas produzem um avanço horizontal (*advance width*), i.e. tem uma largura fixa. É como se marcas espaçadas estivessem inseridas numa “caixa”, e a largura dessa caixa fosse fixa, não permitindo que caracteres adjacentes se sobrepusessem a ela.

A Fig.66, a seguir, ilustra uma marca espaçada (ACUTE ACCENT, U+00B4). As linhas pretas definem a métrica dos caracteres, i.e. as larguras das “caixas”. As áreas delimitadas pelas linhas pontilhadas representam as proteções laterais (*sidebearings*) ou os espaços em branco ao redor do glifo. No exemplo a seguir, a largura total do caractere (*advance width*), de 237 unidades, é definida pela largura do próprio glifo (no caso, o acento agudo) mais a soma dos valores dos espaços em branco à esquerda (L = 50 unidades) e à direita (R = 40 unidades).

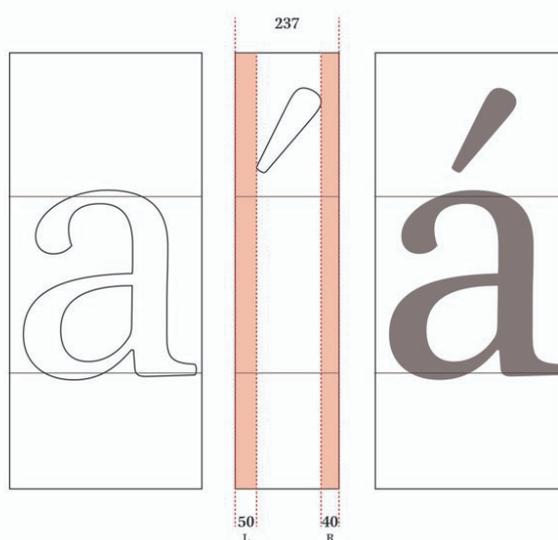


Figura 66: Métrica do acento agudo, U+00B4, do bloco *Spacing Modifier Letters*. Ilustração baseada em caracteres da fonte Brill Roman, de John Hudson. Fonte: do autor.

Marcas combinadas ou marcas não-espaçadas, por outro lado, funcionam de uma forma diferente. Em teoria (e como o próprio nome indica) *non-spacing marks* são marcas de largura igual a zero e que não produzem um avanço horizontal. Isso significa que os caracteres do bloco *Combining Diacritical Marks* devem ter valores negativos para as proteções laterais e um valor igual a zero para a largura total dos caracteres, conforme a Fig.67. É como se as duas linhas que definem a largura do caractere fossem comprimidas, das extremidades para o centro, até se encontrarem na largura zero, conforme simulação da Fig.68.

É importante observar que, em relação ao espaçamento das marcas, o Unicode define somente propriedades num nível mais abstrato. A métrica de um diacrítico (assim como de todos os outros caracteres) é de fato definida no design da fonte, e não pelas propriedades do Unicode.

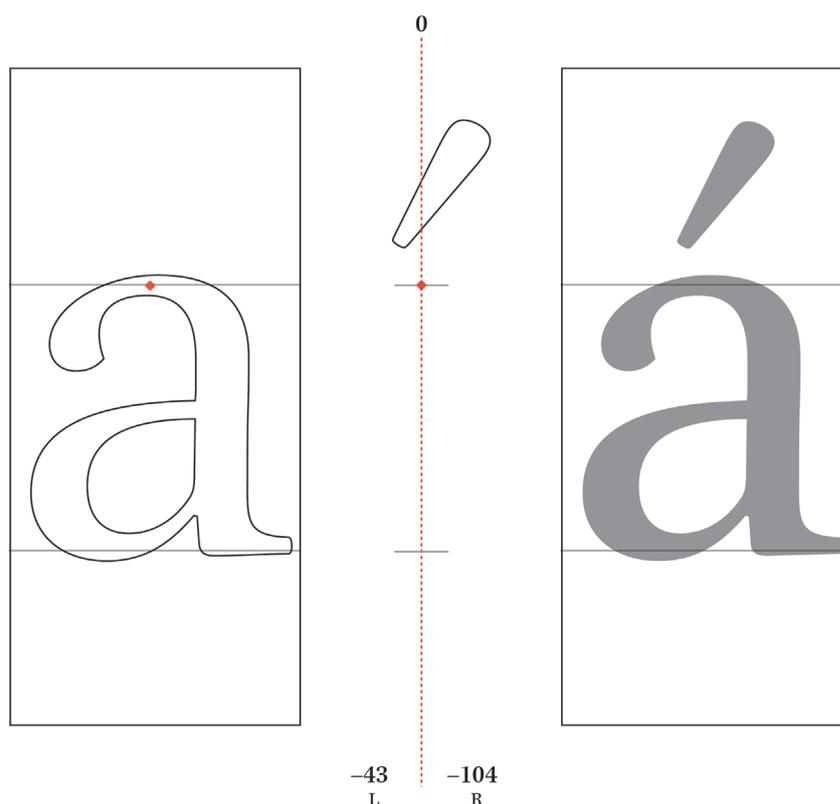


Figura 67: Métrica do acento agudo, U+0301, do bloco *Combining Diacritical Marks*. Ilustração baseada em caracteres da fonte Brill Roman, de John Hudson. Fonte: do autor.

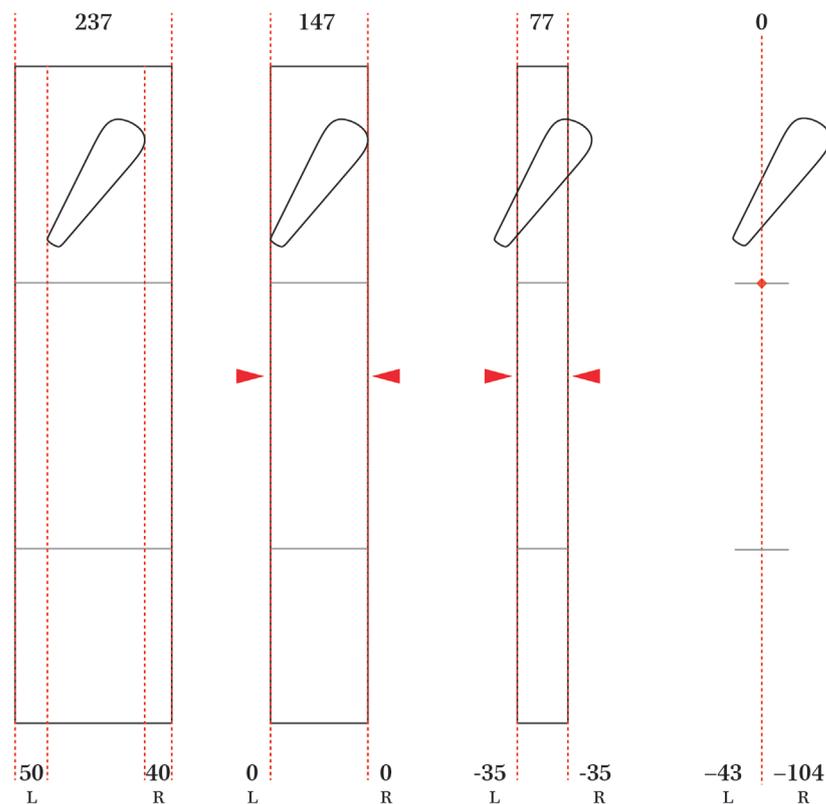


Figura 68: Simulação da diminuição progressiva dos valores das proteções laterais, até o zero. Ilustração baseada em caracteres da fonte Brill Roman, de John Hudson. Fonte: do autor.

O próprio Unicode (2020a, p. 108) indica que processos de renderização devem fazer uso de informações disponíveis sobre práticas tipográficas e convenções específicas para produzir a melhor renderização do texto.

Nesse sentido, para que uma determinada fonte produza o resultado desejado com marcas combinadas, é preciso que o sistema (i.e. dispositivo, programa ou sistema operacional) atenda a alguns requisitos. Entre estes, o principal é a compatibilidade com a especificação OpenType, que será explicada nas seções a seguir.

Uma outra característica importante das marcas combinadas é a presença de âncoras ou pontos de ancoragem (*anchors, attachment points*), representados na Fig.68 pelos pequenos pontos vermelhos à base do diacrítico e ao topo da letra <a>. A função das âncoras é demarcar uma coordenada para conexão das marcas às letras base, por meio de um recurso OpenType, mais bem explicado na seção 8.2.4 *OpenType*, a seguir.

8.2.3.4 *Decomposição*

O conceito de decomposição, de acordo com o Unicode (2020a, p. 114, tradução nossa), consiste em:

[...] uma sequência de um ou mais caracteres equivalente a um caractere decomponível. Uma decomposição completa de uma sequência de caracteres resulta da decomposição de cada um dos caracteres da sequência até que nenhum caractere possa ser mais decomposto.

Ainda segundo o Unicode (2020a, p. 114, tradução nossa):

Caractere decomponível: Um caractere que equivale a uma sequência de um ou mais outros caracteres, de acordo com os mapeamentos de decomposição encontrados no Banco de Dados de Caracteres Unicode [...]. Um caractere decomponível também é conhecido como caractere pré-composto ou caractere composto.

De maneira mais simples, Haralambous (2007, p. 130, tradução nossa) explica:

Caracteres decomponíveis são aqueles em que o glifo consiste em um glifo base e um certo número de marcas diacríticas. Para que um glifo seja decomponível, seus constituintes também precisam ser caracteres Unicode. [...] Praticamente todos os caracteres Unicode com marcas diacríticas são decomponíveis.

Ainda (UNICODE CONSORTIUM, 2020a, p. 62, tradução nossa):

Conceitualmente, uma decomposição implica a redução de um caractere a uma sequência equivalente de partes constituintes, tal como o mapeamento de um caractere acentuado a um caractere base, seguido por uma marca combinada. A grande maioria das decomposições não triviais são, na verdade, um mapeamento de um código de caractere para uma sequência de caracteres.

O Unicode documenta dois principais tipos de composição: canônica e equivalente. Haralambous (2007, p. 130, tradução nossa) explica que “uma decomposição canônica é, por definição, única, e sempre consiste em um ou dois caracteres” e que “quando uma decomposição canônica consiste em dois caracteres, o primeiro é um caractere base e o segundo, uma marca combinada.”

Haralambous (2007, p. 130, tradução nossa) explica ainda:

Há um motivo para chamar essa decomposição de “canônica”, pois [...] também identificamos como “canônica” a forma padrão de combinar caracteres básicos com caracteres combinados. Aplicando a decomposição canônica recursivamente a um caractere, obtemos um caractere base e uma sequência canônica de caracteres combinados.

Entende-se, portanto, que a ordem lógica funciona tanto na composição como na decomposição. Nesse sentido, segue-se um exemplo de como a decomposição é utilizada recursivamente, em rotinas do dia a dia: ao consultar o Google, é comum que se ignore os diacríticos nos termos da pesquisa. Logo, ao se digitar a palavra *zizek*, em vez de *Žižek*, obtém-se o mesmo resultado na pesquisa. Graças ao recurso de decomposição, um banco de dados indexado como o Google (que adota o padrão Unicode) é capaz de decompor o caractere <ž> em uma sequência de <z> + <ˇ> e ignorar os diacríticos na busca, retornando o mesmo resultado tanto para o termo *zizek* como para o termo *Žižek*.

Por último, é importante mencionar que o recurso de decomposição é utilizado também para identificar sequências equivalentes. Por exemplo: o caractere <ž>, mencionado acima, pode ser tanto um caractere pré-composto, com um ponto de código único (U+017E), como uma sequência de letra base e marca combinada, <z> + <ˇ>, com dois pontos de código.

8.2.4 OpenType

Em linhas gerais, esta seção tem como objetivo explicar o funcionamento do formato de fonte OpenType; particularmente, dos recursos OpenType (*OpenType features*), propriedades que descrevem funcionalidades especiais do formato OpenType. Esta seção tem como principal referência a própria especificação do formato, *OpenType Specification* (MICROSOFT CORPORATION; ADOBE INC., 2020, n.p.), em repositório mantido pela Microsoft, em inglês⁶⁰.

60. É importante mencionar que, assim como o Unicode, são escassas as referências sobre o OpenType em português.

8.2.4.1 *Open Type: visão geral*

Compatível com o Unicode e com amplo suporte linguístico e multiplataforma⁶¹, o formato de fonte OpenType foi desenvolvido como uma extensão do formato TrueType e oferece um amplo conjunto de recursos especiais (*OpenType features*). Os *OpenType features* possibilitam uma série de refinamentos tipográficos (Fig. 69 a 73). Alguns desses *features* são indispensáveis para composição complexa⁶². Muitas línguas indígenas brasileiras, particularmente aquelas que utilizam marcas combinadas e diacríticos empilhados, se enquadram nos parâmetros de composição complexa.

Os recursos OpenType (*OpenType features*) definem as funcionalidades de uma fonte e são o meio que os aplicativos usam para requisitar esses recursos. Os recursos são acessados/ativados por meio das tabelas de layout OpenType, ou *OpenType Layout tables*, tornando possível o mapeamento consistente entre caracteres⁶³ e glifos (permitindo ligaduras, formas alternativas e outras substituições), funções de manipulação do posicionamento dos glifos, e especificação do sistema de escrita e idioma (para que o aplicativo possa ajustar seu comportamento de acordo com escrita e idioma do texto). Além disso, é um formato aberto, que permite que os próprios desenvolvedores definam seus próprios *features* (MICROSOFT CORPORATION; ADOBE INC., 2020, n.p.).

61. Fontes OpenType são compatíveis com o Unicode, e uma fonte pode conter até 65,536 caracteres (16 bits) e acomodar mais de uma escrita, e.g. latina, grega e cirílica.

62. Aquelas que requerem um processamento de alto nível para renderização de texto com formas tipográficas apropriadas, e.g. árabe, devanagari etc. Em escritas simples, geralmente a relação direta entre caracteres e glifos em um texto é de um para um (HUDSON, 2002, p. 32).

63. É importante reforçar a distinção entre caracteres e glifos, central para a compreensão do conceito de recursos OpenType. Caractere é a menor componente da linguagem escrita com valor semântico e com um ponto de código, enquanto o glifo é a aparência gráfica de um caractere.

LATIN & PE FEATURES	DEACTIVATED	ACTIVATED
Turkish/Azeri/Crimean Tatar Language feature	fıncan, FİNCAN	fıncan, fıncan
Romanian/Moldavian Language feature	muṭumesc, MULȚUMESC	muṭumesc, MULȚUMESC
Ligatures	baffled financial office	baffled financial office
Small caps	Moskva Μόσχα Μοσκβα (67)	MOSKVA ΜΟΞΧΑ ΜΟСКΒΑ (67)
Case-sensitive punctuation	12-13 ¿Que?	12-13 ¿QUE?
Discretionary ligatures	direct distant cher státý	direct distant cher státý
Proportional lining figures Default style	—	01234567890 \$€£
Proportional oldstyle figures	01234567890	01234567890 \$€£
Tabular lining figures	01234567890	0 1 2 3 4 5 6 7 8 9 0 \$ € £
Tabular oldstyle figures	01234567890	0 1 2 3 4 5 6 7 8 9 0 \$ € £
Numerators/denominators	0123456789 / 0123456789	0123456789 / 0123456789
Fractions	2 3/4 + 193/541	2 3/4 + 193/541
Superiors/inferiors	H ²⁻³ (sup-вepx) Ηδεικτη	H ²⁻³ (sup-вepx) Ηδεικτη
Ordinals	18a 2o 1er	18 ^a 2 ^o 1 ^{er}
Long s Historical forms	Lost paradise Ossian	Loft paradife Offian
Not-descending J Stylistic set 1	Jordan TIJDSCHRIFT	Jordan TIJDSCHRIFT
Alternative forms of a Stylistic set 2	Stamford <i>Stamford</i>	Stamford <i>Stamford</i>
Alternative forms of g Stylistic set 3	dragon <i>dragon</i>	dragon <i>dragon</i>
Alt. forms of ampersand (&) Stylistic set 4	H&M c&A M&S I&Q	H&M c&A M&S I&Q
Alternative forms of at sign (@) Stylistic set 5	@rosettatype	@rosettatype
Alternative forms of K, R Stylistic set 14, uprights only	Karlův Říkonín	Karlův Říkonín

Figura 69: Skolar (type specimen). *OpenType features*. Fonte: BREZINA, 2019, p. 15.

Refinamentos

Ha038

Ha038 HA038 Ha⁰³⁸

numerais de caixa baixa, caixa alta, em versalete, sobrescritos e subscritos, etc. ...

st ct → *ſt cſ*

ligaduras discretionales

QR → **QR**

variantes estilísticas

a-c) → **A-C)**

reposicionamento de pontuação

Ebo → **EBO**

versaletes

Figura 70: *OpenType features*: refinamentos tipográficos. Fonte: SCAGLIONE, 2019, p. 119.

Substituições contextuais



Figura 71: “Substituições e ligaturas contextuais: de um glifo ou de um conjunto de glifos que têm uma forma alternativa de acordo com o lugar da palavra em que se encontram. [...] Cada letra do alfabeto adquire, automaticamente, uma forma diferente, de acordo com as letras que aparecem antes ou depois dela. Da esquerda para a direita: a letra Yeh em sua forma isolada, forma inicial, duas variantes da forma média (utilizada no meio de uma palavra) e forma final. Fonte: AwanZaman, de Mamoun Sakkal (árabe) e Juliet Shen (latino)” (SCAGLIONE, 2019, p. 118). Fonte: SCAGLIONE, 2019, p. 118.

Mudança de ordem, de posição ou de forma



Figura 72: “Reordenamento de caracteres: principalmente em sistemas de escrita da Índia, nos quais alguns glifos são deslocados para o início ou final de uma sílaba, mesmo quando inseridos através do teclado em sua ordem fonética natural. [...] Em sistemas de escrita complexos como o Devanagari, os caracteres são processados em uma ordem diferente da qual foram digitados. O I-Matra, por exemplo, é digitado ao final de uma sequência de caracteres, e automaticamente reposicionado ao início. Mas isso acontece somente a nível de reprodução, já que a sequência original permanece registrada no sistema ou aplicativo. Fonte: Katari Devanagari, de Erin McLaughlin.” (SCAGLIONE, 2019, p. 118). Fonte: SCAGLIONE, 2019, p. 118.

Bidirecionalidade



Figura 73: “Bidirecionalidade: o uso de diferentes sistemas de escrita com direções distintas no mesmo bloco de texto e de sistemas de escrita que são, eles mesmos, bidirecionais. Um exemplo é o árabe, no qual o fluxo de leitura é da direita para a esquerda, mas os numerais são escritos da esquerda para a direita. [...] Os sistemas de escrita com a direção da direita para a esquerda utilizam suporte de bidirecionalidade para suas ortografias. Neste caso, o fluxo de leitura do hebraico muda na metade do caminho, já que o ano deve ser escrito e lido da esquerda para a direita. Fonte: Noam, de Adi Stern.”(SCAGLIONE,2019, p. 118). Fonte: SCAGLIONE, 2019, p. 118.

Enquanto que, por meio dos *features*, os aplicativos requisitam as funcionalidades tipográficas avançadas, os *lookups*⁶⁴ são dados específicos da fonte (i.e. linhas de código) para implementar essas funcionalidades requisitadas pelos *features*. Os *lookups* descrevem ações de substituição ou posicionamento de um glifo que um aplicativo deve realizar para obter o efeito tipográfico desejado. *Lookups* fornecem os dados específicos da fonte usados para implementar esse recurso (MICROSOFT CORPORATION; ADOBE INC., 2020, n.p.).

Como complemento a essas definições iniciais, foram levantadas algumas considerações de John Hudson (2002, p. 30–32, tradução nossa) sobre relações entre codificação e tipografia, uma vez que a especificação OpenType foi pensada para ser usada em conjunto com outras especificações, como o Unicode:

Problemas de incompatibilidade entre diferentes plataformas, ambiguidade textual e suporte inadequado para muitos sistemas de escrita – problemas inerentes ao processamento de texto em oito bits – são exacerbados quando a intenção não é simplesmente a entrada, armazenamento e saída de texto, mas sua

64. Tabela de consulta, ou tabela de pesquisa (*lookup table*), é uma matriz (grupo ou agrupamento) que substitui tempo [de execução] de computação por uma operação de indexação de matriz mais simples.

articulação, utilizando um sistema tipográfico sofisticado. Desenvolvedores de fontes nos primórdios do desktop publishing reconheceram a necessidade de suporte a elementos tradicionais do design tipográfico, como ligaturas, versaltes e letras sobrescritas, mas foram forçados a realizar isso dentro dos limites de um modelo de codificação de caracteres e processamento de texto que se baseava em conjuntos de caracteres de oito bits. [...] O Unicode Standard não aborda diretamente os desafios da sofisticação tipográfica, mas ao prover uma robusta e inequívoca codificação de texto puro, fornece uma firme fundação até mesmo para a mais complexa tipografia. E o faz por meio da divisão entre codificação de texto e visualização de texto. Em vez de tentar codificar todas as variantes tipográficas de caracteres – ligaturas, versaltes, *swash letters*, contextual positioning forms, etc. – que possam ser necessárias para composição de qualquer documento em qualquer escrita e língua, o Unicode codifica somente aqueles caracteres que são necessários para entrada e armazenamento de texto puro que possa subsequentemente ser exibido/visualizado utilizando-se muita ou pouca sofisticação tipográfica, conforme necessário. Para alguns sistemas de escrita, como no uso da escrita latina para a língua inglesa, essa codificação de texto é tipificada por uma relação de um para um (*one-to-one*) entre os caracteres armazenados e os mínimos critérios para a legibilidade humana. A língua inglesa não requer ligaturas ou versaltes para ser legível: precisa de tais recursos para ter uma boa aparência e articular seu significado de forma mais completa. Outros sistemas de escrita, como árabe e muitas escritas do sul da Ásia classificadas como Índicas, requerem um processamento de alto nível para renderizar o texto armazenado com formas tipográficas apropriadas, para que sejam legíveis. Tais sistemas de escrita são comumente referidos, entre os especialistas em processamento de texto e internacionalização, como escritas complexas.

A partir dessas considerações de Hudson, a próxima seção procura explicar um pouco melhor a arquitetura do formato OpenType.

8.2.4.2 Arquitetura do formato OpenType

A arquitetura do formato OpenType utiliza a mesma estrutura tabular de uma fonte *SFNT* do formato TrueType (a especificação OpenType foi desenvolvida como uma extensão do TrueType). Isso quer dizer, em linhas gerais, que o formato *SFNT* é uma coleção de tabelas individuais e que cada uma dessas tabelas carrega um tipo de informação diferente. Por exemplo, a tabela *glyf* carrega informações sobre os contornos vetoriais de cada glifo da fonte; *kern* carrega informações sobre pares de kerning; *cmap* mapeia glifos a seus respectivos pontos de código, e assim por diante. Assim, os aplicativos podem acessar

informações de diferentes tabelas, conforme a necessidade. Acessando a tabela *cmap*, por exemplo, o aplicativo sabe qual glifo deve exibir na tela quando determinado caractere é digitado no teclado. A partir da estrutura tabular do *SFNT*, é relativamente fácil ampliar as capacidades do formato, adicionando novas tabelas⁶⁵. Nessa linha, é importante entender que, entre a variedade de tabelas descritas pela especificação OpenType, duas tabelas de layout OpenType são importantes para a discussão sobre recursos tipográficos avançados (*OpenType features*): *GSUB* e *GPOS*. Essas tabelas contêm informação para, respectivamente, substituição e posicionamento de glifos, que podem ser lidas por aplicativos compatíveis com o OpenType (HUDSON, 2002, p. 36–37).

A sequência de figuras⁶⁶ a seguir (Fig. 74 a 77) tem como objetivo ilustrar essa estrutura do formato OpenType. Nessa linha, Miguel Sousa (2010, p. 3–6) explica que uma fonte é feita de tabelas; *GSUB* e *GPOS* são exemplos delas; as tabelas *GSUB* e *GPOS* contêm *features*; os *features* contêm lookups.

65. Tanto a Apple como a Microsoft/Adobe utilizaram a mesma estratégia para ampliação das capacidades do formato TrueType. Com base na estrutura tabular do *SFNT*, desenvolveram, respectivamente, as especificações AAT (Apple Advanced Typography) e OpenType (HUDSON, 2002, p. 36–37; HARALAMBOUS, 2007, p. 15).

66. Extraído de *OpenType Feature File Syntax*, documento apresentado por Miguel Sousa, desenvolvedor da Adobe, em workshop na conferência da ATypI em Dublin, em 2010.

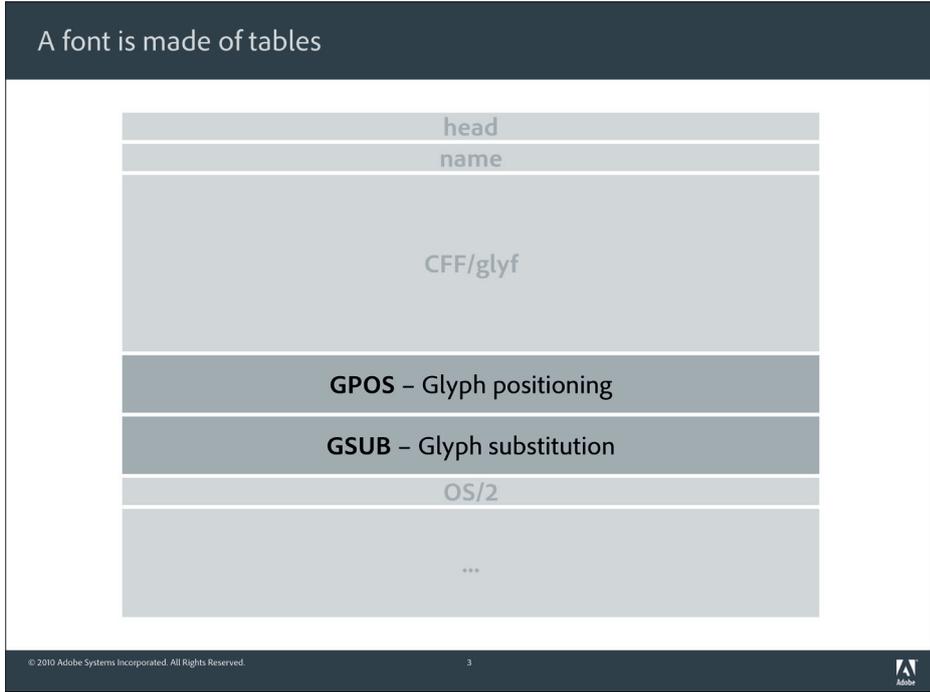


Figura 74: Uma fonte é feita de tabelas. Fonte: SOUSA, 2010, p.3.

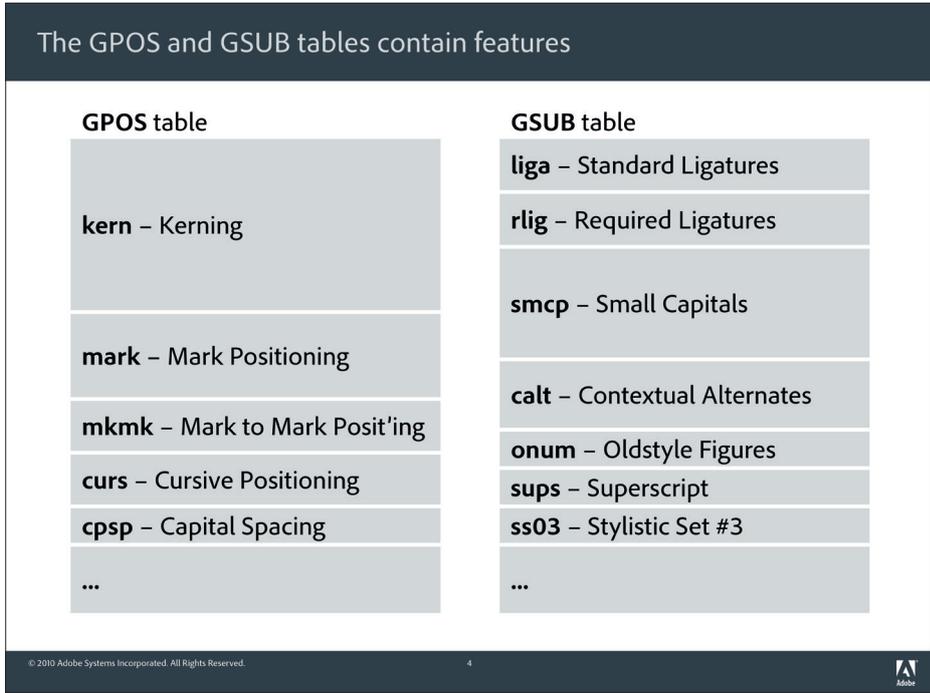


Figura 75: As tabelas GSUB e GPOS contém features. Fonte: SOUSA, 2010, p.4.

The features contain lookups

kern feature

```
pos A V -80;
pos L T -55;
pos @T @a -67;
...
```

smcp feature

```
sub a by a.sc;
sub b by b.sc;
sub @lowercase by @smallcap;
...
```

mark feature

```
pos base x <anchor 125 530>
  mark @above;

pos ligature f_l <anchor 40 0>
  mark @above ligComponent
  <anchor 90 5> mark @above;
...
```

liga feature

```
sub f f i by f_f_i;
sub f i by f_i;
...
```

salt feature

```
sub k from [ k.alt k.alt2 ];
...
```

© 2010 Adobe Systems Incorporated. All Rights Reserved. 5

Figura 76: Os *features* contém *lookups*. Fonte: SOUSA, 2010, p.5.

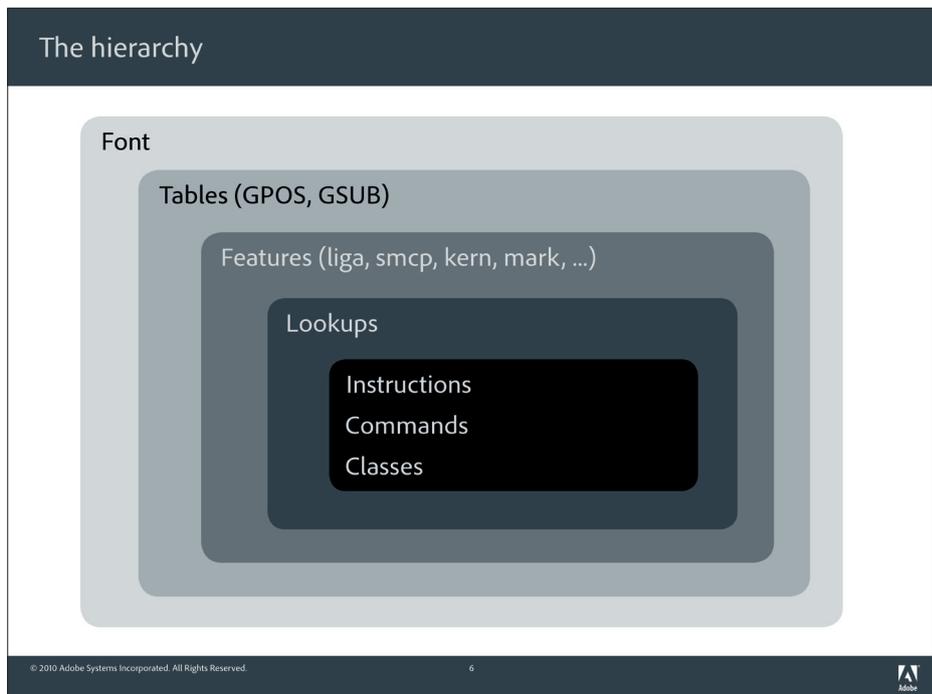


Figura 77: Hierarquia de organização das tabelas na fonte OpenType. Fonte: SOUSA, 2010, p.6.

Por último, é importante entender que muitos *features* são pensados para contextos linguísticos específicos, e um idioma (*language system*) pode modificar as funções ou aparência de glifos em uma escrita para representar um idioma em particular. Nesse sentido, para resultados precisos em determinados contextos linguísticos, existe uma lógica de processamento linguístico no layout OpenType, que segue uma sequência hierárquica: sistema de escrita, idioma (*language system*), *feature* e *lookup*, conforme a Fig.78 a seguir. Logo, o processamento primeiro identifica o sistema de escrita⁶⁷, depois o idioma, para somente depois acessar os *features* e seus respectivos *lookups*.

As próximas seções procuram ilustrar o funcionamento, respectivamente, das tabelas de layout GSUB e GPOS, e seus respectivos *features* e *lookups*.

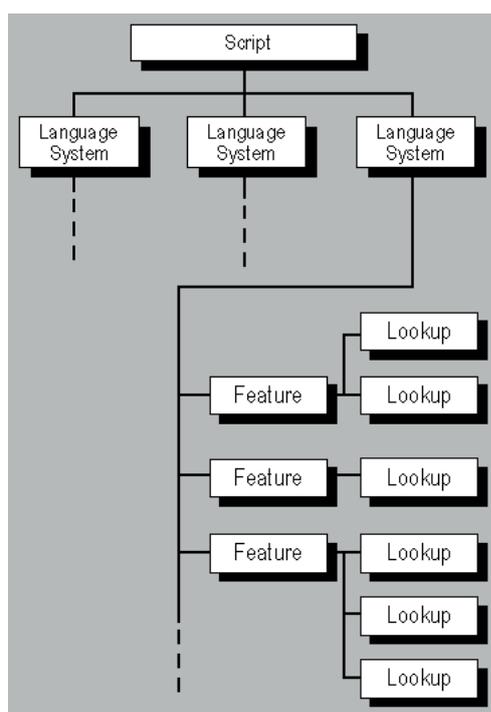


Figura 78: Fluxo do processamento linguístico no OpenType.

Fonte: MICROSOFT CORPORATION; ADOBE INC., 2020, n.p.

67. Com base na tabela *ScriptList*, que identifica sistemas de escrita em uma fonte e de idiomas (*language systems*) que empregam aquele sistema de escrita (MICROSOFT CORPORATION; ADOBE INC., 2020).

8.2.4.3 Tabela de substituição GSUB

A tabela de substituição de glifos (GSUB, *glyph substitution table*), como o próprio nome diz, contém informações sobre substituições de glifo. Muitas escritas e idiomas dependem desses mecanismos de substituição para uma correta renderização; é o caso da escrita árabe, em que a forma do glifo que representa um determinado caractere varia de acordo com sua posição em uma palavra, conforme ilustra a Fig.70 (Substituições e ligaturas contextuais; SCAGLIONE, 2019, p. 118). As substituições de glifo também são um recurso técnico utilizado para ampliar a paleta de glifos, proporcionando uma série de refinamentos tipográficos, como ligaturas, variantes estilísticas, versaletes e outros recursos, conforme ilustrado nas Fig .77 e 78 (BREZINA, 2019, p. 15; SCAGLIONE, 2019, p. 119) acima (MICROSOFT CORPORATION; ADOBE INC., 2020, n.p.).

John Hudson (2002, p. 32) observa que, no Unicode, a divisão entre a codificação e a exibição do texto é expressa pela distinção entre caracteres de glifos. Nas práticas tipográficas da composição manual e da composição mecânica, era típico o uso do termo caractere como referência a qualquer elemento tipográfico individual (um único tipo; uma única peça de metal). Na medida em que a composição tipográfica se casa com o processamento de textos em computador, o termo caractere assume um significado mais preciso, fazendo referência a um elemento de texto codificado.

O OpenType se baseia no padrão de codificação Unicode (que assume distinção entre caracteres e glifos): o texto é codificado como uma sequência de caracteres, e a tabela ‘cmap’ mapeia cada caractere a um único glifo padrão. No entanto, vários caracteres não são mapeados a um único glifo: como no caso do árabe, que pode apresentar quatro formas de glifo diferentes para um mesmo caractere: forma isolada, forma inicial, forma média e forma final. No caso de ligaturas, ocorre o oposto: dois pontos de código diferentes são substituídos por um único glifo. A tabela GSUB fornece uma maneira de descrever essas substituições, para que os aplicativos possam realizar as substituições de glifo em operações de composição e formatação tipográfica e, consequentemente, sua renderização correta (MICROSOFT CORPORATION; ADOBE INC., 2020, n.p.).

Para acessar glifos substitutos, a tabela GSUB usa um esquema de indexação que parte de um *único glifo default*, e os substitutos são acessados a partir do índice desse

glifo default. Por exemplo, se uma fonte tem três variantes do caractere <&>, a tabela ‘cmap’ associará o ponto de código do caractere <&> somente a um dos três glifos. Nas tabelas GSUB, os índices das outras variações do glifo <&> são então referenciados a partir desse índice default (MICROSOFT CORPORATION; ADOBE INC., 2020, n.p.).

Processadores de texto usam os dados das tabelas GSUB para gerenciar as ações de substituições de glifos. A tabela GSUB identifica os glifos que são referência para a entrada e saída em cada substituição, especifica como e onde o processador usa glifos substitutos, e regula a ordem das operações de substituição de glifo (MICROSOFT CORPORATION; ADOBE INC., 2020, n.p.).

A tabela GSUB dá suporte a sete tipos de substituições, conforme a Fig.79 a seguir:

<i>GSUB LookupType Enumeration</i>		
Value	Type	Description
1	Single (format 1.1 1.2)	Replace one glyph with one glyph
2	Multiple (format 2.1)	Replace one glyph with more than one glyph
3	Alternate (format 3.1)	Replace one glyph with one of many glyphs
4	Ligature (format 4.1)	Replace multiple glyphs with one glyph
5	Context (format 5.1 5.2 5.3)	Replace one or more glyphs in context
6	Chaining Context (format 6.1 6.2 6.3)	Replace one or more glyphs in chained context
7	Extension Substitution (format 7.1)	Extension mechanism for other substitutions (i.e. this excludes the Extension type substitution itself)
8	Reverse chaining context single (format 8.1)	Applied in reverse order, replace single glyph in chaining context
9+	Reserved	For future use (set to zero)

Figura 79: Variações de *lookups* de substituição. Fonte: MICROSOFT CORP.; ADOBE INC., 2020

As figuras a seguir (Fig. 80 a 8I) ilustram o funcionamento de alguns tipos de *lookups* de substituição, conforme informado por Miguel Sousa (2010, p. 7–14).

GSUB Lookup Types

- **Type 2 – Multiple substitution** (Replace one glyph with more than one glyph)
`sub <glyph> by <glyph sequence>;`

`sub ampersand by a n d;` **& → and**
`sub f_f_i by f f i;` **ffi → ffi**

© 2010 Adobe Systems Incorporated. All Rights Reserved. ii 

Figura 80: GSUB Lookup. Tipo 1.
Substituição simples (substitui um glifo por outro glifo).
Fonte: SOUSA, 2010, p.7.

GSUB Lookup Types

- **Type 2 – Multiple substitution** (Replace one glyph with more than one glyph)
`sub <glyph> by <glyph sequence>;`

`sub ampersand by a n d;` **& → and**
`sub f_f_i by f f i;` **ffi → ffi**

© 2010 Adobe Systems Incorporated. All Rights Reserved. ii 

Figura 81: GSUB Lookup. Tipo 2.
Substituição múltipla (substitui um glifo por mais de um glifo).
Fonte: SOUSA, 2010, p.11.

GSUB Lookup Types

- **Type 3 – Alternate substitution** (Replace one glyph with one of many glyphs)
`sub <glyph> from <glyph class>;`

`sub ampersand from [ampersand.alt ampersand.alt2 ampersand.alt3];`

& : & & &

`@FleuronAlternates = [fleuron.alt fleuron.alt2 fleuron.alt3];`

`sub fleuron from @FleuronAlternates;`

♥ : ♣ ♠ ♡

© 2010 Adobe Systems Incorporated. All Rights Reserved. 12

Figura 82: GSUB Lookup. Tipo 3.
 Substituição de alternativos (substitui um glifo, por um, entre muitos glifos).
 Fonte: SOUSA, 2010, p.12.

GSUB Lookup Types

- **Type 4 – Ligature substitution** (Replace multiple glyphs with one glyph)
`sub <glyph sequence> by <glyph>;`

`sub f f i by f_f_i;` ffi → ffi

`sub f i by f_i;` fi → fi

`sub s p by s_p;` sp → sp

`sub s t by s_t;` st → st

The order of the substitution commands is very important. The statements with longer glyph sequences should precede the shorter ones.

© 2010 Adobe Systems Incorporated. All Rights Reserved. 13

Figura 83: GSUB Lookup. Tipo 4.
 Substituição de ligatura (substitui múltiplos glifos por um único glifo).
 Fonte: SOUSA, 2010, p.13.

Entre os recursos acima listados, o mais relevante para este trabalho é o *lookup ccmp*, que executa rotinas de composição e decomposição. Os recursos de substituição, aplicados às línguas indígenas brasileiras, serão melhor demonstrados no capítulo *Desenvolvimento*.

8.2.4.4 GPOS

A tabela de posicionamento de glifos (GPOS) contém informações sobre o controle e posicionamento de glifos nas coordenadas X e Y. Fornece dados para lidar com o ajuste de um único glifo, ajuste de pares de glifos, posicionamento de marcas e posicionamento contextual de glifos (MICROSOFT CORPORATION; ADOBE INC., 2020, n.p.).

Enquanto na maioria das línguas que empregam a escrita latina predomina o uso de lookups de substituição, muitas línguas indígenas brasileiras (particularmente aquelas que utilizam marcas combinadas e diacríticos empilhados) requerem o uso de funções GPOS para um posicionamento controlado dos diacríticos.

As fontes OpenType permitem controle e flexibilidade para o posicionamento de um único glifo e para o posicionamento de vários glifos, uns em relação aos outros. É possível ajustar com precisão a posição de um glifo por meio das coordenadas nos eixos X e Y definidas pela tabela GPOS, e também com o recurso do uso de âncoras (*attachment points*).

A tabela GPOS dá suporte a oito tipos de substituições de ações de posicionamento e ancoragem, conforme a Fig.84 a seguir:

The GPOS table supports eight types of actions for positioning and attaching glyphs:

- A *single adjustment* positions one glyph, such as a superscript or subscript.
- A *pair adjustment* positions two glyphs with respect to one another. Kerning is an example of pair adjustment.
- A *cursive attachment* describes cursive scripts and other glyphs that are connected with attachment points when rendered.
- A *mark-to-base attachment* positions combining marks with respect to base glyphs, as when positioning vowels, diacritical marks, or tone marks in Arabic, Hebrew, and Vietnamese.
- A *mark-to-ligature attachment* positions combining marks with respect to ligature glyphs. Because ligatures may have multiple points for attaching marks, the font developer needs to associate each mark with one of the ligature glyph's components.
- A *mark-to-mark attachment* positions one mark relative to another, as when positioning tone marks with respect to vowel diacritical marks in Vietnamese.
- *Contextual positioning* describes how to position one or more glyphs in context, within an identifiable sequence of specific glyphs, glyph classes, or varied sets of glyphs. One or more positioning operations may be performed on "input" context sequences. Figure 4e illustrates a context for positioning adjustments.
- *Chaining contextual positioning* describes how to position one or more glyphs in a chained context, within an identifiable sequence of specific glyphs, glyph classes, or varied sets of glyphs. One or more positioning operations may be performed on "input" context sequences.

Figura 84: Lookups de posicionamento.

Fonte: MICROSOFT CORPORATION; ADOBE INC., 2020, n.p.

Entre os recursos acima listados, dois são mais relevantes para este trabalho: o *lookup* do tipo *Mark-to-base positioning* (posicionamento de marca em relação à base), que conecta uma marca combinada a um glifo base, e o *lookup* do tipo *Mark-to-mark positioning* (posicionamento de marca em relação a outra marca), que conecta uma marca combinada a outra marca. As imagens a seguir (Fig. 85 a 86) ilustram o funcionamento desses dois *lookups* de posicionamento, conforme informado por Miguel Sousa (2010, p. 30–32).

Os recursos de posicionamento, aplicados às línguas indígenas brasileiras, serão melhor demonstrados no capítulo *Desenvolvimento*.

GPOS Lookup Types

- **Type 4 – Mark-to-base positioning** (Attach a combining mark to a base glyph)
`markClass <mark glyph or glyphclass> <anchor> @<mark class name>;`
`pos base <base glyph or glyphclass> <anchor> mark @<mark class name>;`

```
markClass [ringcmb caroncmb] <anchor 0 530> @MarksAbove;  ◦  ˘
markClass dotbelowcmb <anchor 0 -80> @MarksBelow;          ˙  +
feature mark { # Mark Positioning
  pos base x <anchor 240 530> mark @MarksAbove             ˘
  <anchor 225 -80> mark @MarksBelow;                       ˙
} mark;
```

X + ◦ + ˙ → X̂

© 2010 Adobe Systems Incorporated. All Rights Reserved.

30



Figura 85: GPOS Lookup. Tipo 4. Posicionamento de marca em relação à base. Conecta uma marca combinada a um glifo base. Fonte: SOUSA, 2010, p.30.

GPOS Lookup Types

- **Type 6 – Mark-to-mark positioning** (Attach a comb. mark to another mark)
`pos mark <base mark glyph or glyphclass>`
`<anchor> mark @<mark class name>;`

```
markClass [damma fatha] <anchor 0 1000> @MarksAbove;  ˘  ˘
feature mkmk { # Mark to Mark Positioning
  pos mark damma <anchor -30 1200> mark @MarksAbove;  ˘
} mkmk;
```

ل + ˘ + ˘ + | + ˘ → لّ

© 2010 Adobe Systems Incorporated. All Rights Reserved.

32



Figura 86: GPOS Lookup. Tipo 6. Posicionamento de marca em relação a outra marca. Conecta uma marca combinada a outra marca. Fonte: SOUSA, 2010, p.32.

8.2.5 Mecanismo de modelagem de texto (*text-shaping engine*)

Esta seção explica o funcionamento de mecanismos de modelagem de texto (*text-shaping engine*), conforme Esfahbod (2021, n.p., tradução nossa), desenvolvedor do mecanismo de modelagem de código aberto HarfBuzz:

Modelagem de texto (*text shaping*) é o processo de traduzir uma sequência de códigos de caracteres (como pontos de código Unicode) em uma sequência de glifos devidamente organizada que pode ser renderizada em uma tela ou em um formato de saída final para inclusão em um documento.

O processo de modelagem depende da sequência de entrada, da fonte ativa, da escrita (ou sistema de escrita) em que a sequência está e do idioma em que a sequência está.

Os sistemas de software modernos geralmente lidam apenas com sequências no esquema de codificação Unicode (embora os sistemas e documentos legados possam envolver outras codificações).

Existem vários formatos de fonte que um programa pode encontrar, cada um dos quais possui um conjunto de regras de modelagem de texto padrão.

O formato dominante é OpenType. A especificação OpenType define uma série de modelos de modelagem para várias escritas de todo o mundo. Esses modelos de modelagem dependem de a fonte incorporar certos recursos como pesquisas em suas tabelas GSUB e GPOS.

[...]

As sequências de texto geralmente são marcadas com uma escrita e uma tag de idioma que fornecem o contexto necessário para executar a formatação correta do texto. A escrita e as tags de idioma necessários são definidos pelo OpenType.

8.2.6 Teclados e métodos de entrada

Esta seção tem como objetivo explicar o funcionamento do teclado como método de entrada para composição de texto em computadores. Existem dois principais tipos de teclado: os físicos, que são dispositivos externos que podem ser conectados ao computador (ou embutidos neles, no caso dos laptops); e os virtuais, que funcionam por meio de outras interfaces, como nos telefones celulares, que utilizam interface de toque ou *touchscreen* (TECLADO, 2021, n.p.). Este texto se concentra, principalmente, em teclados para computadores desktop e laptop.

Os teclados físicos podem apresentar de 90 a 130 teclas, que podem ser classificadas em algumas categorias (ver Fig.88). Destaca-se aqui as funções de duas categorias, utilizadas para composição de textos: as alfanuméricas (letras, números, pontuação, diacríticos etc.) e as modificadoras (que alteram a função de outras teclas como *Shift*, *Alt* e *Ctrl*). Tipicamente, uma tecla modificadora é pressionada enquanto se digita outra tecla (KEYBOARD LAYOUT, 2021, n.p.; TECLADO, 2021, n.p.).

Sobre as teclas modificadoras, deve-se reforçar aqui que podem ser utilizadas para teclas mortas (*dead keys*) na composição de diacríticos. São assim chamadas porque nada acontece no momento em que são tecladas: o efeito modificador da tecla não pode ser visualizado até que a próxima tecla (uma “viva”) seja digitada (HARALAMBOUS, 2007, p. 167). No caso das marcas combinadas, não se utiliza teclas mortas (mais informações em *Diacríticos: métodos de entrada e características estruturais*, p. 158).

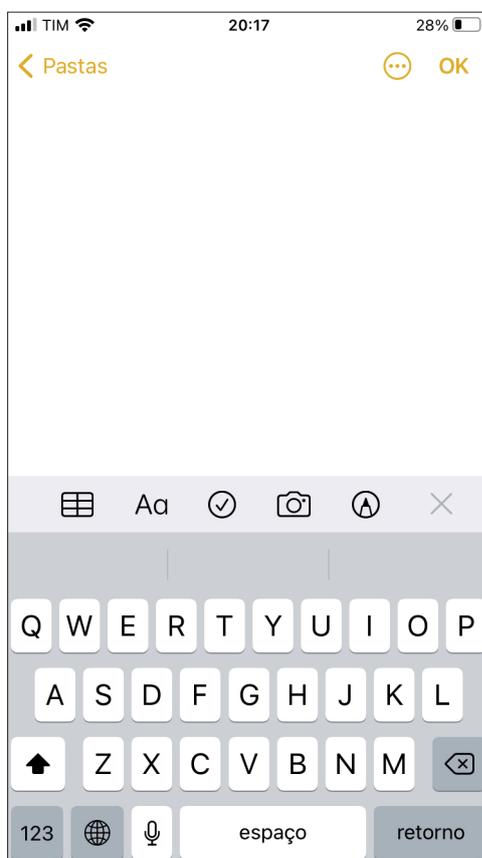


Figura 87: Teclado virtual no iOS 14.6. Fonte: do autor.

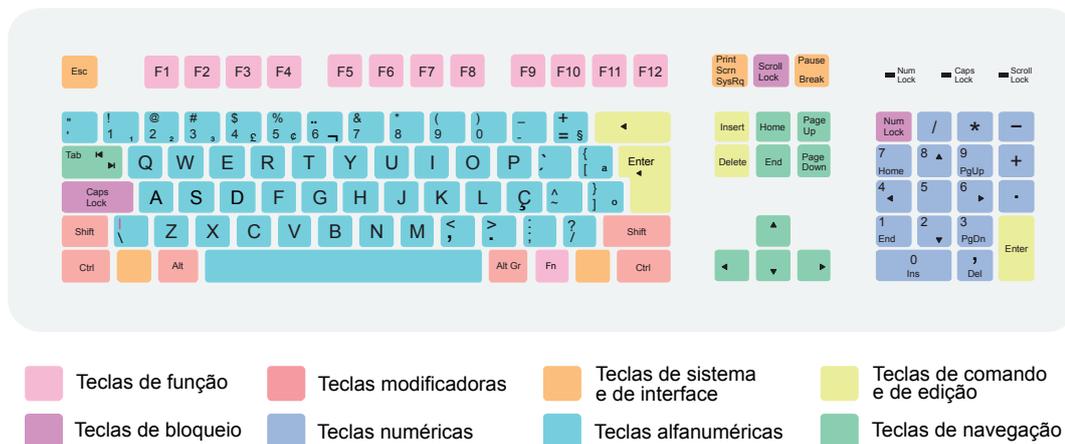


Figura 88: Layout de teclado ABNT. Fonte: TECLADO, 2021.



Figura 89: Layout de teclado ANSI. Fonte: KEYBOARD LAYOUT, 2021.

As teclas de um teclado podem ser organizadas de diferentes formas. A esse arranjo das teclas se dá o nome layout de teclado (*keyboard layout*). Um layout de teclado consiste em um arranjo físico, visual ou funcional específico das teclas, legendas ou associações de significado (respectivamente) de um teclado. O layout físico consiste no posicionamento das teclas de um teclado. O layout visual é o arranjo das legendas (rótulos, marcações) que aparecem nessas teclas. O layout funcional é o esquema de mapeamento dos pontos de código nas teclas de um teclado, associando significado a cada uma delas, i.e. mapeamento entre as teclas físicas e os caracteres a ela associados (KEYBOARD LAYOUT, 2021, n.p.).

Nesse sentido, entende-se que o layout funcional é um programa, é o *driver*⁶⁸ do teclado. É o software que mapeia um ponto de código a cada tecla. Em geral, o termo *layout de teclado* é utilizado como sinônimo de *layout funcional*. Nesta tese, usaremos essa acepção do termo.

Os sistemas operacionais trabalham com algumas configurações padrão de layout de teclado. Por exemplo, um computador com teclado físico no padrão ABNT provavelmente terá um layout de teclado ABNT associado a ele. Esse layout pode ser alterado nas configurações de sistema. Por exemplo, pode-se substituir o layout de teclado brasileiro padrão ABNT pelos layouts padrão utilizados nos Estados Unidos, conhecidos como ANSI ou U.S. International. Pode-se ver a diferença desses dois layouts nas figuras 88 e 89. Nesta tese utiliza-se esses dois layouts (ABNT e ANSI) como as principais referências.

Por último, observa-se que é possível criar versões personalizadas de teclados, adaptadas a necessidades de diferentes línguas. Por exemplo, muitas das línguas indígenas brasileiras utilizam diacríticos e/ou combinações de caracteres que não estão acessíveis por meio dos layouts ABNT e ANSI. Existem alguns programas para criação de teclados, e.g. Ukelele (para MacOS) e Microsoft Keyboard Layout Creator (para Windows). Esse é um ponto melhor discutido nos capítulos *Metodologia e Desenvolvimento* desta tese.

68. Um *driver* é um programa que contém as funções a serem integradas a um sistema operacional para controlar um determinado periférico.

O objetivo das seções *Terminologia e Codificação e Linguagem* é fornecer um arcabouço técnico e teórico para a análise da tipografia das línguas indígenas brasileiras no capítulo *Características tipográficas das línguas indígenas brasileiras*; este último, separado dos dois primeiros, por algumas razões.

Primeiro porque é importante esclarecer as questões terminológicas e os conceitos relativos à codificação de texto necessárias para o entendimento da codificação e tipografia das línguas indígenas brasileiras. Destaca-se que os trechos *Terminologia e Codificação e Linguagem* representam apenas um recorte da documentação técnica, de referências como o padrão Unicode, e a especificação OpenType, imprescindível para a compreensão do capítulo *Características tipográficas das línguas indígenas brasileiras*.

Em segundo lugar, é importante entender que os dois capítulos desta tese que sucedem esta seção (*Padronização e Ortografias*) estão organizados dessa maneira por algumas razões; particularmente pelo fato de que é preciso que o texto *Características tipográficas...* tenha um lastro nas questões linguísticas, sociolinguísticas, históricas e tecnológicas, levantadas nos capítulos *Padronização e Ortografias*.

Dessa maneira, acredita-se que seja possível conectar as questões de codificação com o problema da escrita das línguas indígenas, de uma forma sistêmica e vinculada às áreas acima mencionadas.

9. Padronização

Na introdução de *Language Standardization*, Punya Sloka Ray (1962, p. 91) explica que ao se falar de padronização, normalmente o termo é relacionado a ferramentas. E, de uma ferramenta padronizada, espera-se que seja mais barata, de fácil manutenção, uniforme, confiável. Quando uma ferramenta é raramente usada ou usada somente por poucas pessoas, a padronização é relativamente de pouca importância. Mas se utilizada por um número grande de pessoas, a padronização geralmente é uma vantagem. Ao aplicar esse conceito de padronização a línguas, o autor enfatiza seu caráter de ferramenta, observando que, desse ponto de vista, a língua é um *instrumento de comunicação*; um meio, e não um fim.

Em linhas gerais, este capítulo se concentra em questões relacionadas à padronização da escrita e procura reunir pontos importantes acerca do tema no contexto desta tese. Apesar de representar um recorte temático suficientemente extenso e complexo, a ideia de padronização é algo inerente à escrita e recorrente na literatura de várias áreas, como comunicação visual, tipografia, tecnologia da informação e linguística. Por essa razão, este capítulo busca demonstrar como a padronização permeia os campos supracitados e está organizado conforme descrição a seguir.

Primeiramente, são abordados alguns tópicos relacionados à padronização no campo da linguística. Em segundo lugar, são levantadas algumas referências históricas. Por último, o tema padronização é relacionado ao campo da tipografia. Com isso, pretende-se levantar exemplos e evidências sobre padronização da literatura de distintas áreas (tipografia, linguística, história), demonstrando a importância do tema para a comunicação escrita nas línguas indígenas brasileiras.

No campo da linguística, há considerável discussão acerca da padronização em trabalhos sobre ortografia, sociolinguística, gramática, revitalização de línguas,

planejamento linguístico, entre outros. As relações entre padronização e linguística são abordadas em três seções: *Padronização e linguística*, que faz um recorte de algumas questões gerais do campo; *Padronização e planejamento linguístico*, com explicações resumidas sobre planejamento; e *Motivações e implicações da padronização ortográfica*.

No campo histórico, a referência inicial (e um dos pontos de partida para este capítulo) foi o livro *Linguagens e comunidades nos primórdios da Europa moderna* (2010), do historiador britânico Peter Burke. A obra investiga alguns dos temas centrais da história das línguas na Europa, com um recorte cronológico que vai da invenção da imprensa, no século xv, até a Revolução Francesa, no século xviii, explorando temas significativos das relações entre línguas e comunidades (em sua maioria europeias), assim como relações entre língua e política. Apesar de tratar de um contexto bastante diferente do desta tese (línguas europeias, no continente europeu, nos primórdios do período moderno), Burke explica (2010, p. 108) a padronização das línguas como um resultado “praticamente automático da produção em massa de textos idênticos”. Este é um ponto de discussão relevante deste capítulo e um importante argumento desta tese.

A padronização é, por natureza, uma discussão inerente à tipografia, campo no qual a questão está presente em vários aspectos. Por exemplo, naqueles sobre materialidade (e.g. padronização de materiais para a fabricação de tipos, antes do computador) e unidades de medida; sobre temas relacionados às próprias formas das letras; sobre questões relacionadas às interfaces entre tipografia e codificação, relevantes para esta tese, entre outros.

Uma das referências que se concentra em questões relacionadas à padronização é o trabalho do designer de tipos e pesquisador holandês Frank E. Blokland. Em sua tese de doutorado (2016), Blokland discute a hipótese de que Gutenberg (juntamente a seus contemporâneos) desenvolveu um sistema de padronização para a produção do tipo textura, e que esse sistema foi extrapolado para a produção de tipos romanos na Renascença italiana. Nessa linha, o tipógrafo e educador britânico Peter Burnhill (2003) analisou obras impressas por Aldus Manutius, e suas análises demonstram que havia um sistema padronizado de medidas de corpo, entrelinha, comprimento de linha e mancha gráfica, mostrando evidências que apontam para um sistema de normas internas em sua oficina/editora.

9.1 Linguística e padronização

Esta seção tem como objetivo fazer um recorte de algumas questões gerais sobre padronização linguística, relevantes para essa tese.

Primeiramente, é importante mencionar que a discussão sobre padronização, no campo da linguística, é anterior à imprensa. Burke (2010, p. 32) observa que, ainda na Idade Média, figuras como Dante Alighieri, o rei espanhol Afonso, o Sábio, e o reformador tcheco Jan Hus são exemplos de estudiosos conscientes acerca da necessidade de discussão sobre os vernáculos; provavelmente questões sobre padronização tangenciaram a discussão desses estudiosos.

Nessa linha, Florian Coulmas (1999, p. 477–478, tradução nossa) observa:

No contexto dos vernáculos escritos da Europa medieval, a ortografia por muito tempo permaneceu sujeita aos caprichos de escritores individuais, uma vez que não existiam ortografias. Isso se deve a dois fatores. Havia uma grande variedade nas formas das línguas empregadas (i.e. escritas) por diferentes autores, e havia uma grande variedade nas maneiras de empregar o alfabeto para representá-las. Embora houvesse pouco em termos de princípios para o estabelecimento de regras, impressores tinham um interesse material em eliminar grafias peculiares e promover a uniformidade. Em grande medida, a padronização da grafia andou de mãos dadas com a padronização da língua e a consolidação de uma norma escrita, que, por sua vez, foi reforçada pela disseminação da alfabetização. A escolaridade obrigatória tornou a grafia correta uma parte fundamental da educação, exercendo forte influência na concepção de língua dos leigos. [...] Em sua forma padronizada como uma ortografia, a grafia tornou-se uma das instituições mais conservadoras da sociedade moderna.

Em segundo lugar, é importante reforçar o conceito de ortografia como algo que padroniza a grafia¹ de um sistema de escrita, conforme explicam Meletis (2020, p. 155) e Florian Coulmas (1999, p. 379).

A esse respeito, Meletis (2020, p. 156–157) alerta sobre a questão da adesão às regras estabelecidas em uma comunidade de escritores, destacando uma ou várias ortografias possíveis como corretas, e considera como funções de uma ortografia a

1. em inglês, *spelling*

convencionalização e a padronização. Observa ainda que (2020, p. 164), uma vez definido um padrão, uma norma, passa-se a ter uma referência, e isso possibilita que os usuários consultem a grafia correta a qualquer momento, contribuindo muito para a estabilidade da ortografia.

Sob a ótica da revitalização de línguas ameaçadas, Grenoble e Whaley (2006, p. 129–130) observam que a padronização é uma parte natural do desenvolvimento de qualquer língua escrita, mas nos casos de línguas ágrafas ou de línguas sem uma tradição escrita recente (como no caso de línguas em processo de revitalização), o estabelecimento de um padrão literário representa um passo abrupto e frequentemente controverso. Uma certa quantidade de variação existe em comunidades de falantes de qualquer língua viva, e a padronização da língua escrita não elimina essas diferenças, mas cria uma variedade adicional, que deve ser compreensível para o maior número possível de falantes.

Grenoble e Whaley (2006, p. 130) também observam a padronização como algo crítico para os esforços de revitalização de línguas e para a alfabetização em idiomas locais e enfatizam que, sem a padronização, a escrita se torna idiossincrática e não pode ser interpretada por um grande número de falantes. Nessa linha, destacam que um dos aspectos críticos de um programa de revitalização bem-sucedido é a criação de um corpo ativo de usuários da língua. A língua escrita pode ser uma ferramenta poderosa para atingir esse objetivo se todos, ou um grande número de falantes, puderem ler e entender determinado padrão.

Grenoble e Whaley (2006, p. 130–131, tradução nossa) observam ainda:

A criação de uma forma literária padrão para uma língua ágrafa é tipicamente um processo intencional e explícito, melhor realizado por aqueles com um alto nível de consciência metalinguística ou treinamento especializado em linguística. Dessa forma, é muito diferente da maneira pela qual as formas linguísticas padrão surgiram em línguas amplamente faladas, em que a produção de materiais em e sobre uma língua, bem como os desenvolvimentos tecnológicos, impulsionaram a padronização.

Como exemplo disso, Grenoble e Whaley destacam que a tradução da Bíblia por Lutero (1521–1534) desempenhou um papel crítico na padronização do alemão, assim

como a *King James Bible* (1611) o fez para o idioma inglês. As autoras elencam ainda outras publicações com ampla influência, como o dicionário de Samuel Johnston (1755) e a gramática de Lowth (1762). Peter Burke (passim, 2010) indica ainda iniciativas de diversas figuras que contribuíram para a padronização de vernáculos, e.g. Pietro Bembo, Antonio Nebrija, Robert Estienne etc.

9.2 Padronização e planejamento linguístico

O planejamento linguístico é uma das áreas da linguística em que o tema padronização é bastante recorrente. Segundo Bernard Spolsky (2012, p. 3–4), essa área desenvolveu-se como campo de estudo no processo de reconstrução das sociedades no pós-guerra (c. 1950–1960), com objetivo de entender e resolver problemas linguísticos de novos estados independentes. Em geral, há consenso de que o *planejamento linguístico* produz uma *política linguística*, ou um conjunto de regras para uso e forma da língua dentro de um estado. O autor indica ainda que o processo do planejamento linguístico no pós-guerra se revelou similar às políticas linguísticas de nações recém-independentes no século XIX, contexto em que era comum o entendimento de que uma nação é definida por seu território e língua. Naturalmente, essa visão ignora a existência de minorias, o que ocasionou problemas em diversos locais, e.g. África, União Soviética etc.

Spolsky (2012, p. 3–5) alerta também para o fato de que os linguistas envolvidos em programas de planejamento e políticas na segunda metade do século XX (período caracterizado pelo que o autor chama de *classic language planning*) acreditavam que o planejamento linguístico seria possível, da mesma forma que o planejamento econômico. Ele adverte que essa mentalidade, na prática, raramente funcionou, pois frequentemente os planos ignoravam não somente questões sociolinguísticas complexas, mas também fatores emocionais e simbólicos (e.g. nacionalismo, religião, etnicidade, identidade etc.). Como resultado, muitas atividades de planejamento desse período desapareceram, e a política linguística desenvolvida nas nações do mundo continuou a evoluir, independentemente dos planos originais. A África é um exemplo clássico dessa situação de complexidade linguística, continente sobre o qual foram impostas fronteiras territoriais que antes não existiam (e, dentro dessas fronteiras, a língua do colonizador),

causando inúmeros problemas sociais e políticos. Já as nações que eram parte da União Soviética, libertadas da “russificação” após 1989, puderam reestabelecer antigas línguas nacionais e restaurar suas identidades.

Spolsky (2012, p. 4-6) explica que, para muitos no campo, o planejamento linguístico baseado no “modelo clássico” ainda é a abordagem mais comum. Essa abordagem consiste em um processo que se inicia com um levantamento detalhado da situação da língua, a partir do qual se define outras duas áreas do planejamento linguístico: o planejamento de status (*status planning*) e de corpus (*corpus planning*). Em linhas gerais, o planejamento de status consiste em decisões e escolhas sobre o idioma (ou, ocasionalmente, idiomas) que terá o status de oficial, especialmente no âmbito governamental e educacional. Já o planejamento de corpus consiste em ações que englobam normatização de um sistema de escrita, gramática, ortografia etc.

Assim, é possível dizer que, no campo do planejamento e das políticas linguísticas, a padronização está fortemente associada ao planejamento de corpus. Nessa linha, pode-se indicar como um dos desdobramentos do planejamento de corpus o desenvolvimento de dicionários e gramáticas, entre outras publicações, frequentemente associado às academias de línguas.

Por último, conforme já mencionado no capítulo *Justificativa e relevância*, é importante mencionar a existência de lacunas na literatura do planejamento de corpus, principalmente sobre questões referentes ao modo como as línguas se relacionam com os padrões de codificação de texto, e ao papel da escrita (i.e. a tipografia) como “ponte” entre os campos da linguística e da tecnologia da informação. Nesse sentido, ressalta-se a importância da tipografia dentro do planejamento linguístico.

9.3 Motivações e implicações da padronização ortográfica

Grenoble e Whaley (2006, p. 153–156), em uma seção de sua obra (sobre revitalização de línguas ameaçadas) voltada para a discussão da padronização ortográfica, elencam algumas motivações e implicações da padronização de línguas.

Primeiramente, as autoras destacam (2006, p. 153) o desejo por unidade política ou cultural como uma das razões mais comuns para a promoção de um padrão de escrita.

Nesse sentido, a crença no poder unificador de um padrão comum de escrita motivou muitos governos a impor padrões de escrita e ortografia. Por outro lado, observam a diversidade linguística como um obstáculo à padronização. Nessa linha, Antoine Meillet observa que “[...] na história da língua, duas tendências antagonistas estão em jogo, uma tendendo à diferenciação e a outra, à unificação” (MEILLET, 1921-38, p. 110, apud BURKE, 2010, p. 29)

Grenoble e Whaley (2006, p. 154) elencam ainda algumas vantagens de se ter um sistema de escrita padronizado. Primeiro, um único sistema de escrita padronizado facilita a comunicação de várias maneiras, assim como a distribuição de recursos médicos e o desenvolvimento de infraestrutura, além de ser um recurso de grande valor para a educação. Em segundo lugar, observam que:

[...] uma ortografia padronizada pode contribuir para o aumento dos domínios funcionais de uso de um idioma que, por sua vez, aumenta seu status dentro da comunidade e reforça valores da comunidade. Assim, a padronização pode desempenhar um papel importante no reforço do senso de identidade de um grupo e também pode permitir que um grupo ganhe reconhecimento e status oficial, aumentando ainda mais um senso de identidade e orgulho.” (GRENOBLE; WHALEY, 2006, p. 154, tradução nossa)

É importante também destacar as implicações de se ter um sistema padronizado. Nesse sentido, deve-se ressaltar que o processo de padronização pode ser responsável pela perda linguística de múltiplas maneiras, já que formas escritas padronizadas podem inibir a variabilidade, levando à perda de diversidade da língua. Na comunicação falada, uma gama de variações dialetais pode existir, enquanto na alfabetização são preferíveis as formas padronizadas. Por último, destaca-se que a preferência por um padrão em detrimento a outro pode levar à estratificação social (GRENOBLE; WHALEY, 2006, p. 154–156).

9.4 O papel da imprensa na padronização das línguas

Esta seção se concentra no levantamento de algumas referências históricas e tem como ponto de partida e referência inicial o capítulo *A padronização das línguas* (BURKE, 2010,

p. 105–126), do livro *Linguagens e comunidades nos primórdios da Europa moderna*, do historiador Peter Burke, que explica a padronização das línguas como um resultado “praticamente automático da produção em massa de textos idênticos” (2010, p. 108).

Em resumo, essa obra investiga alguns dos temas centrais da história das línguas na Europa no início do período moderno e explora temas significativos das relações entre línguas e comunidades, assim como relações entre língua e política. Burke aborda temas como o uso do latim como *língua franca* na Europa, e o declínio gradual de seu uso, causado pela ascensão dos vernáculos (e.g. espanhol, alemão etc.) O declínio do uso do latim tem diversos motivos (e.g. Reforma), assim como a ascensão dos vernáculos, que passaram a ganhar importância como línguas literárias e a expandir os seus domínios.

Burke (2010, p. 105) observa que com a expansão do domínio dos vernáculos no início do período moderno, a padronização torna-se cada vez mais necessária, e destaca a ambiguidade do termo, que remete “não somente ao processo de se tornar mais uniforme (*uniformidad*, como caracterizou Nebrija) como também de seguir regras”. O autor destaca ainda que a padronização, além de ser uma medida pragmática (para facilitar a comunicação entre regiões), também tinha como finalidade dar aos vernáculos “uma parte do prestígio ou da dignidade associada ao latim”.

Acerca dos objetivos e finalidades da padronização, Burke (2010, p. 106) destaca ainda um ponto muito importante, que é a questão da estabilidade: a padronização promove a “uniformidade através do espaço” e a “fixidez ao longo do tempo”. O autor observa que naquele período, “para que uma língua conquistasse o nível de prestígio do latim, ela precisava ser estável”, e cita exemplos de autores como John Dryden e Jonathan Swift reclamando da falta de padrão e das variações da língua inglesa, reação que ele chama de “ansiedade da instabilidade”. Uma das consequências dessa “ansiedade” é o surgimento de diversas academias de línguas² por toda a Europa, “todas respon-

2. As mais antigas academias de línguas são as de Florença (Accademia della Crusca, 1582), Paris (Académie Française, 1635) e Madrid (Real Academia Española, 1713). A Academia das Ciências de Lisboa foi fundada somente em 1779, apesar de publicações como *Orthographia da Lingoa Portuguesa* (1576), de Nunes de Leão, e o *Vocabulario Portuguez e Latino* (1712), de Bluteau, serem anteriores a essa data (BURKE, 2010, p. 106; COULMAS, 1999, p. 158; 414; 476).

sáveis por dicionários”. Burke observa ainda que o lema da Real Academia Espanhola é, desde a sua fundação, “*limpia, fija y da esplendor*” (limpa, fixa e dá esplendor), reforçando o papel dessas sociedades científicas.

Burke (2010, p. 107) destaca ainda, nesse período, o surgimento e disseminação de várias gramáticas, assim como a adesão às regras por parte das elites. Com isso, os vernáculos começam a ter maior reconhecimento (em relação ao latim), já que passam a contar com uma norma culta, formalizada e distante da linguagem coloquial. O autor explica ainda que, na Renascença, com base em procedimentos de imitação e emulação, os humanistas tentaram reformar vários vernáculos, com o objetivo de torná-los mais apropriados ao uso literário. Esse fenômeno mostra não somente o poder dos acadêmicos e intelectuais, mas também que “as formas-padrão de vernáculos eram expressões dos valores de novas comunidades”.

Com a criação das academias (e, por sua vez, das gramáticas, dicionários etc.), nota-se a uma crescente preocupação dos Estados com a língua utilizada em seus domínios linguísticos. Nesse sentido, Burke observa ainda (2010, p. 107, grifo nosso) que:

[...] desenrolou-se na Europa nesse período uma tentativa coletiva e cooperativa de **elevar o status dos vernáculos, codificando-os**, enriquecendo-os e, desse modo, **transformando-os em línguas apropriadas para a literatura**. Essa iniciativa foi em grande parte **promovida por um novo meio de comunicação: a imprensa**.

A imprensa exerceu papel fundamental na padronização das línguas no início do período moderno, e muitos historiadores consideraram esse fenômeno como um resultado direto da produção em massa de textos idênticos, já que a padronização das línguas se adequava perfeitamente à lógica da imprensa, de vender um grande número de volumes idênticos para o maior número possível de leitores (BURKE, 2010, p. 108).

Florian Coulmas (1999, p. 419) observa que a padronização não apenas da linguagem, mas de muitos outros aspectos da vida social seja, talvez, o efeito sociocultural mais significativo da impressão com tipos móveis.

Nessa linha, Febvre e Martin (2017, p. 427–428) destacam que a imprensa exerceu importante papel na formação e fixação das línguas nacionais europeias e indicam

que, a partir do século XVII, parece haver uma cristalização dessas línguas nacionais em toda a parte, destacando “a fixação mais ou menos rápida [...] das línguas que são ainda hoje as línguas nacionais”. Os autores (FEBVRE E MARTIN, 2017, p. 427–428) notam ainda que, “procurando atingir a clientela mais vasta possível, os editores são naturalmente levados a favorecer, em muitos domínios, o desenvolvimento das línguas vulgares. De outro lado, a imprensa dá às publicações um caráter estável ... [e] os impressores têm a tendência de eliminar as fantasias ortográficas e as variações dialetais que corriam o risco de tornar o livro menos facilmente acessível a um vasto público”.

Segundo Burke (2010, p. 108), a transição da variedade de formas da mesma palavra (do início do período moderno) para formas padronizadas (nos séculos XIX e XX) sugere que a disseminação dos livros impressos influenciou não somente os hábitos de escrita, mas também da fala. Por outro lado, o autor observa que, em muitos casos, houve uma transição da oralidade para imprensa sem passar pela era do manuscrito. Como exemplo disso, Burke menciona o papel dos missionários (principalmente jesuítas) na fixação das línguas indígenas, com a criação de publicações, como a gramática do Tupi Antigo, de José de Anchieta, de 1595.

É importante observar que a imprensa não foi o único fator que contribuiu para a padronização. Febvre e Martin (2017, p. 427–428) indicam que “existia, nas chancelarias, a preocupação de generalizar usos que, em muitos casos, tornaram-se os das línguas literárias”. Outro fator a ser levado em conta é o surgimento das monarquias nacionais e do Estado centralizador, que favoreciam a unificação linguística, citando a França e a Espanha como exemplos disso.

Nessa linha, Burke (2010, p. 108–110) adverte: “[...] não podemos presumir que a imprensa tenha sido o único personagem nesse enredo. Quatro observações gerais devem servir para complicar ainda mais a trama”. Primeiramente, do ponto de vista conceitual, Burke (2010, p. 108–110) observa “que o conceito de ‘padronização’ é menos simples do que parece [...]. A codificação formal de uma língua não é a mesma coisa que a ascensão à dominância de um dialeto ou socioleto em particular. [...] A padronização na prática, seja na escrita, seja na fala, é uma questão de grau”; ou seja, é uma questão de referencial, já que algumas línguas são mais padronizadas que outras (e.g. o espanhol

é mais padronizado que o alemão). Em segundo lugar, do ponto de vista cronológico, Burke constata que o processo de padronização é mais antigo que a imprensa; como exemplo, cita o papel de personagens (e.g. Jan Hus), de monarquias e das chancelarias na fixação de línguas literárias. Em terceiro lugar, Burke (2010, p. 108–110) destaca “que a imprensa era uma faca de dois gumes, que poderia ser utilizada para promover padrões rivais”. Nesse sentido, ele apresenta fatores que influenciaram a fixação de padrões. Na França e na Inglaterra, por exemplo, havia um padrão dominante difundido pelas gráficas concentradas, principalmente, em Londres e Paris. Já na Espanha, a imprensa estava espalhada em várias cidades; e apesar disso, o padrão do castelhano era estável. Na Alemanha, “o período antes de 1550 tem sido descrito como uma época de ‘línguas tipográficas’ (*Druckersprachen*) concorrentes em Augsburg, Frankfurt, Leipzig, Basileia, Wittenberg e outras regiões” (BURKE, 2010, p. 108–110). Já na Itália, cidades como Roma, Florença, Milão e Veneza promoviam diferentes variedades da língua, e o toscano só se tornou o padrão tipográfico no início do século XVII. Por último, Burke destaca o problema do veículo, apontando para autores que defendem uma tese da “revolução da imprensa” (e.g. Eisenstein), enquanto outros (e.g. Harold Innis) entendem a que “a imprensa pode ser vista como um catalisador, ajudando as mudanças em vez de iniciá-las [...]” (BURKE, 2010, p. 108–110).

9.5 Tipografia e padronização

A ideia de padronização e normatização é intrínseca à tipografia. Gerrit Noordzij (2000, p. 3) define tipografia como “escrita com letras pré-fabricadas”. Anthony Froshaug (apud BURNHILL, 2003, p. 8) indica que “a palavra tipografia significa escrever ou imprimir utilizando elementos padronizados”. Na obra *Entre parágrafos: fundamentos tipográficos*, Cyrus Highsmith (2017, p.13) ilustra esses conceitos de forma muito simples e clara (conforme mencionado no capítulo *Codificação e linguagem*), mostrando que ao colocar as letras dentro de “caixas”, Gutenberg cria um conceito-chave dos tipos móveis, que é a padronização dos espaços em branco ao redor das letras.

Esta seção tem como objetivo fazer um levantamento acerca da padronização no campo tipográfico e trata de algumas questões como a padronização dos espaços e de

sistemas de unidades e medidas. Dois autores principais foram utilizados como referência para esta seção: Peter Burnhill e Frank E. Blokland.

Em *Type spaces: in-house norms in the typography of Aldus Manutius*, Peter Burnhill (2003) analisa algumas obras³ impressas por Aldus Manutius entre 1495 e 1522. Suas análises demonstram que havia um sistema padronizado de medidas de corpo, entrelinha, comprimento de linha e mancha gráfica, mostrando evidências que apontam para um sistema de normas internas utilizadas por Manutius e Francesco Griffo, cortador de punções de Manutius. Por se tratar de um trabalho tecnicamente complexo, serão levantadas aqui algumas ideias relevantes sobre trabalho, numa tentativa de síntese.

Burnhill (2003, p. 12) explica que seu interesse primário é demonstrar que as normas tipográficas aldinas se baseavam em um sistema de medidas unificado, que poderia ser utilizado como referência em ambos os eixos da página, em todos os níveis de ordem. Observa também que, naquele período, Manutius despertou grande interesse do público, em especial pela realização da série padronizada de 47 títulos no formato *octavo*⁴.

De maneira resumida, o método de Burnhill (2003, p. 22) consiste em:

1. Ampliar (e.g. 8x) as fontes primárias em fotocópias.
2. Medir, a partir das ampliações, a distância entre as linhas de base do texto.
3. Fracionar essa medida em doze partes iguais.
4. Construir, a partir dessas medições, uma malha quadriculada de linhas finas sobre material transparente, com base na unidade encontrada na etapa anterior (1/12 da distância entre as linhas de base).
5. Utilizar a malha sobreposta a diferentes páginas para checar a veracidade da noção de que o sistema de referência dimensional utilizado por Manutius se baseava em múltiplos da duodécima parte do corpo do tipo em questão.

3. *Erotemata* (1495); *De Aetna* (1495); *Metamorphoses* (1502); *Quaestiones Naturales* (1522).

4. Manutius também é pioneiro na edição de uma coleção (ou série) de publicações de temas relacionados. Os 47 títulos da série tinham um projeto gráfico e formato semelhante.

Computer-drawn base-lines and machine calculation of line-increments (mould sizes)

COMPENDIUM OCTO ORATIONIS
TIVMET ALIO^Q VOR VNDAM NE^{ce}
EDITVM A CONSTANTINO
LASCARI BYZANTIO.

Lascaris, *Erotemata*, 1495.
First casting.
Line-increment: 6.15 mm
+ 2.0% = 6.27
or + 2.5% = 6.30
nearest 0.1 mm = 6.3

¶ De diuisione litterarum.

Liber primus:

¶ I littera est pars minima uo

pedemontana regione passim uisuntur:
alibi per summa ora exundans incen-
dium faxis fluentibus rotas plagas in-
occupauit; alibi arenarum campi ma-
gnam in longitudinem, et latitudi-
nem extenduntur. In supremo cra-
tères duo sunt, quorum alterum, qui
minus altissimus est, ipsi uidimus in pu-
tei rotunditatem angustum, emissis ue-

De Aetna, roman, 1495/6.
Line-increment: 5.62 mm
+ 2.0% = 5.73
or + 2.5% = 5.80
nearest 0.1 mm = 5.8

dia uado si maris, cum effugerimus procellosos de super montes, per
quos praecipit in nauigantes uentus impingit, cum inuolutis nubilo
dies, et nimbis ac tonitribus horredas noctes, cum turbinibus diuisa
nauigna, quis erit huius laboris ac metus fructus? quis nos fessos tot
malis portus excipiet? Bellum scilicet, et obuius in litore hostis, et ex-
cidenda gens, tractura magna ex parte uictore, et antiquarum
urbium flammam. Quid in arma cogimus populos? Quid exercitus
scribimus directores ac in medijs fluctibus? Quid maria inque-
ramus? Quorum uidelicet ad mortem nostram terra late patet, nimis
delicate fortuna nos tractat, nimis dura dedit nobis corpora, felicem
uacuitudinem, non depopulatur nos casus incurrens, emetiri cuique annos
suos ex comodo licet, et ad senectutem decurrere. Itaque eamus in pe-

Seneca, italic, 1522.
Line-increment: 3.92 mm
+ 2.0% = 3.998
or + 2.5% = 4.017
nearest 0.1 mm = 4.0

Figura 90: Exemplos de medições dos livros de Manutius. Fonte: BURNHILL, 2003, p. 23.

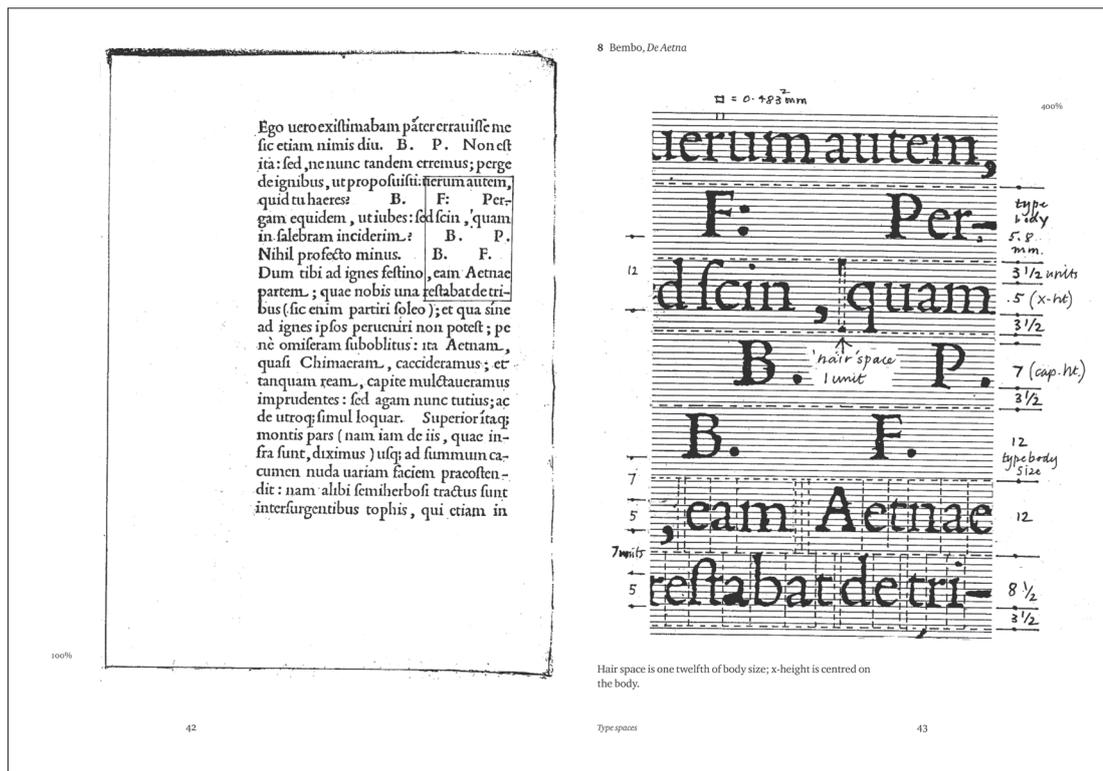


Figura 91: Exemplos de medições dos livros de Manutius. Fonte: BURNHILL, 2003, p. 42–43.

Burnhill (2003, p. 25) explica que as larguras dos materiais utilizados para espaçamento com tipos móveis de metal eram familiares aos tipógrafos que trabalhavam com essa tecnologia. No entanto, historiadores da tipografia nunca se interessaram muito pelas funções sintáticas do espaço da maneira como demonstraram interesse pelo desenvolvimento do molde de fundição de tipos móveis. Mas ninguém investigou o molde como um mecanismo para o estabelecimento de uma escala de valores espaciais.

Neste ponto, é importante que se entenda que a questão dos espaços é bastante relacionada à ferramenta do molde (Fig.92) e das matrizes de fundição (Fig.93). O que determina o espaço de um tipo móvel é a largura da matriz. As matrizes de um determinado corpo têm alturas fixas (i.e. corpo), mas larguras variáveis (e.g. a letra *i* é estreita, a letra *m* é larga). E o molde se ajusta às diferentes larguras das matrizes. Smeijers (2005, p. 55–56) explica que o molde é formado por “duas partes que, juntas, se encaixam, deslizando para segurar matrizes de diferentes larguras”, conforme ilustra a Fig.94.

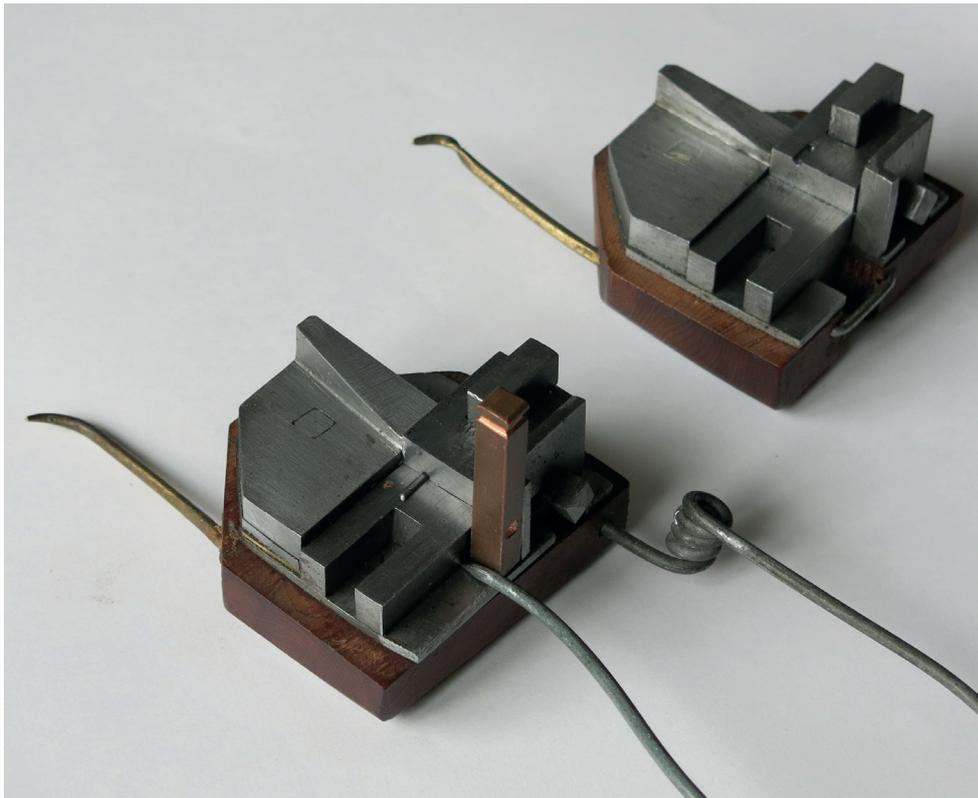
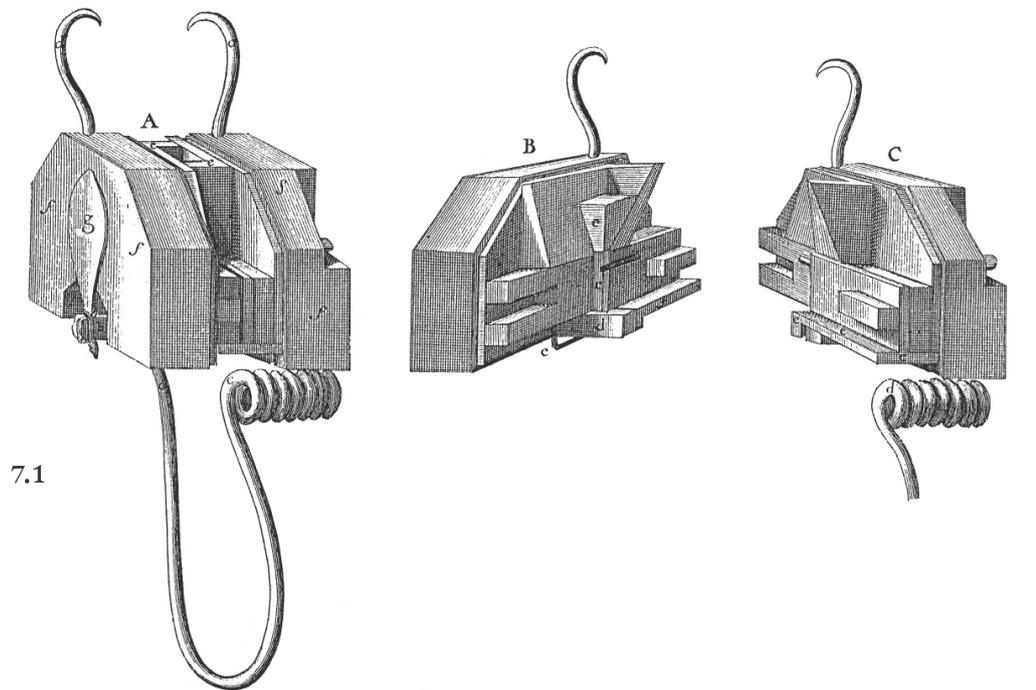


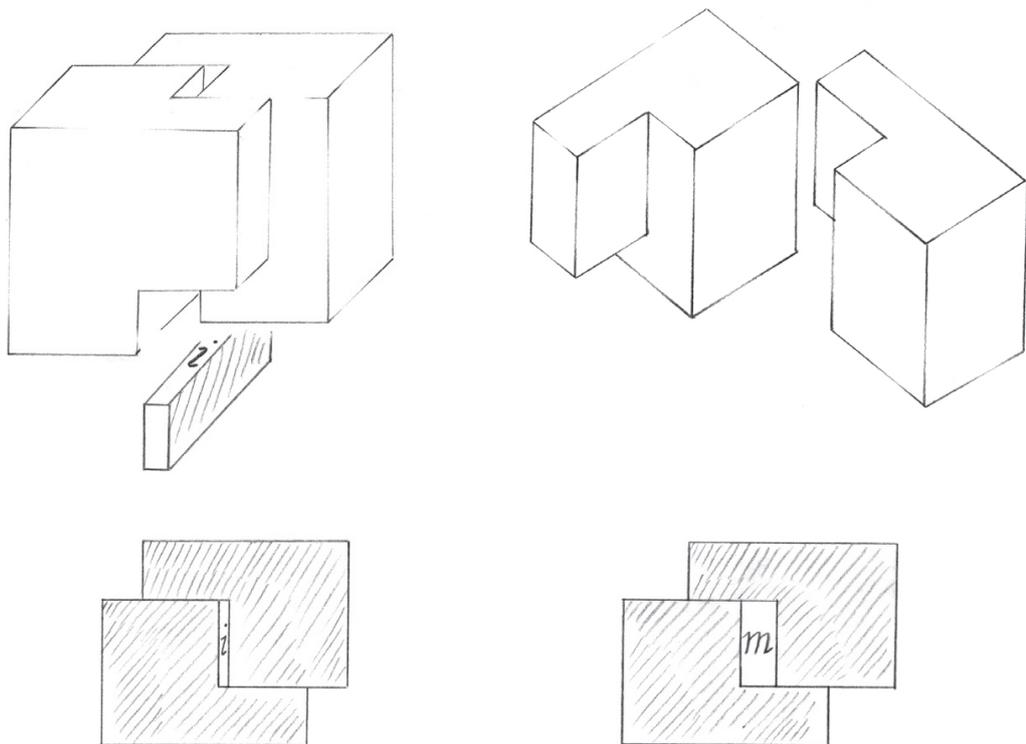
Figura 92: Molde de fundição de tipos, reconstruído por Alan May a partir de descrições de Joseph Moxon. No canto inferior direito pode-se ver a mola, que funciona como mecanismo de trava, fechando o molde e deixando-o pronto para receber o chumbo. Fonte: MAY, 2018.



Figura 93: Matrizes originais de Claude Garamond, século XVI. Fotografado no Museu Plantin-Moretus, Antuérpia, em março de 2012. Fonte: do autor.



7.1



7.2

Figura 94: A parte de cima da imagem mostra o molde conforme gravura da série feita para a Académie Royale des Sciences, na França, na última década do século XVII. Os desenhos na parte inferior da imagem mostram os princípios do funcionamento do molde.

Fonte: SMEIJERS, 2015, p. 56.

Burnhill (2003, p. 25) explica que a largura dos espaços corresponde a frações do quadratim (*em/em-space*; eme/espço eme/quadrado eme) de uma fonte. Normalmente, se utiliza algumas frações desse quadratim: *en-space* (ene/espço ene) = 1/2 quadratim; *thick space* (espço largo/grosso/gordo) = 1/3 quadratim; *middle space* (espço médio) = 1/4 quadratim; *thin space* (espço fino) = 1/5 quadratim; *hair space* (espço ultrafino/espço-cabelo) \leq 1/6 quadratim⁵ (Fig.95). O autor faz uma revisão bibliográfica apontando para autores que fizeram menção à largura dos espaços, como Harry Carter, Moxon, Jacobi, Fertel e Fournier.

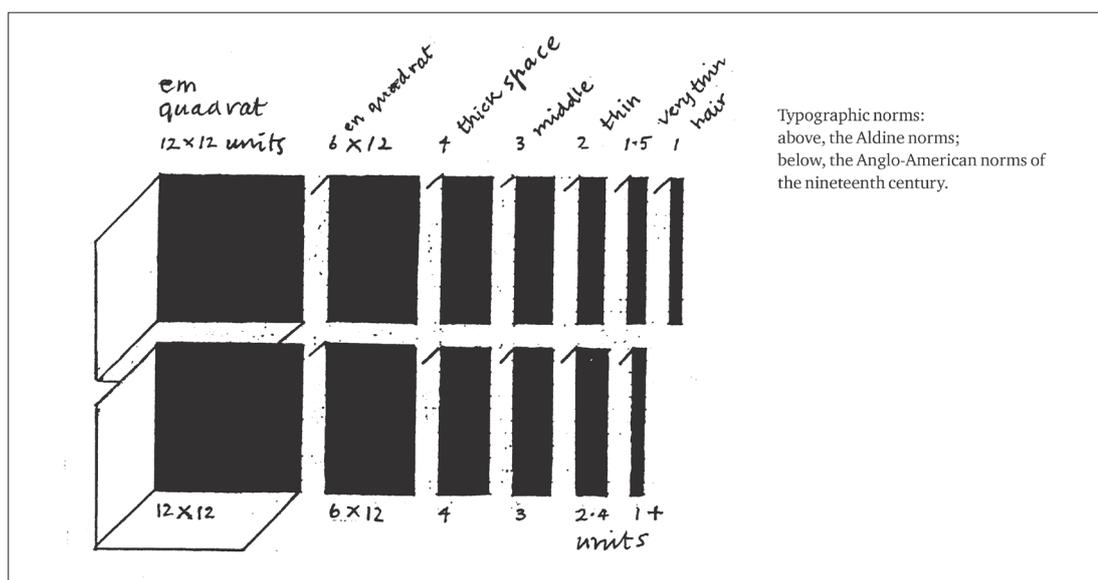


Figura 95: Exemplos de medições das frações do quadratim. Acima, as normas aldinas; abaixo, normas anglo-americanas do século XIX. Fonte: BURNHILL, 2003, p. 36.

Fertel (apud BURNHILL, 2003, p. 26) ressalta que a justificação é a principal razão para a existência de espaços de larguras diferentes (Fig.96). Na composição com tipos móveis de metal, os espaços são importantes para a fixação da composição em determinada medida de largura. Isso quer dizer que os espaços são importantes para fazer com que a letras se ajustem à largura da linha, ficando bem justas e presas à composição.

5. Sobre nomenclatura dos espaços tipográficos em português: ver BRINGHURST, 2005, p. 356.

Nesse sentido, Burnhill (2003, p. 26) destaca um ponto de tensão entre o significado do texto e sua tecnologia de produção: de longa data, um possível ponto de diferença entre editores e tipógrafos, preocupados com a estrutura linguística do texto, e impressores, preocupados com a mecânica de sua produção.



Figura 96: Na parte inferior da gravura é possível ver os tipos e espaços em perspectiva isométrica (Fig.4), o componedor (Fig.5) e um bloco justificado (Fig.6).Fonte: GOUSSIER, c.1700.

Em resumo, pode-se dizer que a fração de 1/12 do corpo é uma unidade-chave em ambos os eixos (x e y). A Fig.97, a seguir, mostra uma análise de Burnhill, apontando para a medição da menor fração do quadratim (*hair space*) no bloco de texto.

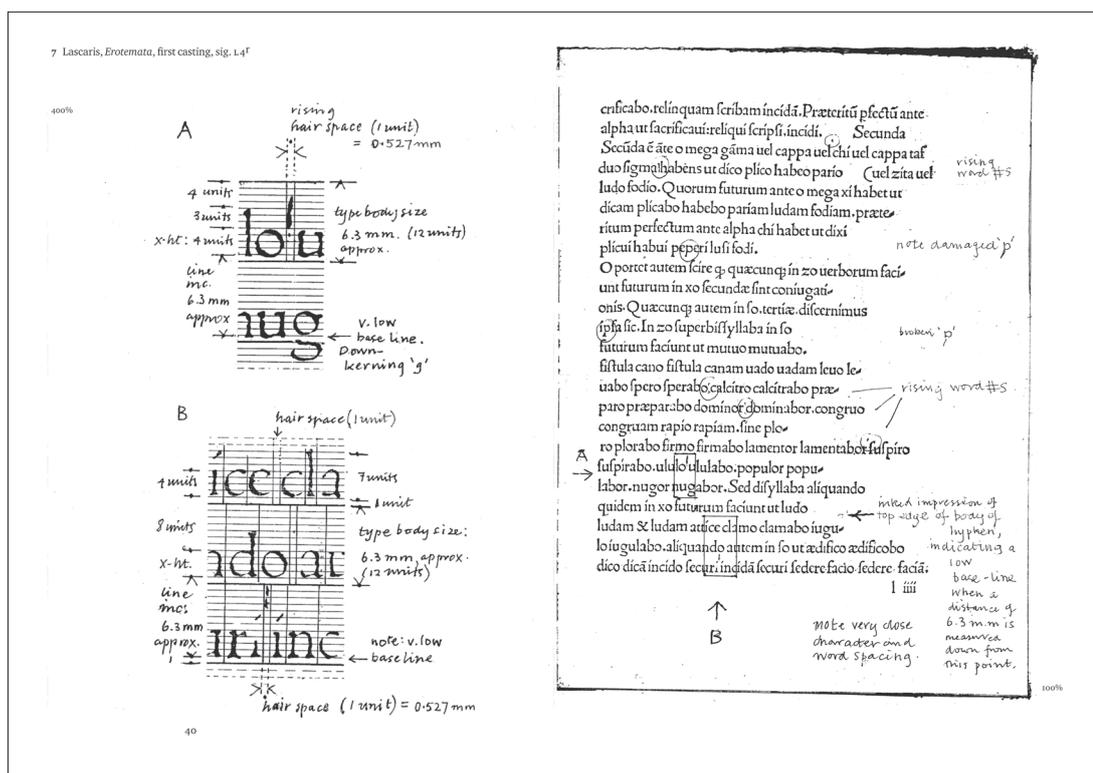


Figura 97: Exemplos de medições dos livros de Manutius. Observa-se a medição da menor fração do quadratim (*hair space*). Fonte: BURNHILL, 2003, p. 40–41.

Além de Peter Burnhill, Frank E. Blokland é outro autor que investigou a ideia de padronização na tipografia. Em sua tese de doutorado, *On the origin of patterning in movable Latin type: Renaissance standardisation, systematisation, and unitisation of textura and roman type*, Blokland (2016) trabalha com a hipótese de que Gutenberg e seus pares desenvolveram um sistema padronizado para a produção de seus tipos góticos, e que esse sistema foi extrapolado para a produção dos tipos romanos na Itália. Para tal, a escrita humanista foi utilizada como modelo dos tipos, que tinham proporções padronizadas (em especial de largura e espaçamento). Sua pesquisa tenta provar que aspectos técnicos da produção de tipos na Renascença tiveram influência direta nas

proporções de tipos romanos e itálicos e, com isso, no condicionamento a ele relacionado, que forma a base de nossa percepção (BLOKLAND, 2017).

Nos trechos a seguir, Blokland explica seu ponto de vista sobre os arquétipos de tipos de letras, considerando dois principais momentos históricos: a Renascença (primeiro na Itália e depois na França e em Flandres) e o Iluminismo:

O tipo empregado por Nicolas Jenson em 1470 no tratado *De Præparatione Evangelica* é geralmente considerado o primeiro tipo romano de alta qualidade. Não foi o primeiro tipo romano: precedido pelo tipo de Sweynheym e Pannartz utilizado na obra *De Oratore* de Cícero em 1465 [...] o tipo romano de Jenson formou a base para os tipos romanos de Francesco Griffo de 1495 e 1499, que, por sua vez, formaram as bases para os tipos da Renascença francesa de Claude Garamont, Robert Granjon e Hendrik van den Keere [...]. Desde o século dezesseis, os tipos de Garamont, em particular, foram de grande influência para os trabalhos de seus sucessores e continuam sendo uma base para os tipógrafos de hoje. (BLOKLAND, 2016, p. 9, tradução nossa)

Na França, no séc. XIX, durante o reinado de Louis XIV, um comitê formado pela Académie des Sciences desenvolveu padrões [esquemas de construção] geométricos para a construção de uma nova série de tipos para uso exclusivo da Imprimerie Royale, conhecidos pelo nome de “Romain du Roi”. Essa tentativa de racionalizar, padronizar e unitizar o design e a produção de tipos com o uso de geometria e grids é geralmente considerada uma mudança de direção em relação aos desenvolvimentos [de tipos] anteriores. A teoria é que esse novo tipo rompeu com uma tradição desenvolvida sob a influência de cortadores de punções como Nicolas Jenson (c.1404–1480), Claude Garamont (c.1510–1561) e Robert Granjon (1513–c.1590). Segundo a literatura, até a criação do Romain du Roi, a produção de tipos era meramente uma forma mecanizada e disciplinada de caligrafia, sem nenhuma forma de sistematização. Isso coloca uma ênfase no olho do cortador de punções e desconsidera a possibilidade de que a sistematização pudesse ser uma parte intrínseca da produção de fontes. No entanto, essa ênfase na estética não tem lastro em evidências históricas. Nesta dissertação, eu questiono essa crença por meio do uso de artefatos históricos; minhas hipóteses são o resultado do questionamento de suposições comumente aceitas, como esta, sobre as origens dos tipos romanos e itálicos. (BLOKLAND, 2016, p. 45, tradução nossa)

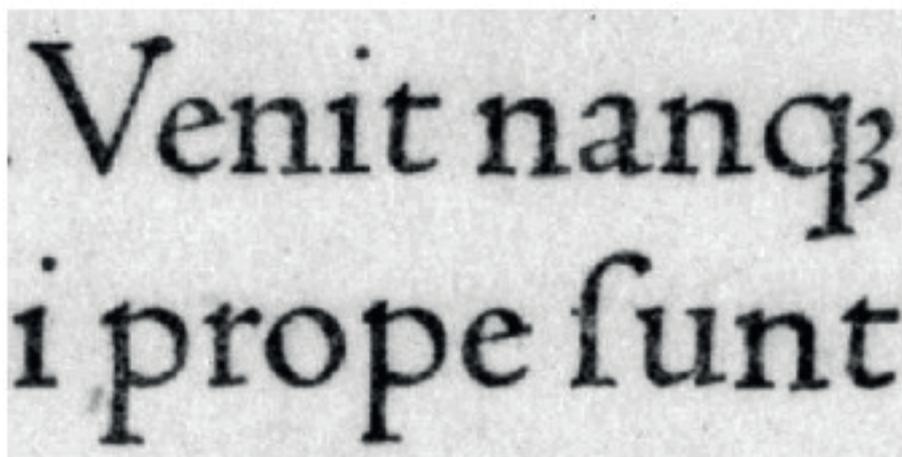


Figura 98: Modelo arquetípico (para o tipo romano). Fonte: BLOKLAND, 2017

Blokland (2016, p. 52), ao relacionar suas hipóteses, considerando a existência prévia dos modelos de escrita que precedem os tipos góticos e romanos, questiona: é possível que, em alguma medida, as proporções e detalhes dos tipos romanos sejam resultado de padronização? Blokland explica que, no campo da tipografia, sempre existiu uma crença que os cortadores de punções julgavam as proporções corretas somente pela observação, “no olho”⁶. No entanto, considerando o fato de que a fabricação de tipos e o corte de punções na Renascença eram processos de produção complexos e sofisticados, qual a probabilidade de que tenham sido realizados sem uma extensiva estruturação das formas de letras manuscritas? A questão é: os modelos arquetípicos da Renascença, de Jenson, Griffo e Garamond, foram feitos com base em padrões? E se esse for o caso, seriam a harmonia e estética em tipos, que estão embutidos em convenções tipográficas, não somente o resultado de preferências ópticas que precedem a invenção dos tipos móveis, mas também da padronização na produção de tipos da Renascença? Essas perguntas levaram o autor à principal hipótese de sua tese:

A criação dos tipos romanos foi influenciada tanto por considerações estéticas quanto por considerações técnicas.

6. Fournier: “the eye should rule”. Carter, *Fournier on Typefoundry*, p. 9. v. também Smeijers, Fred, *Contrapunção*.

E, como suporte para essa hipótese, ele levanta duas sub-hipóteses:

Os tipos romanos são o resultado da padronização da minúscula humanista ao processo de produção de tipos na Renascença. Esta é uma analogia à padronização que ocorreu quando a escrita gótica, já bastante “unitizada”, foi usada como base para o tipo textura.

Preferências estéticas nos tipos romanos ainda são condicionadas pela padronização do início da produção de tipos romanos.

Em suporte à primeira sub-hipótese, na primeira parte de sua tese, Blokland (2016) examina a ligação entre os tipos romanos e góticos e entre os tipos romanos e suas origens manuscritas, tentando provar que a regularidade da *textura quadrata* facilitou o trabalho de Gutenberg e seus parceiros de sistematização e padronização dos tipos de góticos, diretamente baseados em seu modelo manuscrito. Com isso funcionando para o tipo textura, era natural que se aplicasse o mesmo sistema aos tipos romanos. Dessa forma, o autor tenta demonstrar a evidência de padronização em tipos romanos e propor um arcabouço em que essa padronização estava baseada.

Como tentativa de provar a sua segunda sub-hipótese, Blokland (2016) discute as falhas na teoria comumente aceita de que os tipos romanos são o resultado de considerações estéticas, questionando se o que se acha opticamente [visualmente] interessante, pelo menos parcialmente, na tipografia latina é resultado de processo de padronização.

Blokland (2016) observa ainda que não há evidências ou documentação sobre os métodos de produção de cortadores de punções da Renascença. Hoje, os materiais mais antigos (moldes, tipos, matrizes) são de cerca do começo do século XVI. Exemplos desses materiais estão no acervo do Museu Plantin-Moretus, em Antuérpia, Bélgica. Sobre como foram produzidos, há pouca, para não dizer nenhuma, documentação. Nesse sentido, muito do que se sabe sobre o período dos incunáveis vem de descrições dos séculos XVII e XVIII. Por essa razão, Blokland baseia seu trabalho em artefatos históricos (moldes, tipos, matrizes).

Em resumo, reforça-se aqui que a padronização é uma parte natural do desenvolvimento de qualquer língua escrita, está fortemente associada ao planejamento de corpus e é um ponto crítico para os esforços de revitalização linguística.

Considerando a revisão histórica dos impactos das tecnologias empregadas nas práticas de veiculação da escrita (apontando-se os prós e contras), assinala-se a consciência das implicações da padronização no contexto desta pesquisa, já que a padronização é um aspecto que pode proporcionar uma possível ampliação das práticas de escrita em línguas indígenas.

O próximo capítulo desta tese trata de questões que têm relação direta com a padronização da grafia de um sistema de escrita.

10. Ortografias

O objetivo desta seção é mostrar fatores envolvidos no desenvolvimento de ortografias para línguas ágrafas, a fim de ilustrar algumas questões envolvidas na comunicação escrita nas línguas indígenas brasileiras. Reúne-se também uma breve revisão de literatura sobre o assunto.

10.1 Referências acerca do desenvolvimento de ortografias para línguas ágrafas

Neste trabalho, foram consultadas algumas referências sobre o desenvolvimento de ortografias para línguas ágrafas; em particular do campo da linguística, no qual existe uma literatura considerável sobre o assunto. Além de Kenneth Pike (já mencionado na seção *Terminologia*), bastante citado no contexto do desenvolvimento de metodologias específicas para esse propósito, o parágrafo a seguir destaca algumas referências contemporâneas de trabalhos publicados nos últimos 15 anos.

Em *Orthographies in Papua New Guinea through the Years*, Ray Stegeman (2019, p. 272–274) faz uma breve revisão de literatura sobre o desenvolvimento de ortografias e apresenta um estudo de caso bastante atual no país que tem a maior diversidade linguística do mundo. A revisão de literatura do capítulo sobre ortografias de *Saving Languages: an Introduction to Language Revitalization*, de Lenore A. Grenoble e Lindsay J. Whaley (2006, p. 137–159) aborda distintas referências e apresenta o tema sob a ótica da preservação e revitalização de línguas ameaçadas. Em *Non-Linguistic Factors in Orthographies* (2011), Michael Cahill (também membro do SIL) lista uma série de fatores não linguísticos envolvidos no desenvolvimento de ortografias e, em *Factors in designing effective orthographies for unwritten languages*, em coautoria com Elke Karan (2008), aborda fatores para o desenvolvimento de ortografias efetivas para línguas ágrafas. Sob o ponto

de vista da codificação digital, Deborah Anderson, Rick McGowan e Ken Whistler (2005) descrevem, em *Unicode Technical Note #19: Recommendations for creating new orthographies*, uma série de recomendações para a criação de novas ortografias, com base no padrão Unicode de codificação de caracteres. Em *The Nature of Writing: A Theory of Grapholinguistics*, Dimitrios Meletis (2020) dedica um capítulo ao desenvolvimento de ortografias sob o ponto de vista da grafolinguística, i.e. o estudo linguístico da escrita. Por último, pode-se mencionar algumas publicações de referência que trazem informações sobre o assunto: *The Cambridge Handbook of Language Policy* (SPOLSKY, 2012), sob a ótica do planejamento linguístico; *The Cambridge Handbook of Endangered Languages* (AUSTIN; SALLABANK, 2011), sob a ótica das línguas ameaçadas; e o já mencionado *The Blackwell Encyclopedia of Writing Systems* (COULMAS, 1999), de caráter enciclopédico, citado em diversos trechos desta tese.

Além dos trabalhos supracitados, foram consultadas algumas obras sobre o desenvolvimento de ortografias para línguas indígenas brasileiras. Em *A guerra dos alfabetos: os povos indígenas na fronteira entre o oral e o escrito*, a professora Bruna Franchetto (2008, p. 59) relata “[...] experiências de pesquisa lingüística entre os Taurepáng, Macuxi, Wapichana e Kuikuro, em tempos, regiões e situações distintos, [...] [e] trata do confronto entre oralidade e escrita, quando a ‘ortografização’ de uma língua indígena, falada por uma sociedade de tradição oral, transforma e cristaliza sons e ditos em folhas de papel”. Em *A ortografia da língua Awetí*, Sebastian Drude, Waranaku Awete e Awajatu Aweti (2008) descrevem e fundamentam a ortografia da língua awetí. Sob a ótica do planejamento linguístico, no artigo *O que quer a linguística e o que se quer da linguística – a delicada questão da assessoria linguística no movimento indígena*, Gilvan Müller de Oliveira (1999, p. 26) observa “[...] para o fato de que o compromisso do lingüista envolvido com sociedades indígenas não pode ser, ao contrário da crença generalizada, o de elaboração de descrições das línguas indígenas, já que as descrições não são essenciais ao processo de criação de tradições escritas para línguas antes ágrafas”.

A consulta à literatura foi importante para o entendimento de fatores e critérios envolvidos no planejamento de uma ortografia. No entanto, é importante destacar que

o planejamento e desenvolvimento de ortografias não são o objetivo principal deste trabalho por algumas razões. Primeiramente, por se tratar de uma área que exige conhecimentos de fonética e fonologia, entre outros campos da linguística sobre os quais este autor não tem instrução adequada. Em segundo lugar, é importante lembrar que o escopo deste estudo abrange a análise de *línguas indígenas para as quais já existe uma (ou mais de uma) ortografia estabelecida e em uso*; logo, parte-se do princípio, aqui, de que as ortografias já existem, para que se possa trabalhar ferramentas para elas. Por último, mas não menos importante, deve-se salientar o fato de que, apesar da interdisciplinaridade, este é um trabalho da comunicação (mais especificamente, da comunicação visual, tipografia); assim, apesar das referências da linguística serem extremamente relevantes, o cerne da questão está na comunicação visual.

Em todos os trabalhos supracitados, é recorrente a atenção a questões e temas além dos fatores linguísticos, como fatores socioculturais, sociolinguísticos, políticos, tecnológicos e educacionais. Também são recorrentes as discussões acerca dos seguintes temas e termos: sub e super-representação; aceitação pelas partes envolvidas; transferibilidade; facilidade de aprendizagem; identidade; legibilidade e leiturabilidade; transparência; fatores tecnológicos etc. Esses temas serão discutidos nas próximas seções deste trabalho.

10.2 Critérios essenciais

Além dos autores supracitados, há ainda outros com diferentes opiniões sobre quais são os critérios essenciais para a avaliação de sistemas de escrita, assunto sobre o qual há diferentes abordagens. Dimitrios Meletis (2020, p. 183) faz uma análise e classificação mais sistemática de algumas dessas diversas abordagens e mostra, em uma tabela, as listagens de critérios de avaliação de sistemas de escrita utilizadas por diferentes autores. Os critérios variam de autor pra autor, apesar de haver interseções entre eles. Por essa razão, elegeu-se como fio condutor desta seção uma das abordagens mencionadas por Meletis, a de Michael Cahill, por tratar especificamente de fatores essenciais no desenvolvimento de ortografias (e não de escritas e sistemas de escritas de uma forma mais ampla) e por sintetizar muitos dos temas/critérios recorrentes encontrados na literatura.

Em *Factors in designing effective orthographies for unwritten languages*, Michael Cahill e Elke Karam (2008, p. 3) informam que recentemente houve um aumento do interesse sobre o tema, possivelmente impulsionado por fatores como o aumento da conscientização e interesse acerca das línguas ameaçadas, maior atenção aos direitos humanos e também da influência do padrão Unicode. A esses fatores, pode-se acrescentar também o aumento considerável, nas últimas duas décadas, ao acesso a computadores e dispositivos móveis. Além disso, Cahill e Karam destacam também os efeitos positivos que a alfabetização e uma cultura literária podem ter para a preservação dessas línguas. Ressaltam, no entanto, que o simples fato de uma ortografia existir não vai resolver os problemas de comunicação e que o desenvolvimento de uma ortografia é apenas uma parte dos programas de revitalização linguística. Por último, observam que, para sua efetividade, é preciso que uma ortografia seja linguisticamente correta, aceitável para todas as partes envolvidas, ensinável¹ e fácil de reproduzir. Esses quatro aspectos podem ser enquadrados, grosso modo, como os fatores científicos, políticos, educacionais e técnicos, e que frequentemente entram em conflito uns com os outros.

Em linhas gerais, os fatores acima listados por Cahill e Karam também aparecem em outros trabalhos, desde os textos mais antigos (e.g. PIKE, 1947, p. 208) até a literatura mais recente, acima mencionada. Fazendo uma interpretação mais aberta e subjetiva da tabela comparativa de Meletis (2020, p. 183), pode-se dizer que há um consenso de que, entre vários outros conceitos, a simplicidade, precisão, consistência e praticidade, além do equilíbrio entre fatores linguísticos e não linguísticos, são alguns daqueles que permeiam esses desenvolvimentos.

Conforme mencionado anteriormente, este texto procura fazer um resumo dos fatores mais importantes a serem considerados no desenvolvimento de ortografias para línguas ágrafas; as próximas seções deste texto tentarão descrever um pouco esse panorama no contexto desta tese.

1. Tradução literal do termo em inglês, *teachable*. Ou seja, que se pode ensinar, de fácil didática.

10.3 Fatores científicos: o ideal fonêmico

A expressão “fatores científicos”, no contexto do desenvolvimento de ortografias, está ligada aos fatores linguísticos propriamente ditos. Em linhas gerais, pode-se dizer que as análises linguísticas (científicas ou não científicas) são o seu ponto de partida.

Nesse sentido, conforme apontado por Cahill e Karam (2008, p. 4–5), por muitas décadas prevaleceu uma premissa bastante comum em trabalhos linguísticos, que consiste na representação fonêmica como o ideal ortográfico, promovido por linguistas como Kenneth Pike e Sarah C. Gudschinsky. Essa premissa demonstra a relação de proximidade entre análise fonêmica e desenvolvimento de ortografias, com base no princípio ideal de correspondência de um símbolo por fonema, um fonema por símbolo². Atualmente, entende-se que a análise fonológica, apesar de fundamental, não é o único fator determinante para o sucesso de uma ortografia, assim como a melhor solução não necessariamente é aquela em que há a correspondência de um símbolo por fonema, um fonema por símbolo. Os autores descrevem outros fatores linguísticos a serem levados em conta, como divisão de palavras, marcação de características suprasegmentais (e.g. marcação de tom), e ainda a importância da intuição do falante nativo no processo.

Por último, é importante reforçar o fato de que uma ortografia cientificamente correta não garante a sua efetividade; é preciso um equilíbrio aceitável entre os fatores linguísticos e os outros fatores envolvidos (CAHILL; KARAM, 2008, p. 2; COULMAS, 1999, p. 381; PIKE, 1947, p. 208).

10.4 Fatores políticos: regulação e restrições governamentais

Cahill e Karam (2008, p. 4) observam que há países nos quais há políticas regulatórias que, em geral, restringem as opções no desenvolvimento de ortografias (i.e. restrição ao repertório de grafemas) ou de uma escrita (e.g. não se permite o uso da escrita árabe). Nesse sentido, observam que, além de linguisticamente correta, uma ortografia deve estar de acordo com as leis de um país e ressaltam que, em muitos casos, as ortografias devem ser aprovadas pelas autoridades governamentais. Os autores reúnem alguns

2. Ver Pike, 1947, p. 208–209.

exemplos de como políticas linguísticas podem influenciar as ortografias das línguas indígenas: em Gana, o Bureau of Ghana Languages não permite o uso de marcação de tom em nenhuma ortografia; em 1979, Camarões estabeleceu um alfabeto unificado, e novas ortografias devem estar em conformidade com esse alfabeto; na República Centro-Africana, o governo precisa aprovar novas ortografias; enquanto na Etiópia, a aprovação depende de uma das várias agências do país.

No Brasil, enquanto há instituições que regulam a norma e a codificação da língua portuguesa, não há nada de específico sobre as línguas indígenas, de imigração, crioulas e afro-brasileiras. A iniciativa mais próxima disso é o Inventário Nacional da Diversidade de Linguística (INDL), instrumento oficial de identificação, documentação, reconhecimento e valorização das línguas faladas pelos diferentes grupos formadores da sociedade brasileira (GARCIA et al., 2016).

É importante mencionar ainda que todas as ortografias são políticas, apesar de algumas serem mais politizadas que outras. Às vezes, uma ortografia pode ficar associada a um determinado grupo (e.g. católico x evangélico), e essas associações podem ser de caráter religioso, político³, entre outras. Essas afiliações devem ser levadas em conta desde o início do desenvolvimento ou em revisão ou reforma de uma ortografia, já que a introdução de uma mudança posterior, em um sistema em uso, pode ser extremamente complicado. É necessário incluir todas as partes interessadas no planejamento ortográfico e nas decisões desde o início. Por último, deve-se considerar não somente os fatores linguísticos, mas também examinar todas as ortografias anteriores, quem as desenhou, quem as apoia e usa (CAHILL; KARAM, 2008, p. 10-11)

10.5 Aceitação pelas partes envolvidas

Entre os autores citados nesta seção sobre ortografias, há consenso de que a aceitação pelas partes envolvidas é um fator preponderante para a efetividade de uma ortografia (CAHILL, 2011, p. 3; CAHILL; KARAM, 2008, p. 12; COULMAS, 1999, p. 12; FRANCHETTO, 2008, p. 38–39; LÜPKE, 2011, p. 330; MELETIS, 2020, p. 352; STEGEMAN, 2019, p. 273).

3. Bruna Franchetto (2008), por exemplo, reporta o uso de grafemas como emblema/insígnia político.

Cahill e Karam (2008, p. 12) observam que o envolvimento de uma comunidade no desenvolvimento de uma ortografia é crucial para sua aceitação. É muito mais provável que uma ferramenta da qual a comunidade participou ativamente da criação seja implementada e de fato utilizada pela comunidade.

Por último, Cahill observa que uma ortografia é uma expressão de identidade de um povo, e que este pode aceitá-la ou rejeitá-la com base em fatores sociolinguísticos. Se o grupo não a aceita, não importa o quão linguisticamente correta ela seja, ela não será usada. Nesse sentido, “o que o povo quer” não é somente mais um fator; é o fator mais crítico na aceitação de uma ortografia (CAHILL, 2011, p. 3).

10.6 Facilidade de aprendizagem

Do ponto de vista educacional, Cahill e Karam (2008, p. 7-9) destacam que um dos critérios mais relevantes para sistemas de escrita é a facilidade de aprendizagem. Quando há um descompasso entre a linguagem falada e a escrita, tornam-se mais complicados o ensino e a aprendizagem da leitura e da escrita, levando aprendizes a uma condição de frustração, de falta de motivação e a um tempo de aprendizado elevado. Observam ainda (2008, p. 12) a relevância de ferramentas pedagógicas apropriadas, ressaltando que somente uma boa ortografia não é suficiente para que a leitura exista de forma efetiva; é preciso instrução e materiais didáticos adequados.

10.7 Transferibilidade

Uma outra questão importante está relacionada à transferibilidade. Na literatura em inglês, é comum o uso do termo *bridging* (fazer a ponte; no caso, entre duas línguas), usado nesse contexto.

A transferibilidade ou, grosso modo, a transferência das habilidades de escrita e leitura da língua indígena para a língua oficial (e vice-versa) também são extremamente importantes no processo, visto que, além de suas línguas originárias, as pessoas frequentemente querem aprender outras línguas. Além disso, os governos também costumam querer incorporar a população indígena à vida da nação, o que pode exigir o conhecimento de um idioma nacional ou regional (CAHILL; KARAM, 2008, p. 8-9).

Pensando na transferibilidade, pode-se levantar como requisito básico a especificação de símbolos e convenções que facilitem a transferência para outras línguas. A premissa é que, ao harmonizar dois sistemas de escrita, minimiza-se o esforço necessário para transferir habilidades de leitura e escrita de um idioma para outro (CAHILL; KARAM, 2008, p. 8). Um exemplo disso é o uso dos grafemas <c> e <qu> para representar o fonema /k/, na América Latina (CAHILL; KARAM, 2011, p. 3). Em um contexto no qual o contato é com a língua inglesa, entre os grafemas <qu> e <k>, provavelmente <k> seria o escolhido.

A pressão para a harmonização com o idioma de comunicação mais ampla pode resultar em sub ou super-representação, entre outros problemas. Nesse sentido, é importante reforçar que questões de preservação de identidade frequentemente têm precedência sobre as considerações de transferência, inclinando a balança para convenções que diferem das línguas de prestígio (CAHILL; KARAM, 2008, p. 8–9).

10.8 Profundidade ortográfica, sub-representação e super-representação

A literatura utiliza-se da expressão “profundidade ortográfica” (*orthographic depth*) para determinar o grau de abstração inerente a um sistema de escrita. Em linhas gerais, ortografias rasas (*shallow orthographies*) têm uma boa correspondência entre fonemas e grafemas⁴, enquanto em ortografias profundas (*deep orthographies*) ocorre justamente o oposto (CAHILL; KARAM, 2008, p. 7). Florian Coulmas (1999, p. 379-380) explica que em ortografias rasas, a correspondência grafema–fonema é menos polivalente; logo, mais transparente. Ele explica ainda que o fato de algumas ortografias serem mais rasas não significa que a elas se aplica o princípio de “um som, um símbolo”. Como na maioria das línguas os fonemas são mais numerosos que as letras, é comum, em sistemas alfabéticos, o uso de dígrafos e trígrafos como forma de aumentar o repertório gráfico.

Ortografias profundas são o resultado de fenômenos como sub-representação (ou subdiferenciação) e super-representação (ou superdiferenciação), inconsistências entre

4. Pode-se assim associar estas dualidades entre profundo–raso e transparente–opaco como similares ou análogas, conforme descrição de Coulmas (1999, p. 11) e Cahill e Karam (2008, p. 7–9).

som e símbolo (fonema–grafema), entre outros fatores. As ortografias profundas carregam consigo um custo para os aprendizes e educadores. Aprendizes podem vivenciar um alto grau de frustração, levando à falta de motivação, ao fracasso e a um tempo de aprendizado elevado. E, para o sistema educacional, pode representar um aumento nos investimentos de tempo, recursos humanos e materiais (CAHILL; KARAM, 2008, p. 7).

Em linhas gerais, a sub-representação ocorre quando há menos grafemas que fonemas. Isso pode resultar em dificuldades para a leitura, pois um leitor pode ter dificuldade para determinar a pronúncia de um grafema, já que este mesmo grafema é usado para mais de um fonema (CAHILL; KARAM, 2008, p. 7). Na língua portuguesa, a sub-representação ocorre com os grafemas <e> e <o> em finais de palavras paroxítonas ou proparoxítonas, pois estas assumem o mesmo som das letras <i> e <u>, respectivamente. O mesmo acontece com a letra <a> no inglês, que pode ter diversos sons.

No caso da super-representação ocorre justamente o oposto: há mais grafemas que fonemas. Assim sendo, há maior dificuldade para escrita, porém maior facilidade para o aprendizado da leitura (CAHILL; KARAM, 2008, p. 7–8). Um bom exemplo de super-representação ocorre na língua portuguesa, na qual o som [s] (como na primeira letra da palavra sapo) pode ser representado graficamente de seis formas diferentes: <s>, <c>, <ss>, <sc>, <xc> e <ç>.

Do ponto de vista linguístico, a sub-representação está longe do ideal. No entanto, pode haver fatores sociolinguísticos que levem a ela. Por exemplo: pode haver pressão da língua nacional/dominante para que haja uma maior semelhança entre as duas ortografias. Pode ocorrer também que as partes interessadas não queiram grafemas ou caracteres adicionais, que não ocorrem na língua nacional/dominante, por uma variedade de motivos (CAHILL; KARAM, 2008, p. 7).

Outro caso clássico de sub-representação ocorre quando há a omissão da marcação de tom. Considera-se que a marcação de tom tem a importante função de estabelecer contraste. No entanto, entende-se também que contrastes tonais podem ser percebidos pelo contexto, em especial por leitores mais experientes, o que justifica a omissão de marcas de tom. Observa-se que não se pode simplesmente presumir a omissão; essa questão precisa ser testada (CAHILL; KARAM, 2008, p. 8).

10.9 Legibilidade e leiturabilidade

Em relação à influência da aparência visual dos caracteres no processo de leitura, Cahill e Karam (2008, p. 9) levantam algumas questões que podem influenciar a legibilidade e leiturabilidade. Nesse sentido, é importante mencionar que a bibliografia sobre tipografia tem muito a contribuir para essa discussão; para citar dois exemplos, os livros *Enquanto você lê* (2016) e *Theory of type design* (2018), ambos de autoria do professor Gerard Unger.

Em primeiro lugar, para que uma ortografia seja efetiva, é importante que o repertório de caracteres seja visualmente discernível; caracteres muito parecidos podem dificultar a leitura. Segundo Cahill e Karam (2011, p. 4) Pesquisas indicam que letras espelhadas ou rotacionadas não são ideais para leitores em fase de aprendizagem. Além disso, a superposição ou empilhamento de diacríticos, apesar de tecnicamente possível, é desaconselhada, já que a reprodução do texto em certas condições técnicas (e.g. em tamanhos pequenos ou em baixa resolução) pode afetar bastante a legibilidade (i.e. a facilidade de distinção dos caracteres). Recomenda-se, como boa prática, o uso de apenas um diacrítico por letra. Apesar de algumas línguas (como o vietnamita) utilizarem diacríticos empilhados, do ponto de vista da leiturabilidade e legibilidade, essa não é considerada uma boa prática.

Readability – aim for good *visual discrimination*, little “cluttering”
Extreme examples:

n n̂ n̄ n̆ n̈	too many symbols resemble each other
á â ã ä	too many stacking diacritics

Figura 99: Critérios de legibilidade para ortografias. Fonte: CAHILL; KARAM, 2011, p. 4

10.10 Identidade: todas as ortografias são políticas

Entre os fatores sociolinguísticos, talvez a questão da identidade seja a de maior relevância para esta discussão. As frases “todas as ortografias são políticas” e “uma ortografia é uma expressão de identidade de um povo” são recorrentes nos trabalhos de Cahill.

Hoje, o multilinguismo faz parte do ambiente social de muitas línguas ao redor do mundo, no qual poucas comunidades linguísticas ainda existem isoladamente. Nesse contexto, frequentemente há influência das línguas nacionais ou oficiais, ou de línguas vizinhas, sobre as quais pode haver diversas atitudes (CAHILL; KARAM, 2008, p. 10).

A população local pode querer que a ortografia se pareça com a ortografia da língua majoritária ou oficial, ou que não se pareça com a língua de um grupo rival, conforme os exemplos a seguir. Em Gana, por exemplo, o grafema <r> ocorre na maioria das ortografias locais, contrastantes ou não, sob a influência do inglês, língua oficial do país. Ainda em Gana, os konkombas usam o grafema <ln> para um som lateral nasalizado, enquanto os kombas, um grupo vizinho e linguisticamente próximo, utilizam o grafema <nl> para o mesmo som, ficando claro que essas escolhas representam marcas distintivas de identidade. Um terceiro exemplo está relacionado ao fenômeno de hiperfragmentação em duas cidades de etnia Trique, em Oaxaca, no México: embora ambas as cidades falem o mesmo dialeto, cada uma tem uma ortografia diferente como forma de afirmação de identidade (CAHILL; KARAM, 2008, p. 10).

Bruna Franchetto (2008, p. 43–52) também descreve com detalhes questões relacionadas ao desenvolvimento das ortografias de algumas línguas; em especial do wapihana, em Roraima, e do kuikuro, no Xingu, das quais participou ativamente do desenvolvimento junto às comunidades. No caso do wapihana, havia uma ortografia em uso, criada por missionários protestantes da Guiana e levada posteriormente para Roraima. Franchetto relata não somente inconsistências na ortografia dos missionários anglófonos, como também explica a dificuldade de transferibilidade para com a língua portuguesa. Nesse sentido, ela explica o processo de reforma, substituindo grafemas comuns ao inglês, como <sh> e <k>, por outros, comuns ao português, como <ch>, <c> e <qu>, respectivamente. Obviamente, não se trata somente de uma escolha de grafemas, mas também da não associação a um determinado grupo religioso.

10.11 Dialectos

A grande maioria das línguas tem dialetos, variações da língua de acordo com a localidade ou população, frequentemente com diferenças fonético-fonológicas. Ao tratar de

um grupo linguístico com diferentes dialetos, um dos desafios é conseguir desenvolver uma ortografia que sirva a todos. Apesar de uma ortografia unificada ser desejável, nem sempre isso é possível (CAHILL; KARAM, 2008, p. 10).

Existem duas principais abordagens para uma mesma ortografia. Na abordagem unidialetal⁵, toma-se como referência o dialeto de maior prestígio (ou mais usado), e outros se adaptam a ele. Na abordagem multidialetal, toma-se como referência elementos de vários dialetos, de forma que um único grupo não seja favorecido (CAHILL; KARAM, 2008, p. 10).

A relação entre ortografias e dialetos tem particular relevância para o estudo da tipografia das línguas indígenas brasileiras por dois motivos. Primeiramente, há variações dialetais em diversas línguas brasileiras, com suas respectivas variações ortográficas; é o caso da língua Kaingang⁶, com as variantes ortográficas do Kaingang e Kaingang Paulista. Aqui, o termo *dialeto gráfico*, conforme proposto por Robert Bringhurst (BRINGHURST, 2002, p. 12), tem uma aplicação apropriada. Em segundo lugar, dado o recente estado de desenvolvimento escrito de muitas línguas brasileiras, é comum que exista variação e instabilidade ortográfica.

10.12 Tecnologia: fontes, teclados e dispositivos

Uma questão recorrente nas discussões sobre desenvolvimento de ortografias está relacionada ao uso da tecnologia de escrita, em particular no que se refere à codificação, às fontes e aos teclados. Esse assunto será melhor abordado em outras partes deste trabalho, especialmente nos capítulos *Codificação e linguagem, Metodologia e Desenvolvimento*.

Dessa maneira, questões como a forma de codificação e dos métodos de entrada apropriados, assim como da disponibilidade de fontes com conjuntos de caracteres compatíveis com determinada(s) línguas, serão melhor abordadas nos capítulos mencionados no parágrafo anterior.

5. Cahill utiliza os termos *uni-lectal* e *multi-lectal*. Meletis (2020, p. 369–372) emprega, como sinônimos, os termos *unidialectal* e *multidialectal*.

6. Conforme descrição do Prof. Wilmar da Rocha D'Angelis, da Unicamp, no trabalho de levantamento de dados desta pesquisa.

Com a conclusão deste capítulo, conclui-se também o que, de certa forma, se considera o aporte teórico desta tese, formado pelos capítulos *Codificação e linguagem, Padronização e Ortografias*.

A partir das informações reunidas até aqui (desde os capítulos introdutórios até os três últimos capítulos supracitados), o próximo capítulo amplia um pouco esta discussão, examinando, a partir de uma perspectiva tipográfica, as condições técnicas e problemas relativos à composição de texto, codificação das ortografias e uso de diacríticos nas línguas indígenas brasileiras.

11. Características tipográficas das línguas indígenas brasileiras

O objetivo deste capítulo é examinar, a partir de uma perspectiva tipográfica, as condições técnicas e problemas relativos à composição de texto, codificação das ortografias e uso de diacríticos nas línguas indígenas brasileiras.

Este capítulo é dividido em três seções. A primeira delas delimita o escopo da análise. A segunda explica como as ortografias das línguas indígenas brasileiras utilizam o padrão Unicode. E a terceira tenta ilustrar alguns problemas de composição tipográfica em línguas indígenas brasileiras.

11.1 Escopo

Este capítulo tem como referência inicial a dissertação de mestrado deste autor¹. Em resumo, a dissertação faz uma análise das características tipográficas e propõe algumas diretrizes técnico-projetuais para composição de texto e design de tipos para línguas indígenas brasileiras. Nesse sentido, grande parte desta seção consiste em uma versão atualizada, em português, do texto *Typographic analysis of Brazilian languages*, capítulo 3 da dissertação supracitada. A decisão pela tradução e atualização deste conteúdo tem algumas razões.

Primeiramente, porque este conteúdo ainda não foi devidamente traduzido para o português. Segundo, porque quase não há análises das características tipográficas das línguas indígenas brasileiras nos moldes propostos por este autor; o único trabalho conhecido de análise das características tipográficas das línguas indígenas brasileiras é

1. DIETZSCH, Rafael. **Typefaces for Brazilian indigenous languages**. Dissertation (Master of Arts in Typeface Design), University of Reading, Reading, UK, 2012.

o artigo pioneiro *Notas sobre tipografias para línguas indígenas do Brasil*, de autoria de Kollontai Cossich Diniz (2007). Além disso, as análises são essenciais para a compreensão desta tese, visto que é por meio delas que são explicados e ilustrados os principais problemas de codificação das línguas indígenas brasileiras. Por último, porque o trabalho de análise realizado na dissertação ainda é representativo; alguns problemas de composição e de codificação ainda persistem, mesmo quase após uma década.

As análises do referido capítulo se baseiam em fontes primárias coletadas durante o curso de mestrado: publicações do período de c. 1990–2010, muitas delas financiadas pelo governo brasileiro e, em alguns casos, cofinanciadas pelo setor privado. Essas publicações são, em sua grande maioria, material educativo (e.g. cartilhas, livros escolares) e livros de literatura. É importante mencionar que no trabalho de catalogação ortográfica desta tese (v. capítulo *Metodologia*), analisou-se, pela segunda vez, quase todas as publicações que já haviam sido analisadas na pesquisa de mestrado,

As fontes primárias da análise compõem um cenário muito heterogêneo: muitas línguas para uma amostra relativamente pequena de publicações, dificultando a classificação das fontes primárias por línguas, famílias ou outros tipos de agrupamento². Tampouco se pensou em fazer uma tabulação dos dados analisados. Naquele momento, o que mais fazia sentido era mostrar um panorama geral da situação, ilustrando os problemas tipográficos mais recorrentes.

Em praticamente todas as publicações analisadas, observa-se a presença da língua portuguesa, de formas variadas. De maneira geral, elementos pré e pós-textuais (e.g. ficha catalográfica, créditos, ficha técnica) são compostos em português, ocasionalmente em línguas indígenas ou com tradução nas línguas indígenas. Pode-se dizer que muitas dessas edições são integralmente bilíngues, pois apresentam o mesmo conteúdo de forma paralela (i.e. em português e uma língua indígena). Há trabalhos compostos integralmente em línguas indígenas e, ainda, casos nos quais a presença da língua portuguesa se resume aos textos de introdução e/ou apresentação.

2. A metodologia do artigo de Diniz (2007, *passim*) é consistente nesse sentido e classifica as publicações em agrupamentos por família ou língua.

A grande maioria das publicações analisadas foi originada em aplicativos digitais de editoração eletrônica e fontes tipográficas digitais. Há, no entanto, exemplos de publicações que indicam o uso de outras tecnologias de composição tipográfica, como máquinas de escrever e fotocomposição.

Por último, é importante observar a delimitação do corpus desta tese, uma vez que a dissertação de mestrado deste autor é baseada no mesmo recorte.

11.2 Codificações e conjuntos de caracteres nas ortografias das línguas indígenas brasileiras

Em linhas gerais, pode-se dizer que o uso intensivo de diacríticos é uma característica marcante da ortografia de muitas línguas indígenas brasileiras. Uma das razões disso é que o alfabeto latino não é suficiente para codificar o abundante número de sons de muitas dessas línguas (por exemplo, muitas vogais). Conforme já explicado na seção *Terminologia*, diacríticos são um recurso muito usado para a ampliação de alfabetos.

A nota técnica *Recommendations for creating new orthographies* (ANDERSON et al., 2005), do Unicode, sugere que, no processo de criação de novas ortografias, é particularmente vantajoso eger caracteres que possam ser facilmente encontrados em um computador, visando à manipulação de documentos eletrônicos sem a necessidade de fontes ou softwares especiais. A recomendação para a utilização de caracteres já codificados pelo Unicode é enfatizada, e o contrário, fortemente desencorajado. A nota ressalta ainda que, quando se projeta uma ortografia com caracteres completamente novos (não codificados pelo Unicode), são muito baixas as chances de que haja suporte em softwares e fontes existentes para a ortografia completa. A aprovação e implementação de novos caracteres no Unicode Consortium pode levar anos; sem contar com possíveis atrasos nesse processo. Ademais, leva-se um certo tempo até que novos caracteres sejam incorporados às fontes. E, por último, nada garante que a proposta de implementação de um novo caractere seja aprovada. Além do uso de caracteres já codificados pelo Unicode, há ainda orientações para que se evite a mistura de escritas diferentes (e.g. grega e latina) e para que se considere a semântica apropriada (não apenas a forma certa).

A partir dessas recomendações do Unicode, este trabalho faz uma breve análise³ de como as ortografias das línguas indígenas brasileiras se enquadram no padrão Unicode. Para efeitos de análise, serão considerados, principalmente, os caracteres acentuados utilizados em cada língua e o modo como estão distribuídos nos blocos do Unicode, partindo dos casos mais comuns para os mais complexos.

Os casos mais comuns envolvem as línguas que utilizam-se somente de caracteres pré-compostos do conjunto de 256 caracteres da codificação ISO-8859-1, ou Latin-1 (Fig.100), que é equivalente aos dois primeiros blocos do Unicode (Basic Latin e Latin-1 Supplement)⁴ e que compreende todos os caracteres necessários para composição de textos em português e da maioria das línguas da Europa Ocidental. Nesses casos, há milhares de fontes com as quais se pode trabalhar, visto que esse conjunto de caracteres é considerado o mínimo para fontes de texto. As figuras 101 e 102 mostram exemplos de ortografias que utilizam somente caracteres dos blocos Basic Latin e Latin-1 Supplement.

Neste ponto, é importante que se entenda que os blocos do Unicode são compostos de grandes conjuntos de caracteres (e.g. 96, 128 ou 256 caracteres). No entanto, a utilização de somente um ou outro caractere de um bloco é prática comum. Em outras palavras, isso significa que, ao se utilizar um caractere de determinado bloco, não é obrigatório que se utilize o bloco inteiro.

3. Maiores detalhes dos conjuntos de caracteres utilizados em cada língua serão apresentados nos capítulos *Metodologia e Desenvolvimento*.

4. Descrição completa destes blocos: Co Controls and Basic Latin (UNICODE CONSORTIUM, 2020b); CI Controls and Latin-1 Supplement (UNICODE CONSORTIUM, 2020c). Ver seção *O surgimento do padrão ISO 8859: transição para o padrão de oito bits*, no capítulo *Codificação e linguagem*.

CURRENT CHARACTER SET 256															EXPORT
NUL	SOH	STX	ETX	EOT	ENQ	ACK	BEL	BS	TAB	EOL	VT	FF	CR	SO	SI
DLE	DC1	DC2	DC3	DC4	NAK	SYN	ETB	CAN	EOM	SUB	ESC	FS	GS	RS	US
SP	!	"	#	\$	%	&	'	()	*	+	,	-	.	/
0	1	2	3	4	5	6	7	8	9	:	;	<	=	>	?
@	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
P	Q	R	S	T	U	V	W	X	Y	Z	[\]	^	_
`	a	b	c	d	e	f	g	h	i	j	k	l	m	n	o
p	q	r	s	t	u	v	w	x	y	z	{		}	~	DEL
PAD	HOP	BPH	NBH	IND	NEL	SSA	ESA	HTS	HTJ	VTS	PLD	PLU	RI	SS2	SS3
DC5	PU1	PU2	STS	CCH	MW	SPA	EPA	SOS	SGC	SCI	CSI	ST	OSC	PM	APC
NB SP	ı	ç	£	¤	¥	¦	§	¨	©	ª	«	¬	SHY	®	¯
°	±	²	³	´	µ	¶	·	,	¹	º	»	¼	½	¾	¿
À	Á	Â	Ã	Ä	Å	Æ	Ç	È	É	Ê	Ë	Ì	Í	Î	Ï
Ð	Ñ	Ò	Ó	Ô	Õ	Ö	×	Ø	Ù	Ú	Û	Ü	Ý	Þ	ß
à	á	â	ã	ä	å	æ	ç	è	é	ê	ë	ì	í	î	ï
ð	ñ	ò	ó	ô	õ	ö	÷	ø	ù	ú	û	ü	ý	þ	ÿ

Figura 100: Co Controls and Basic Latin. CI Controls and Latin-1 Supplement. Amostra gerada no aplicativo online Charset Buidar. Fonte: EIGENDORF; ACSANY, 2016.

Apawali apaita tsiáli lia liboyatá depiatóa ikoli irsio. Lia, pida ita iyo, até liokawa apanhaa kalita liko, yaka lidzakale iyodza.

Neeni pida linoa liñanawa iiniri. Makhewipe pida nha liitsale, lipeko kadana nha liitsale. Liapida lhira likapa, lhima kilikili meka linoka ikoli makadali pida, lhireta pidani lipara karo ni ettiyo. Není pida idza ino likalhe, lídia khapani pida lioka oñairhe lhita naixa nhá liwíni, lhínikoita lidzana, matsiakena liorokoita lhita líñawa kadzo, apaita itsaletaka pia.

Figura 101: Amostra de texto na língua baniwa-coripaco, utilizando somente caracteres do bloco Latin-1. Fonte: FONTES; DINIZ; TRUJILLO, 2005, p.58.

Onok tu agu ompan yen? Werengo angrī (*micróbio*) kenanpa tupi, keni, agulī yen.

Mantan eret (*bactéria*). Angrī wīnperan meganopnang werengo, angrī ruman angnang gwen, ketpotke menaptxan wīripne poryan tumogamo, tupi mīragri.

Otīt ñlon wīmtagri, wot, ungwo, yay wilī, menaptup mapyanketman oen.

Araru kenang kurotxepyankepra kutkerup koga? Kurotximtagrikenang, kerup inangman wīmtagri gwen apyam, wīnkinangman, tīmrongmo arawīremkom kenang man werengo angrī, gwen aktowoge tīngkit kutketkenentam.

Figura 102: Amostra de texto na língua ikpeng, utilizando somente caracteres do bloco Latin-1. Fonte: WÜRKER, 2003c.

O primeiro nível de limitação para composição tipográfica está relacionado ao uso de alguns caracteres pré-compostos dos blocos Latin Extended-A (Fig.103), e.g. <ũ> ou <ĩ>; Latin Extended-B (Fig.104), e.g. <ç>; e Latin Extended Additional (Fig.105), e.g. <ẽ>. Caracteres desses blocos são mais comumente encontrados em fontes padrão de sistemas operacionais, como Windows (e.g. Arial, Calibri, Lucida, Times New Roman), MacOS (e.g. San Francisco), Android (e.g. Roboto, Noto) e em fontes que fazem parte de pacotes de aplicativos, como Adobe InDesign (e.g. Minion Pro, Myriad Pro). Além destas, há outras alternativas de fontes, tanto gratuitas como pagas, em que se pode encontrar grande parte desses caracteres. Há, inclusive, boas alternativas de famílias, com abrangentes conjuntos de pesos e estilos. A Fig.106 mostra exemplo de ortografia que utiliza caracteres de mais de um dos blocos listados acima.

do advento da composição dinâmica⁶. Além de ser praticamente impossível encontrar fontes que contenham esses caracteres pré-compostos, há também diversos problemas relacionados às dificuldades técnicas para este tipo de composição, demonstrados na próxima seção deste capítulo.

Há quem afirme que caracteres como < ã > são *não codificados* ou que não têm um ponto de código, afirmações que não são exatamente corretas. O que ocorre é que o < ã > não tem uma forma pré-composta codificada; i.e. não tem um ponto de código como caractere pré-composto. Por isso, é comum que se diga que o < ã > não é codificado, o que não é verdade. O < ã > é formado por uma sequência de dois pontos de código: U+0067 < g > e U+0303 < ~ >. Neste capítulo, assumiremos que a expressão *não codificado(s)* é equivalente a caractere que não tem uma forma pré-composta codificada.

2. Jeikove porã ha umi tembiapo, tesãi, tekombo'e
Tetãygua omomorãva, pytyvõpópe ha oñondivepa,
péa va'erã hi'ãva ñepyrumby ojeheko mopu'ã ikatu
haãuã ohupyty opaite rupi umi ha'ekuéra oiko hápe.
Umi tembiaporã jeiko hapegua ñemopu'arãva avei
osẽ va'erã umi tetãyguágui ikatu haãuã oipytyvo,
oiko porãve haãuã.

Figura 108: Amostra de texto na língua guarani kaiowá. Pode-se perceber o caracteres dos blocos Latin-I, Latin Extended-A e Latin Extended Additional. Destaca-se o uso de marcas combinadas no < ã >. Fonte: DE LIMA, 2012, p.19.

6. V. *Diacríticos: métodos de entrada e características estruturais*, no capítulo *Codificação e linguagem*.

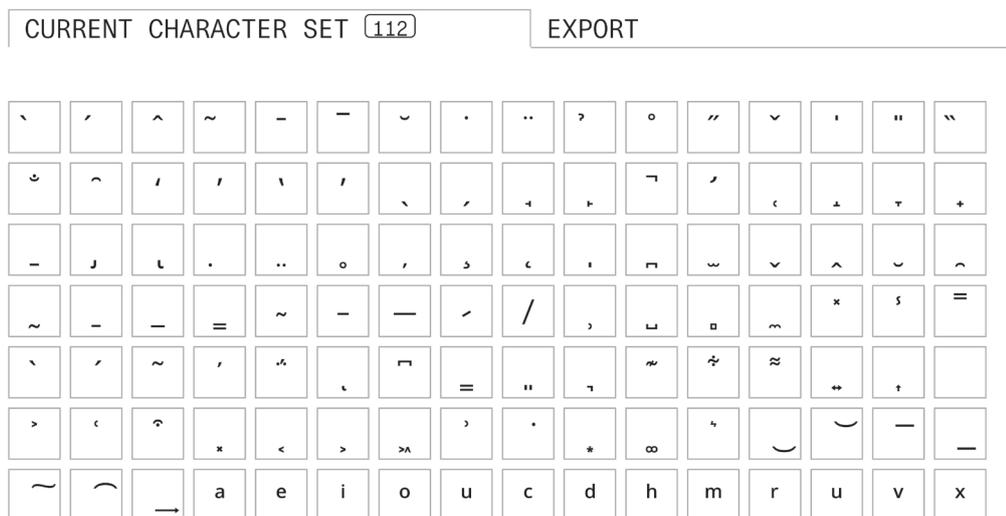


Figura 109: Combining Diacritical Marks. Amostra gerada no aplicativo online Charset Builder. Fonte: EIGENDORF; ACSANY, 2016.

Uma vez levantados todos os blocos do Unicode, é importante mencionar que há idiomas que utilizam caracteres de todos os blocos acima mencionados, além de marcas combinadas, conforme ilustra a Fig. 110 a seguir.

Tiatopure wimagu utinañio tiri akūwionekohã tigu kū paku.
 Wisopeputo kugã utinukū tiritabere wimagu sukuberupe atigu watī. Tora kure we oma wahoīgu. Kure terora kutaputi wiō kutuagura tigu kū ekeāputogã tū tū tū tū hīri.
 Tero tiarisiro basuka hīra sa. Watīarã tero tiya hī iñanunuse tihīra. Watiã sañari kope buaira. Titoanuko bia sikapoka newa tihīra buseira. Busemi pastiro metaragã witiati sugera. Kuãre pasiakuhã tira. Biana kuã buse tiri witiatirare. Kuã dia petiarisiro witiatigu buku saha. Kū nituogura nigu bayiro basuka yarigu. Teroti siāpetihaīra.

Figura 110: Amostra de texto na língua tuyuka. Caracteres dos blocos Latin-1, Latin Extended-A, Latin Extended-B, IPA Extensions e Latin Extended Additional, além caracteres acentuados com marcas combinadas. 70% do tamanho real. Fonte: CABALZAR, 2002, p.14.

Antes de finalizar esta seção, é importante enfatizar que a renderização correta, especialmente nos casos de glifos formados por letra base e marca(s) combinada(s), depende de alguns fatores principais, que serão decisivos no resultado final:

- Pontos de código com a semântica apropriada, organizados de acordo com os princípios do texto puro e ordem lógica.
- Teclados preparados para marcas combinadas; aplicativos e sistemas operacionais que reconheçam essa forma de entrada.
- Fontes com recursos OpenType apropriados (GPOS).
- Posicionamento e configuração corretos das âncoras nas letras base e nos diacríticos.
- Diacríticos com relações harmônicas de proporção e peso com o restante da fonte.
- Mecanismo de modelagem atuando no nível do sistema operacional ou no aplicativo.

Alguns desses fatores serão discutidos com detalhes nos capítulos *Metodologia e Desenvolvimento*.

11.3 Problemas tipográficos e de codificação em línguas indígenas brasileiras

Esta seção tem como objetivo ilustrar alguns problemas de codificação e de composição tipográfica em línguas indígenas brasileiras, problemas estes particularmente relacionados ao uso de diacríticos. A ausência de alguns caracteres em fontes digitais e a codificação incorreta, entre outros fatores, podem levar a resultados inconsistentes, tanto do ponto de vista da codificação como do ponto de vista tipográfico. Tais problemas poderiam ser evitados com o uso de tecnologias e métodos apropriados.

Destaca-se que, apesar de muitos dos exemplos serem de publicações da década passada, é importante a atenção a duas questões: Primeiro, observa-se que muitos desses problemas são uma questão superada, já não existem, tendo em vista as tecnologias atuais. No entanto, muitos deles persistirão se não tratados de maneira correta (codificação correta, semântica apropriada, fontes adequadas etc.).

Os casos mais problemáticos documentados estão relacionados ao uso de letras acentuadas para as quais não existe uma forma pré-composta codificada no Unicode, i.e. caracteres não codificados. Como essas pré-combinações de letra e diacrítico não

existem em praticamente nenhuma fonte (pelo menos na forma de caractere pré-composto), pode parecer que sua renderização de forma correta é impossível, o que não é verdade. Conforme já explicado anteriormente (v. *Codificação e linguagem*), o método correto de composição dessas combinações é por meio da utilização de um teclado personalizado, com o qual é possível combinar uma letra base com uma marca combinada.

Devido aos problemas e inconsistências encontrados na pesquisa, acredita-se na hipótese de que, no Brasil, os profissionais do campo da comunicação visual e do meio editorial utilizam pouco esse método, uma vez que não se trata de algo essencial para composição do português e outras línguas europeias mais usadas. Logo, não se trata de uma habilidade obrigatória para a atuação no mercado. Esses fatores, em conjunção com a falta de conhecimento sobre tipografia para línguas indígenas, resultam em diversos tipos de soluções improvisadas para a composição de texto nessas línguas.

É muito difícil descrever com precisão a origem dos métodos usados para a composição dos caracteres com diacríticos nas publicações analisadas. A partir de algumas evidências de problemas de alinhamento (entre letra e diacrítico), pode-se presumir que há algum tipo de ajuste de posicionamento manual⁷, improvisado (e.g. arrastar o diacrítico, à força, até o topo da letra), e também de substituição de caracteres ausentes por outros caracteres de forma similar.

Tendo em vista as questões levantadas até aqui, esta seção tenta ilustrar, a seguir, uma série de problemas tipográficos encontrados na pesquisa; a maioria deles relacionados ao uso de caracteres não codificados.

11.4 Substituições de caracteres e diacríticos

Para profissionais que já desenvolveram um projeto gráfico de publicação, deparar-se com algum caractere faltando em uma determinada fonte não é algo inédito. Muitas vezes, para contornar esse problema, a solução é a escolha de uma outra fonte ou família que ofereçam um conjunto de caracteres adequado. No entanto, em algumas das

7. Procurou-se evitar o termo *gambiarra* neste trabalho, apesar de ser difícil de achar outro mais adequado.

publicações analisadas, em vez de se procurar uma outra fonte, optou-se por uma solução improvisada para a substituição do glifo faltante.

É importante ressaltar que, ao se substituir um caractere por outro ou ao se realizar ajustes manuais de posicionamento, o texto não estará codificado corretamente, terá a semântica incorreta, e o resultado visual provavelmente será inconsistente. Por exemplo: ao trocar o til <~> pelo til ASCII <~>, um caractere de forma similar, pode ser que se consiga ler o texto mesmo com a forma incorreta do diacrítico. No entanto, ao copiar e colar esse texto em outro programa ou ao substituir a fonte, as chances de a combinação funcionar são pequenas.⁸

Há casos em que a semântica da codificação é correta, mas a aparência é errada. Por exemplo: a sequência digitada é correta, mas a fonte utilizada no texto não contém o caractere digitado, podendo levar a dois resultados. Um deles é o *fallback*, um processo de substituição automática, pelo mesmo caractere, só que de uma fonte diferente⁹. Há casos, no entanto, em que o aplicativo ou sistema operacional substitui o caractere faltante por um substituto; tipicamente um retângulo branco, comumente chamado de *tofu*¹⁰.

Os exemplos a seguir tentam ilustrar algumas das questões acima descritas.

Nesta cartilha na língua mehinaku (Fig. 111), o caractere destacado não possui um ponto de código. A forma do til é, na verdade, o til ASCII de outra fonte. Além disso, o til não tem um bom alinhamento com o glifo base, indicando que o diacrítico pode ter sido adicionado manualmente sobre a letra base. A amostra à direita mostra escala e posicionamento corretos.

8. O til ASCII (*asciitilde*, U+007E) ou travessão ondulado <~> é um caractere utilizado em lógica e matemática como sinal para similaridade; e.g. a ~ b (BRINGHURST, 2005, p. 348).

9. A essa função de substituição automática dá-se o nome de *fallback*; uma opção de contingência a ser tomada se a escolha principal não estiver disponível.

10. Por sua similaridade com o formato branco e retangular do queijo de soja asiático, *tofu*. O caractere substituto pode ter outra forma, como um ponto de interrogação ou algum outro grafismo que indique que não há o caractere na fonte ou sistema.



Figura 111: Amostra de texto na língua mehinaku. Fonte: PROFESSORES MEHINAKU, 2002.

No livro de saúde na língua suiá (Fig. 112), há um exemplo de substituição do til pelo til ASCII no caractere <Ë>. O til ASCII em questão é da fonte Arial Regular, enquanto o caractere base <E> é da Arial Bold. Versões mais recentes da Arial dispõem do caractere pré-composto <Ë>, conforme a amostra em Arial Bold 26pt à direita da imagem. Segundo a documentação da Microsoft (2020b), em versões mais antigas, o caractere <Ë> não fazia parte do conjunto de caracteres. No entanto, em outras partes do livro, o <Ë> está codificado corretamente.



Figura 112: Amostra de texto na língua suiá. Fonte: WÜRKER, 2003b.

Nas amostras na língua mehinaku (Fig. 113 e 114), compostas em Verdana, encontra-se três formas diferentes para o til. A forma correta do diacrítico é do caractere <ã>. Observa-se, sobre o <i>, uma forma do til incompatível com o design da Verdana; provavelmente o til ASCII da fonte Times New Roman. No detalhe da (Fig. 113), além dos problemas de alinhamento, observa-se a forma do til ASCII da fonte Verdana sobre o caractere <ü>.

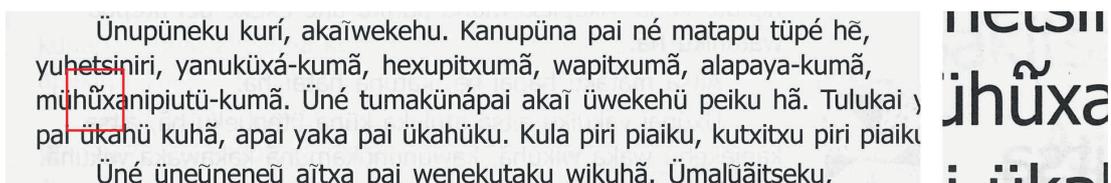


Figura 113: Amostra de texto na língua mehinaku. Fonte: PROFESSORES MEHINAKU, 2002.

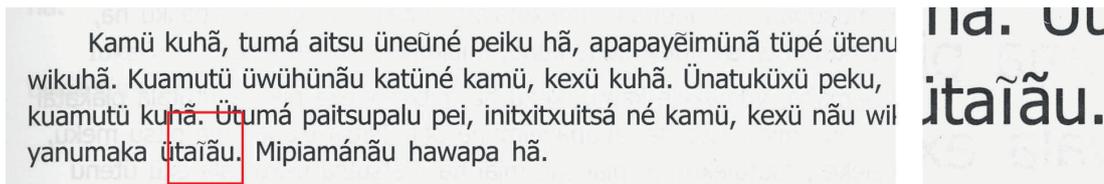


Figura 114: Amostra de texto na língua mehinaku. Fonte: PROFESSORES MEHINAKU, 2002.

11.5 Uso dos recursos de tachado e sublinhado

Uma forma documentada como alternativa aos diacríticos adequados, quando estes não estão disponíveis em uma determinada fonte, consiste no uso das funções de tachado (*strikethrough*) e sublinhado (*underline*), disponíveis na grande maioria dos editores de texto, programas de paginação e editoração, e até mesmo ferramentas de edição de texto online (e.g. GoogleDocs). Em geral, a espessura e largura desses traços não são harmônicos e proporcionais (i.e. peso e largura) em relação à letra base. Além disso, o alinhamento vertical desses traços, em relação à letra base, dificilmente funcionará bem com os parâmetros padrão dos programas. Por último, é importante lembrar que uma letra <i> ou <u> com tachado é muito diferente de <̄i> ou <̄u>; as letras com o tachado não têm o ponto de código correto, são simplesmente a mesma letra <i> ou <u> com um traço criado pelo aplicativo. Os exemplos a seguir ilustram alguns casos de uso dos recursos de tachado e sublinhado.

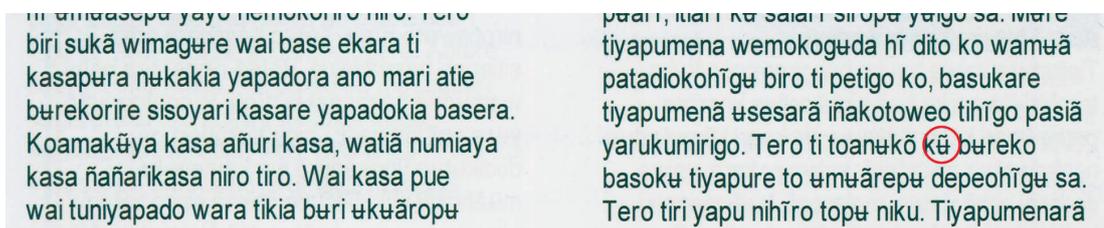


Figura 115: Amostra de texto na língua tuyuka. Fonte: CABALZAR et al., 2009.

Na cartilha na língua tuyuka (Fig.116), o texto principal é composto em Verdana; no entanto, para representar o caractere <̄ũ>, utiliza-se o <u> e o til <~> de outra fonte,

com o recurso do tachado/riscado aplicado sobre a letra. Para minimizar o problema, o caractere <ũ> da Verdana poderia ter sido utilizado, já que, de acordo com a documentação da Microsoft (2021c), a codificação WGL4, que contém o caractere <ũ>, é padrão desde 1997.

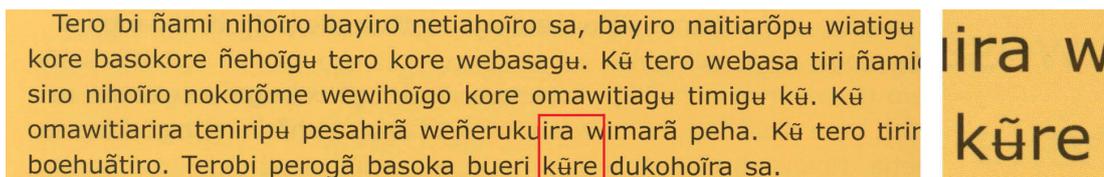


Figura 116: Amostra de texto na língua tuyuka. Fonte: CABALZAR, 2002.

No livro de alfabetização na língua waujá (Fig. 117), observa-se um tipo de grafismo que parece ter sido gerado por meio da função de sublinhado de um programa de editoração. Do ponto de vista tipográfico, o caminho natural seria utilizar uma marca combinada como o mácron subscrito, como no caractere <̣>. O mácron <̄> é um sinal que tem peso e largura similares ao travessão (traço ene, *en-dash*) ou sinal de subtração. No bloco Combining Diacritical Marks há dois pontos de código para esse sinal: U+0331 COMBINING MACRON BELOW (subscrito); e U+0304 COMBINING MACRON (sobrescrito). Existe a possibilidade de o texto ter sido composto com o caractere U+0332 COMBINING LOW LINE, o que seria incorreto, já que a documentação do bloco Combining Diacritical Marks indica que esse diacrítico é utilizado quando se quer conectar outras marcas à esquerda e à direita dele. O resultado é uma linha muito fina e muito próxima à letra, resultado este que poderia ser tipograficamente mais consistente com o uso da marca correta.

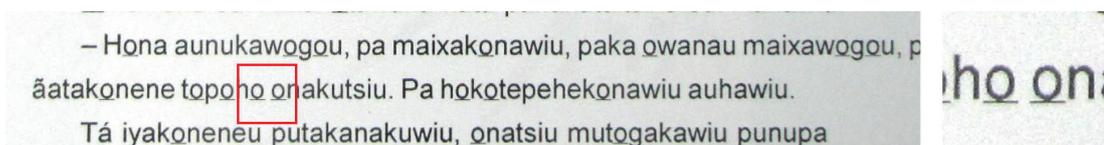


Figura 117: Amostra de texto na língua waujá. Fonte: TRONCARELLI, 2004.

11.6 Diacríticos empilhados

Algumas línguas indígenas brasileiras empregam mais de um diacrítico numa mesma letra base, um empilhado sobre o outro. A grande maioria dessas combinações só pode ser composta por meio da composição dinâmica com marcas combinadas. Praticamente nenhuma combinação de diacríticos empilhados é codificada como caractere pré-composto.

Esses casos merecem atenção especial e dependem de vários recursos para um bom funcionamento. Além da codificação correta, as letras e os diacríticos da fonte devem ter as âncoras posicionadas no local correto, além das funções de posicionamento OpenType devidamente programadas.

Gaultney (2002, p. 12) observa que a composição com esses caracteres requer atenção especial, pois a altura geral dos diacríticos empilhados pode exigir entrelinhamento extra.

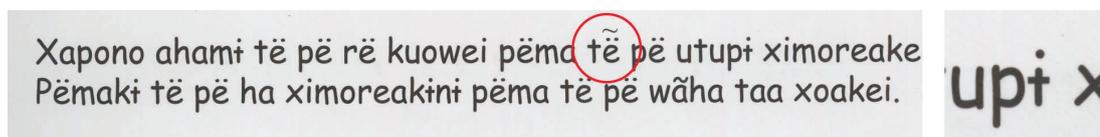


Figura 118: Amostra de texto na língua yanomami. Fonte: PROFESSORES YANOMAMI, 2007.

11.7 Problemas de harmonia de largura e espaçamento

Além de questões relacionadas à ausência de caracteres nas fontes, há casos em que estes estão disponíveis, mas exigem atenção especial, devido a questões específicas das línguas brasileiras. Essas questões têm relação com as proporções de largura entre a letra base e os diacríticos. Nesse sentido, o caso mais notável diz respeito às combinações da letra <i> com os diacríticos mais largos, como o til e o circunflexo. Como em um grande número de fontes (especialmente as sem serifas) diacríticos como o til e o circunflexo são mais largos que a letra <i>, os diacríticos acabam se chocando com letras adjacentes ou com diacríticos das letras adjacentes. Os capítulos *Metodologia* e *Desenvolvimento* descrevem algumas possíveis soluções para esse problema.

Outro problema está associado a um caractere que é amplamente utilizado para representar o fonema da parada glotal, caractere de forma similar a um apóstrofo. Esse caractere não é um diacrítico. Pelo contrário, costuma ter valor de fonema e ocupa um espaço entre duas letras. No entanto, é comum que esse caractere se choque com outras letras adjacentes a ele, problema que pode ser solucionado, principalmente, com ajustes de proporções no glifo, assim como em seu espaçamento.

Os exemplos a seguir procuram ilustrar as questões acima descritas, sobre relações de harmonia de largura e espaçamento.

O *Livro de saúde na língua Yudja* (Fig. 119 e 120) é um dos poucos exemplos de diacríticos empilhados que aparentemente foram codificados de forma correta, tendo em vista a boa harmonização de proporções de alinhamento e peso em relação à letra base. No entanto, na Fig. 121, pode-se perceber um problema de espaçamento ao lado esquerdo dessa letra, que se repete em outros trechos do texto. Na Fig. 121, os diacríticos na letra <i> se chocam com caracteres adjacentes.

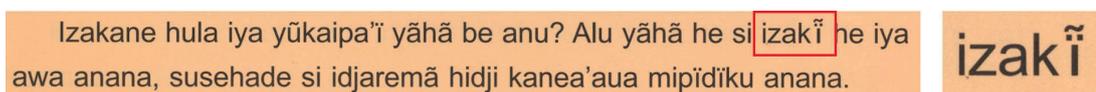


Figura 119: Amostra de texto na língua yudja. Fonte: WÜRKER, 2005.

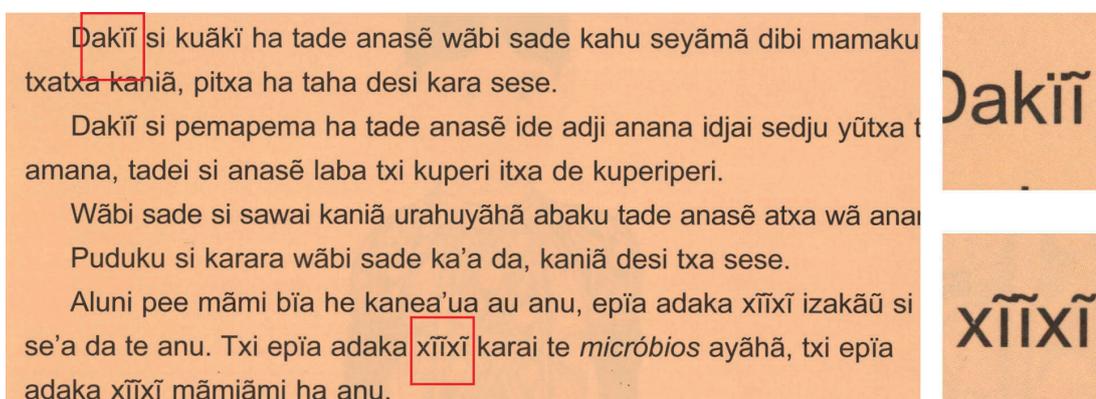


Figura 120: Amostra de texto na língua yudja. Fonte: WÜRKER, 2005.

Na Fig.121, o caractere representado pelo apóstrofo choca-se com o til da letra i.

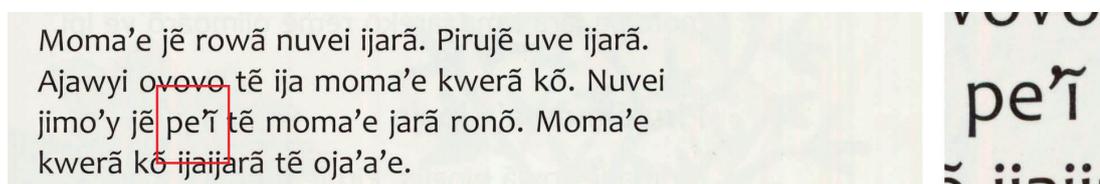


Figura 121: Amostra de texto na língua wajãpi. Fonte: PROFESSORES WAJÃPI, 2005.

Com esses últimos exemplos conclui-se esta seção de análises de problemas de composição tipográfica em línguas indígenas brasileiras.

Em resumo, este capítulo fornece um panorama das condições técnicas e problemas relativos à composição de texto, codificação das ortografias e uso de diacríticos nas línguas indígenas brasileiras.

Há, certamente, outros problemas de composição tipográfica não documentados aqui. Acredita-se que os exemplos ilustrados acima deixem claro que é necessário um método próprio para a construção de uma especificação técnica para a tipografia das línguas indígenas brasileiras, assunto que será mais bem discutido nos capítulos *Metodologia e Desenvolvimento*.

12. Metodologia

Este capítulo tem como propósito a descrição da metodologia utilizada para o levantamento de dados desta pesquisa. Esta metodologia foi aplicada, em especial, no trabalho de catalogação ortográfica, que tem como objetivo principal a especificação de uma lista de requisitos para o suporte linguístico de fontes para línguas indígenas brasileiras. Este método pode ser replicado e utilizado em outros sistemas e bancos de dados, e também em outros contextos (e.g. outros países da América Latina).

Este capítulo tem início com uma breve revisão de literatura sobre a metodologia de desenvolvimento de design de tipos, com algumas indicações para alguns métodos mais relevantes para esta tese. Em seguida, dá-se a descrição do desenvolvimento dessa metodologia, das diretrizes iniciais e de suas finalidades.

É importante mencionar que outros trechos deste trabalho levantam referências teórico-metodológicas importantes para o desenvolvimento da tese; em particular, os capítulos *Codificação e Linguagem*, *Padronização e Ortografias*.

Por último, deve-se entender que, apesar deste trabalho compreender alguns desenvolvimentos práticos (fontes, teclados), esta metodologia não é específica para esses tipos de desenvolvimentos (design de fontes, design de mapas de teclados). No entanto, esta metodologia é indispensável para o suporte a projetos de fontes e teclados para línguas indígenas brasileiras.

12.1 Metodologia: referências iniciais

Assim como em outras áreas (e.g. design automotivo, editorial, de mobiliário etc.), o desenvolvimento de um projeto de design de tipos parte, em geral, de um problema, de uma pergunta específica. Grosso modo, a procura de uma solução para um problema, de

uma resposta para uma pergunta, é premissa comum para o desenvolvimento de qualquer projeto de design. Nessa linha, Gerard Unger (2018, p. 20, tradução nossa) observa:

O design de tipos pode começar com um plano, como uma lista de propriedades ou uma descrição do propósito do design. Você pode começar com formas de letras angulares ou arredondas, largas ou estreitas, discretas ou vívidas, e uma infinidade de outras possibilidades. Finalidades de aplicações podem variar da sinalização a livros de literatura, e de material para ajudar leitores iniciantes a criação de uma atmosfera com chamadas em um website de moda. Pensar sobre um conceito é algo frequentemente combinado à formulação de ideias visuais provisoriamente. Esboçar levanta questões e leva à pesquisa – esboçar é pesquisa. Apenas raramente um design de tipos é iniciado do zero, com uma abordagem completamente nova, como se o designer de tipos fosse uma *tabula rasa* (folha em branco), sem formas de letras gravadas em sua memória. Quase sempre um plano se desenvolve a partir da busca entre designs e temas existentes [...].

Nesse sentido, é importante explicar que não existe um método específico para o desenvolvimento de fontes para línguas indígenas ou para línguas ágrafas. Na área do design de tipos, assim como em outras áreas da comunicação visual, as referências que descrevem metodologias de projeto tendem a ser mais generalistas.

Entre as poucas obras que tentam explicar como se projeta fontes digitais, destacam-se os trabalhos *O MECOTipo: Método de Ensino de Desenho Coletivo de Caracteres Tipográficos* (2018), do recifense Leonardo Buggy; *Como criar tipos: do esboço à tela* (2019), trabalho originalmente editado em espanhol, em 2012, de autoria da espanhola Laura Meseguer, do argentino José Scaglione e do mexicano Cristóbal Henestrosa; e *Designing Type* (2006), da americana Karen Cheng. Nessa linha, também é importante mencionar o livro *Theory of type design* (2018), do holandês Gerard Unger. A discussão sobre as metodologias de design de tipos, por si só, renderia assunto para mais de uma tese. Por essa razão, será feita uma tentativa de recorte, com foco no desenvolvimento de tipografia para línguas indígenas.

Entre as referências relevantes para este trabalho, é preciso reconhecer o pioneirismo de Marina Garone Gravier na pesquisa de tipografia e línguas indígenas na América Latina. Em *Tipografía e identidad lingüística*, a autora (2013, p. 94–95) esboça uma proposta metodológica para o desenvolvimento de projetos de tipografia para línguas

ágrafas, levando em conta quatro principais aspectos: históricos, etnográficos, linguísticos e tecnológicos. Cada um desses quatro tópicos lista alguns temas de desenvolvimento, com maior destaque para questões sociolinguísticas e menor destaque para questões tecnológicas. Não se trata exatamente de uma metodologia, mas, como propõe a autora, de “uma série de tarefas e tópicos que seria desejável que os designers considerassem, embora a lista não esteja fechada”. Em resumo, parece uma lista de possíveis temas de pesquisa.

Por estar se dirigindo, principalmente, a uma audiência do campo do design, Gravier (2013, p. 94–95) enfatiza a importância das questões culturais e observa que os designers não devem menosprezar essas questões e se concentrar somente na forma. No entanto, algumas das sugestões de Gravier parecem muito amplas e, de certa maneira, menos relevantes para o desenvolvimento de fontes tipográficas digitais. A autora sugere, por exemplo, um aprofundamento em questões históricas, etnográficas e linguísticas que, sob a ótica desta tese, não são tão essenciais para o processo de codificação e desenvolvimento de fontes.

Nessa linha, o texto *Unicode Technical Note #19: Recommendations for creating new orthographies* (ANDERSON et al., 2005) é um trabalho que consegue relacionar melhor os campos da tipografia, linguística e tecnologia da informação. Não é uma proposta de metodologia, mas reúne ideias mais pertinentes sobre o tema e trata diretamente de questões relacionadas a línguas ágrafas.

Trabalhos acadêmicos que têm como tema o design de tipos, em especial aqueles que têm um caráter descritivo (i.e. documentam um desenvolvimento de projeto de design de tipos), também são válidos como fontes de referência metodológica. Nessa linha, destacam-se alguns trabalhos de dissertações de mestrado e relatórios de projetos de programas de pós-graduação em tipografia, como o da Universidade de Reading (Inglaterra), o da Universidade de Buenos Aires (Argentina) e o da Real Academia de Arte (Holanda). Alguns projetos desenvolvidos por alunos serão listados na seção *Bibliografia Consultada*.

Talvez uma das maiores referências metodológicas para o desenvolvimento de design de tipos ainda sejam os relatos de profissionais do campo, apesar de estes ainda

serem raros na forma de publicações científicas por algumas razões. Entre elas, destaca-se o fato de que a tipografia é um negócio, do qual os processos de desenvolvimento são parte. Em muitos casos, o jeito de se projetar fontes (i.e. o método) é o segredo de sucesso do negócio. Nesse sentido, pode-se fazer uma analogia com o ramo de alimentos: em geral, vende-se o hambúrguer ou a pizza (o produto), e não o modo de fazê-lo.

Por último, é importante destacar a atuação de John Hudson como um dos maiores (senão o maior) especialistas em design de tipos para sistema de escritas complexas da atualidade. A seção a seguir descreve dois de seus projetos que são as principais referências para metodologia desta tese.

12.2 John Hudson e o uso de bancos de dados no design de tipos

Entre os relatos de Hudson, destacam-se um texto¹ para uma publicação da ATypI (HUDSON, 2000), sobre o desenvolvimento do projeto da fonte Sylfaen para a Microsoft, e uma palestra na conferência TypoBerlin (HUDSON, 2014), em que apresentou o desenvolvimento de uma família de fontes para editora acadêmica holandesa Brill², projeto de alta complexidade que durou cerca de seis anos.

Em ambos os projetos, Hudson mostra um lado do design de tipos que tem muito a ver com esta pesquisa e que é pouco mostrado e discutido. Trata-se de uma parte do trabalho que não tem nada a ver com desenho, mas com processos tediosos de levantamento de dados cuja finalidade é uma codificação consistente, tanto do ponto de vista da engenharia de software como também do ponto de vista tipográfico e linguístico.

Para ilustrar isso, no projeto para a editora Brill, Hudson inicia seu argumento mostrando algumas belas imagens de um caderno de rascunhos do designer de tipos britânico Jeremy Tankard e destaca que está mostrando aquilo por duas razões. Primeiro,

1. O artigo *Sylfaen: Foundations of multiscript typography* foi publicado online no site da ATypI, já que o número 3 da revista *Type* (para a qual o texto foi escrito) não saiu, e a revista parou de ser editada. Hudson também replicou a publicação do texto em seu próprio site, versão referenciada neste texto: HUDSON, 2000.

2. Em 2012, na ocasião do lançamento das fontes projetadas por John Hudson, a editora comemorou 330 anos de existência. Hudson disponibilizou o documento com as imagens e textos da apresentação no site de sua empresa, Tiro Typeworks.

pela beleza daquelas páginas e da qualidade daqueles desenhos, que refletem os estágios iniciais de desenvolvimento de Tankard. Segundo, porque é um bom exemplo de um processo de design de tipos que tem início com ideias sobre a forma das letras, processo este que se desenvolve, sobretudo, por meio do desenho e exploração formal. Por último, Hudson destaca que aquilo representa o que muitos imaginam a respeito do desenvolvimento de design de tipos (se é que alguém, além dos designers de tipos, se dá ao trabalho de imaginar isso) (HUDSON, 2014, p. 9).



Figura 122: Esboços para o tipo Kingfisher, de Jeremy Tankard. Fonte: HUDSON, 2014, p. 9.

Em seguida, Hudson explica que sua abordagem é totalmente diferente e que seus projetos geralmente começam por uma planilha, que é a sua maneira de documentação preliminar das necessidades do cliente. Durante o desenvolvimento, as planilhas são utilizadas para o planejamento e estruturação das fontes. À medida que o projeto progride, as planilhas se tornam um repositório central de informações sobre os glifos e codificação, que serão utilizadas em seu fluxo de produção. Hudson explica ainda que, com efeito, essas planilhas são a especificação técnica da fonte (HUDSON, 2014, p. 10).

192	0x00E0	agrave	à	LATIN SMALL LETTER A WITH GRAVE					ccmp	I.ccmp\NFD	a gravecmb -> agrave	2	
193	0x00E1	acute	á	LATIN SMALL LETTER A WITH ACUTE					ccmp	I.ccmp\NFD	a acutecmb -> acute	2	
194	0x00E2	acircumflex	â	LATIN SMALL LETTER A WITH CIRCUMFLEX					ccmp	I.ccmp\NFD	a circumflexcmb -> acircumflex	2	
195	0x00E3	atilde	ã	LATIN SMALL LETTER A WITH TILDE					ccmp	I.ccmp\NFD	a tildecmb -> atilde	2	
196		atilde_acutecmb	ã			ccmp	I.ccmp\NFC	atilde acutecmb -> atilde_acutecmb	2	ccmp	I.ccmp\NFD	a tildecmb acutecmb -> atilde_acutecmb	3
197	0x00E4	adieresis	ä	LATIN SMALL LETTER A WITH DIAERESIS					ccmp	I.ccmp\NFD	a dieresiscmb -> adieresis	2	
198	0x0101	amacron	ā	LATIN SMALL LETTER A WITH MACRON					ccmp	I.ccmp\NFD	a macroncmb -> amacron	2	
199		amacron_gravecmb	ā			ccmp	I.ccmp\NFC	amacron gravecmb -> amacron_gravecmb	2	ccmp	I.ccmp\NFD	a macroncmb gravecmb -> amacron_gravecmb	3
200		amacron_acutecmb	ā			ccmp	I.ccmp\NFC	amacron acutecmb -> amacron_acutecmb	2	ccmp	I.ccmp\NFD	a macroncmb acutecmb -> amacron_acutecmb	3
201		amacron_tildecmb	ā			ccmp	I.ccmp\NFC	amacron tildecmb -> amacron_tildecmb	2	ccmp	I.ccmp\NFD	a macroncmb tildecmb -> amacron_tildecmb	3
202		amacron_brevecmb	ā			ccmp	I.ccmp\NFC	amacron brevecmb -> amacron_brevecmb	2	ccmp	I.ccmp\NFD	a macroncmb brevecmb -> amacron_brevecmb	3
203	0x0103	abreve	ǎ	LATIN SMALL LETTER A WITH BREVE					ccmp	I.ccmp\NFD	a brevecmb -> abreve	2	
204	0x00E5	aring	å	LATIN SMALL LETTER A WITH RING ABOVE					ccmp	I.ccmp\NFD	a ringcmb -> aring	2	
205	0x00E6	ae	æ	LATIN SMALL LETTER AE									
206		b_dieresisbelowcmb	ḃ						ccmp	I.ccmp\NFD	b dieresisbelowcmb -> b_dieresisbelowcmb	2	
207	0x010D	ccaron	č	LATIN SMALL LETTER C WITH CARON					ccmp	I.ccmp\NFD	c caroncmb -> ccaron	2	
208	0x00E7	cedilla	ç	LATIN SMALL LETTER C WITH CEDILLA									
209	0x1E0D	ddotbelow	ḍ	LATIN SMALL LETTER D WITH DOT BELOW					ccmp	I.ccmp\NFD	d dotbelowcmb -> ddotbelow	2	
210		d_dieresisbelowcmb	ḍ						ccmp	I.ccmp\NFD	d dieresisbelowcmb -> d_dieresisbelowcmb	2	
211	0x1E0F	dlinebelow	ḏ	LATIN SMALL LETTER D WITH LINE BELOW					ccmp	I.ccmp\NFD	d macronbelowcmb -> dlinebelow	2	
212	0x00E8	egrave	è	LATIN SMALL LETTER E WITH GRAVE					ccmp	I.ccmp\NFD	e gravecmb -> egrave	2	
213	0x00E9	eacute	é	LATIN SMALL LETTER E WITH ACUTE					ccmp	I.ccmp\NFD	e acutecmb -> eacute	2	
214	0x00EA	ecircumflex	ê	LATIN SMALL LETTER E WITH CIRCUMFLEX					ccmp	I.ccmp\NFD	e circumflexcmb -> ecircumflex	2	
215	0x1E8D	etilde	ẽ	LATIN SMALL LETTER E WITH TILDE					ccmp	I.ccmp\NFD	e tildecmb -> etilde	2	

Figura 123: Exemplo de planilha utilizada para levantamento de dados.

Fonte: HUDSON, 2014, p. 10.

Nesse sentido, Hudson explica que a Brill se revelou o tipo de cliente ideal para ele, posto que, logo no início do projeto, o primeiro documento enviado pela editora Brill foi uma planilha gigantesca, algo que ratificava sua concepção de metodologia. A equipe da editora trabalhou na compilação de uma enorme lista de itens aos quais eles se referem como “entidades” (*entities*); elementos distintos que ocorrem nas publicações da editora³. Para tal, a equipe consultou não somente caracteres e glifos de diversas fontes específicas⁴ de periódicos e livros recentes como também examinou a fundo publicações mais antigas. Essa equipe, sempre que possível, associou pontos de código do Unicode para as entidades, assim como estabeleceu sua correspondência com os glifos das fontes específicas utilizadas nas publicações, e

3. O conceito de entidade é similar ao conceito de glifo. A principal diferença é que o termo glifo é utilizado em meio digital e, na editora Brill, entidade se refere também a marcas impressas no papel em publicações mais antigas.

4. Hudson usa a expressão “*various ad hoc fonts*”, dando a entender que a editora utilizava fontes personalizadas (*custom fonts*) para tarefas específicas, e.g. matemática, egiptologia etc.

ainda documentou os dados bibliográficos das publicações em que as ditas entidades foram encontradas (HUDSON, 2014, p. 10–12).

Hudson já apontava para tal tipo de abordagem em projetos anteriores. Como na ocasião de sua colaboração com a Microsoft, por volta de 1997, quando trabalhou no desenvolvimento do *Web Resource for International Typography* (WRIT), uma base de dados indexada que reúne informações sobre glifos e caracteres. No WRIT, as informações sobre glifos e caracteres eram associadas a línguas, sistemas de escritas e recursos tipográficos que, por sua vez, podiam ser mapeados a regiões geográficas, países, organizações geopolíticas e a codificações existentes (e.g. codificações legadas). Por meio de uma interface gráfica (Fig. 124), era possível realizar pesquisas no banco de dados do WRIT, que compilava os resultados na forma de conjuntos de caracteres personalizados em resposta aos critérios definidos pelo usuário. A partir da combinação de mais de um critério de pesquisa, podia-se gerar especificações de conjuntos de caracteres com suporte para grupos de línguas, para regiões e países, para línguas oficiais e minoritárias etc. Apesar de a iniciativa ter sido descontinuada, Hudson declarou seu desejo de que um dia fosse retomada e destacou a serventia de uma ferramenta como o WRIT, especialmente no contexto de internacionalização no desenvolvimento de software (HUDSON, 2000, p. 1–2).



[1] Entering search criteria in WRIT, top, and viewing query results, bottom.

Figura 124: Interface do WRIT. Fonte: HUDSON, 2000, p. 2.

Hudson também informa que, logo no início do projeto, percebeu a necessidade de uma fonte para exibir as formas normativas (i.e. a forma básica, arquetípica) de cada glifo do banco de dados, disponibilizando para o usuário uma forma gráfica básica de cada caractere, e não somente informações técnicas (ponto de código, nome PostScript). A fonte Sylfaen, desenvolvida por John Hudson e Ross Mills, ambos da Tiro Typeworks, e Geraldine Wade, da Monotype Typography, tinha como objetivo representar visualmente todos os caracteres e glifos documentados no banco de dados em um estilo único⁵ e razoavelmente neutro (HUDSON, 2000, p. 3).

5. Família formada somente pelo estilo regular (romano), sem itálico ou variações de peso.



Figura 125: Amostra de caracteres da fonte Sylfaen, de John Hudson, Ross Mills e Geraldine Wade. Fonte: do autor.

Hudson informa⁶ que o WRIT foi completamente abandonado no final dos anos 1990, quando ele e a equipe pararam de trabalhar no projeto. Foi escrito para funcionar na plataforma WindowsNT e, por isso, ele nem sabe dizer se funcionaria em versões recentes do Windows. Apesar disso, ele acredita que a base de dados ainda seria útil.

Apesar da interrupção no desenvolvimento do WRIT, versões da fonte Sylfaen ainda persistiram em distribuições do Windows posteriores, com conjuntos de caracteres para os sistemas de escrita latino, grego monotônico (i.e. grego moderno), cirílico, armênio e georgiano. O projeto original, de 1998, era composto por 3842 glifos, cobrindo seis sistemas de escrita (latino, grego, cirílico, armênio, georgiano e etíope) e IPA. Além da extensa cobertura para várias línguas e escritas, dispunha ainda de recursos OpenType, variantes de caracteres, entre outras funcionalidades tipográficas (HUDSON, 2000, p. 3–5).

A metodologia de desenvolvimento da Sylfaen contemplava dois cenários que apresentam similaridades com o das línguas indígenas brasileiras, que são os dos conjuntos

6. Em correspondência pessoal por e-mail. Ver apêndices.

Durante o desenvolvimento da Sylfaen, Hudson reporta que, depois de analisar uma amostra de mais de 200 línguas africanas (i.e. apenas cerca de 10% do total), teve que acrescentar mais 570 glifos ao conjunto de caracteres (sem contar versaletes, ligaturas e outras variantes) (HUDSON, 2000, p. 6–7). Assim como no Brasil, essa grande diversidade representa um obstáculo para o desenvolvimento de fontes, tendo em vista que para o correto suporte a um determinado grupo linguístico, é preciso levantar informações (i.e. preencher as planilhas), e essas informações nem sempre estão acessíveis, ou são de difícil acesso, ou estão dispersas entre pesquisadores espalhados pelo continente.

A partir desses dois projetos, fica clara a importância das bases de dados para a codificação e desenvolvimento de fontes. As informações de metadados compiladas nessas bases não abordam diretamente questões sobre a qualidade das formas dos glifos. No entanto, fornecem informações para o mapeamento correto entre caracteres e glifos, e sobre o suporte linguístico de um determinado conjunto de caracteres.

A seção a seguir descreve algumas ferramentas atualmente utilizadas no desenvolvimento de fontes para a especificação de suporte linguístico, conjuntos de caracteres, mapeamento correto entre caracteres e glifos etc.

12.3 Ferramentas atuais de bancos de dados para codificação e tipografia

Durante esta pesquisa, identificou-se algumas ferramentas e repositórios de informações relevantes para o desenvolvimento de conjuntos de caracteres e suporte linguístico em projetos de design de tipos. Entre essas ferramentas, que utilizam bases metodológicas similares àquelas propostas por John Hudson, destacam-se o Hyperglot (BREZINA, 2021, n.p.) e o CharsetBuilder (EIGENDORF; ACSANY, 2016, n.p.). Ambas podem ser acessadas pelo navegador, com interfaces bastante intuitivas.

O Hyperglot é uma ferramenta desenvolvida para ajudar na identificação do suporte linguístico em fontes. Com a ferramenta, é possível detectar se uma determinada fonte X pode ser usada para a composição de textos numa determinada língua Y,

além de identificar um conjunto de caracteres padrão para a ortografia⁷ de cada língua. Idealizado pelo desenvolvedor e designer de tipos tcheco David Březina, da Rosetta Type, o Hyperglot é um projeto *open-source* e, atualmente, sua base de dados contém informações de mais de 640 línguas. É um trabalho em andamento e planejado para expansão. Além do aplicativo web (que funciona no navegador), o projeto também está disponível publicamente em repositório do GitHub⁸, como ferramenta de linha de comando, pacote Python e base de dados (BREZINA, 2021, n.p.).

A interface do Hyperglot é constituída por duas colunas. A coluna à esquerda apresenta um menu que contém uma lista de fontes padrão do aplicativo e um campo de busca para consulta das línguas da base de dados. Abaixo do campo de busca, consta uma listagem completa de todos os sistemas de escrita. As línguas são listadas dentro de cada sistema de escrita e podem ser selecionadas com um clique. A coluna à direita mostra os conjuntos de caracteres para as línguas selecionadas e algumas informações básicas sobre aquela língua (Fig.127). Por último, destaca-se que o Hyperglot possui uma funcionalidade de *mouse-over* para identificação dos caracteres: ao passar o mouse por cima de algum caractere, este se amplia em um retângulo de fundo escuro, fornecendo ao usuário uma imagem mais detalhada do glifo, assim como seu ponto de código e algumas outras de suas propriedades (i.e. categoria, direcionalidade).

7. A ferramenta contempla línguas que apresentam mais de uma ortografia (e.g. servo-croata). Nesses casos, o sistema mostra as duas ortografias quando essa língua estiver selecionada.

8. O Git é um tipo de sistema de controle de versão utilizado no desenvolvimento de softwares. O GitHub é um repositório online que hospeda uma ferramenta Git e pode ser acessada simultaneamente por diferentes desenvolvedores, e é a forma mais popular de tornar público o código e a documentação de um projeto open-source.

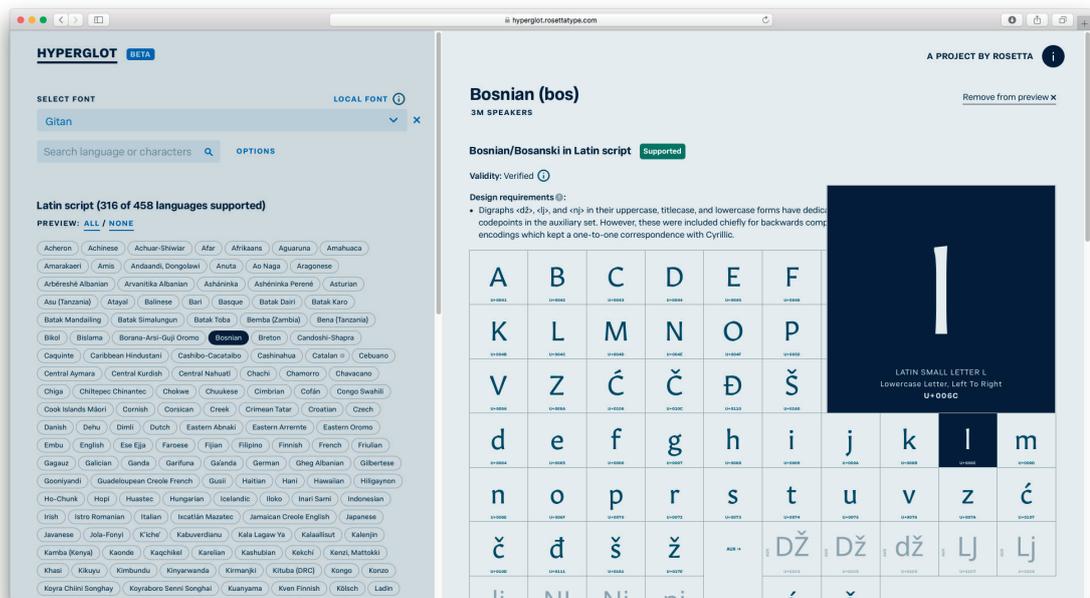


Figura 127: Captura de tela do Hyperglot. Fonte: BREZINA, 2021, n.p.

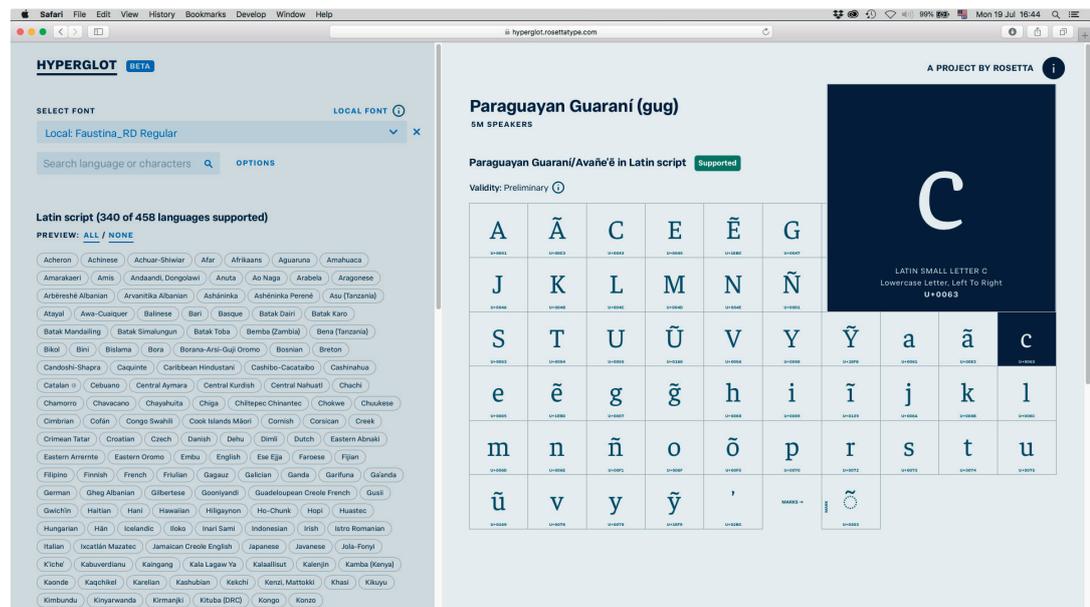


Figura 128: Captura de tela do Hyperglot. Fonte: BREZINA, 2021, n.p.

Além da lista de fontes padrão, pré-instaladas no sistema, o Hyperglot também permite ao usuário adicionar e analisar suas próprias fontes. Para adicionar outras fontes, deve-se arrastar o arquivo de fonte do computador para a janela do navegador. O aplicativo analisa o arquivo de fonte e fornece uma previsão dos sistemas de escrita e

do número de línguas para a qual a fonte tem suporte (Fig.128). O número de línguas é uma estimativa parcial, pois só conta com línguas já documentadas pelo aplicativo.

Por se tratar de uma ferramenta de testes, recomenda-se o uso do Hyperglot somente para verificar se uma determinada fonte é adequada para composição de textos em um determinado idioma. A ferramenta não diz nada sobre a qualidade do design de uma fonte (BREZINA, 2021, n.p.).

Enquanto o Hyperglot funciona, principalmente, para procedimentos de testes, o Charset Builder é uma ferramenta online para construção de conjuntos de caracteres personalizados. Desenvolvido por Andreas Eigendorf e Philipp Acsany, da empresa alemã Alphabet Type, o CharSet Builder utiliza blocos construtivos padronizados a partir de informações fornecidas por instituições como o Unicode, ISO, Microsoft e Apple, e possibilita ao usuário criar seus próprios conjuntos de caracteres por meio de combinações modulares de diferentes blocos. Os autores partem da premissa de que, assim como a forma individual de cada letra, o conjunto de caracteres (i.e. codificação) é uma importante decisão de design (EIGENDORF; ACSANY, 2016, n.p.).

A interface do Charset Builder apresenta dois campos de busca principais. O primeiro faz uma filtragem por blocos e aceita, como entrada, nomes de línguas (e.g. português), sistemas de escrita (e.g. cirílico), blocos do Unicode (e.g. Latin Extended-A) ou codificações existentes (e.g. MacOS Roman, ISO 8859-1, MS Windows 1250). Ao digitar qualquer termo de uma dessas categorias, a ferramenta filtra todos os resultados da base de dados associados ao termo, conforme demonstrado na busca pelo termo “Latin”, na Fig.129. O segundo campo de busca é utilizado para buscas por caracteres específicos (e.g. <ñ>) e mostra as codificações em que aquele caractere aparece⁹.

9. Tanto no Hyperglot como no Charset Builder, os resultados da busca por caracteres específicos são pouco consistentes. Por exemplo: ao buscar pelo caractere <ã>, o Hyperglot indica que este é utilizado somente em uma única língua (idioma ga ou gã, falado em Gana), enquanto o Charset Builder indica que o caractere é utilizado em vários outros sistemas de escrita (e.g. árabe, armênio, devanagari) dos quais o <ã> não faz parte.

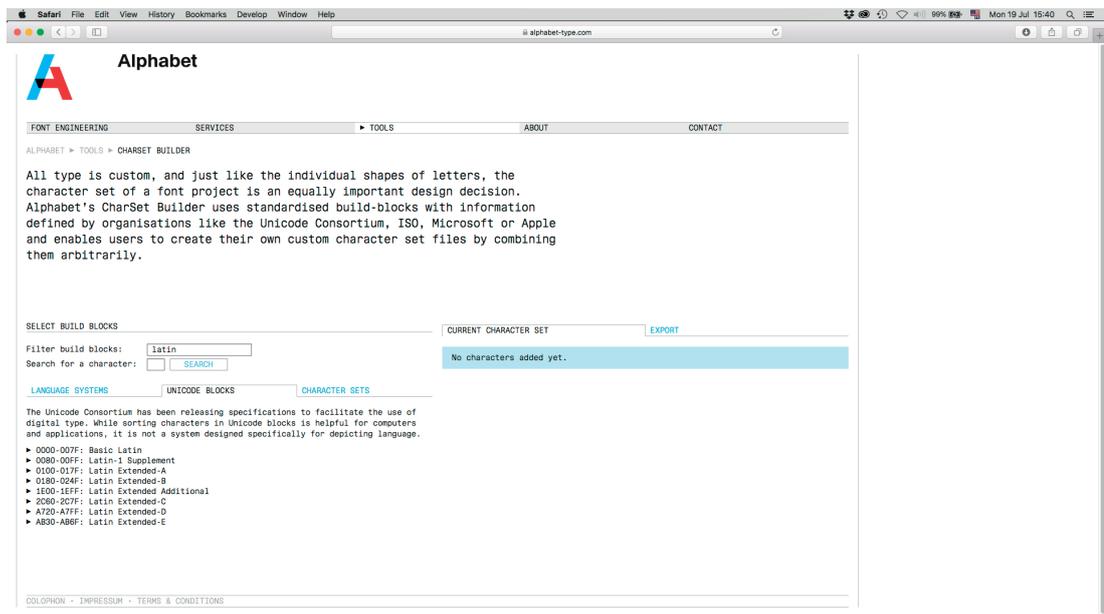


Figura 129: Captura de tela do Charset Builder. Fonte: EIGENDORF; ACSANY, 2016.

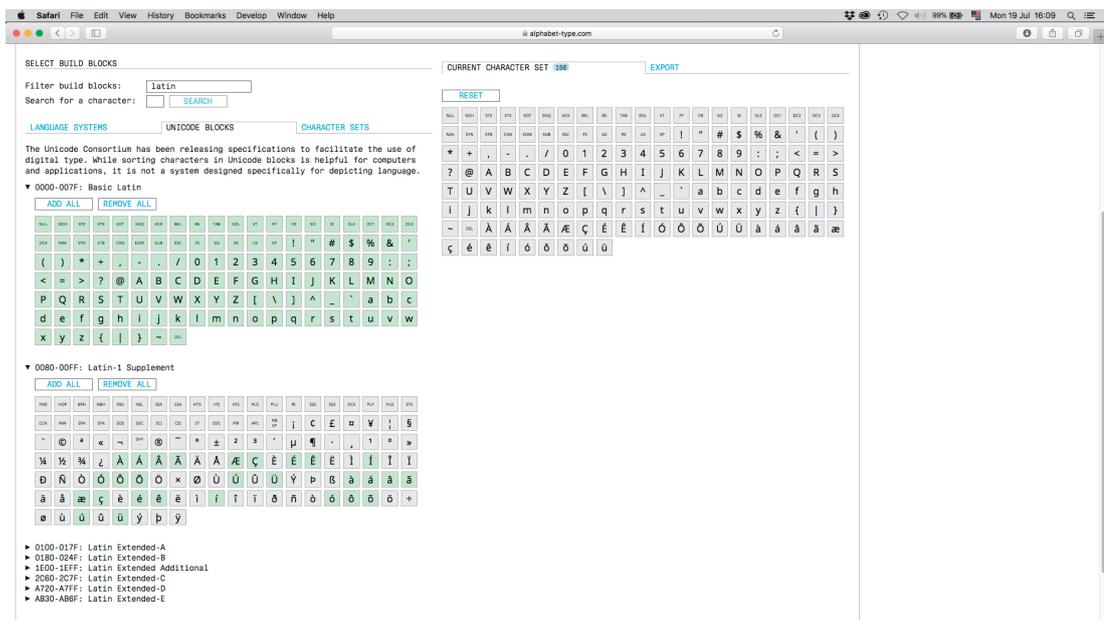


Figura 130: Captura de tela do Charset Builder. Fonte: EIGENDORF; ACSANY, 2016.

Ao clicar-se no nome de algum dos blocos listados, na coluna à esquerda, abre-se, logo abaixo do nome, uma listagem de todos os caracteres daquele bloco. Pode-se adicionar, ao conjunto de caracteres a ser criado, um bloco inteiro (clicando no botão “ADD

ALL”) ou somente caracteres individuais. A cor verde indica que o caractere já foi adicionado ao conjunto, que é mostrado na coluna à direita da página, como mostra a Fig.130.

Ao finalizar a seleção, clica-se no botão “EXPORT” na coluna da esquerda. Ao realizar essa ação, o site disponibiliza a lista dos caracteres, em forma de texto puro, de três formas diferentes: a representação do caractere em si (*Character Representation*), o ponto de código Unicode (*Unicode*) e o nome do glifo (*Glyphname*). Essas listas podem ser copiadas diretamente para os programas de design e edição de fontes (comandos copy/paste), para a criação de conjuntos de caracteres.

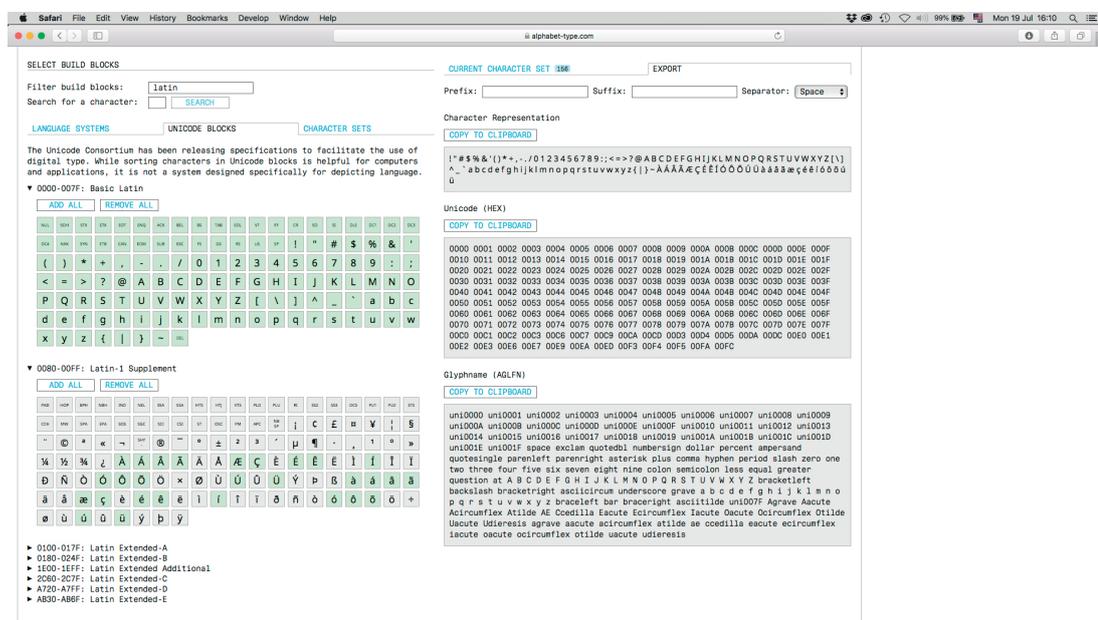


Figura 131: Captura de tela do Charset Builder. Fonte: EIGENDORF; ACSANY, 2016.

A Fig.132 mostra uma captura de tela do editor de fontes Glyphs, na qual a janela em destaque contém a lista de nomes de glifos copiados do Charset Builder. Ao clicar-se no botão *Generate*, o programa cria uma posição para cada caractere dentro do arquivo do Glyphs e automaticamente associa o ponto de código Unicode correspondente a cada caractere. Nota-se que os caracteres ainda não existem de fato; todas as posições para os caracteres foram criadas, mas ainda é preciso adicionar os contornos.

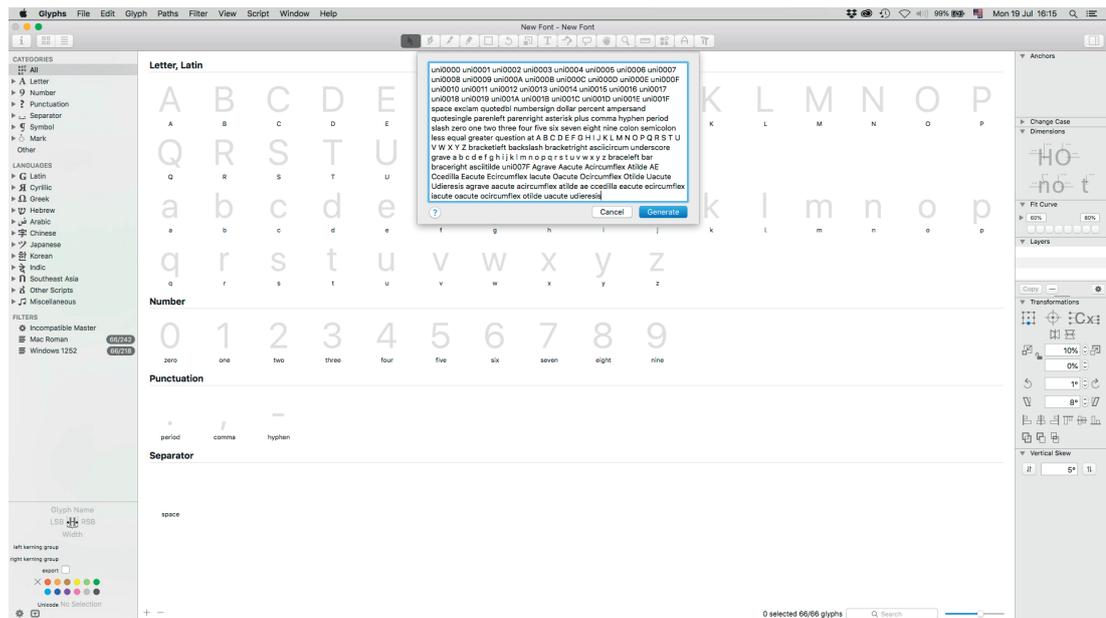


Figura 132: Captura de tela da janela de diálogo para adição de novos caracteres no software Glyphs. Fonte: do autor.

A partir das descrições acima, pode-se dizer que tanto o Hyperglot como o Charset Builder fornecem respostas a muitas das perguntas sobre o suporte linguístico em desenvolvimentos de fontes e famílias de tipos. No entanto, no contexto do desenvolvimento de fontes com suporte para línguas indígenas brasileiras, ambas as ferramentas apresentam deficiências; a principal delas é a falta de informações sobre os caracteres utilizados na maioria dessas línguas. Uma forma de corrigir alguns desses problemas é alimentando os bancos de dados do Hyperglot e do Charset Builder com informações sobre as línguas indígenas do Brasil. Por essa razão, decidiu-se realizar o levantamento de dados desta pesquisa conforme descrição a seguir.

12.4 Desenvolvimento do método de levantamento de dados

A partir das referências acima mencionadas, esta seção procura descrever a metodologia empregada no levantamento de dados desta pesquisa. Como produto final desse levantamento, esperava-se, desde o início do projeto, uma catalogação abrangente e sistemática das ortografias das línguas indígenas brasileiras, capaz de fornecer respostas a uma série de perguntas sobre o suporte linguístico em projetos de design de tipos

para essas línguas. Somente a partir dessas respostas é que será possível construir uma especificação técnica consistente para a codificação das línguas indígenas brasileiras.

Em resumo, o objetivo do levantamento de dados é fornecer informações-chave para a especificação de conjuntos de caracteres adequados para línguas ou grupos de línguas indígenas brasileiras.

Destaca-se, aqui, que foram utilizadas duas formas de coleta de dados: análise de publicações e pesquisa por meio de formulários online com a contribuição de pesquisadores da área. Os dois processos serão explicados, respectivamente, nessa ordem.

Neste ponto, ressalta-se que essa forma de catalogação não nasceu pronta. Para uma melhor compreensão de como se desenvolveu esta metodologia, é importante voltar ao início desta pesquisa, mostrando como se construiu este percurso metodológico.

Nos estágios iniciais desta pesquisa¹⁰, pouco antes do ingresso deste autor no Programa de Pós-graduação em Comunicação da UnB, iniciou-se uma primeira tentativa de catalogação, com base na análise de fontes primárias provenientes de um acervo físico de publicações em línguas indígenas do Ministério da Educação, em Brasília, e também do site do Laboratório de Educação e Material Didático da Universidade de São Paulo, que disponibiliza uma série de livros digitalizados¹¹.

No caso dos livros impressos do acervo do Ministério da Educação, o processo se iniciava com a documentação fotográfica da capa, da folha de rosto e de uma amostra todos os diacríticos utilizados naquele determinado idioma, para fins de referência e arquivamento. No caso dos materiais digitalizados, o processo se iniciava com o arquivamento de uma cópia do documento em formato PDF. As fotos e arquivos em formato PDF foram organizadas em pastas e classificadas por línguas.

10. Por volta de meados de 2016.

11. Em sua grande maioria, livros impressos que foram digitalizados. Nesse sentido, pode-se considerar estas como fontes primárias impressas.

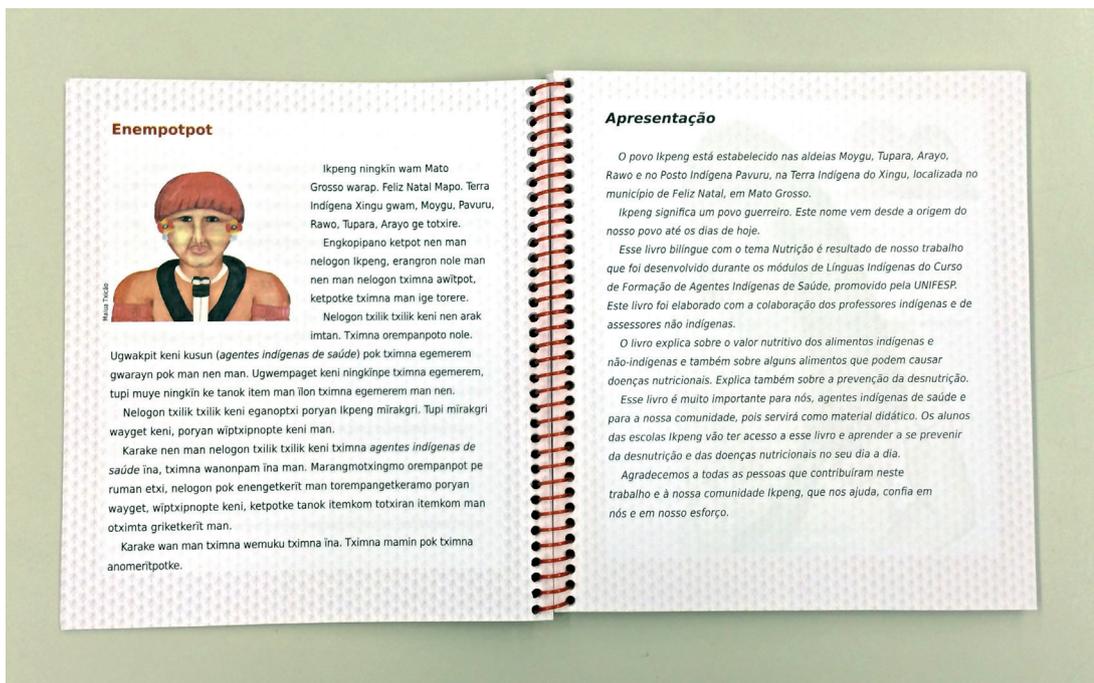


Figura 133: Exemplo de documentação fotográfica realizada no acervo do Ministério da Educação. Publicação em língua ikpeng. Fonte: do autor.

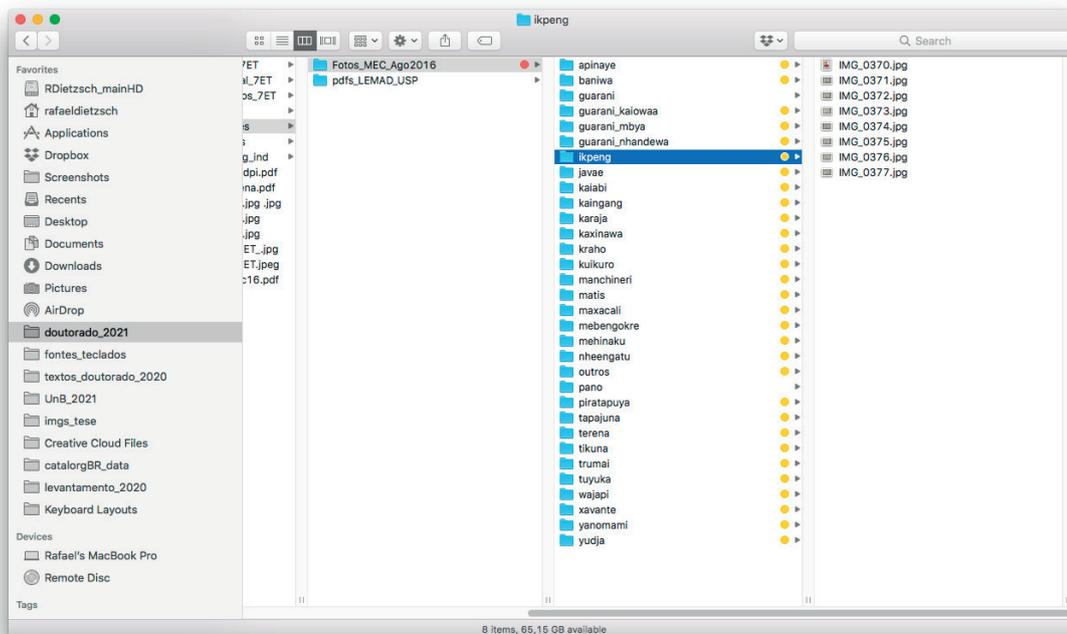


Figura 134: Sistema de classificação das fontes primárias. Fonte: do autor.

substituídas por anotações em arquivos digitais, utilizando os próprios caracteres. Ao criar-se um documento eletrônico (e.g. arquivo .txt) para reproduzir aqueles caracteres em meio digital, um ponto de código Unicode automaticamente é associado ao glifo encontrado na fonte primária. Dessa forma, cria-se uma informação padronizada, com um correto mapeamento entre glifos e pontos de códigos.

Em segundo lugar, a única tabela produzida não foi estruturada de forma correta. Naquele momento, percebeu-se que havia algo de errado, mas não se conseguia identificar o problema. Por essa razão, realizou-se uma reunião¹² com professor e pesquisador Eduardo Gomes, do Departamento de Estatística da UnB, que sugeriu duas mudanças principais na forma de construção das tabelas. Primeiramente, Gomes indicou que era importante pensar em cada fonte primária como uma tabela separada, pois posteriormente isso facilitaria a análise de cada amostra individualmente e a sua classificação de uma maneira global. Isso facilitaria também o processamento das tabelas em bloco, para gerar estatísticas do cenário completo. Além disso, tabelas independentes podem acomodar mais informações sobre cada caractere, como ponto de código, representação gráfica (i.e. grafema, glifo), nome do glifo etc. Gomes sugeriu ainda que fosse adotada uma sintaxe para a nomenclatura das tabelas, utilizando um prefixo único para cada língua e uma numeração sequencial para as publicações dessas línguas (e.g. POR_001, POR002 etc.). Essa sugestão do estatístico levou a resultados interessantes, com melhorias técnicas significativas, conforme descrições a seguir.

Primeiro, a sugestão do prefixo na nomenclatura dos arquivos acabou conduzindo ao uso da norma ISO 639-3 para classificação das línguas, posteriormente adotada, de uma forma geral, em toda a pesquisa. O padrão ISO 639-3 é uma norma internacional que define códigos de três letras para a representação de nomes de línguas, utilizado principalmente em contextos de localização de software. Por exemplo: [por] é o código para português, [tuo] é o código para tukano. A norma foi publicada pela International Organization for Standardization em 2007, e atualmente tem como autoridade de registro o SIL International, por meio de sua publicação *Ethnologue: Languages*

12. Por volta da época em que se realizou o exame de qualificação para esta tese, no início de 2019.

of the World (EBERHARD; SIMONS; FENNIG, 2021; INTERNATIONAL ORGANIZATION FOR STANDARDIZATION, 2007; SIL INTERNATIONAL, 2021).

Em segundo lugar, sua sugestão para que cada tabela representasse a ortografia de uma única publicação levou a um entendimento ainda mais importante para o processo. A partir desse momento percebeu-se que, com a documentação das análises de cada publicação em tabelas independentes, seria possível associar também metadados bibliográficos específicos de cada publicação, como autor, título, editor, ano de publicação etc. Além disso, as tabelas independentes possibilitariam uma análise comparativa, importante para a compreensão da estabilidade da ortografia de algumas línguas, em casos específicos.

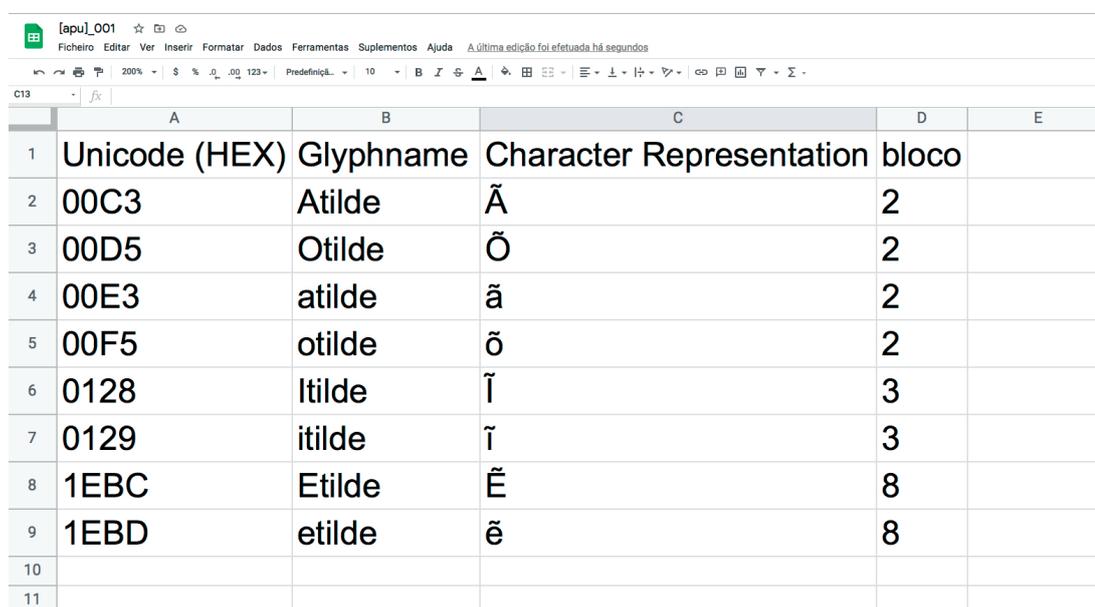
Em terceiro lugar, com tabelas independentes torna-se possível rastrear a origem da instabilidade, da inconsistência ou de problemas técnicos relativos à ortografia de uma determinada língua. Por exemplo: pode ser que duas publicações na mesma língua apresentem ortografias diferentes e, ao se documentar cada publicação separadamente, torna-se possível mapear, se necessário, as origens dessas diferenças.

Por último, é importante ressaltar que essa é uma forma de se documentar a existência de um caractere que foi utilizado em uma publicação que, de fato, existe. Esse ponto é bastante relevante, pois resalta o fato de que a metodologia foi pensada para a análise e documentação de dados concretos, que fornecem um lastro para a discussão sobre as tipografias das línguas indígenas em vários aspectos. Esse lastro representa uma referência importante não somente para a discussão no campo da linguística, mas também para profissionais de comunicação visual.

Uma vez compreendida a necessidade de se estruturar as tabelas como arquivos independentes, era necessária a criação de um padrão para a formatação dos dados nas tabelas. Para isso, utilizou-se o Charset Builder como referência dos parâmetros que deveriam constar nas tabelas.

Conforme mencionado na seção anterior, o Charset Builder lista três principais parâmetros para a construção de um conjunto de caracteres: a representação do caractere em si (*Character Representation*), o ponto de código Unicode (*Unicode*) e o nome do glifo (*Glyphname*). A partir desses três parâmetros é possível codificar caracteres

tipográficos de forma consistente. Por essa razão, determinou-se que esses eram os parâmetros básicos que deveriam constar nas tabelas. Além disso, a título de referência, também decidiu-se especificar o bloco do Unicode em que cada caractere estava localizado. A Fig. 136 mostra a estrutura básica das tabelas no aplicativo online Google Sheets, com cada coluna representando um parâmetro, e o nome do arquivo seguindo a sintaxe do código ISO 639-3, acompanhado pela ordem da numeração, e.g. [apu]_001.



	A	B	C	D	E
1	Unicode (HEX)	Glyphname	Character Representation	bloco	
2	00C3	Atilde	Ã	2	
3	00D5	Otilde	Õ	2	
4	00E3	atilde	ã	2	
5	00F5	otilde	õ	2	
6	0128	ltilde	ĩ	3	
7	0129	itilde	ï	3	
8	1EBC	Etilde	Ë	8	
9	1EBD	etilde	ë	8	
10					
11					

Figura 136: Estrutura básica das tabelas do levantamento no aplicativo online Google Sheets. Fonte: do autor.

Para simplificação das tabelas e pelo fato de a pesquisa se basear em poucos dos blocos do Unicode, decidiu-se criar uma lista com uma numeração para esses blocos. Constam na lista somente os blocos em que se encontrou algum caractere no levantamento preliminar¹³, nos estágios iniciais desta pesquisa. A numeração dos blocos é a seguinte:

1. Basic Latin
2. Latin-1 Supplement
3. Latin Extended-A

¹³. Posteriormente foi adicionada uma 11ª categoria, com o bloco General Punctuation.

4. Latin Extended-B
5. IPA Extensions
6. Spacing Modifier Letters
7. Combining Diacritical Marks
8. Latin Extended Additional
9. Latin Extended-D
10. Caracteres formados pela combinação de uma letra base e uma marca combinada

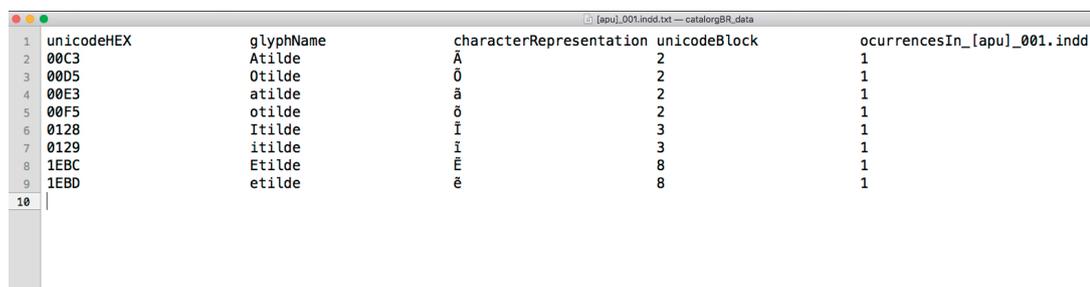
Uma vez decidido o padrão de formatação para as tabelas, havia um outro problema a ser resolvido. Depois da primeira tentativa de se preencher uma dessas tabelas, percebeu-se que o processo de inserção de caractere por caractere, manualmente, seria um trabalho muito demorado e altamente sujeito a erros. Uma alternativa seria copiar os dados das tabelas de referência de cada bloco (ver Fig.138), algo também bastante trabalhoso, embora um pouco menos sujeito ao erro.

	A	B	C	D	E	F	G	H
1	2. LATIN-1 SUPPLEMENT							
2	00A8	dieresis	¨					
3	00AA	ordfeminine	ª					
4	00AF	macron	¯					
5	00B0	degree	°					
6	00B2	uni00B2	²					
7	00B3	uni00B3	³					
8	00B4	acute	´					
9	00B7	periodcentered	·					
10	00B8	cedilla	¸					
11	00B9	uni00B9	¹					
12	00BA	ordmasculine	º					
13	00C0	Agrave	À					
14	00C1	Aacute	Á					
15	00C2	Acircumflex	Â					
16	00C3	Atilde	Ã					
17	00C4	Adieresis	Ä					
18	00C5	Aring	Å					
19	00C6	AE	Æ					

Figura 137: Tabela de referência do Bloco Latin-1 Supplement. Fonte: do autor.

Logo, percebeu-se que a inserção de dados manual seria algo impraticável, não somente pelo grande número de publicações, mas também pelo número de caracteres encontrado em cada uma delas. Assim, procurou-se uma forma de automatizar esse processo. Para isso, foi feito contato com Gustavo Soares, um colega especialista em tipografia, design de tipos e automatização de fluxos de trabalho em sistemas de design de informação.

Com base nos 10 blocos do Unicode utilizados na pesquisa, Soares programou um script para o Adobe InDesign, que processa amostras de texto simples e detecta qualquer incidência de diacrítico dessa amostra dentro do arquivo de InDesign ativo. Ao rodar o script, detecta-se todas as incidências de caracteres acentuados, estes listados num arquivo formato .txt que é gravado na área de trabalho do computador, com o mesmo nome do arquivo do InDesign analisado. O arquivo .txt emprega a mesma sintaxe, notação e formatação das tabelas; logo, é possível copiar e colar esses dados produzidos pelo script diretamente no arquivo da tabela. Esse processo cria uma grande consistência no levantamento e diminui significativamente a margem de erro.



The image shows a screenshot of a text file named [apu]_001.indd.txt. The file contains a table with five columns: unicodeHEX, glyphName, characterRepresentation, unicodeBlock, and occurrencesIn_[apu]_001.indd. The table lists ten rows of data, each corresponding to a specific diacritical character and its occurrence count in the source file.

unicodeHEX	glyphName	characterRepresentation	unicodeBlock	occurrencesIn_[apu]_001.indd
00C3	Atilde	Ã	2	1
00D5	Otilde	Õ	2	1
00E3	atilde	ã	2	1
00F5	otilde	õ	2	1
0128	Itilde	İ	3	1
0129	itilde	ı	3	1
1EBC	Etilde	Ë	8	1
1EBD	etilde	ë	8	1

Figura 138: Arquivo produzido pelo script ContadorDiacriticos de Gustavo Soares. Fonte: do autor.

A implementação dessa ferramenta foi um divisor de águas no processo. Ainda que fosse necessário digitar os caracteres em um arquivo de InDesign enquanto se realizava as análises, linha a linha, página a página, o uso do script diminuiu muito a margem de erros no processo. Além disso, poderá ser utilizado por terceiros (e.g. bolsistas, outros pesquisadores), e não somente por este pesquisador, com resultados consistentes e padronizados.

Por último, também é importante mencionar a forma de catalogação de dados bibliográficos. Para isso, foi utilizado o Zotero¹⁴, um software gerenciador de referências em software livre e de código aberto. Suas características principais são a integração com navegadores, a sincronização online, a geração de citações em texto, rodapés e bibliografias, bem como a integração com os processadores de texto. O uso dessa ferramenta possibilitou a padronização dos metadados bibliográficos na pesquisa.

A Fig.140 mostra a interface do Zotero. A inserção de dados é feita por meio de um formulário, na coluna à direita. Uma vez inseridos os dados de cada documento, é possível gerar relatórios (Fig.141) de cada documento individualmente ou de um conjunto de documentos. Finalmente, é importante mencionar que esses metadados foram, posteriormente, integrados às tabelas das análises das ortografias de cada publicação.

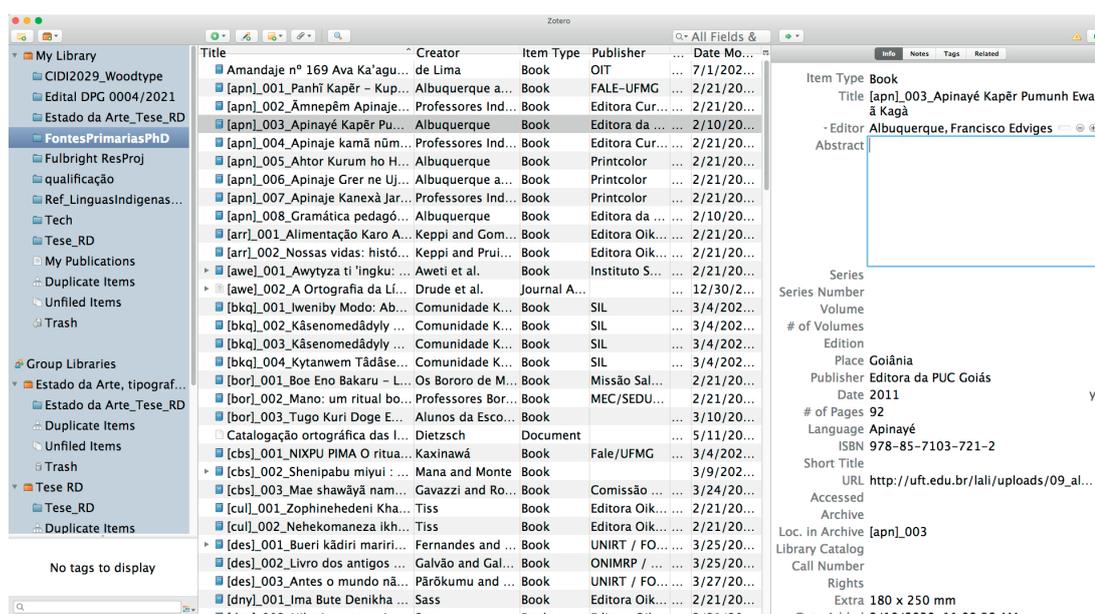


Figura 139: Captura de tela do software Zotero. Fonte: do autor.

14. O Zotero também foi utilizado para o gerenciamento de referências bibliográficas da tese.

[apn]_003_Apinayé Kapēr Pumunh Ewa ã Kagà

Type	Book
Editor	Francisco Edviges Albuquerque
Date	2011
Language	Apinayé
Loc. in Archive	[apn]_003
URL	http://uft.edu.br/lali/uploads/09_alfabetizacao2.pdf
Extra	180 x 250 mm
Place	Goiânia
Publisher	Editora da PUC Goiás
ISBN	978-85-7103-721-2
# of Pages	92
Date Added	2/10/2020, 11:09:58 AM
Modified	2/10/2020, 11:37:50 AM

Figura 140: Exemplo de relatório gerado pelo Zotero. Fonte: do autor.

Entre as fontes primárias analisadas, é importante destacar aqui o documento *Projeto 914BRZ4019 Contrato SA-3993/2019*, um relatório do desenvolvimento de dicionários para o Projeto de Documentação de Línguas no Brasil (PRODOCLIN) do Museu do Índio/FUNAI, do linguista Helder Perri Ferreira (2021). A segunda parte do relatório descreve as ortografias de quinze línguas do projeto, com uma lista completa de grafemas e seus respectivos pontos de código. Trata-se do único documento analisado nesta pesquisa que fornece informações técnicas completas (i.e. grafema e ponto de código) sobre conjuntos de caracteres utilizados em línguas indígenas brasileiras. A Fig.141 mostra uma página da publicação, na qual se pode ver um fragmento da sistematização de Ferreira, por meio de tabelas e quadros, com resultados similares aos utilizados nesta tese.



Consultor: Helder Perri Ferreira_SA-3993/2019_Produto 01_01/03/2020

1.18. Alfabeto Korubo

O alfabeto Korubo é composto de 17 letras: [a, e, ê, i, k, l, m, n, o, p, s, t, u, v, w, x, y] e suas correspondentes maiúsculas. Não existem diacríticos combináveis na ortografia da língua. A ordem alfabética do Korubo é a seguinte:

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
A	E	E	I	K	L	M	N	O	P	S	T	U	V	W	X	Y
a	e	e	i	k	l	m	n	o	p	s	t	u	v	w	x	y

Incluindo letras maiúsculas e minúsculas, são necessários portanto 34 caracteres ou combinações de caracteres para se escrever Korubo. Desse total, 32 caracteres fazem parte do bloco Latino Básico (Basic Latin) do mapa Unicode: [A, E, Ê, I, K, L, M, N, O, P, S, T, U, V, W, X, Y, a, e, ê, i, k, l, m, n, o, p, s, t, u, v, w, x, y]. Dois caracteres [Ê, ê] estão mapeados no bloco Latino Suplementar 1 (Latin-1 Supplement). A tabela abaixo resume a distribuição dos grafemas Korubo nos diferentes blocos do mapa Unicode.

Bloco Unicode	Grafemas Korubo
Basic Latin	32
Latin-1 Supplement	2
Total	34

Abaixo encontra-se a lista completa de grafemas utilizados na ortografia Korubo, incluindo suas possíveis combinações diacríticas. A tabela ainda contém o código Unicode, o código numérico HTML, o código CSS, e a cadeia de caracteres de escape (scape sequence) de cada grafema e dígrafo combinado, assim como o bloco Unicode em que está mapeado.

Caractere	Nome do Caractere	HTML	Unicode	CSS	Sequence	Bloco Unicode
A	Letra Latina Maiúscula A	A	U+0041	U0041	u0041	Basic Latin
a	Letra Latina Minúscula A	a	U+0061	U0061	u0061	Basic Latin
E	Letra Latina Maiúscula E	E	U+0045	U0045	u0045	Basic Latin
e	Letra Latina Minúscula E	e	U+0065	U0065	u0065	Basic Latin
I	Letra Latina Maiúscula I	I	U+0049	U0049	u0049	Basic Latin
i	Letra Latina Minúscula I	i	U+0069	U0069	u0069	Basic Latin
K	Letra Latina Maiúscula K	K	U+004B	U004B	u004B	Basic Latin
k	Letra Latina Minúscula K	k	U+006B	U006B	u006B	Basic Latin
L	Letra Latina Maiúscula L	L	U+004C	U004C	u004C	Basic Latin
l	Letra Latina Minúscula L	l	U+006C	U006C	u006C	Basic Latin
M	Letra Latina Maiúscula M	M	U+004D	U004D	u004D	Basic Latin
m	Letra Latina Minúscula M	m	U+006D	U006D	u006D	Basic Latin
N	Letra Latina Maiúscula N	N	U+004E	U004E	u004E	Basic Latin
n	Letra Latina Minúscula N	n	U+006E	U006E	u006E	Basic Latin
O	Letra Latina Maiúscula O	O	U+004F	U004F	u004F	Basic Latin



Consultor: Helder Perri Ferreira_SA-3993/2019_Produto 01_01/03/2020

1.19. Alfabeto More-Kuyubim

O alfabeto More-Kuyubim é composto de 17 letras: [a, e, f, i, k, m, n, ñ, o, p, r, s, t, u, w, x, y, z] e suas correspondentes maiúsculas, à exceção da consoante <ç> cuja contraparte maiúscula é igual à forma minúscula. Não existem diacríticos combináveis na ortografia da língua. A ordem alfabética do More-Kuyubim é a seguinte:

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
A	E	F	I	K	M	N	Ñ	O	P	R	S	T	U	W	X	Y	Z	.
a	e	f	i	k	m	n	ñ	o	p	r	s	t	u	w	x	y	z	.

Incluindo letras maiúsculas e minúsculas, são necessários portanto 37 caracteres ou combinações de caracteres para se escrever More-Kuyubim. Desse total, 35 caracteres fazem parte do bloco Latino Básico (Basic Latin) do mapa Unicode: [A, E, F, I, K, M, N, O, P, R, S, T, U, W, X, Y, Z, a, e, f, i, k, m, n, o, p, r, s, t, u, w, x, y, z] e dois caracteres [Ñ, ñ] estão mapeados no bloco Latino Suplementar 1 (Latin-1 Supplement). A tabela abaixo resume a distribuição dos grafemas More-Kuyubim nos diferentes blocos do mapa Unicode.

Bloco Unicode	Grafemas More-Kuyubim
Basic Latin	35
Latin-1 Supplement	2

Figura 141: Exemplo de sistematização das ortografias na publicação do PRODOCLIN. Fonte: FERREIRA, 2021, p. 39–40.

Em seu relatório, Ferreira (2021, p. 23–24) enumera as tarefas descritivas da ortografia e alfabeto das línguas:

- 1) Inventário dos grafemas básicos e diacríticos combinatórios que compõem o alfabeto nativo;
- 2) Apresentação da ordem alfabética regular da língua, considerando tanto os grafemas básicos como seu uso combinado com diacríticos e seu possível efeito na ordem regular;
- 3) Inventário dos dígrafos relevantes para o ordenamento alfabético, quando for o caso;
- 4) Inventário dos blocos de caracteres Unicode em que estão mapeados os grafemas e dígrafos da ortografia nativa;
- 5) Número de caracteres mapeados em cada bloco Unicode;
- 6) Indicação das codificações relevantes para a correta representação em meio digital de todos os grafemas e dígrafos combinados da língua, a saber, o código Unicode, o código numérico Html, o código css, e a cadeia de caracteres de escape (scape sequence).

Finalmente, é importante ressaltar que os resultados completos do levantamento com base nas fontes primárias podem ser consultados nos apêndices.

Além do trabalho de levantamento por meio da análise de publicações, foi realizado, paralelamente, uma coleta de dados por meio de formulários. Conforme mencionado anteriormente no capítulo *Delimitação do corpus*, essa forma de levantamento consiste em uma pesquisa online, na qual cada colaborador fornece uma listagem dos caracteres acentuados utilizados na língua (ou línguas) com a qual tem familiaridade, contato, envolvimento. É possível fazer múltiplas contribuições por meio desse formulário. A ideia era fazer uma campanha de divulgação online, convidando pesquisadores envolvidos com línguas indígenas brasileiras para colaborar com o levantamento.



Figura 142: Imagem usada para divulgação do projeto de catalogação ortográfica. Fonte: do autor.

O dado mais importante coletado pelo formulário é a lista dos caracteres acentuados utilizados em cada língua. A entrada se dá por meio da digitação da lista completa dos caracteres acentuados de cada língua num campo específico do formulário, em que cada caractere deve ser separado somente por uma vírgula ou espaço. Pode-se incluir caracteres somente em caixa alta, somente em caixa baixa ou em ambos. O importante é que conste a listagem completa de todas as letras acentuadas (vogais e consoantes) em determinada língua e, em um outro campo, a indicação do código ISO 639-3

12.5 Critérios para classificação dos dados

Esta seção descreve os critérios para classificação dos dados coletados. Somente a partir destes critérios é possível analisar e classificar corretamente as línguas pesquisadas.

Esse tipo de levantamento de dados pode resultar em um grande volume de informações, algo que pode ser difícil de analisar. Essas informações podem ser utilizadas em vários contextos: e.g. documentação de línguas, especificação de conjuntos de caracteres etc. No entanto, para que essas informações (i.e. listas de caracteres e seus respectivos códigos) tenham aplicações práticas no desenvolvimento de ferramentas (i.e. fontes, teclados), é preciso estabelecer uma relação entre essas informações e os blocos do Unicode e/ou outras codificações existentes (e.g. MacOS Roman, ISO 8859-1, MS Windows 1250), uma vez que o suporte linguístico de muitas ferramentas é pensado a partir das codificações supracitadas. Ou seja, não basta apenas realizar o mapeamento entre caracteres e glifos; é preciso contextualizar este mapeamento.

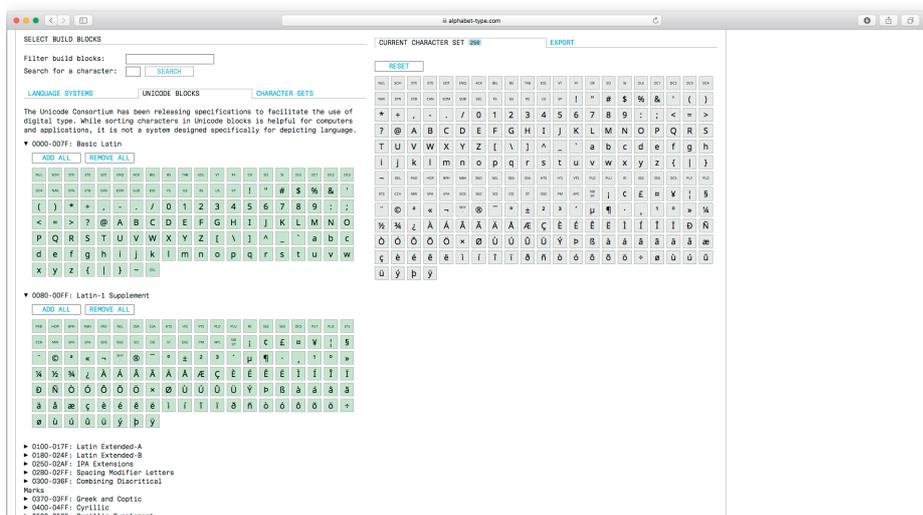


Figura 144: Basic Latin e Latin-1 Supplement (Latin-I). Amostra gerada no aplicativo online Charset Builder. Fonte: EIGENDORF; ACSANY, 2016

Nesse sentido é importante reforçar que para todos os desdobramentos desta pesquisa (em especial, novas fontes), os blocos Basic Latin e Latin-1 Supplement (i.e. Latin-1 ou ISO 8859-1) são obrigatórios. Desde a década de 1990, este é um conjunto

de caracteres padrão, considerado como o mínimo para desenvolvimento de fontes de texto pela esmagadora maioria dos desenvolvedores. Mesmo que algumas línguas indígenas brasileiras utilizem somente o bloco Basic Latin¹⁵ (i.e. ASCII), é preciso levar em conta a transferibilidade com a língua portuguesa, que utiliza vários caracteres da codificação Latin-1.

A estratégia encontrada para sistematizar essas relações foi a criação de três grandes grupos linguísticos, com base nos conjuntos de caracteres utilizados por cada língua. O grupo 1 é formado por línguas que utilizam somente caracteres dos dois primeiros blocos do Unicode (Basic Latin e Latin-1 Supplement, ver Fig.144) e podem ser compostas com layouts de teclado comuns, como ABNT ou ANSI. O grupo 2 é formado por línguas que utilizam caracteres dos dois primeiros blocos do Unicode, e também de outros blocos do Unicode, como Latin Extended-A, e Latin Extended Additional, mas que dependem de layouts de teclado personalizados, já que muitos dos caracteres encontrados nesses blocos complementares ao Latin-1 não podem ser compostos com teclados comuns ABNT ou ANSI. E o grupo 3, o mais complexo, é formado por línguas que utilizam caracteres não codificados pelo Unicode, i.e. caracteres que são formados por mais de um ponto de código. Para esse terceiro grupo é preciso teclados específicos, que utilizam marcas combinadas. Nesse sentido, é importante reforçar aqui que as línguas dos grupos 1 e 2 utilizam exclusivamente caracteres pré-compostos, enquanto as línguas do grupo 3 utilizam tanto caracteres pré-compostos como marcas combinadas.

Uma vez estabelecidos os critérios de análise das informações, é possível iniciar de fato o processo de catalogação ortográfica. Grosso modo, a catalogação ortográfica consiste na análise e classificação das informações do levantamento de dados e tem como objetivo estabelecer uma série de diretrizes de desenvolvimento para a tipografia de línguas indígenas brasileiras. O processo de catalogação ortográfica será descrito no próximo capítulo, *Desenvolvimento*.

15. A língua baniwa, por exemplo, não utiliza as seguintes letras do alfabeto: C, F, G, J, L, Q, U, V, Z (FERREIRA, 2021, p. 26–28). Por razões práticas e questões de transferibilidade, é impensável uma fonte sem essas letras.

Neste capítulo, procurou-se explicar o desenvolvimento do método utilizado para especificação do suporte linguístico de fontes para línguas indígenas brasileiras. Pelo método, é possível se estabelecer quais caracteres são utilizados para cada língua e, assim, trabalhar de forma consistente no desenvolvimento de fontes e teclados.

Reforça-se o fato de que esta metodologia é específica para o levantamento de dados para o suporte linguístico, e não consiste em um método para design de tipos. O objetivo principal é a criação de uma lista de requisitos (i.e. quais caracteres são necessários) para determinado grupo linguístico.

A partir desses requisitos, é possível trabalhar, de fato, no desenvolvimento de aplicações práticas, como fontes digitais e mapas de teclados. O capítulo *Desenvolvimento*, a seguir, demonstra como aplicar esta metodologia na customização de fontes e teclados para línguas indígenas brasileiras.

13. Desenvolvimento

Este capítulo tem como objetivo descrever o desenvolvimento de projeto das ferramentas de escrita desta tese, especialmente da catalogação ortográfica e das diretrizes de desenvolvimento, e também do desenvolvimento das fontes digitais e layouts de teclado.

13.1 Catalogação ortográfica

Esta seção descreve o desenvolvimento da catalogação ortográfica, um resultado prático da aplicação da metodologia, e tem como objetivo demonstrar os princípios de análise dos dados e como se deu a classificação das ortografias.

Grosso modo, a catalogação ortográfica consiste na análise e na classificação das informações do levantamento de dados e tem como objetivo estabelecer uma série de diretrizes de desenvolvimento para a tipografia de línguas indígenas brasileiras. A partir dessas diretrizes é possível desenvolver especificações consistentes para conjuntos de caracteres, recursos OpenType, teclados etc.

13.1.1 Análise de dados

O primeiro passo da catalogação ortográfica consiste numa análise preliminar de todos os dados coletados, em todas as planilhas. Para isso, o professor e pesquisador Eduardo Gomes, do Departamento de Estatística da UnB, que já havia recomendado que o banco de dados fosse construído com tabelas individuais, gerou uma compilação de todas as tabelas em um único documento, utilizando o software LaTeX.

O processamento automatizado com o LaTeX permite a combinação de duas variáveis por vez; e.g. línguas e blocos, línguas e caracteres, línguas e Unicode. Assim, é possível gerar uma tabela que agrupa todas as línguas no eixo y (linhas) e todos os

caracteres utilizados em cada uma delas no eixo x (colunas). Pode-se alterar também a variável do eixo y.

Além disso, Gomes gerou um documento, que ele chamou de *tabelaCombinada*, que consiste na compilação de todas as 239 entradas em uma única tabela, contendo todos os dados: código Unicode, nome do glifo, representação do caractere e bloco. Nessa tabela (Fig. 145) é possível visualizar as quatro principais variáveis, além do nome do arquivo e da língua. Essa tabela possibilita uma visão global de todos os caracteres utilizados no levantamento.

Para a análise global da amostra, utilizou-se o script *ContadorDiacriticos*, que indica não somente quais os caracteres representados naquela determinada amostra, mas também o número total de ocorrências¹ de cada caractere na amostra analisada. A partir da análise global, é possível gerar uma lista de caracteres necessários para cobrir todas as línguas.

Também é possível selecionar e copiar informações específicas de uma língua ou grupo de línguas da *tabelaCombinada* para análise e para uso em outros contextos. Por exemplo: as informações da seleção em azul (nomes de glifos) na Fig. 145 podem ser copiadas e utilizadas no software Glyphs. As figuras 146 e 147 mostram como são adicionados, com essas informações, novos caracteres a uma fonte. Por meio desse procedimento é possível adicionar todos os caracteres necessários para a ortografia de uma língua ao conjunto mínimo obrigatório (i.e. Latin-1).

1. O resultado do número total de ocorrências de cada caractere de uma determinada amostra (e.g. todos os caracteres da *tabelaCombinada*) pode ser relevante para alguns propósitos. Por exemplo: pode ser que alguns caracteres sejam usados em poucas línguas, mas às vezes essas línguas apresentam um grande número de usuários. Para fins de análise, todas as línguas e todos os caracteres foram considerados de igual importância.

	A	B	C	D	E	F	G	H	I	J	K	L	M
476	[cax]_003.xlsx	cax	2019	quoteright	'	11. General Punctuation							
477	[cbs]_001.xlsx	cbs	00C1	Aacute	Á	2. Latin-1 Supplement							
478	[cbs]_001.xlsx	cbs	00C3	Atilde	Ã	2. Latin-1 Supplement							
479	[cbs]_001.xlsx	cbs	00E1	aaacute	á	2. Latin-1 Supplement							
480	[cbs]_001.xlsx	cbs	00E3	atilde	ã	2. Latin-1 Supplement							
481	[cbs]_001.xlsx	cbs	0128	titilde	ı̃	3. Latin Extended-A							
482	[cbs]_001.xlsx	cbs	0129	itilde	ı̄	3. Latin Extended-A							
483	[cbs]_001.xlsx	cbs	0168	Utilde	Ũ	3. Latin Extended-A							
484	[cbs]_001.xlsx	cbs	0169	utilde	ũ	3. Latin Extended-A							
485	[cbs]_001.xlsx	cbs	1EBC	Etilde	Ê	8. Latin Extended Additional							
486	[cbs]_001.xlsx	cbs	1EBD	etilde	ê	8. Latin Extended Additional							
487	[cbs]_002.xlsx	cbs	00C3	Atilde	Ã	2. Latin-1 Supplement							
488	[cbs]_002.xlsx	cbs	00E3	aaacute	á	2. Latin-1 Supplement							
489	[cbs]_002.xlsx	cbs	0128	titilde	ı̃	3. Latin Extended-A							
490	[cbs]_002.xlsx	cbs	0129	itilde	ı̄	3. Latin Extended-A							
491	[cbs]_002.xlsx	cbs	0168	Utilde	Ũ	3. Latin Extended-A							
492	[cbs]_002.xlsx	cbs	0169	utilde	ũ	3. Latin Extended-A							
493	[cbs]_002.xlsx	cbs	1EBC	Etilde	Ê	8. Latin Extended Additional							
494	[cbs]_002.xlsx	cbs	1EBD	etilde	ê	8. Latin Extended Additional							
495	[cbs]_003.xlsx	cbs	00C1	Aacute	Á	2. Latin-1 Supplement							
496	[cbs]_003.xlsx	cbs	00C3	Atilde	Ã	2. Latin-1 Supplement							
497	[cbs]_003.xlsx	cbs	00E1	aaacute	á	2. Latin-1 Supplement							
498	[cbs]_003.xlsx	cbs	00E3	atilde	ã	2. Latin-1 Supplement							
499	[cbs]_003.xlsx	cbs	0128	titilde	ı̃	3. Latin Extended-A							
500	[cbs]_003.xlsx	cbs	0129	itilde	ı̄	3. Latin Extended-A							
501	[cbs]_003.xlsx	cbs	0168	Utilde	Ũ	3. Latin Extended-A							
502	[cbs]_003.xlsx	cbs	0169	utilde	ũ	3. Latin Extended-A							
503	[cbs]_003.xlsx	cbs	1EBC	Etilde	Ê	8. Latin Extended Additional							
504	[cbs]_003.xlsx	cbs	1EBD	etilde	ê	8. Latin Extended Additional							
505	[cin]_001.xlsx	cin	0027	quotingsingle	'	1. Basic Latin							
506	[cin]_001.xlsx	cin	00C1	Aacute	Á	2. Latin-1 Supplement							

Figura 145: tabelaCombinada. Documento gerado por Eduardo Gomes. Fonte: do autor.

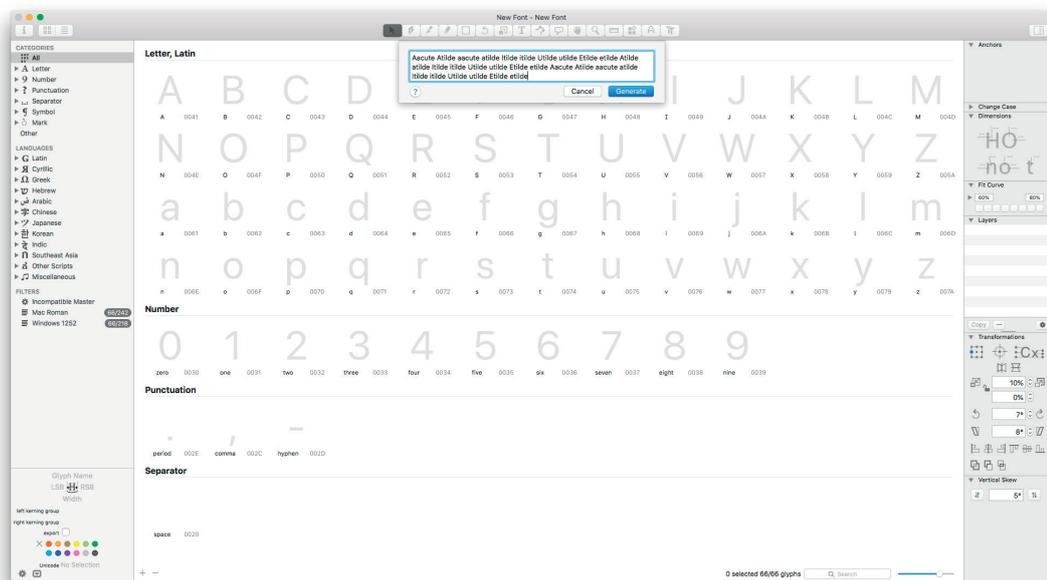


Figura 146: Captura de tela do software Glyphs mostrando janela de diálogo para adição de novos glifos. No topo da imagem pode-se ver a janela de diálogo aberta, com uma lista de nomes de glifos que serão adicionados à fonte ao se clicar o botão *Generate*. Fonte: do autor.

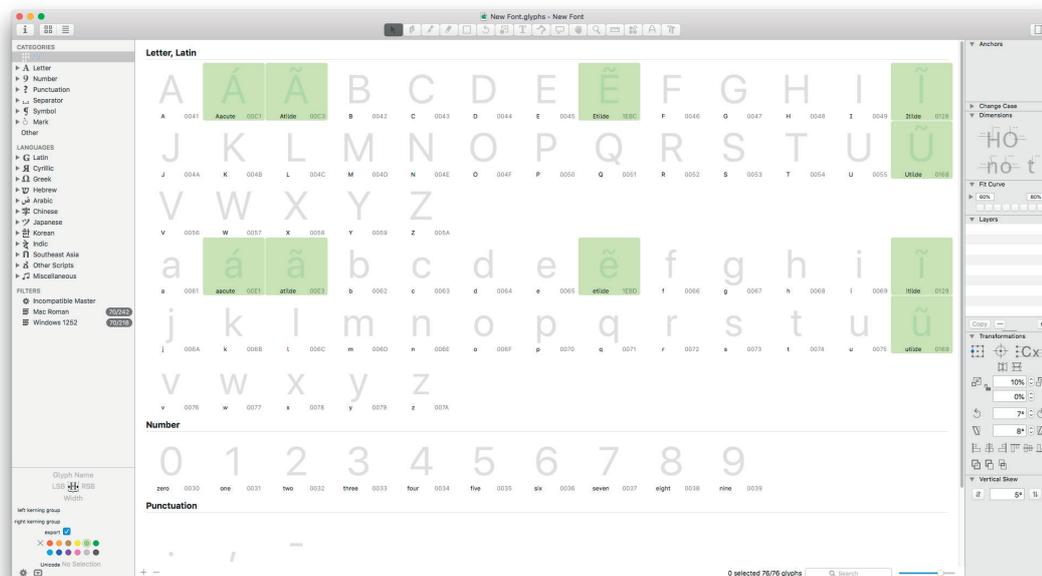


Figura 147: Captura de tela do software Glyphs, mostrando novos caracteres adicionados. Os quadrados marcados em verde mostram os novos caracteres. Fonte: do autor.

Finalmente, foi possível utilizar a *tabela Combinada* para classificar as línguas em três grandes grupos linguísticos, conforme descrição a seguir.

13.1.2 Classificação das ortografias

Analisando as relações das línguas por blocos, foi possível detectar quais delas estavam no grupo 1 (línguas que usam somente a codificação básica Latin-1), grupo 2 (línguas que utilizam a codificação Latin-1, mas também caracteres de outros blocos do Unicode, como Latin Extended-A, e Latin Extended Additional) e grupo 3 (línguas que utilizam caracteres não codificados pelo Unicode)

O resultado da classificação das 239 entradas e 111 línguas são os 3 grupos linguísticos listados a seguir:

grupo 1 (27 línguas)

[aoc] Taurepang, [asu] Asuriní do Tocantins, [atr] Waimiri Atroari, [bkq] Bakairi, [bwi] Baniwa-Koripako, [cul] Kulina, [dny] Deni, [gub] Guajajara, [ite] More-Kuyubim, [hix] Hixkaryána, [jaa] Jamamadi, [kay] Kamayurá, [knm] Kanamari, [kpc] Baniwa Coripaco, [mcf] Matses, [mch] Ye'kwana, [mdz] Suruí do Pará, [mpd] Machinere, [mpq] Matis, [pab] Paresí, [pad] Paumari, [pak] Parakanã, [plu] Palikur, [taf] Tapirapé, [ter] Terena, [txi] Ikpeng, [xor] Korubo.

grupo 2 (50 línguas)

[apn] Apinaye, [apu] Apurinã, [apy] Aparai, [ash] Ashaninka, [atx] Arutani, [awe] Aweti, Tapayuna (beic1238)², [bor] Bororo, [cax] Chiquitano, [cbs] Kaxinawá, [cni] Ashaninka, [gae] Werekena, [gta] Guató, [gun] Guarani Mbya, [gvp] Parkatêjê, [jbt] Djeoromitxí (Jabuti), [jua] Kawahiva (Juma), [kay] Kamaiurá, [kqp] Kaingang, [kmv] Kheuól Galibi-Marworno, [kmv] Kheuól Karipuna, [knt] Katukina Pano, [kpj] Karajá, [kre] Panará, [kui] Kuikuro, [kuq] Kawahiva_Karipuna, [kyz] Kaiabi, [mbl] Maxakalí, [mzr] Marubo, [nhd] Guarani Nhandeva, [oym] Wajãpi, [pah] Kawahiva, [pto] Zoé, [pur] Puruborá, [ram] Canela, [rkb] Rikbaktsa, [skf] Sakurabiat, [suy] Suyá, [swo] Xinanenawa, [tpr] Tuparí, [txu] Kayapó, [urb] Ka'apor, [urz] Kawahiva_Uru Eu Wau Wau, [xav] Xavante, [xer] Xerente, [xok] Xokleng, [xra] Kraho, [xri] Krikati_Gavião, [yrl] Nheengatu, [zcp] Kaingang Paulista.

grupo 3 (34 línguas)

[ark] Arikapú, [arr] Karo, [cax] Chiquitano (Bolívia)*, [cin] Cinta Larga*, [cub] Kubeo*, [des] Desana**, [gug] Guarani Paraguaio, [gvc] Kotiria (Wanano)**, [jup] Hupdê*, [jur] Yudja, [kbc] Kadiweu, [kgk] Guarani-Kaiowa, [ktn] Karitiana, [Int] Latundê, [mav] Sateré-Mawé, [mbj] Nadëb, [mmh] Mehinaku, [myu] Mundurukú, [nab] Nambikwara, [pah] Tenharim, [pir] Piratapuyo**, [shb] Ninam*, [sru] Suruí, [tae] Tariana*, [tba] Aikanã*, [tca] Ticuna, [tpy] Trumai, [tue] Tuyuka**, [tuo] Tukano* **, [wau] Wauja, [wca] Yanomami*, [wmd] Mamaindé, [xsu] Sanöma, [xwa] Kwazá*.

* línguas que usam o caractere <i>.

** línguas que usam o caractere <u>.

2. Não há um código ISO para a língua. O pesquisador indicou o Glottocode beic1238.

grupo1 (27 línguas)

teclado ABNT ou Internacional padrão

[aoc] Taurepang
 [asu] Asuriní do Tocantins
 [atr] Waimiri Atoari
 [bkq] Bakairi
 [bwi] Baniwa-Koripako
 [cul] Kulina
 [dny] Deni
 [gub] Guajajara
 [ite] More-Kuyubim
 [hix] Hixkaryána
 [jaa] Jamamadi
 [kay] Kamayurá
 [knm] Kanamari
 [kpc] Baniwa_Coripaco
 [mcf] Matses
 [mch] Ye'kwana
 [mdz] Suruí do Pará
 [mpd] Machinere
 [mpq] Matis
 [pab] Paresí
 [pad] Paumari
 [pak] Parakanã
 [plu] Palikur
 [taf] Tapirapé
 [ter] Terena
 [txi] Ikpeng
 [xor] Korubo

grupo 2 (50 línguas)

teclado ABNT modificado

[apn] Apinaye
 [apu] Apurinã
 [apy] Aparai
 [ash] Ashaninka
 [atx] Arutani
 [awe] Aweti
 [] Tapayuna (beic1238)
 [bor] Bororo
 [cax] Chiquitano
 [cbs] Kaxinawá
 [cni] Ashaninka
 [gae] Werekena
 [gta] Guató
 [gun] Guarani Mbya
 [gvp] Parkatêjê
 [jbt] Djeoromitxí (Jabuti)
 [jua] Kawahiva (Juma)
 [kay] Kamaiurá
 [kqp] Kaingang
 [kmv] Kheuól Galibi-Marworno
 [kmv] Kheuól Karipuna
 [knt] Katukina Pano
 [kpj] Karajá
 [kre] Panará
 [kui] Kuikuro
 [kuq] Kawahiva_Karipuna
 [kyz] Kaiabi
 [mbl] Maxakalí
 [mzr] Marubo
 [nhd] Guarani Nhandeva
 [oym] Wajãpi
 [pah] Kawahiva
 [pto] Zoé
 [pur] Puruborá
 [ram] Canela
 [rkb] Rikbaktsa
 [skf] Sakurabiat
 [suy] Suyá
 [swo] Xinanenawa
 [tpr] Tuparí
 [txu] Kayapó
 [urb] Ka'apor
 [urz] Kawahiva_Uru Eu Wau Wau
 [xav] Xavante
 [xer] Xerente
 [xok] Xokleng
 [xra] Kraho
 [xri] Krikati_Gavião
 [yrl] Nheengatu
 [zkp] Kaingang Paulista

grupo3 (34 línguas)

teclado para caracteres complexos

[ark] Arikapú
 [arr] Karo
 [cax] Chiquitano (Bolívia) *
 [cin] Cinta Larga *
 [cub] Kubeo *
 [des] Desana °
 [gug] Guarani Paraguaio
 [gvc] Kotiria (Wanano) °
 [jup] Hupdë *
 [jur] Yudja
 [kbc] Kadiweu
 [kgk] Guarani-Kaiowa
 [ktn] Karitiana
 [lnt] Latundê
 [mav] Sateré-Mawé
 [mbj] Nadëb
 [mmh] Mehinaku
 [myu] Mundurukú
 [nab] Nambikwara
 [pah] Tenharim
 [pir] Piratapuyo °
 [shb] Ninam *
 [sru] Suruí
 [tae] Tariana *
 [tba] Aikanã *
 [tca] Ticuna
 [tpy] Trumai
 [tue] Tuyuka °
 [tuo] Tukano * °
 [wau] Wauja
 [wca] Yanomami*
 [wmd] Mamaindê
 [xsu] Sanõma
 [xwa] Kwazá*

* línguas que usam o caractere <ï>

° línguas que usam o caractere <ɥ>

Aqui, neste ponto, deve-se fazer um parêntese: sobre a lista das línguas, é importante mencionar que este é um trabalho em desenvolvimento e que dados incorretos podem (e devem) ser revisados. As causas dos dados incorretos variam desde inconsistências nas fontes primárias a erros na interpretação (por parte deste autor) dessas fontes. O importante é que a origem da entrada dos dados (fonte primária ou coleta de dados por formulário) foi documentada, sendo possível rastreá-la. Além disso, destaca-se que a catalogação ortográfica pode ser integrada a outros bancos de dados linguísticos e que o cruzamento dos dados com informações de outras bases pode levar a novas pesquisas.

A Fig. 148 mostra os caracteres necessários para cada um dos três grupos. Nos apêndices pode-se consultar as tabelas específicas de cada um deles, em que constam também os pontos de código e nomes dos glifos. Além disso, pode-se consultar, separadamente, os conjuntos de caracteres necessários para cada língua. As tabelas estão disponíveis no link a seguir:

<https://github.com/RafaelDietzsch/LinguasIndigenasBrasileiras>

Fazendo uma breve avaliação dos resultados, pode-se dizer que, a partir dessa catalogação, é possível desenvolver uma especificação para cada grupo de línguas, especialmente para o grupo 3, que envolve casos de composição mais complicados.

Reforça-se também que, do ponto de vista da codificação, as línguas do grupo 1 são relativamente simples, já que utilizam somente caracteres da codificação Latin-1 e não dependem de teclados especiais. Além disso, conforme explicado anteriormente, a grande maioria das fontes profissionais apresenta esse conjunto mínimo obrigatório. Nesse sentido, as línguas do grupo 1 não dependem de fontes e teclados especiais; com o que há disponível hoje se pode produzir tipografia de qualidade para essas línguas.

As próximas seções mostram como se deu o desenvolvimento do projeto: da especificação de diretrizes até o resultado final.

13.2 Diretrizes de desenvolvimento

A partir das informações reunidas ao longo da pesquisa foi possível estabelecer uma série de diretrizes de desenvolvimento para a tipografia de línguas indígenas brasileiras.

Primeiramente, é importante reforçar alguns dos principais fatores para renderização correta, que serão decisivos no resultado final, especialmente para as línguas do grupo 3:

- Pontos de código com a semântica apropriada, organizados de acordo com os princípios do texto puro e ordem lógica.
- Teclados preparados especificamente para caracteres pré-compostos ou para marcas combinadas; aplicativos e sistemas operacionais que reconheçam essas formas de entrada.
- Diacríticos com relações harmônicas de proporção e peso com o restante da fonte.
- Posicionamento e configuração corretos das âncoras nas letras base e nos diacríticos.
- Fontes com recursos OpenType apropriados.
- Mecanismo de modelagem atuando no nível do sistema operacional ou no aplicativo.

Além dos fatores supracitados, destaca-se a seguir algumas características das línguas indígenas brasileiras que devem ser levadas em conta na esfera do desenvolvimento de fontes digitais e teclados.

13.2.1 Especificação do conjunto de caracteres acentuados

No planejamento de uma fonte com suporte linguístico para línguas indígenas brasileiras, uma das primeiras coisas a se avaliar é o conjunto de caracteres acentuados necessários para uma língua ou grupo de línguas. Para isso, utiliza-se a catalogação ortográfica como ferramenta auxiliar. A partir de uma consulta rápida na catalogação, é possível determinar quais os caracteres acentuados necessários para aquela(s) língua(s) e também em qual dos três grupos ela(s) se encaixa(m). Em seguida, pode-se começar a adicionar³ os diacríticos e os caracteres pré-compostos às fontes.

A Fig.149 mostra o conjunto completo de caracteres acentuados detectados no levantamento de dados e organizados conforme o sistema de classificação da catalogação ortográfica. Essa amostra também pode ser utilizada para identificação, a partir de uma análise visual, de quais são os caracteres acentuados necessários para determinada

3. Caso os diacríticos ainda não existam na fonte com a qual se está trabalhando. As diretrizes podem ser aplicadas tanto a fontes existentes quanto a novos projetos.

13.2.2 Planejamento e organização dos conjuntos de diacríticos

A seguir, explica-se a abordagem utilizada para o planejamento e desenvolvimento dos diacríticos e de sua organização dentro do conjunto de caracteres de uma fonte.

Antes de explicar essa abordagem, deve-se ressaltar que existe hoje uma bibliografia considerável sobre diacríticos⁴ na literatura da tipografia, especialmente por parte de autores e designers de tipos do leste europeu, que abordam não somente questões de design, mas também aspectos históricos, linguísticos e tecnológicos. Por esse motivo, esta tese não se aprofundará em questões básicas do design de diacríticos.

Primeiramente, é importante definir alguns parâmetros básicos para o planejamento dos diacríticos. Nessa linha, a amostra da Fig. 150 apresenta um grupo básico de diacríticos. A inclusão desse grupo básico é considerada uma boa prática⁵ no design de tipos.



Figura 150: Conjunto básico de diacríticos. *grave, acute, circumflex, tilde, macron, breve, dotaccent, dieresis, ring, hungarumlaut, caron, cedilla, ogonek*. Fonte: do autor.

A partir dessas formas básicas, recomenda-se a criação de dois grupos de diacríticos dentro das fontes: as marcas espaçadas (*spacing marks*, i.e. default) e as marcas combinadas. Isso representa pouco trabalho a mais, já que uma é praticamente uma cópia da outra. A principal diferença é que, apesar de parecerem idênticos, esses diacríticos

4. Referências: BLAŽEK, 2004, 2020; BREZINA, 2009; BURIAN, 2021; DIETZSCH, 2012; GAULTNEY, 2002, 2006; GIOVAGNOLI, 2011; HUDSON, 2014; JACQUERYE, 2013; JÓB, 2013; KAPLAN, 2011; LÓPEZ ROCHA, 2017; OSLISLO; MAŁECKA, 2016; SAMUŽENKOVA, 2016, 2019; TRUONG, 2018; TWARDOCH, 1999, 2002; UNGER, 2018. Além destes, destaca-se o meticuloso trabalho dos manuais e tutoriais do Glyphs, com valiosas informações sobre o design de diacríticos (SCHEICHELBAUER; SEIFERT, 2016, 2021).

5. Adota-se o termo “boa prática” como sinônimo de prática padrão da indústria, uma obrigatoriedade. E não somente para as línguas indígenas, mas de maneira geral.

Uma abordagem proposta pela designer argentina Carolina Giovagnoli (2011, p. 36) é a de criar variações mais estreitas¹⁰ dos diacríticos para a letra <i>, conforme se pode ver na Fig.154. Conforme explicado no Capítulo 11, diacríticos largos em caracteres como o <ĩ> tendem a se chocar com outros caracteres adjacentes. Por essa razão, fazer versões dos diacríticos mais estreitos para o <i> pode ser uma forma de se minimizar o problema. A Fig.155 mostra esses caracteres adicionados ao conjunto, destacados em laranja. Esses glifos são nomeados com o sufixo *comb.narrow* (*narrow*, estreito).

Tres set de diacríticos: minúsculos, mayúsculos, i.

ó í Ó Í Õ õ ï

Alineación media.

Oñeha'ã apópe rupi.

Figura 154: Três variações da forma do mesmo diacrítico. Andada, de Carolina Giovagnoli.

Fonte: GIOVAGNOLI,2011, p.36.

ā ī ñ ñ ã ã ô ô

Figura 155: Variações da forma dos diacríticos. comb, comb.cap e comb.narrow. Fonte: do autor.

10. As versões alternativas dos diacríticos para a letra <i> não têm ponto de código. São glifos dentro da fonte que podem ser usados para gerar caracteres pré-compostos. Também podem ser acessados por meio de recursos de substituição OpenType (GSUB).

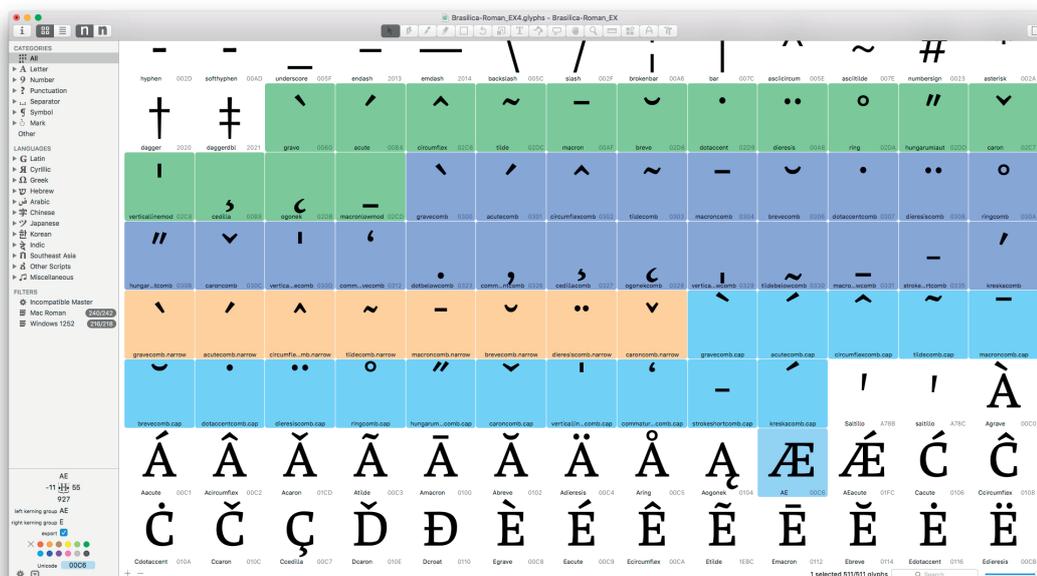


Figura 156: Captura de tela do software Glyphs mostrando quatro conjuntos de diacríticos: *default*, *comb*, *comb.narrow*, *comb.cap*. Fonte: do autor.

Além das quatro variações supracitadas, é possível criar ainda versões para os casos de línguas que utilizam diacríticos empilhados. Esses diacríticos são mais achataados e têm um alinhamento mais baixo para evitar que, em um texto, as descendentes da linha de cima se choquem com os diacríticos (Fig. 157 a 159). Esses glifos são nomeados com o sufixo *comb.low* (baixo), destacados em rosa, e os maiúsculos *comb.cap.low*, destacados em amarelo (Fig.160).

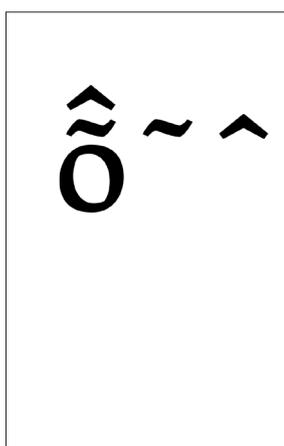


Figura 157: Variações dos diacríticos *comb.low*.
Fonte: do autor.

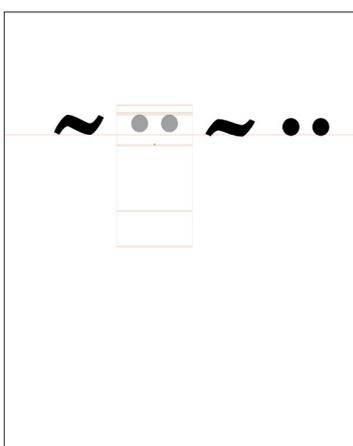


Figura 158: Comparação entre as variações *comb* (esq.) e *comb.low* (dir.). Fonte: do autor.

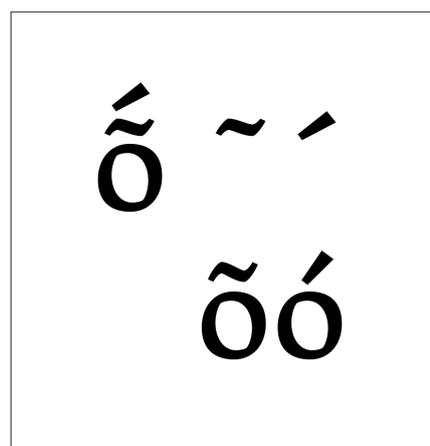


Figura 159: Comparação entre as variações *comb.low* (topo) e *comb* (abaixo). Fonte: do autor.

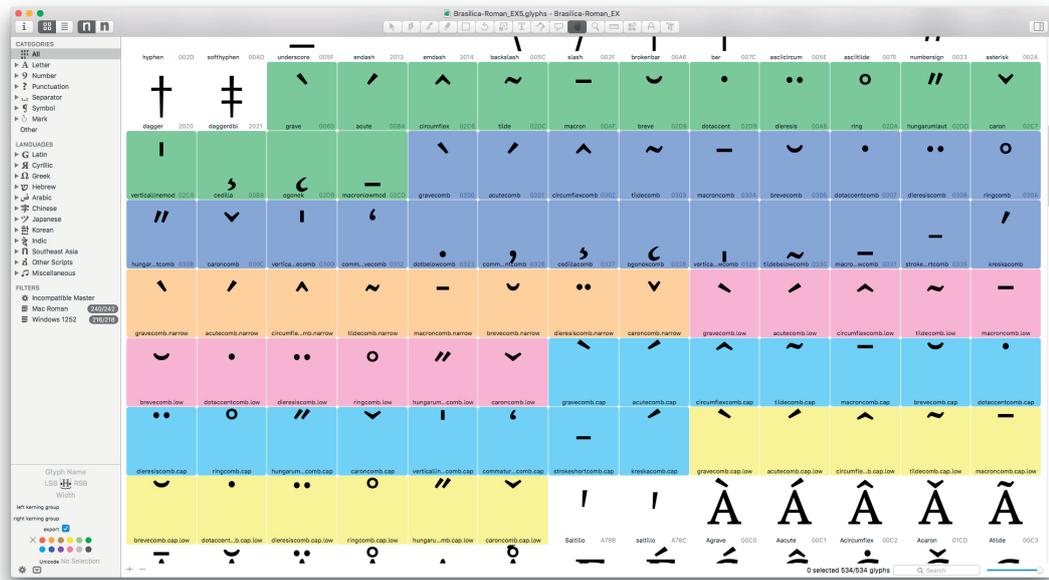


Figura 160: Captura de tela do software Glyphs mostrando seis conjuntos de diacríticos: *default*, *comb*, *comb.narrow*, *comb.low*, *comb.cap*, *comb.cap.low*. Fonte: do autor.

Por último, pode-se adicionar ainda uma última variação para empilhamento na letra <i>, com o sufixo *comb.narrow.low*, destacada em verde claro (Fig.160). Essas versões são mais baixas e mais estreitas (Fig.161).



Figura 161: Comparação entre as variações *comb.narrow.low* (topo) *comb.narrow* (abaixo).

Fonte: do autor.

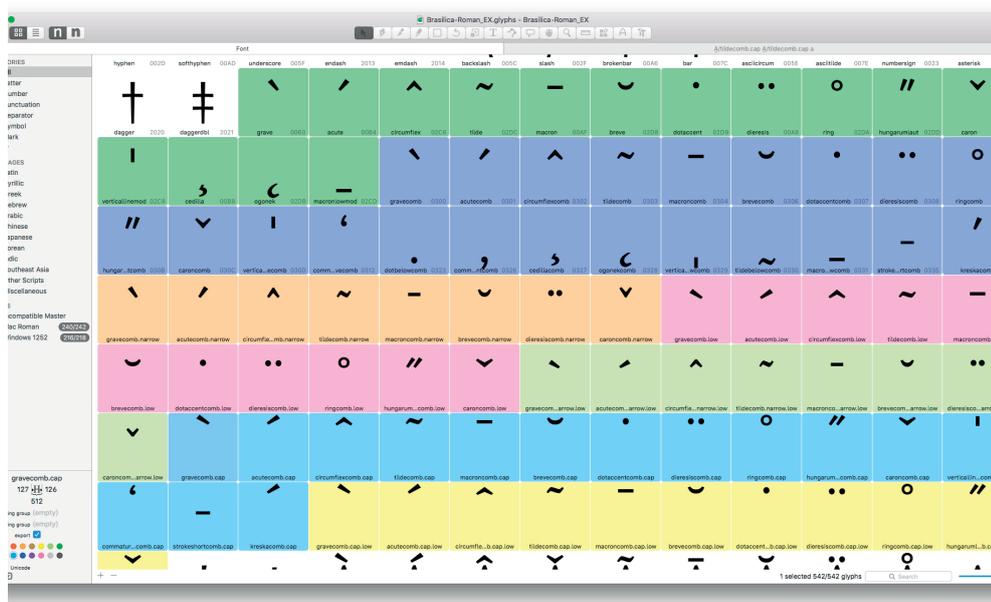


Figura 162: Captura de tela do software Glyphs mostrando sete conjuntos de diacríticos: *default*, *comb*, *comb.narrow*, *comb.low*, *comb.narrow.low*, *comb.cap*, *comb.cap.low*. Fonte: do autor.

Em resumo, é importante mencionar que essa maneira de se planejar os conjuntos de caracteres é só uma proposta, uma abordagem utilizada neste projeto, e não é obrigatória. A intenção aqui é a criação de um arcabouço que possa resolver até mesmo os problemas mais complexos. Trata-se de uma proposta experimental, já que atualmente a grande maioria das fontes não é construída dessa maneira. Em raros casos será preciso incluir os sete conjuntos de diacríticos. Além disso, na prática, só se utiliza partes deles, normalmente (i.e. alguns diacríticos somente). Aqui, a título de demonstração, desenvolveu-se um conjunto completo. Utilizou-se para o experimento a fonte Brasilica, projetada por este autor, com contribuições de Kalapi Gajjar-Bordawekar (programação Open Type).

Essa forma de se trabalhar pode ter como resultado fontes com um grande número total de glifos. Sem contar que, se um novo estilo (e.g. bold ou itálico) for adicionado, esse número pode dobrar. Lidar com um conjunto como esse exige um trabalho cuidadoso de controle de versões. Por essa razão, é preciso ponderar em que casos é realmente fundamental a aplicação dessa abordagem.

Por exemplo: acredita-se que, em fontes sem serifa, uma forma mais estreita dos diacríticos para a letra <i> seja mais harmônica, podendo melhorar a qualidade de leitura. Essa condensação é de aproximadamente 80% da largura da forma padrão.

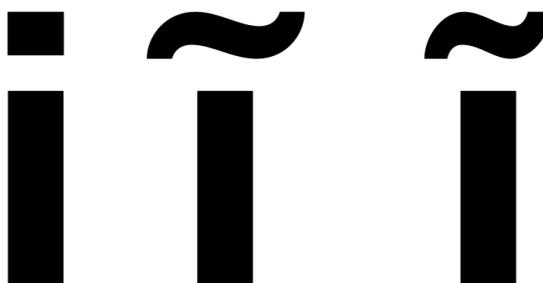


Figura 163: Variações na largura do til. O til à direita foi modificado a partir da forma básica do til (ao centro) da fonte Chivo Bold, de Hector Gatti. Fonte: do autor.

Já no caso de fontes serifadas, é possível planejar, desde o início, formas dos diacríticos que sejam mais versáteis, i.e. nem tão largas, nem tão estreitas, e que possam se adaptar mais facilmente a todas as condições. Essa não é, no entanto, uma tarefa fácil, e pode levar a resultados com diacríticos muito pequenos ou visualmente muito leves em relação ao peso visual das letras.

Esta seção reuniu uma série de informações de caráter *macro*, i.e. relacionadas ao comportamento do conjunto como um todo. A partir dessas informações, de como se planeja os conjuntos de diacríticos de uma fonte para línguas indígenas brasileiras, as próximas seções se concentram em aspectos *micro*, i.e. propriedades intrínsecas ao caractere, particularidades de sua estrutura interna.

13.2.3 Âncoras e componentes

Esta seção tem como objetivo descrever alguns parâmetros para construção de âncoras e componentes, explicar suas funções e funcionamento e, também, demonstrar como podem ser utilizados para a ampliação do suporte linguístico de uma fonte.

Além de suas características formais (contornos, espaçamento), os caracteres dispõem de informações de metadados e de alguns mecanismos que garantem seu bom funcionamento em determinadas funções e/ou contextos. Dentre esses mecanismos, é preciso explicar aqui como são configuradas as âncoras, tanto dos diacríticos como das letras. Conforme observado anteriormente, no capítulo *Codificação e linguagem*, a função das âncoras ou pontos de ancoragem (*anchors, attachment points*) é demarcar uma

coordenada (i.e. demarcar o centro óptico) para conexão das marcas às letras base. O ato de demarcar o centro óptico das letras e dos diacríticos garante o alinhamento correto de qualquer diacrítico, em qualquer letra. Incluir as âncoras é um processo trabalhoso, mas garante uma consistência ao longo do desenvolvimento.

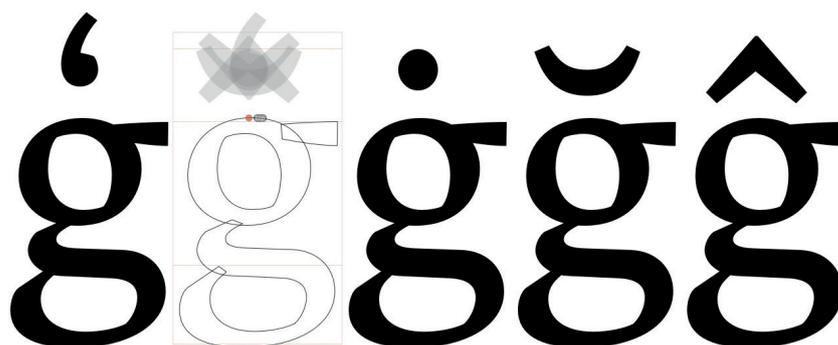


Figura 164: Representação do centro óptico da letra <g>. A nuvem cinza acima da letra demonstra que a âncora é ponto comum de alinhamento de todos os diacríticos. Fonte: do autor.

Outro recurso essencial para a construção de conjuntos de caracteres acentuados é o uso de componentes. A principal função dos componentes é a reutilização de formas para a construção de glifos pré-compostos, sempre tendo como referência um glifo base. Por exemplo, para estruturar um caractere como <ñ>, basta somar os dois componentes <n> e <~>. Se ambos os componentes (letra base e diacrítico) tiverem âncoras, o alinhamento é automático. A grande vantagem de se utilizar componentes é que se houver uma modificação em algum glifo base, ele se modifica automaticamente em todos os caracteres pré-compostos nos quais foi utilizado.

Uma segunda funcionalidade dos componentes é a possibilidade de criação de caracteres pré-compostos automaticamente por meio de fórmulas. Para isso, a maioria dos editores de fontes aceita, como forma de entrada, listas em formato de texto para geração dos componentes. Por exemplo, para criar os caracteres <à>, <á>, <â>, <ã>, <ä>, <ā>, <ǎ>, utiliza-se o seguinte texto como entrada: a+gravecomb=agrave a+acutecomb=acute a+circumflexcomb=acircumflex a+tildecomb=atilde a+dieresiscomb=

b=adieresis a+macroncomb=amacron a+brevecomb=abreve¹¹, tendo como resultado os caracteres ilustrados na Fig.165.

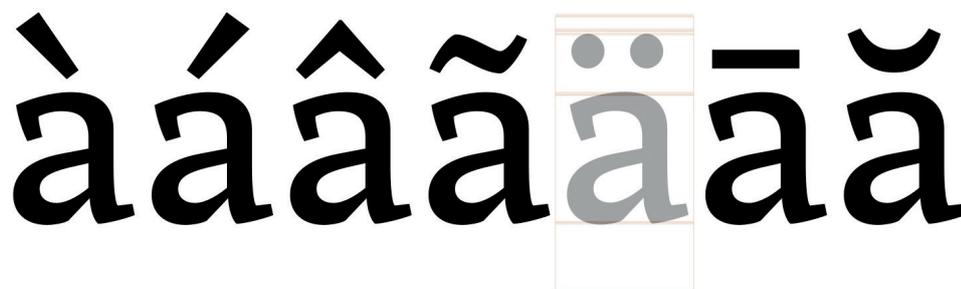


Figura 165: Caracteres criados com o uso de componentes, representados em cinza. Fonte: do autor.

No contexto da tipografia das línguas indígenas brasileiras, pode-se considerar o uso de âncoras e componentes como algo essencial. Especialmente para as línguas do grupo 3, para o qual as âncoras são indispensáveis. Isso quer dizer que, sem âncoras, as línguas do grupo 3 não podem ser compostas corretamente. Nessa linha, adotou-se algumas diretrizes e padrões para o desenvolvimento e implementação das âncoras.

Primeiramente, é importante mencionar que todos os diacríticos têm a mesma largura da letra <0>, e todos os glifos e âncoras são centralizados nessa largura (Fig.167). A única exceção são os diacríticos assimétricos (e.g. agudo e grave), que têm a âncora deslocada para o centro óptico do diacrítico. Além disso, as âncoras são sempre posicionadas em alguma linha da métrica vertical (e.g. linha de base, altura-x).

Segundo, adotou-se como diretriz a inclusão de âncoras em todas as vogais (Fig.166), maiúsculas e minúsculas. Exceto pelo <y>, todas as vogais podem receber âncoras no topo e na base. E no caso do <i> e do <u>, há ainda uma terceira âncora, no meio da altura-x. Ainda, deve-se observar que, no caso do <i>, há um componente específico, o *idotless* (i sem pingão), que é usado para gerar todos os caracteres pré-compostos relacionados a essa vogal.

11. Sintaxe adotada pelo programa Glyphs.

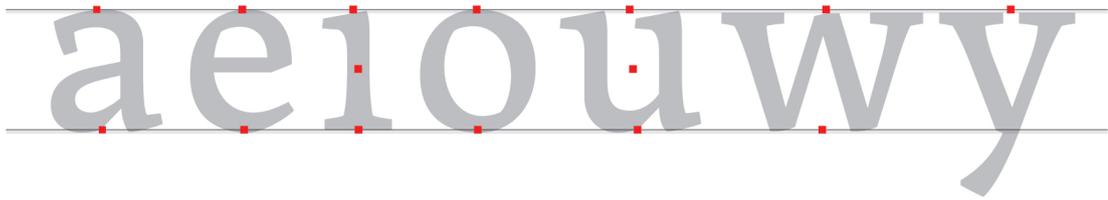


Figura 166: Ilustração mostrando o posicionamento das âncoras nas vogais. Fonte: do autor.

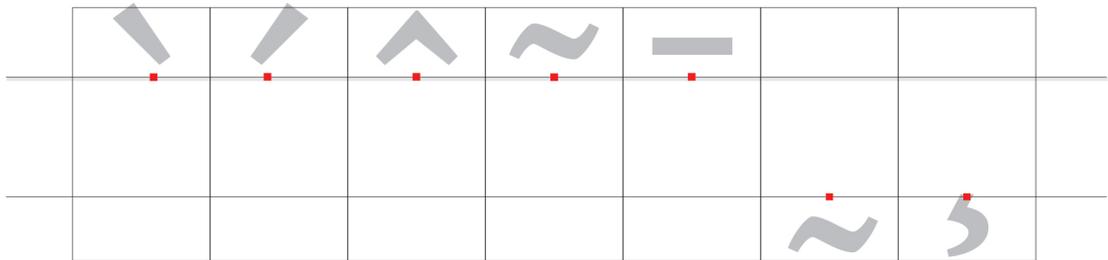


Figura 167: Ilustração mostrando o posicionamento das âncoras em alguns diacríticos.
Fonte: do autor.

No início da seção anterior definiu-se como boa prática e como parâmetro de desenvolvimento um conjunto de diacríticos básico, com o qual é possível gerar os caracteres pré-compostos necessários para pelo menos 2/3 das 111 línguas catalogadas. Esse mesmo raciocínio vale para as âncoras e componentes: se houverem âncoras em todos os componentes, ampliam-se exponencialmente as possibilidades de cobertura linguística, com caracteres gerados automaticamente por meio de fórmulas (e.g. a+gravecomb=agrave), como na Fig.168.

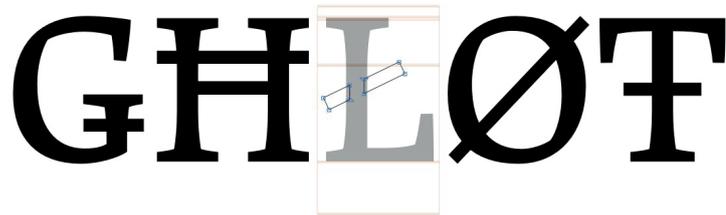


Figura 169: Caracteres pré-compostos com diacríticos sobrepostos . Fonte: do autor.

A outra particularidade é que, enquanto a maiúscula é uma forma fácil de ser resolver e apresenta uma certa padronização em relação à posição do diacrítico, na minúscula ocorrem três formas diferentes de apresentação, conforme se pode ver na Fig.170. A diretriz para o caso das minúsculas é de tentar adaptar o traço de acordo com o design da fonte. No entanto, observa-se que a forma da letra com “dois andares” representa alguns desafios, especialmente nos estilos mais pesados (e.g. bold), já que a contraforma inferior é muito pequena e tem a tendência de se fechar quando o diacrítico é sobreposto à letra.

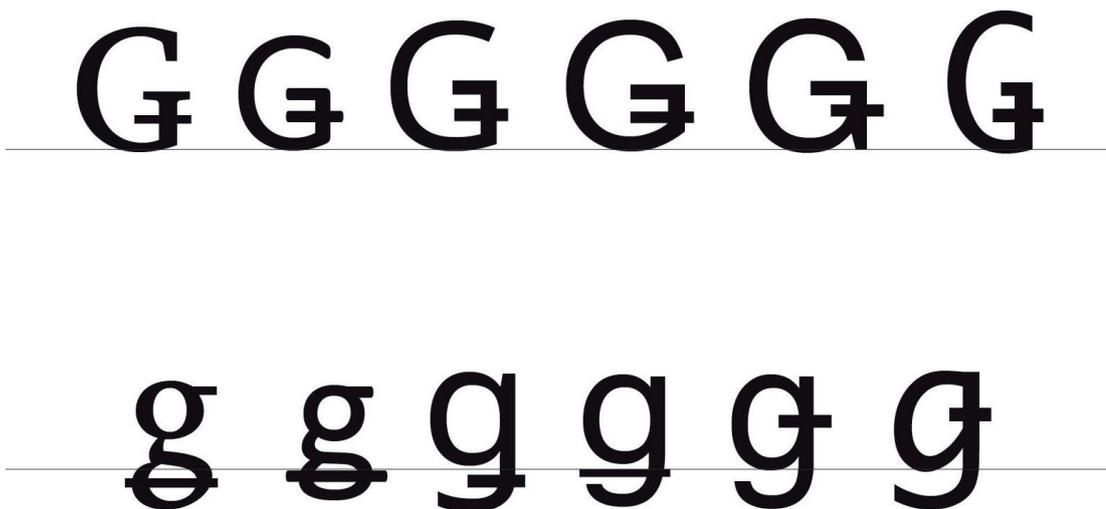


Figura 170: Variantes dos caracteres <g> e <G> em diferentes fontes. Da esq. para a dir.: Charis, Calibri, Noto Sans, Arial, Helvetica Neue, Monaco. Fonte: do autor.

É importante mencionar que no caso do <g> e <G>, a grande maioria das fontes possuem formas que deixam a desejar. Isso acontece porque, muitas das vezes, o caractere é incluído na fonte como parte de uma política de cobertura linguística que, em geral, compreende um número muito grande de caracteres. Isso leva, naturalmente, a abordagens que comprometem a integridade do caractere.

O segundo caso a se relatar é o do <u> e <U>, *ubar* e *Ubar* (Fig.171), detectados em cinco das línguas do levantamento. Esses caracteres não representam grandes problemas de codificação, já que ambos são formas pré-compostas codificadas pelo Unicode. Apesar disso, é interessante observar que, enquanto a forma minúscula está no bloco IPA Extensions, a forma maiúscula está no bloco Latin Extended-B.

O design desses caracteres é relativamente fácil de se resolver. Deve-se, no entanto, observar a necessidade de inclusão de âncoras no topo das letras, já que a maioria das línguas que utilizam esses caracteres também utilizam diacríticos sobre eles. As âncoras no meio da altura da letra podem ser utilizadas para posicionamento de componentes durante o desenvolvimento da fonte.

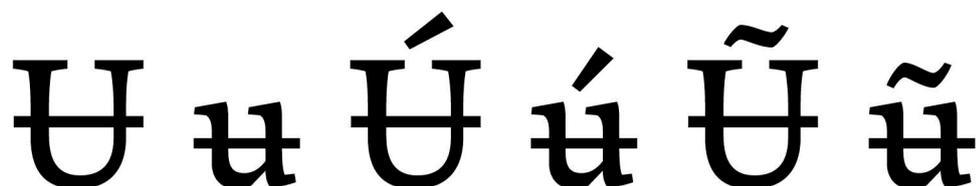


Figura 171: Amostras dos caracteres <u> e <U>, assim como de seu uso com marcas combinadas. Fonte: do autor.

Por último, e o mais complicado deles, é o caso dos caracteres <i> e <I>, *istroke* e *Istroke*, detectados em dez línguas no levantamento. Esses caracteres, na sua configuração básica (sem marcas combinadas no topo) não representam grandes problemas de codificação, já que ambos são formas pré-compostas codificadas pelo Unicode. Apesar disso, é interessante observar que, enquanto a forma minúscula está no bloco IPA Extensions, a forma maiúscula está no bloco Latin Extended-B.

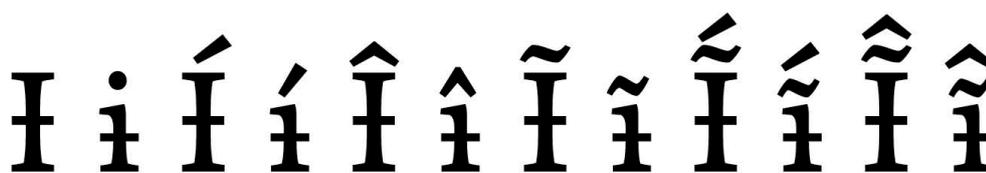


Figura 172: Amostras dos caracteres <i> e <I>, assim como de seu uso com marcas combinadas. Fonte: do autor.

O design desses caracteres é relativamente fácil de se resolver. Deve-se, no entanto, observar a necessidade de inclusão de âncoras no topo das letras, já que a maioria das línguas que utilizam esses caracteres também utilizam diacríticos sobre ele. No caso das minúsculas, é necessário um glifo, que não tem ponto de código, chamado *idotlessstroke* (i cortado sem pingo), utilizado para as variantes desse caractere com marcas combinadas. Por se tratar de um glifo não codificado, depende de recursos OpenType para funcionar corretamente.

Nesse sentido, são muito poucas as fontes que dão suporte para o <i> e <I> associados a marcas combinadas. Isto porque a fonte deve conter uma programação OpenType que, em linhas gerais, decompõe o glifo inteiro e o compõe novamente, de acordo com a entrada digitada. Para isso, a fonte deve contar não somente com os recursos OpenType apropriados, mas também com todos os componentes separados, com as devidas âncoras. John Hudson (2014, p. 15–18) relata um problema semelhante no desenvolvimento da fonte para editora Brill, em que mostra um caractere formado por 4 pontos de código, para o qual descreve 14 diferentes formas de entrada (Fig.173).

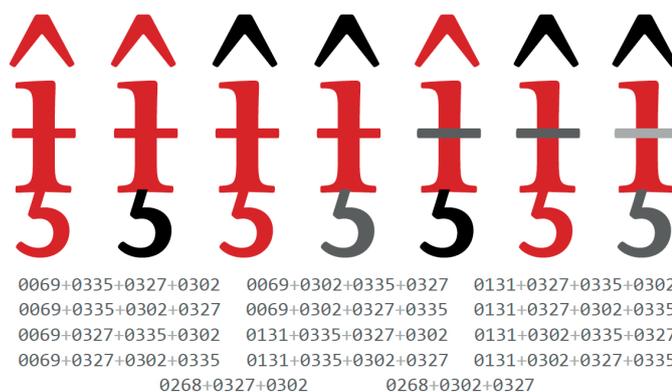


Figura 173: Diferentes formas de entrada para o mesmo caractere. Fonte: HUDSON, 2014, p. 16.

13.2.5 Diacríticos empilhados

Esta seção descreve diretrizes para estruturação de uma fonte capaz de renderizar corretamente diacríticos empilhados. Conforme explicado anteriormente, é possível pensar em formas mais achatadas, compactas e com um alinhamento mais baixo, para evitar que, em um texto, as descendentes da linha de cima se choquem com os diacríticos. Além disso, é fundamental que os diacríticos planejados para empilhamento apresentem duas âncoras: uma abaixo do diacrítico, para conexão com a letra base, e outra acima do diacrítico, para o empilhamento da próxima marca.

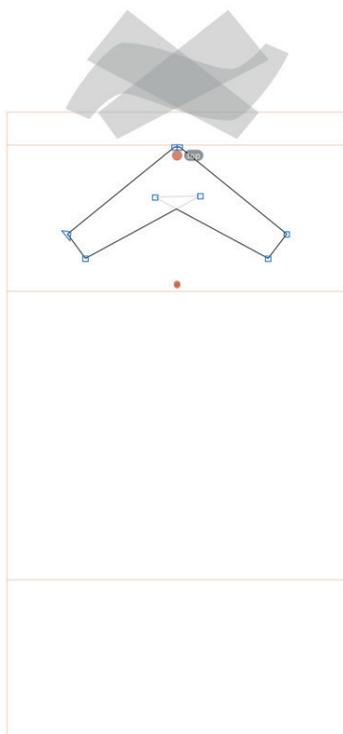


Figura 174: Exemplo de marca combinada com duas âncoras, para renderização de diacríticos empilhados. Fonte: do autor.

13.2.6 Recursos Open Type

Esta seção descreve alguns dos recursos OpenType necessários para o correto funcionamento de caracteres que não apresentam uma forma pré-composta e dependem das marcas combinadas para uma renderização correta. Destacam-se três *lookups* utilizados: um de substituição (GSUB) e dois de posicionamento (GPOS).

Entre os *lookups* de posicionamento, destacam-se o do tipo *mark* (*Mark-to-base positioning*, posicionamento de marca em relação à base), que conecta uma marca combinada a um glifo base, e o do tipo *mkmk* (*Mark-to-mark positioning*, posicionamento de marca em relação a outra marca), que conecta uma marca combinada a outra marca.

Para os *lookups* do tipo *mark* e *mkmk* não é preciso escrever o código. Hoje em dia, editores de fontes como o Glyphs geram o código desses recursos OpenType automaticamente. Para tal, é preciso que o designer coloque as âncoras na posição correta, tanto nos diacríticos como nas letras. Nesse sentido, praticamente todas as marcas combinadas são renderizadas automaticamente por meio desses recursos.

A exceção são os caracteres <i>, <i> e <j>, que precisam de um tratamento especial por causa do pingo. Para a renderização correta, é necessário programar o *lookup* de substituição *ccmp* (*Composition/Decomposition*, composição e decomposição). Esse recurso, em linhas gerais, faz uma decomposição do caractere da seguinte maneira: quando se digita o <i> seguido por uma marca combinada (e.g. *circumflexcomb*), o *ccmp* é o recurso que decompõe o glifo, substituindo o glifo base pela sua versão sem pingo (*idotlessstroke*), para que este possa receber marcas combinadas, conforme exemplo da Fig.172.

O código da programação OpenType das fontes desenvolvidas durante este projeto pode ser consultado nas próprias fontes, disponíveis no repositório digital¹² do projeto.

Por último, mas não menos importante, é preciso creditar e reconhecer o trabalho do designer de tipos indiano Kalapi Gajjar-Bordawekar, colega deste autor, da época do curso de mestrado (MA Typeface Design) na Universidade de Reading, Inglaterra. Kalapi escreveu o código OpenType das fontes e supervisionou a engenharia de software do processo. Além disso, as discussões com o colega trouxeram inestimáveis contribuições para a pesquisa, especialmente para as diretrizes descritas neste capítulo. Por último, Kalapi ajudou ainda na especificação das diretrizes dos teclados específicos para as línguas indígenas, assunto da próxima seção.

12. <https://github.com/RafaelDietzsch/LinguasIndigenasBrasileiras>

13.2.7 Teclados

Esta seção tem como objetivo explicar as diretrizes de planejamento dos layouts de teclado. Esta abordagem se concentra, principalmente, em teclados para computadores desktop e laptop. As diretrizes de desenvolvimento foram pensadas a partir dos requisitos dos três grupos linguísticos da catalogação ortográfica.

Nesse sentido, reforça-se que para as línguas do grupo 1 não é necessário nenhum teclado especial; qualquer teclado brasileiro padrão servirá ao propósito. Teclados com layout norte-americano também são funcionais para esse grupo.

Para o grupo 2, que utiliza somente caracteres pré-compostos, foi pensado um layout de teclado personalizado. A principal característica desse teclado é que ele é capaz de colocar qualquer diacrítico (do grupo básico) em qualquer vogal, por meio de teclas mortas. Além disso, funciona também para entrada das consoantes acentuadas detectadas na catalogação. A principal característica desse desenvolvimento é amplificação das possibilidades de combinação das teclas mortas dos teclados padrão ABNT e ANSI. Por exemplo: os teclados padrão brasileiros não permitem a digitação do til em todas as vogais, somente no <ã> e <õ>. Com a reconfiguração das teclas mortas, é possível dar entrada em caracteres como <ẽ>, <ĩ>, <ũ>, <ỹ>.

Para o grupo 3, que utiliza marcas combinadas, utilizou-se a mesma premissa do teclado acima mencionado, capaz de colocar qualquer diacrítico em qualquer vogal. A diferença é que, por se tratar de marcas combinadas, não se utiliza nenhuma tecla morta neste layout. Também é importante mencionar que, por se tratar de marcas combinadas, esse teclado permite o empilhamento de diacríticos.

Ainda, é importante mencionar que a principal diretriz de desenvolvimento dos teclados é que eles respeitassem os layouts de teclado já conhecidos e de maior acesso no Brasil: ABNT e ANSI. Dessa forma, o usuário vai encontrar o diacrítico onde ele sempre esteve. A exceção são algumas marcas combinadas que foram acrescentadas (*tildebelowcomb* <_>, *macronbelowcomb* <_>, *macroncomb* <^->) para as línguas do grupo 3 e que não existem nos teclados ABNT e ANSI.

Por último, destaca-se a diretriz que indica a necessidade de que todos os layouts de teclado sejam disponibilizados para as plataformas MS Windows e MacOS.

13.3 Desenvolvimento de fontes

Esta seção tem como objetivo descrever o desenvolvimento das famílias de fontes que foram customizadas especificamente para este projeto, com o suporte do designer de tipos Kalapi Gajjar-Bordawekar.

Primeiramente, é importante explicar como se deu a escolha das fontes para o desenvolvimento. Por se tratar de um projeto que tem como finalidade a inclusão e a promoção das línguas indígenas brasileiras no meio digital, é importante que suas ferramentas sejam de caráter *open-source*. Por essa razão, a pesquisa teve como ponto de partida a exigência de que se trabalhasse somente com fontes livres.

A partir dessa premissa, iniciou-se uma pesquisa nos repositórios de fontes *open-source*; particularmente, no site Google Fonts. Com a explosão das *webfonts* na última década e com o recente crescimento da demanda por *webfonts* para aplicativos de celular e interfaces de smart TV, a demanda por fontes para web também aumentou proporcionalmente. Por essa razão, o Google (entre outras empresas) passou a investir mais fortemente em fontes *open-source*; em especial, em algumas iniciativas que têm como finalidade a expansão do suporte linguístico, levando ao crescimento dessa biblioteca de fontes, tanto quantitativamente como qualitativamente.

Um grande entusiasta e colaborador deste projeto do Google é o professor e designer de tipos argentino Pablo Cosgaya. À frente da fundidora Omnibus-Type, publicou uma série de famílias de fontes de qualidade, em parceria com vários designers latino-americanos. As famílias da Omnibus-Type são distribuídas livremente no GoogleFonts e também em seu próprio site. As versões originais e os arquivos editáveis das fontes também estão disponíveis em seu repositório público no GitHub.

A ideia inicial era escolher algumas famílias da Omnibus-Type para customização. No entanto, só se produziu, de fato, duas famílias funcionais com resultados consistentes. As famílias são a serifada Faustina, de autoria de Alfonso Garcia (Fig.176) e a sem serifa Chivo, de Hector Gatti (Fig.175).

My profile

BLACK

Engraving

EXTRABOLD

Chamorro

BOLD

Eventually

MEDIUM

Lithuanian

REGULAR

Prototypes

LIGHT

Mysterious

THIN

CHIVO THIN, CHIVO LIGHT, CHIVO REGULAR, CHIVO MEDIUM, CHIVO BOLD, CHIVO EXTRABOLD, CHIVO BLACK (WITH ITALICS)

Figura 175: Chivo (specimen). Fonte: GATTI; OMNIBUS-TYPE, 2019

FAUSTINA
SPECIMEN



François Hollande **INTERNATIONAL**

Life Inside Hong Kong's 'Coffin Cubicles'

664 GLYPHS
PER FONT

La servante écarlate

L'élection de Donald Trump en a fait un livre prophétique

Gather the Daughters

Two new Harry Potter 'History of Magic' books coming this October

Geschäftsführergehälter

Luc Besson adapta ao cinema "*Valerian e a cidade dos mil planetas*"

FAUSTINA
SPECIMEN



REGULAR
29 PT

Democracy and free markets are intimately

MEDIUM
41 PT

connected to organised crime

SEMIBOLD
43 PT

notes Federico Varese in his

BOLD
55 PT

exploration of mafias

ITALIC
31 PT

A fossil older than Adam and Eve: how early

MEDIUM ITALIC
35 PT

geologists discovered the secret of life

SEMIBOLD ITALIC
43 PT

Two centuries ago, geology has

BOLD ITALIC
45 PT

enjoyed fashionable interest

Figura 176: Faustina (specimen). Fonte: GARCÍA; OMNIBUS-TYPE, 2017

A partir das diretrizes de desenvolvimento de projeto e da especificação do conjunto de caracteres para línguas indígenas brasileiras, as versões romanas das duas famílias foram customizadas em todos os pesos (de light a black), com todos os caracteres e recursos OpenType necessários para composição de texto nas 111 línguas indígenas catalogadas no projeto. É importante mencionar que o fato de as famílias Chivo e Faustina já terem um sólido histórico de desenvolvimento facilitou muito o trabalho de customização das fontes.

Nesse sentido, foram acrescentados apenas alguns caracteres, ao conjunto já existente. Além disso, a programação OpenType necessária também foi incorporada à fonte.

As duas famílias completas, assim como as instruções de instalação das fontes, estão disponíveis no repositório digital do projeto.

Thin	Light
Light	Regular
Regular	Medium
Medium	Bold
Bold	Bold
ExtraBold	ExtraBold
Black	

Figura 177: Variações de pesos das fontes Chivo e Faustina.
Fonte: GATTI; OMNIBUS-TYPE, 2019; GARCÍA; OMNIBUS-TYPE, 2017

À Á Â Ã Ä È É Ê Ë Ì Í Î Ï Ñ Ò
 Ó Ô Õ Ö Ù Ú Û Ü Ý à á â ã
 ä è é ê ë ì í î ï ñ ò ó ô õ ö ù ú
 û ü ý Ā ā Ć ć Č č Ē ē Ī ī Ī ī Ō
 ō Ũ ũ Ū ū Ħ Ħ Ğ ğ Ŧ ŧ Ÿ ŷ Ƨ Ƨ
 Ě ě Ỳ ỳ Ỹ ỹ Ƨ Ƨ Ǝ ɛ ǁ ǁ ǁ ǁ
 Ĵ ĵ Ğ ğ Ȧ ȧ Ȯ ȯ Ȧ ȧ Ǝ ɛ ǁ ǁ
 ǁ ǁ Ā ā Ě ě Ī ī Ā ā Ě ě Ī ī Ō ō
 Ũ ũ Ā ā Ě ě Ī ī Ō ō Ũ ũ Ȧ ȧ Ě
 ě Ī ī Ū ũ Ā ā Ě ě Ī ī Ō ō Ũ ũ
 Ó ó Í í Î î Ī ī Ƨ Ƨ Ƨ Ƨ Ī ī

Figura 178: Amostra de caracteres acentuados da fonte Faustina.

Fonte: GARCÍA; OMNIBUS-TYPE, 2017

À Á Â Ã Ä È É Ê Ë Ì Í Î Ï Ñ
 Ò Ó Ô Õ Ö Ù Ú Û Ü Ý à á â
 ã ä è é ê ë ì í î ï ñ ò ó ô õ ö
 ù ú û ü ý Ā ā Ĉ ĉ Ċ ċ Ē ē Ī
 ī Ī ī Ō ō Ū ū Ū ū ĩ Ğ ğ Ṭ ṭ Ÿ
 ŷ Ƨ Ƨ Ĕ ĕ ÿ ÿ Ŷ ŷ Ƨ Ƨ Ƨ
 Ĵ ĵ Ĝ ĝ Ḃ ḃ Ɔ Ɔ Ḅ ḅ Ǝ Ǝ Ĭ Ĭ
 Ɔ Ɔ Ʈ Ʈ Ā ā Ę ę Ĩ ĩ Ą ą ě ę
 Ĩ Ĩ Ō ō Ū ū Ą ą ě ę Ĩ Ĩ Ō ō Ū
 ū Ą ą Ą ą ě ę Ĩ Ĩ Ō ō Ū ū Ó
 ó Ĩ Ĩ Ĩ Ĩ Ĩ Ĩ Ƨ Ƨ Ƨ Ƨ Ĩ Ĩ Ĩ Ĩ Ĩ

Figura 179: Amostra de caracteres acentuados da fonte Chivo.

Fonte: GATTI; OMNIBUS-TYPE, 2019

ninam

Xinaroia pikinē ninam pik arama xeeyo hakeerahe, peheri e harireh rinē txaropik aramam rakenmahkii txaropik xira mah rareha txaohik pata tha hakeere iha thiwēnē ethē wari rarehakire hap naha txaro wapik xiramah pari mah rareha txarowahik pata pih tha reh hii e kitxopēha e kitxohakire kamahik kirikaowii txahik thai ei naha epata kiha keere tharik thik pata hiraraahakeere iha thēk hatha hiraranē kama peheri e pixiyap e einaha iha kip xēt̄xot re hakire iha kip araxoterarewii pehe iham pikwai roko hai waro txotekehakire, iha pat kinokai rērēt̄xot txo hakire

wanano

Phanop̄ure P̄a khamono ñoakap̄u hi siniatia. Toi hiro sã ñũhchũ Turku wamatiriro hiatia. Tiro to pahko mahk̄unore Keneine ñũ tiniatia, õp̄u Yuhka phadu w̄u'ũ sotoai hori w̄u'ũ khotoi hirirore. Ayoa to pahko mahk̄uno Kenei nia tirore “pahko mahk̄u wahp̄uti biari m̄u” nia tirore. “Kha'aka kure m̄u hichũ noaboa” nia tirore. To anichũ thũ'oro sã ñũhchũ yoha ta'a pa b̄uhse khoai Wĩhsoana phitai tiro wi'i taka. Toi wi'i taka'a, tiro to pahko mahk̄unore ñitiniatia. Ayoaro õ Menekoana ñoaka ti kh̄arich̄ari ku'tu hiatia. Ayoa sã ñihchĩ kh̄iro phã'a wi'i d̄uhka hi d̄uhkatia õ Menekoana ñoakare. Mip̄ure to phanamana hiturina hira.

arutani

pari pari yãn / oo wasiwaka isi / oo wasiwaka isi paripari asinhõ sako asinhõ sahaw asinhõ / asinhõ / kaipsa apopon, kaipsa apopon, aay kani arutani kiwa, oo, oo kohay sinho isi / ããrika ni, mata kiwa, arutani kiwa / saki, sako kaya kinimaw ã kas kontani kontanhi asinho / kasikontanki kaya mantama / ãrikani kēa arutani kiwa / oo wasiwakaisi / oo wasiwaka isi kontanki, korayki kisinho, korayki kasinho / ãrika ni / wasiwaka kanika / wasiwaka kanika oo kontansi / korayki sinho sã koray sinho sã kãã / kaya matanma, kaya matanma / kaya matampki / kakwakati wahap naki / xinho / ããrikani arutani kiwa / yama kosa waitanki kēa asinho / yama kosa waita / sinho katopin / topinki sinho

yanomami

Awei, hwei tēhē kami yamaki hiramayu kōa, hapēnaha kure thēãha yamaki hiramayoma: Puu thēãha õniha yamaki pihi xaariramayu kōa. Kama Estevã axē yamaki pree komi kiai kōa. Kami yamaki pihi riã xaaririmapē yamaki hiramai, puu thēã õni xirõha yamaki kiãi. Papeo sikihã yama thēã õnimai koõma. Puu thēã wãri yapai koõ tēhē, ñhi tēhē, yama e thēã tratusimoma, kama ani thēã wãri kuapēnaha, yanomae thēãha yama e thēã thai xoaoma. Ei naha, kami pesquisadores yamakini yama thēã thama. ñhi kama puu pē piriowi thēãha yamaki hiramayu. Kama puu pēni huu tihi pēni he yai puuwihã, yama thēã thaa xoaoma, kama puu pēni huu tihi pē nēhē yai.

Figura 180: Amostra de texto em línguas indígenas na fonte Faustina.

Fonte: GARCÍA; OMNIBUS-TYPE, 2017

ninam

Xinaroia pikinē ninam pik arama xeeyo hakeerahe, peheri e harireh rinē txaropik aramam rakenmahkii txaropik xira mah rareha txahik pata tha hakeere iha thiwēnē ethē wari rarehakire hap naha txaro wapik xiramah pari mah rareha txarowahik pata pih tha reh hii e kitxopēha e kitxohakire kamahik kiri-kaowii txahik thai ei naha epata kiha keere tharik thik pata hiraraahakeere iha thēk hatha hiraranē kama peheri e pixiyap e einaha iha kip xētxot re hakire iha kip aratxoterarewii pehe iham pikwai roko hai waro txotekehakire, iha pat kinokai rērētxot txo hakire

wanano

Phanopure Pɛ khamono ñoakapɛ hi siniatia. Toi hiro sã ñɛhchɛ Turku wamatiro hiatia. Tiro to pahko mahkɛnore Keneine ñɛ tiniatia, ɔpɛ Yuhka phadɛ wɛ'ɛ sotoai hori wɛ'ɛ khoti hirirore. Ayoa to pahko mahkɛno Kenei nia tirore “pahko mahkɛ wahpɛti biari mɛ” nia tirore. “Kha'aka kure mɛ hichɛ noaboa” nia tirore. To anichɛ thɛ'oro sã ñɛhchɛ yoha ta'a pa bɛhse khoai Wihsoana phitai tiro wi'i taka. Toi wi'i taka'a, tiro to pahko mahkɛnore ñitiniatia. Ayoaro ɔ Menekoana ñoaka ti khārichɛri ku'tu hiatia. Ayoa sã ñihchɛ khiro phā'a wi'i dɛhka hi dɛhkatia ɔ Menekoana ñoakare. Mipure to phanamana hiturina hira.

arutani

pari pari yān / oo wasiwaka isi / oo wasiwaka isi paripari asinhō sako asinhō sahaw asinhō / asinhō / kaipsa apopon, kaipsa apopon, aay kani arutani kiwa, oo, oo kohay sinho isi / āārika ni, mata kiwa, arutani kiwa / saki, sako kaya kinimaw ā kas kontani kontanhi asinho / kasikontanki kaya mantama / ārikani kēa arutani kiwa / oo wasiwakaisi / oo wasiwaka isi kontanki, korayki kisinho, korayki kasinho / ārika ni / wasiwaka kanika / wasiwaka kanika oo kontansi / korayki sinho sã koray sinho sã kãã / kaya matanma, kaya matanma / kaya matampki / kakwakati wahap naki / xinho / āārikani arutani kiwa / yama kosa waitanki kēa asinho / yama kosa waita / sinho katopin / topinki sinho

yanomami

Awei, hwei tēhē kami yamakī hiramayu kōa, hapēnaha kure thēāha yamakī hiramayoma: Puu thēāha ōniha yamakī pihi xaariramayu kōa. Kama Estevão axē yamakī pree komi kiaī kōa. Kami yamakī pihi riā xaariramapē yamakī hiramai, puu thēā ōni xirōha yamakī kiāi. Papeo sikiha yama thēā ōnimai koōma. Puu thēā wāri yapai koō tēhē, ĩhi tēhē, yama e thēā tratusimoma, kama ani thēā wāri kuapēnaha, yanomae thēāha yama e thēā thai xoama. Ei naha, kami pesquidores yamakini yama thēā thama. ĩhi kama puu pē pirīowi thēāha yamakī hiramayu. Kama puu pēni huu tihī pēni he yai puuwihī, yama thēā thaa xoama, kama puu pēni huu tihī pē nēhē yai.

Figura 181: Amostra de texto em línguas indígenas na fonte Chivo.

Fonte: GATTI; OMNIBUS-TYPE, 2019

13.4 Desenvolvimento de teclados

A última seção deste capítulo tem como objetivo descrever o desenvolvimento dos teclados. Conforme as diretrizes de desenvolvimento, foram desenvolvidos teclados específicos para o grupo 2 e grupo 3, para as plataformas MacOS e Windows.

Em todos os casos, contemplou-se a principal diretriz de desenvolvimento: respeitar os layouts de teclado já conhecidos, mantendo os diacríticos em suas teclas usuais. A exceção são algumas marcas combinadas que foram acrescentadas (*tildebelowcomb* <~>, *macronbelowcomb* <_>, *macroncomb* <¯>) e também os caracteres <i> e <u>, para as línguas do grupo 3, que não existem nos teclados ABNT e ANSI.

Para o desenvolvimento dos teclados, foram utilizados dois programas: Ukelele (para MacOS) e Microsoft Keyboard Layout Creator (para Windows). Esses dois programas só geram layouts de teclado para seu sistema operacional nativo. As diretrizes de desenvolvimento foram mantidas intactas, independentemente do sistema operacional utilizado.

Antes de explicar como funcionam os teclados, é importante explicar os esquemas de organização e nomenclatura adotados no desenvolvimento. Primeiramente, os teclados são classificados em dois grandes grupos, de acordo o layout e mapeamento das teclas: *ABNT* ou *Internacional (INT)*. Segundo, utilizou-se os nomes *Standard* para os teclados que têm como alvo o grupo 2 e *AcentosCombinados* para os teclados que têm como alvo o grupo 3. Por último, há uma nomenclatura que diferencia os teclados específicos para línguas que utilizam o caractere <i>, *iSTROKE* e, <u>, *UBAR*.

Sobre esses últimos dois casos, dos caracteres <i> e <u>, são necessárias algumas explicações. Como regra geral, esses caracteres são produzidos por meio da combinação Option + <i> (MacOS) e Ctrl + Alt + <i> (Windows), resultando no <i>, e da combinação Option + <u> (MacOS) e Ctrl + Alt + <u> (Windows), resultando no <u>. No entanto, no MacOS, há alguns conflitos de teclas. Tradicionalmente, no layout U.S. International do MacOS, a combinação Option + <i> tem como resultado o acento circunflexo <^>, e a combinação Option + <u> tem como resultado o trema <¨>.

Por essa razão, foram criados layouts separados para o padrão U.S. International (*INT*), com os sufixos *iSTROKE* e *UBAR*. Com a criação desses dois layouts separados, evita-se os conflitos de teclas com o trema e o acento circunflexo.

No caso dos layouts *INT_Acentos Combinados_iSTROKE*, o caractere <i> pode ser digitado com a combinação de Option + <i>. O circunflexo pode ser acessado pela combinação do Shift + <6>, e o trema se mantém na mesma posição, Option + <u>.

No caso dos layouts *INT_Acentos Combinados_UBAR*, o caractere <u> pode ser digitado com a combinação de Option + <u>. O trema pode ser acessado pela combinação do Shift + Alt + <6>, e o circunflexo se mantém na mesma posição, Option + <i>.

Ainda, observa-se que as línguas que utilizam o caractere <i> não utilizam trema, e as línguas que utilizam o caractere <u> não utilizam acento circunflexo.

Em todos os teclados para marcas combinadas, para MacOS e Windows, pode-se acessar algumas marcas que não são encontradas nos layouts ABNT e ANSI, conforme as instruções a seguir:

- Option + B (MacOS) ou Ctrl + Alt + B (Windows) = <~>
U+0330 COMBINING TILDE BELOW
- Option + X (MacOS) ou Ctrl + Alt + X (Windows) = <_>
U+0331 COMBINING MACRON BELOW
- Option + V (MacOS) ou Ctrl + Alt + V (Windows) = <¯>
U+0304 COMBINING MACRON

A Fig.182, a seguir, mostra o mapeamento dos caracteres em um teclado ABNT.

Todos os teclados, assim como as instruções de instalação, estão disponíveis no repositório digital¹³ do projeto.

13. <https://github.com/RafaelDietzsch/LinguasIndigenasBrasileiras>

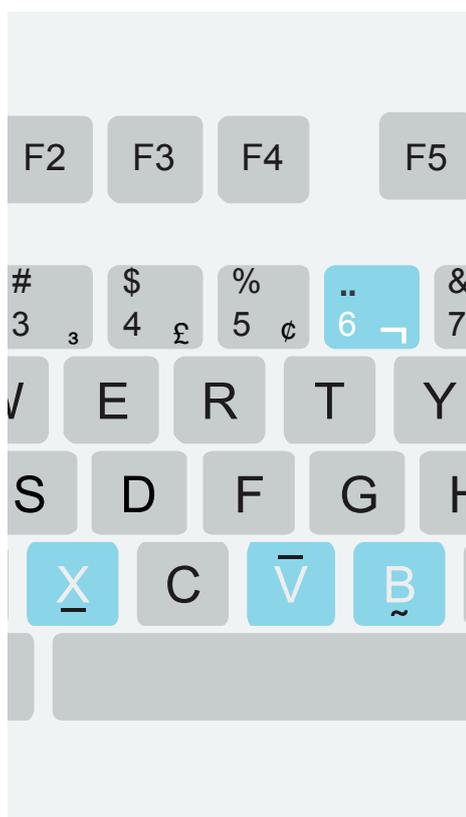


Figura 182: Esquema de mapeamento do layout de teclado ABNT modificado. Fonte: TECLADO, 2021

14. Conclusão

Esta tese é uma tentativa de contribuição para a promoção e fortalecimento dos usos sociais das línguas indígenas brasileiras nos meios de comunicação e para a comunicação digital dos povos indígenas brasileiros. Nesse sentido, o trabalho procura estabelecer mecanismos para que a tipografia se torne uma ferramenta de comunicação indispensável para os falantes de línguas indígenas brasileiras em meios digitais. Para isso, apresenta métodos e técnicas que demonstram como a tipografia pode beneficiar a manutenção e representação gráfica correta dessas línguas originalmente ágrafas. Em resumo, a ideia desta pesquisa é construir conhecimento e ferramentas para viabilizar a comunicação textual em meios digitais no cotidiano dos povos indígenas em seus idiomas originários.

Ao descrever o estado da arte das línguas indígenas brasileiras e propor um método de desenvolvimento de ferramentas de escrita para essas línguas, este trabalho preenche algumas lacunas no campo da comunicação visual, da linguística e da tecnologia da informação.

Um dos principais objetivos propostos para a pesquisa foi a realização de um levantamento de dados abrangente para identificação do maior número possível de ocorrências de diacríticos. Para isso, desenvolveu-se uma metodologia que teve como principal resultado a catalogação ortográfica, fundamental para a especificação de uma lista de requisitos para o suporte linguístico de fontes e teclados para línguas indígenas brasileiras. Além disso, a metodologia descreve critérios para classificação dos dados coletados, fundamentais para análise e classificação das línguas pesquisadas.

Ainda, esta tese demonstra como aplicar a metodologia na customização de fontes e teclados para línguas indígenas brasileiras e descreve o desenvolvimento de projeto

dessas ferramentas. Além disso, explica como funciona a catalogação ortográfica – que reúne, ao todo, 239 tabelas, com informações sobre 111 línguas – e como utilizá-la.

Como resultado prático, foram produzidas versões de duas famílias de fontes *open-source*: a serifada Faustina e a sem serifa Chivo. A partir das diretrizes de desenvolvimento de projeto e da especificação do conjunto de caracteres para línguas indígenas brasileiras, as versões romanas das duas famílias foram customizadas em todos os pesos (de *light* a *black*), com todos os caracteres e recursos OpenType necessários para composição de texto nas 111 línguas indígenas catalogadas no projeto. Um segundo resultado prático é o desenvolvimento de layouts de teclados para as plataformas MacOS e Windows específicos para línguas indígenas brasileiras. Todas as fontes e teclados, assim como suas respectivas instruções de instalação, estão disponíveis no repositório digital do projeto: <https://github.com/RafaelDietzsch/LinguasIndigenasBrasileiras>.

Dessa forma, acredita-se que o ato de disponibilizar ferramentas apropriadas, capazes de amplificar os usos e a prática da escrita, representa uma medida pragmática para facilitar a comunicação, podendo levar a um impacto muito grande para essas comunidades, já que a atividade coletiva e cotidiana da escrita é algo indispensável para o seu próprio desenvolvimento. Por essa razão, enfatiza-se aqui o caráter da tipografia como ferramenta.

Além disso, deve-se ressaltar a contribuição teórico-metodológica deste trabalho, que descreve um tipo de abordagem ainda pouco utilizada na pesquisa em comunicação digital e tipografia relacionada a línguas indígenas. Inspirada no trabalho de John Hudson, essa abordagem descreve métodos de levantamento, tratamento e manipulação da informação para o desenvolvimento de ferramentas de escrita, apontando para a utilização dessas informações em diferentes contextos: e.g. comunicação visual, tipografia, tecnologia da informação e linguística. No entanto, para que essas informações representem de fato um lastro para o campo teórico-metodológico e para o desenvolvimento de ferramentas, é preciso estabelecer uma relação entre essas informações e a codificação digital, uma vez que o suporte linguístico de muitas ferramentas é pensado a partir de padrões de codificações existentes. Reforçando: não basta apenas realizar um enorme levantamento de dados; é preciso contextualizar esse levantamento.

A partir desse lastro, acredita-se que é possível pensar de fato na inserção das línguas indígenas brasileiras na comunicação textual cotidiana de cidadãos que não têm o português como língua materna. Além disso, com a disponibilidade dessas ferramentas, abrem-se possibilidades efetivas para a inserção e promoção de línguas indígenas brasileiras em produtos de comunicação, como revistas, jornais, publicações, redes sociais etc.

Esta tese, apesar de cobrir muitos assuntos, deixa ainda algumas lacunas. Especialmente por se tratar de um projeto de pesquisa em desenvolvimento.

A primeira dessas lacunas refere-se ao fato de que, apesar dos esforços para que o levantamento de dados pudesse cobrir a maior amostra possível, as línguas documentadas nesta pesquisa não representam nem a metade do total das línguas indígenas listadas no recenseamento do IBGE de 2010. Provavelmente há informações sobre muitas dessas línguas não documentadas na pesquisa, especialmente nos acervos físicos em que não se pôde pesquisar por conta da pandemia do COVID-19. Por essa razão, há ainda problemas tipográficos e de codificação ainda não conhecidos.

Em segundo lugar, é importante reforçar que esta pesquisa não se encerra com esta tese. Nesse sentido, há planos (iniciados no ano de 2019) de se formalizar um Acordo de Cooperação Técnica entre a UnB e o IPHAN, para que se possa estreitar a colaboração com o INDL. Novamente, em razão dos problemas causados pelo COVID-19 (nesse caso, questões logísticas e administrativas), esses planos acabaram sendo adiados para depois do término desta tese.

Terceiro, é preciso reconhecer que há uma parte da documentação das ferramentas que ainda precisa ser desenvolvida. Em especial, materiais de suporte ao usuário. Um dos planos era que esses materiais fossem construídos no contexto do Acordo de Cooperação Técnica com o IPHAN e, por isso, ficaram em segundo plano.

Esta tese, no entanto, abre caminhos para pesquisas relevantes acerca do tema. A partir da catalogação ortográfica e do ferramental existente, é possível pensar em expandir o suporte linguístico para os teclados de dispositivos móveis, como telefones celulares e tablets. Nesse sentido, há um interessante texto de Fiona Ross (2019) sobre o impacto dos dispositivos móveis para o design de tipos. É importante aqui um parêntese, destacando a longa atuação da professora e pesquisadora: com formação em

linguística, participou ativamente do desenvolvimento de dezenas de fontes para árabe e escritas indianas. Além disso, inventou, em 1978, junto a Mike Fellows, o teclado fonético para escritas indianas (ROSS, 2012).

Outra possibilidade está ligada ao fato de ainda haver toda uma discussão em aberto sobre as línguas indígenas no campo da comunicação. Este trabalho servirá como base para a estruturação de outras pesquisas e também de produtos como revistas, jornais, sites etc.

Um dos assuntos que ainda precisa ser mais bem estudado é relativo ao uso do apóstrofo, muito utilizado para marcar a oclusão glotal. Há algumas referências de designers de tipos (particularmente, do México, Argentina e Paraguai) que discutiram mais a fundo essa questão. O fato é que utiliza-se hoje diferentes sinais para a representação desse caractere: apóstrofo <'>, aspas simples <'>, aspas retas <'>, além do saltillo <'> e de sua versão maiúscula <'>, que é o sinal recomendado pelos colegas hispanófonos. Diferentemente do apóstrofo e das aspas, o saltillo não é codificado como sinal de pontuação. Além disso, tem valor fonético, e não suprasegmental. Na palavra *avañe'ẽ*, em guarani, pode-se ver o uso do saltillo, que tem um espaçamento diferente dos sinais supracitados e, em geral, é maior e mais profundo. Apesar de se ter incluído esse caractere em todas as fontes desenvolvidas, também não se chegou uma conclusão de como tratar esse caractere nos teclados.

Por último, é importante mencionar que o professor e pesquisador Eduardo Gomes, do Departamento de Estatística da UnB, aponta para a possibilidade de um estudo estatístico a partir dos dados da catalogação ortográfica, observando padrões recorrentes em certos grupos de caracteres e blocos do Unicode e de sua relação com certas línguas.

Além das possibilidades de pesquisa acima mencionadas, há ainda alguns interesses, por parte deste autor. Por se tratar de um trabalho em desenvolvimento, é importante ter em vista a possibilidade de colaboração com outros desenvolvedores e instituições.

Primeiramente, destaca-se que há interesse, por parte dos desenvolvedores do Glyphs, em prover suporte para línguas indígenas da América do Sul. Em abril de 2020, por intermédio da designer argentina Sol Matas (que reside em Berlim e tem maior proximidade com os desenvolvedores do Glyphs, Rainer Erich Scheichelbauer e Georg

Seifert), Pablo Cosgaya (2020) iniciou uma documentação para essas implementações ao programa, da qual o autor desta tese participou e tem interesse. Participaram também dessa discussão as argentinas Marcela Romero e Carolina Giovagnoli, o mexicano José Manuel López Rocha e o venezuelano Yorlmar Campos. Em resumo, a primeira listagem a seguir contempla os idiomas com o maior número de falantes, com universo de cobertura de mais de 18.400.000 falantes:

- Quechua [que]: > 7.700.000 falantes
- Guaraní [grn]: > 6.000.000 falantes
- Aymara [aym]: > 1.600.000 falantes
- Náhuatl [nhn]: > 1.700.000 falantes
- Maya [yua]: > 800.000 falantes
- Mapudungun [arn]: > 250.000 falantes

Em segundo lugar, é importante destacar interesses de pesquisa e desenvolvimento na área do design de tipos. Pretende-se continuar o trabalho iniciado junto ao colega Kalapi Gajjar-Bordawekar no desenvolvimento de novas fontes e também na customização de fontes open-source existentes. Nessa linha, Pablo Cosgaya também sinalizou a possibilidade de participação em futuros desenvolvimentos em parceria com sua equipe.

Terceiro, destaca-se o interesse de colaboração e contribuição com outras ferramentas e bases de dados, como Hyperglot, Charset Builder e CLDR (Unicode). A integração com bancos de dados linguísticos é de suma importância para o avanço da comunicação digital em línguas indígenas brasileiras.

Por último, mas não menos importante, ressalta-se a necessidade de uma interface visual para o banco de dados. Isso não somente facilitará novas contribuições de terceiros, como também tornará mais clara a visualização das contribuições existentes.

Finalmente, é importante observar a relevância deste trabalho e destacar que é muito pouco provável que um projeto com este tivesse sido desenvolvido em outro contexto, diferente deste: i.e. em uma universidade pública, em um programa de pós-graduação. Esse tipo de desenvolvimento dificilmente despertaria interesse financeiro para empresas de desenvolvimento de software. Por essa razão, reforça-se mais uma vez a importância do desenvolvimento contínuo da pesquisa em tipografia e línguas indígenas brasileiras.

Referências bibliográficas

Bibliografia citada

- ACADEMIA BRASILEIRA DE LETRAS. **Vocabulário Ortográfico da Língua Portuguesa**. 5. ed. [s.l.] Global, 2010.
- ADOBE INC. **OpenType User Guide for Adobe Fonts**. Adobe Inc., 2008. Disponível em: <<https://blog.typekit.com/wp-content/uploads/2008/06/otguide.pdf>>. Acesso em: 21 jun. 2021
- ANCHIETA, J. DE. **Arte de grammatica da lingoa mais vsada na costa do Brasil**. Coimbra: Antonio de Mariz, 1595.
- ANDERSON, D. et al. **Unicode Technical Note #19: Recommendations for creating new orthographies** Unicode, 5 jan. 2005. Disponível em: <<http://www.unicode.org/notes/tn19/>>. Acesso em: 20 jan. 2019
- AUSTIN, P. K.; SALLABANK, J. (EDS.). **The Cambridge Handbook of Endangered Languages**. Cambridge: Cambridge University Press, 2011.
- BARROS, M. C. D. M. A missão Summer Institute of Linguistics e o indigenismo latino-americano: história de uma aliança (décadas de 1930 a 1970). **Revista de Antropologia**, v. 47, n. 1, p. 45–85, 2004.
- BATISTA, R. DE O. Descrição de línguas indígenas em gramáticas missionárias do Brasil colonial. **DELTA: Documentação de Estudos em Lingüística Teórica e Aplicada**, v. 21, n. 1, p. 121–147, jun. 2005.
- BERRY, J. D. (ED.). **Language culture type: international type design in the age of Unicode**. Redhill, UK; New York, NY: ATypI : Graphis, 2002.
- BINNQUÏST, R. C. L. **The endurance of Mexican Amate paper: exploring additional dimensions to the sustainable development concept**. PhD thesis – Enschede, NL: University of Twente, 2003.
- BLAŽEK, F. Accents. **Typo Magazine**, n. 10, Agosto 2004.
- BLAŽEK, F. **Diacritics Project**. Disponível em: <<http://diacritics.typo.cz/index.php?id=12>>. Acesso em: 20 jan. 2019.
- BLOKLAND, F. E. **On the origin of patterning in movable Latin type : Renaissance standardisation, systematisation, and unitisation of textura and roman type**. Leiden: Leiden University, 11 out. 2016.

BLOKLAND, F. E. **On the Origin of Patterning in Movable Latin Type**. Disponível em: <<https://www.lettermodel.org/>>. Acesso em: 28 jun. 2021.

BODONI, G. **Manuale tipografico**. Parma: Margherita Dall’Aglia (viúva de Bodoni), 1818.

BŘEZINA, D. **On diacritics**. Disponível em: <<https://ilovetypography.com/2009/01/24/on-diacritics/>>. Acesso em: 20 jan. 2019.

BŘEZINA, D. **Coherence in typeface design: visual similarity of characters in Cyrillic, Devanagari, and Latin**. Tese de doutorado – Reading, UK: The University of Reading, set. 2018.

BŘEZINA, D. **Skolar (type specimen)** Rosetta Type Foundry, 2019. Disponível em: <https://www.rossettatype.com/microsite/current/Skolar/files/Skolar_specimen.pdf>. Acesso em: 23 jun. 2021

BŘEZINA, D. **Hyperglot: a database and tools for detecting language support in fonts**. Disponível em: <<https://hyperglot.rossettatype.com/>>. Acesso em: 19 jul. 2021.

BRINGHURST, R. Voices, languages and scripts around the world. In: BERRY, J. D. (Ed.). **Language culture type: international type design in the age of Unicode**. Redhill, UK; New York, NY: ATypI : Graphis, 2002. p. 3–23.

BRINGHURST, R. **Elementos do estilo tipografico: versão 3.0**. Tradução: André Stolarski. São Paulo: Cosac Naify, 2005.

BUGGY, L. A. DA C. **O MECOTipo: Método de Ensino de Desenho Coletivo de Caracteres Tipográficos**. 2. ed. Brasília: Serifa Fina / Estereográfica, 2018.

BURIAN, V. **Understanding diacritics | TypeTogether**. Disponível em: <<https://www.type-together.com/understanding-diacritics>>. Acesso em: 30 jul. 2021.

BURKE, P. **Linguagens e comunidades nos primórdios da Europa moderna**. São Paulo: Editora Unesp, 2010.

BURNHILL, P. **Type spaces: in-house norms in the typography of Aldus Manutius**. London: Hyphen Press, 2003.

CABALZAR, F. D. (ED.). **Kiti wederira tuhoarira**. [s.l.] ISA, 2002.

CABALZAR, A. et al. (EDS.). **Bureko watotire, wametire 1**. São Gabriel da Cachoeira; São Paulo: ISA / FOIRN / AEITY, 2009.

CAHILL, M. **Non-Linguistic Factors in Orthographies**. Linguistic Society of America Annual Meeting, Pittsburgh, Jan. 6-9, 2011. Symposium on Developing Orthographies for Unwritten Languages. **Anais...** In: LINGUISTIC SOCIETY OF AMERICA ANNUAL MEETING. Pittsburgh: jan. 2011.

CAHILL, M.; KARAN, E. Factors in designing effective orthographies for unwritten languages. **SIL Electronic Working Papers**, SIL Electronic Working Papers. n. 2008– 001, 2008.

Caractere. Michaelis On-Line. [s.l.] Melhoramentos, 2015. Disponível em: <<https://michaelis.uol.com.br/moderno-portugues/busca/portugues-brasileiro/caractere/>>. Acesso em: 14 jul. 2020

Character. Glossary of Unicode Terms. [s.l.: s.n.]. Disponível em: <<https://www.unicode.org/glossary/#character>>. Acesso em: 14 jul. 2020

CHENG, K. **Designing Type**. New Haven: Yale University Press, 2006.

Codificação. Michaelis On-Line. [s.l.] Melhoramentos, 2015. Disponível em: <<https://michaelis.uol.com.br/moderno-portugues/busca/portugues-brasileiro/codifica%C3%A7%C3%A3o/>>. Acesso em: 19 jun. 2020

Codificar. Michaelis On-Line. [s.l.] Melhoramentos, 2015. Disponível em: <<https://michaelis.uol.com.br/moderno-portugues/busca/portugues-brasileiro/codificar/>>. Acesso em: 19 jun. 2020

Código. Michaelis On-Line. [s.l.] Melhoramentos, 2015. Disponível em: <<https://michaelis.uol.com.br/moderno-portugues/busca/portugues-brasileiro/c%C3%B3digo/>>. Acesso em: 17 jun. 2020

COELHO NETTO, J. T. **Semiótica, informação e comunicação: diagrama da teoria do signo**. São Paulo, Brasil: Editora Perspectiva, 1980.

CONSTABLE, P. **Character set encoding basics**. Disponível em: <https://scripts.sil.org/cms/scripts/page.php?item_id=IWS-Chapter03>. Acesso em: 11 jun. 2020.

COSGAYA, P. Reseña de la tipografía argentina. **tipoGráfica**, n. 56, p. 20–27, 2003.

COSGAYA, P. **Lenguas de la región**, abr. 2020. Disponível em: <<https://docs.google.com/document/d/17Xb7dir8obhFcaSfaUKOB2vcl8Jqxffmil9-AmvtJw/edit>>. Acesso em: 2 ago. 2021

COULMAS, F. **The Writing Systems of the World**. Reprint edition ed. Oxford, UK Cambridge, Massachusetts, USA: Wiley-Blackwell, 1991.

COULMAS, F. **The Blackwell Encyclopedia of Writing Systems**. Malden, Mass.: Wiley-Blackwell, 1999.

COULMAS, F. **Writing systems: an introduction to their linguistic analysis**. Cambridge, U.K.; New York, NY: Cambridge University Press, 2003.

CRYSTAL, D. Toward a typographical linguistics. **Type: A Journal of the ATypI**, n. 2, p. 7–23, 1998.

CUNHA, R. B. Políticas de línguas e educação escolar indígena no Brasil. **Educar em Revista**, n. 32, p. 143–159, 2008.

DE LIMA, T. B. (ED.). **Amandaje nº 169 Ava Ka’aguygua rehegua Tetã Sãsoyguakuéraguáva ha Ñemobohechapyrã ojejapo va’erã pe OITre ñemoñe’epyre Ava ka’aguygua rehe. Tradução da Convenção nº 169 sobre Povos Indígenas e Tribais e Resolução referente à ação da OIT para a língua indígena brasileira Guarani Kaiowá**. Brasília: OIT, 2012.

DESCHANEL, A. P. Electric Telegraphs. **The Popular Science Monthly**, v. 3, p. 401–418, 1873.

DIETZSCH, R. **Typefaces for Brazilian indigenous languages**. Dissertation (Master of Arts in Typeface Design) – Reading, UK: University of Reading, set. 2012.

DINIZ, K. C. Notas sobre tipografias para línguas indígenas do Brasil. **InfoDesign - Revista Brasileira de Design da Informação**, v. 4, n. 1, p. 36–46, 2007.

DRUDE, S.; AWETI, W.; AWETI, A. [awe]_002_A Ortografia da Língua Awetí. v. 19, n. e019014, p. 1–23, 2019.

DUNCAN, B. C.; GARVIN, D. **NBS TECHNICAL NOTE 820: Complete clear text representation of scientific documents in machine-readable form**. Washington, D.C.: National Bureau of Standards, 1974.

- EBERHARD, D. M.; SIMONS, G. F.; FENNIG, C. D. (EDS.). **Ethnologue: Languages of the World. Twenty-fourth edition**. 24. ed. Dallas: SIL International, 2021.
- EIGENDORF, A.; ACSANY, P. **CharSet Builder**. Disponível em: <<https://www.alphabet-type.com/tools/charset-builder/>>. Acesso em: 29 jun. 2021.
- EISENSTEIN, E. L. **A Revolução da Cultura Impressa. Os Primórdios da Europa Moderna**. São Paulo: Ática, 1998.
- Escrita. Michaelis On-Line**. [s.l.] Melhoramentos, 2015. Disponível em: <<https://michaelis.uol.com.br/moderno-portugues/busca/portugues-brasileiro/escrita/>>. Acesso em: 25 jun. 2020
- ESFAHBOD, B. **What is text shaping?** Disponível em: <<https://harfbuzz.github.io/what-is-harfbuzz.html#what-is-text-shaping>>. Acesso em: 22 jun. 2021.
- FARIAS, P. L. **Notas para uma normatização da nomenclatura tipográfica**. Anais P&D Design 2004 - 6º Congresso Brasileiro de Pesquisa e Desenvolvimento em Design. **Anais...** In: P&D DESIGN 2004 - 6º CONGRESSO BRASILEIRO DE PESQUISA E DESENVOLVIMENTO EM DESIGN. São Paulo: FAAP, 2004.
- FEBVRE, L.; MARTIN, H.-J. **O Aparecimento Do Livro**. São Paulo: Edusp, 2017.
- FERREIRA, H. P. **Portal Japiim | ProDoclin**. Disponível em: <<https://japiim.linguasyanomami.com>>. Acesso em: 25 jul. 2021.
- FERREIRA, H. P. **Projeto 914BRZ4019 Contrato SA-3993/2019**. Brasília: Museu do Índio - FUNAI, 2021.
- FIGUEIRA, L. **Arte da lingua brasilica**. Lisboa: Manuel da Silva, 1621.
- FIGUIER, L. **Les merveilles de la science, ou description populaire des inventions humaines.** Paris: Furne, Jouvet et Cie., Editeurs, 1867.
- FISCHER, E. **The Evolution of Character Codes, 1874-1968**, 20 jun. 2000. Disponível em: <<https://archive.org/details/enf-ascii>>
- FLUSSER, V. **O mundo codificado: por uma filosofia do design e da comunicação**. São Paulo: Cosac Naify, 2007.
- FONTES, A.; DINIZ, L. L.; TRUJILLO, T. P. (EDS.). **Ikadzekatakadapha**. São Paulo: ISA, 2005.
- FOURNIER, P. S. **Manuel typographique, utile aux gens de lettres, & à ceux qui exercent les différentes parties de l'art de l'imprimerie**. Paris: Fournier, 1764.
- FRANCHETTO, B. A guerra dos alfabetos: os povos indígenas na fronteira entre o oral e o escrito. **Mana**, v. 14, n. 1, p. 31–59, abr. 2008.
- FRUTIGER, A. **Sinais e Símbolos: desenho, projeto e significado**. Sao Paulo, SP: Martins Fontes, 1999.
- GARCÍA, A.; OMNIBUS-TYPE. Faustina (fonte tipográfica). Argentina: Omnibus-Type, 2017.
- GARCIA, M. V. C. et al. (EDS.). **Guia de pesquisa e documentação para o INDL: patrimônio cultural e diversidade linguística**. Brasília: IPHAN, 2016.

- GATTI, H.; OMNIBUS-TYPE. Chivo (fonte tipográfica). Argentina: Omnibus-Type, 2019.
- GAULTNEY, V. J. **Problems of diacritic design for Latin script text faces**. Dissertation (Master of Arts in Typeface Design) – Reading, UK: University of Reading, 2002.
- GAULTNEY, V. J. **Multitudinous Alphabets: The challenges of extended Latin typeface design**. High Wycombe, UK, 2006.
- GIOVAGNOLI, C. **Andada Ht: tipografia para texto latinoamericanos**, 2011.
- GITTINGS, J. Chinese gloss on the king’s Korean outrages patriots. **The Guardian**, 2 dez. 1999.
- Glossary of Unicode Terms**. Disponível em: <<https://www.unicode.org/glossary/>>. Acesso em: 25 jun. 2020.
- Glyph. Glossary of Unicode Terms**. [s.l: s.n.]. Disponível em: <<https://www.unicode.org/glossary/#glyph>>. Acesso em: 14 jul. 2020
- GOUSSIER. **Planche de l’Encyclopédie de Diderot et d’Alembert. Pl. 1. / Imprimerie en Lettres, L’Opération de la casse | Paris Musées**, c 1700. Disponível em: <<https://www.parismuseescollections.paris.fr/fr/musee-carnavalet/oeuvres/planche-de-l-encyclopedie-de-diderot-et-d-alembert-pl-1-imprimerie-en>>. Acesso em: 15 jun. 2021
- Grapheme. Glossary of Unicode Terms**. [s.l: s.n.]. Disponível em: <<https://www.unicode.org/glossary/#grapheme>>. Acesso em: 22 jul. 2020
- GRAVIER, M. G. **Historia de la tipografía colonial para lenguas indígenas**. Tipos tópicos textos y contextos: ponências. **Anais...** In: PRIMER CONGRESO DE TIPOGRAFÍA. Valencia: Asociación de Diseñadores de la Comunidad Valenciana, 2004.
- GRAVIER, M. G. Tipografia e identidade lingüística. In: CONSOLO, C. (Ed.). **Tipografia en Latinoamérica**. São Paulo: Blucher, 2013a.
- GRAVIER, M. G. Letras para una lengua indomable. Tipografía y edición colonial en otomí. **Estudios de Historia Novohispana**, v. 49, p. 119–168, jul. 2013b.
- GRENOBLE, L. A.; WHALEY, L. J. **Saving Languages: An Introduction to Language Revitalization**. Cambridge: Cambridge University Press, 2006.
- GROEGER, L. **Berthold Datype disc**, 2013. Disponível em: <https://commons.wikimedia.org/wiki/File:Berthold_Datype_disc_Futura_-_PencilToPixel_exhibition.jpg>
- HARALAMBOUS, Y. **Fonts & Encodings: From Advanced Typography to Unicode and Everything in Between**. Tradução: P. Scott Horne. Beijing ; Sebastopol, Calif: O’Reilly Media, 2007.
- HEIDELBACH, W. **Metal movable type**, 2004. Disponível em: <https://commons.wikimedia.org/wiki/File:Metal_movable_type.jpg>. Acesso em: 15 jun. 2021
- HENESTROSA, C.; MESEGUER, L.; SCAGLIONE, J. **Como criar tipos: do esboço à tela**. 2. ed. Brasília: Estereográfica, 2019.
- HIGHSMITH, C. **Entre parágrafos: fundamentos tipográficos**. Tradução: Bruna Moreira. Brasília: Estereográfica, 2017.
- HOLZMANN, G. J.; PEHRSON, B. The First Data Networks. **Scientific American**, v. 270, n. 1, p. 124–129, jan. 1994.

HUDSON, J. **Sylfaen: Foundations of multiscript typography**Tiro Typeworks, 2000. Disponível em: <http://www.tiro.com/Articles/sylfaen_article.pdf>. Acesso em: 25 ago. 2020

HUDSON, J. Unicode, from text to type. In: BERRY, J. D. (Ed.). **Language culture type: international type design in the age of Unicode**. Redhill, UK; New York, NY: ATypI : Graphis, 2002. p. 24–44.

HUDSON, J. **The Brill Types**, 16 maio 2014. Disponível em: <<http://www.tiro.com/John/Hudson-Brill-DECK.pdf>>. Acesso em: 25 ago. 2020

IBGE. **O Brasil Indígena**Instituto Brasileiro de Geografia e Estatística, 2010a. Disponível em: <<https://indigenas.ibge.gov.br/estudos-especiais-3/o-brasil-indigena/download>>. Acesso em: 19 jan. 2019

IBGE. **População Indígena total**, 2010b.

INTERNATIONAL ORGANIZATION FOR STANDARDIZATION. **ISO 639-3:2007: Codes for the representation of names of languages – Part 3: Alpha-3 code for comprehensive coverage of languages**. Disponível em: <<https://www.iso.org/cms/render/live/en/sites/isoorg/contents/data/standard/03/95/39534.html>>. Acesso em: 20 jul. 2021.

INTERNATIONAL PAPER COMPANY. **Pocket pal: a graphic arts production handbook**. New York: International Paper Co., 1983.

INTERNATIONAL PHONETIC ASSOCIATION. **Handbook of the International Phonetic Association: A Guide to the Use of the International Phonetic Alphabet**. Cambridge: Cambridge University Press, 1999.

IPHAN. **Inventário Nacional da Diversidade Linguística (INDL)**. Disponível em: <<http://portal.iphan.gov.br/pagina/detalhes/140>>. Acesso em: 26 jul. 2021.

ISA. **Povos Indígenas no Brasil: Quem são?** Disponível em: <https://pib.socioambiental.org/pt/Quem_s%C3%A3o>. Acesso em: 20 jan. 2019.

JACKSON, J. **How Alan Turing set the rules for computing**. Disponível em: <<https://www.computerworld.com/article/2504774/how-alan-turing-set-the-rules-for-computing.html>>. Acesso em: 18 set. 2020.

JACQUERYE, D. **Comments on cedilla and comma below (revision 2)**Unicode Consortium, 30 jul. 2013. Disponível em: <<http://www.unicode.org/L2/L2013/13155r-cedilla-comma.pdf>>

JENNINGS, T. **Annotated history of character codes**. Disponível em: <<https://web.archive.org/web/20121103151213/http://www.wps.com:80/projects/codes/index.html#NOTE2>>. Acesso em: 15 set. 2020.

JERNUDD, B.; NEKVAPIL, J. History of the field: a sketch. In: SPOLSKY, B. (Ed.). **The Cambridge Handbook of Language Policy**. Cambridge Handbooks in Language and Linguistics. Cambridge: Cambridge University Press, 2012. p. 16–36.

JÓB, O. **Context of Diacritics | A research project**. Disponível em: <<https://www.setuatype.com:443/x/cod/>>. Acesso em: 30 jul. 2021.

KAPLAN, M. S. **The history of messing up Romanian on computers**. Disponível em: <<http://archives.miloush.net/michkap/archive/2011/08/24/10199324.html>>. Acesso em: 30 jul. 2021.

KAROW, DR. P. **Digital Typography & Artificial Intelligence**. Amsterdam: Adobe / Dutch Type Library, 2013.

Keyboard layout. Disponível em: <https://en.wikipedia.org/w/index.php?title=Keyboard_layout&oldid=1034476113>. Acesso em: 24 jul. 2021.

KINROSS, R. **Modern typography: an essay in critical history**. 2. ed. London: Hyphen Press, 2010.

KORPELA, J. K. **Unicode Explained: Internationalize Documents, Programs, and Web Sites**. 1 edition ed. Sebastopol, CA: O'Reilly Media, 2006.

KURY, L. **Viagens CientíficasBNDigital**, [s.d.]. Disponível em: <<http://bndigital.bn.gov.br/dossies/rede-da-memoria-virtual-brasileira/ciencias/viagens-cientificas/>>. Acesso em: 20 jan. 2019

LEGROS, L. A.; GRANT, J. C. **Typographical printing-surfaces: the technology and mechanism of their production**. London, New York [etc.]: Longmans, Green and Co., 1916.

LEHNI, J. **Typeface As Programme**. Disponível em: <https://www.typotheque.com/articles/typeface_as_programme>. Acesso em: 31 jul. 2021.

LEONIDAS, G. Foreword. In: **Theory of type design**. Rotterdam: nai010 publishers, 2018. p. 7–8.

LÉRY, J. DE. **Histoire d'un voyage fait en la terre du Bresil, autrement dite Amerique**. La Rochelle: Antoine Chuppin, 1578.

Letra. Michaelis On-Line. [s.l.] Melhoramentos, 2015. Disponível em: <<https://michaelis.uol.com.br/moderno-portugues/busca/portugues-brasileiro/letra/>>. Acesso em: 14 jul. 2020

Letter. Glossary of Unicode Terms. [s.l.: s.n.]. Disponível em: <<https://www.unicode.org/glossary/#letter>>. Acesso em: 2 jul. 2020

Língua. Michaelis On-Line. [s.l.] Melhoramentos, 2015. Disponível em: <<https://michaelis.uol.com.br/moderno-portugues/busca/portugues-brasileiro/1%C3%ADngua/>>. Acesso em: 17 jun. 2020

LÓPEZ ROCHA, J. M. **Type design for american native languages: four case studies: ATypI 2017 Montréal**. Montreal, 16 set. 2017. Disponível em: <<https://www.youtube.com/watch?v=IDFKz95E9Cs>>. Acesso em: 25 ago. 2020

LUNA, P. **Typography: A Very Short Introduction**. Oxford: Oxford University Press, 2018.

LÜPKE, F. Orthography development. In: AUSTIN, P. K.; SALLABANK, J. (Eds.). **The Cambridge handbook of endangered languages**. Cambridge: Cambridge University Press, 2011. p. 312–336.

MAIA, J. **Towards A Typographical Linguistics: The Semantics-Pragmatics of Typographic Emphasis in Discourse**. Tese de doutorado – Columbia: University of South Carolina, 2018.

MAMIANI DELLA ROVERE, L. V. **Arte de grammatica da lingua brasilica da naçam Kiriri**. Lisboa: Officina de Miguel Deslandes, 1699.

MAY, A. **Moxon's mould**, 2018. Disponível em: <<https://makerpress.co.uk/moxons-mould/>>. Acesso em: 15 jun. 2021

MEGGS, P. B.; PURVIS, A. W. **História do Design Gráfico**. São Paulo: Cosac & Naify, 2009.

MELETIS, D. **The Nature of Writing: A Theory of Grapholinguistics**. Brest: Fluxus Editions, 2020.

MICROSOFT CORPORATION. **Archive of OpenType versions - Typography**. Disponível em: <<https://docs.microsoft.com/en-us/typography/opentype/opentypeversions>>. Acesso em: 18 jun. 2021a.

MICROSOFT CORPORATION. **Arial font family**. Disponível em: <<https://docs.microsoft.com/en-us/typography/font-list/arial>>. Acesso em: 3 jul. 2021b.

MICROSOFT CORPORATION. **Verdana font family**. Disponível em: <<https://docs.microsoft.com/en-us/typography/font-list/verdana>>. Acesso em: 3 jul. 2021c.

MICROSOFT CORPORATION; ADOBE INC. **OpenType specification version 1.8.4**. Disponível em: <<https://docs.microsoft.com/en-us/typography/opentype/spec/>>. Acesso em: 8 fev. 2021.

MORIN, E. **O método: 1. a natureza da natureza**. 2. ed. Mem Martins: Europa-America, 1977.

MOSLEY, J. French academicians and modern typography: designing new types in the 1690s. **Typography papers**, n. 2, p. 5–29, 1997.

MOXON, J. **Mechanic exercises: or, The doctrine of handy-works, applied to the art of printing**. London: Printed for Joseph Moxon on the Westside of Fleet-ditch, at the Sign of the Atlas, 1683.

MURATOV, V. **Linotype machine**, 2012. Disponível em: <https://commons.wikimedia.org/wiki/File:Linotype_machine.jpg>. Acesso em: 15 jun. 2021

NELSON, S. Mould Making, Matrix Fitting, and Hand Casting. **Visible Language**, v. XIX, n. 1, p. 106–120, 1985.

NEWMAN, E. “De letra de índios” – cultura escrita e memória indígena nas reduções guaranis do Paraguai. **Varia Historia**, v. 25, n. 41, p. 177–196, jun. 2009.

NIMUENDAJÚ, C. U. **Mapa Etno-histórico do Brasil e Regiões Adjacentes**, 2017.

NOORDZIJ, G. **Letterletter: an inconsistent collection of tentative theories that do not claim any other authority than that of common sense**. Point Roberts, WA; Vancouver, BC: Hartley & Marks, 2000.

NOORDZIJ, G. **The stroke: theory of writing**. Tradução: Peter Enneson. London: Hyphen Press, 2005.

NOORDZIJ, G. **O traço: teoria da escrita**. Tradução: Luciano Cardinali. São Paulo: Blucher, 2013.

O que é Unicode? Disponível em: <<https://www.unicode.org/standard/translations/portuguese.html>>. Acesso em: 8 jul. 2020.

OLIVEIRA, G. M. DE. O que quer a lingüística e o que se quer da lingüística - a delicada questão da assessoria lingüística no movimento indígena. **Cadernos CEDES**, v. 19, n. 49, p. 26–38, dez. 1999.

OLIVEIRA, R. F. DE. **Tape with baudot code [fotografia]**, 2010. Disponível em: <https://commons.wikimedia.org/wiki/File:Baudot_Tape.JPG>. Acesso em: 7 jan. 2021

ORR, R. **Slug**, 15 set. 2010. Disponível em: <<https://flic.kr/p/8BMB8w>>

ORTH, A. **Željuša road sign**, 24 fev. 2016. Disponível em: <<https://englishbulgaria.net/2016/02/beautiful-ligatures-in-serbian-cyrillic/>>. Acesso em: 7 jan. 2021

OSLISLO, Z.; MAŁECKA, A. (EDS.). **The Insects Project: Problems of Diacritic Design for Central European Languages**. 2. ed. Katowice: Academy of Fine Arts in Katowice, 2016.

PERRY, A. J. After 37 years in US, Korean adoptee speaks out about imminent deportation. **the Guardian**, 28 out. 2016.

PIKE, K. L. **Phonemics: A Technique for Reducing Languages to Writing**. Ann Arbor: The University of Michigan Press, 1947.

Polybius square. Disponível em: <https://en.wikipedia.org/w/index.php?title=Polybius_square&oldid=1020727709>. Acesso em: 26 jul. 2021.

POSSELT, N. **Apple Macintosh that runs Ikarus M software, keyboard and mouse, letterform master drawings marked for digitization using an Aristo tablet and a sensor device, from Erik Spiekermann's collection, Berlin**, 2018. Disponível em: <https://miro.medium.com/max/4724/1*Uf734eU4pLQ6DFeuDaAqAQ.jpeg>. Acesso em: 18 jun. 2021

PROFESSORES MEHINAKU. **Imiehünaku layaka : livro para alfabetização na língua mehinaku**. Brasília: ISA / ATIX, 2002.

RAMÍREZ, A. DE G.; LEÓN, N. **Breve compendio de todo lo que debe saber, y entender el christiano, para poder lograr, ver, conocer, y gozar de Dios nuestro señor en el cielo eternamente**. México: Imprenta nueva Madrileña de los Herederos del Lic. D. Joseph de Jauregui, 1785.

RAY, P. S. Language Standardization. In: RICE, F. A. (Ed.). **Study of the Role of Second Languages in Asia, Africa, and Latin America**. Washington: Center for Applied Linguistics of the Modern Language Association of America, 1962. p. 91–104.

RIBEIRO, D. **O povo brasileiro: a formação e o sentido do Brasil**. 2. ed. São Paulo: Companhia das Letras, 1995.

RODRIGUES, A. D. **Línguas ameríndias**. Rio de Janeiro: Editora Delta, 1970. (Nota técnica).

RODRIGUES, A. D. **Línguas brasileiras: para o conhecimento das línguas indígenas**. São Paulo: Edições Loyola, 1986.

RODRIGUES, A. D. **A Originalidade das Línguas Indígenas Brasileiras**. Disponível em: <<http://www.comciencia.br/dossies-1-72/reportagens/linguagem/ling13.htm>>. Acesso em: 19 jan. 2019.

RODRIGUES, A. D. Sobre as línguas indígenas e sua pesquisa no Brasil. **Ciência e Cultura**, v. 57, n. 2, p. 35–38, jun. 2005.

ROSS, F. **Non-Latin scripts: from metal to digital type**. London: St. Bride Foundation, 2012.

ROSS, F. **A pocketful of type: the impact of mobile phones on global type design**. Disponível em: <<https://www.typemag.org/post/a-pocketful-of-type>>. Acesso em: 2 ago. 2021.

ROSS, S. **From Glyphs to Letters: the Mayan Writing Reform**. Dissertation (Master of Arts in Typeface Design) – Reading, UK: University of Reading, 2010.

RUIZ DE MONTOYA, A.; PLATZMANN, J. **Arte, Bocabulario, Tesoro y Catecismo de la lengua gvarani por Antonio Ruiz de Montoya. Publicado nuevamente sin alteracion alguna por Julio Platzmann**. Leipzig: B.G. Teubner, 1876.

- SAMUŁENKOVA, A. **Diacritics as a Means of Self-Identification: the Case of Latvia** Warsaw, Poland, 25 out. 2016. Disponível em: <<https://www.youtube.com/watch?v=uBqReCzyqRg>>. Acesso em: 20 jan. 2019
- SAMUŁENKOVA, A. **List of resources on designing diacritics**, 2019. Disponível em: <<https://www.alphabettes.org/resources-on-designing-diacritics/>>. Acesso em: 30 jul. 2021
- SANG-HUN, C. South Korea's Latest Export: Its Alphabet. **The New York Times**, 9 nov. 2009.
- SAUSSURE, F. DE. **Curso de lingüística geral**. Tradução: Antônio Chelini; Tradução: José Paulo Paes; Tradução: Izidoro Blikstein. 27. ed. São Paulo: Cultrix, 2006.
- SAVOIE, A. **International cross-currents in typeface design: France, Britain and the US in the phototypesetting era**. PhD thesis – Reading, UK: The University of Reading, jan. 2014.
- SCAGLIONE, J. Fuentes Inteligentes. **tipoGráfica**, n. 70, p. 06–07, 2006.
- SCAGLIONE, J. Tipografia como software. In: HENESTROSA, C.; MESEGUER, L.; SCAGLIONE, J. (Eds.). **Como criar tipos: do esboço à tela**. Tradução: Priscila Farias. 2. ed. Brasília: Estereográfica, 2019. p. 112–119.
- SCHEICHELBAUER, R. E. **Unicode | Glyphs**. Tutorial. Disponível em: <<https://glyphsapp.com/tutorials/unicode>>. Acesso em: 15 jul. 2020.
- SCHEICHELBAUER, R. E.; SEIFERT, G. **Glyphs 2.3 Handbook: Create–Produce–Release Font Editing for Everyone**. Viena: Glyphs GmbH, 2016.
- SCHEICHELBAUER, R. E.; SEIFERT, G. **Learn – Search results for: ‘diacritics’**. Disponível em: <<https://glyphsapp.com/learn?q=diacritics>>. Acesso em: 31 jul. 2021.
- SCIENCE MUSEUM GROUP. **Murray keyboard perforator model No. 5-PM, 1914-1916. 1976-366**, [s.d.]. Disponível em: <<https://collection.sciencemuseumgroup.org.uk/objects/co33634/murray-keyboard-perforator-model-no-5-pm-1914-1916-telegraph>>. Acesso em: 8 jan. 2021
- Script. Glossary of Unicode Terms**. [s.l.: s.n.]. Disponível em: <<https://www.unicode.org/glossary/#script>>. Acesso em: 25 jun. 2020
- SEKI, L. A linguística indígena no Brasil. **DELTA: Documentação de Estudos em Lingüística Teórica e Aplicada**, v. 15, n. SPE, p. 257–290, 1999.
- SHANNON, C. E.; WEAVER, W. **A teoria matemática da comunicação**. Tradução: Orlando Agueda. São Paulo: Difel, 1975.
- SIL INTERNATIONAL. **ISO 639-3**. Disponível em: <<https://iso639-3.sil.org/>>. Acesso em: 20 jul. 2021.
- SMEIJERS, F. **Contrapunção: fabricando tipos no século dezesseis, projetando tipos hoje**. Tradução: Gustavo Ferreira. Brasília: Estereográfica, 2015.
- SOUSA, M. **OpenType Feature File Syntax**. Adobe, 2010.
- SOUTHALL, R. **Printer's type in the twentieth century: manufacturing and design methods**. London; Newcastle: The British Library; Oak Noll Press, 2005.
- SPOLSKY, B. (ED.). **The Cambridge Handbook of Language Policy**. Cambridge: Cambridge University Press, 2012a.

SPOLSKY, B. What is language policy? In: SPOLSKY, B. (Ed.). **The Cambridge Handbook of Language Policy**. Cambridge Handbooks in Language and Linguistics. Cambridge: Cambridge University Press, 2012b. p. 3–15.

STEGEMAN, R. Orthographies in Papua New Guinea through the Years. **Graphemics in the 21st Century. Brest, June 1315, 2018. Proceedings**, Grapholinguistics and Its Applications. v. 1, p. 269–292, 2019.

STEINBERG, S. H. **Five Hundred Years of Printing**. 4. ed. London: The British Library; Oak Noll Press, 1996.

STRAAIJER, R. **Language Standardization**. Disponível em: <<https://www.oxfordbibliographies.com/view/document/obo-9780199772810/obo-9780199772810-0250.xml>>. Acesso em: 15 mar. 2021.

Teclado. Disponível em: <[https://pt.wikipedia.org/w/index.php?title=Teclado_\(inform%C3%A1tica\)&oldid=61229018](https://pt.wikipedia.org/w/index.php?title=Teclado_(inform%C3%A1tica)&oldid=61229018)>. Acesso em: 24 jul. 2021.

THEVET, A. **La cosmographie universelle d'Andre Thevet cosmographe du roy**. Paris: Chez Pierre L'Huillier : Chez Guillaume Chardiere, 1575.

TRONCARELLI, M. C. (ED.). **Wauja Ogatakojataku: livro para alfabetização na língua Wauja**. São Paulo: ISA / ATIX, 2004.

TROUILLE, INGÉNIEUR. **Plan, coupe et élévation du clocher de St Pierre d'Oléron**, 1807. Disponível em: <<https://gallica.bnf.fr/ark:/12148/btv1b7741800r>>. Acesso em: 7 jan. 2021

TRU'ONG, D. **Vietnamese Typography**. Disponível em: <<https://vietnamesetypography.com/>>. Acesso em: 30 jul. 2021.

TURING, A. M. On Computable Numbers, with an Application to the Entscheidungsproblem. **Proceedings of the London Mathematical Society**, v. s2-42, n. 1, p. 230–265, 1 jan. 1937.

TWARDOCH, A. **Polish Diacritics: how to?** Disponível em: <<http://www.twardoch.com/download/polishhowto/intro.html>>. Acesso em: 20 jan. 2019.

TWARDOCH, A. Pickled herring and strawberry ice cream: designing Polish diacritics. In: **Language culture type: international type design in the age of Unicode**. Redhill, UK; New York, NY: ATypI : Graphis, 2002. p. 148–157.

ULRICH, F. P. A brief overview of developments in digital type design. In: **Yearbook of Type III**. Karlsruhe: Slanted Publishers, 2018.

UNGER, G. The Design of a Typeface. **Visible Language**, v. XIII, n. 2, p. 134–149, 1979.

UNGER, G. **Enquanto você lê**. Tradução: Maíra M. Galvão. Brasília: Estereográfica, 2016.

UNGER, G. **Theory of type design**. Rotterdam: nai010 publishers, 2018.

UNICODE CONSORTIUM. **The Unicode standard: version 13.0 - core specification**. Mountain View, CA: Unicode Consortium, 2020a.

UNICODE CONSORTIUM. **C0 Controls and Basic Latin**. Unicode Consortium, 2020b. Disponível em: <<https://www.unicode.org/charts/PDF/U0000.pdf>>. Acesso em: 12 jan. 2021

UNICODE CONSORTIUM. **C1 Controls and Latin-1 Supplement**. Unicode Consortium, 2020c. Disponível em: <<https://www.unicode.org/charts/PDF/U0080.pdf>>. Acesso em: 12 jan. 2021

UNICODE CONSORTIUM. **Combining Diacritical Marks**. Unicode Consortium, 2020d. Disponível em: <<https://www.unicode.org/charts/PDF/U0300.pdf>>. Acesso em: 12 jan. 2021

UNICODE CONSORTIUM. **Spacing Modifier Letters**. Unicode Consortium, 2020e. Disponível em: <<https://www.unicode.org/charts/PDF/U02B0.pdf>>. Acesso em: 12 jan. 2021

UNNA, J. DE B. Impresiones de la tipografía mexicana. **tipoGráfica**, n. 59, p. 26–31, mar. 2004.

WHISTLER, K.; DAVIS, M.; FREYTAG, A. **Unicode Technical Report #17: Unicode Character Encoding Model**. Disponível em: <<http://www.unicode.org/reports/tr17/tr17-7.html>>. Acesso em: 24 jun. 2020.

WOLF, M. **Teorias da Comunicação**. Tradução: Maria Jorge Vilar De Figueiredo. Lisboa: Editorial Presença, 1987.

WRIGHT, S. Language policy, the nation and nationalism. In: SPOLSKY, B. (Ed.). **The Cambridge Handbook of Language Policy**. Cambridge Handbooks in Language and Linguistics. Cambridge: Cambridge University Press, 2012. p. 59–78.

Writing System. Glossary of Unicode Terms. [s.l.: s.n.]. Disponível em: <https://www.unicode.org/glossary/#writing_system>. Acesso em: 2 jul. 2020

WÜRKER, E. (ED.). **Nana ae jeangui jera’au wi: livro de saúde Kaiabi**. São Paulo: Instituto Socioambiental / ATIX, 2003a.

WÜRKER, E. (ED.). **Wa mbeto sujareni: livro de saúde Kisêdjê (Suiá)**. São Paulo: ISA / ATIX, 2003b.

WÜRKER, E. (ED.). **Ikpeng agingpinpe itowo: livro de saúde Ikpeng**. São Paulo: ISA / ATIX, 2003c.

WÜRKER, E. (ED.). **Wábiseha: livro de saúde na língua Yudja**. São Paulo: ISA / ATIX, 2005.

YAPUGUAY, N. **Sermones y exemplos en lengua Guaraní [facsimile]**. Buenos Aires.: Editorial Guaranía, 1953.

ZENTGRAF, D. C. **What Every Programmer Absolutely, Positively Needs to Know About Encodings and Character Sets to Work With Text**. Disponível em: <<https://kunststube.net/encoding/>>. Acesso em: 11 jun. 2020.

Bibliografia analisada no levantamento de dados

As tabelas correspondentes a cada publicação podem ser encontradas pelo código que antecede o nome do documento.

ALBUQUERQUE, F. E. (ED.). [apn]_005_Ahtor Kurum ho Hkrax ne ma te ne Xore Kamã **Hatuji: inicia em Jaó e finaliza em Raposa**. Fortaleza: Printcolor, 2007.

ALBUQUERQUE, F. E. (ED.). [apn]_003_Apinayé Kapêr Pumunh Ewa ã Kagã. Goiânia: Editora da PUC Goiás, 2011a.

ALBUQUERQUE, F. E. [apn]_008_Gramática pedagógica da língua Apinajé_. Goiânia: Editora da PUC Goiás, 2011b.

ALBUQUERQUE, F. E. (ED.). [xra]_002_Mê Ipê Krahô Catêjê Te Amj Ton Xà: **Arte e Cultura do Povo Krahô**. Belo Horizonte: FALÉ-UFMG, 2012.

ALBUQUERQUE, F. E. (ED.). [xra]_003_Krahô Jô Ikhãhhôc xa kat nã Carô (Alfabeto krahô **ilustrado**). Campinas: Pontes Editores, 2013.

ALBUQUERQUE, F. E. (ED.). [xra]_006_Geografia Krahô: Pjê Kãm ampo itajê nã carô. Campinas: Pontes Editores, 2014a.

ALBUQUERQUE, F. E. (ED.). [xra]_007_História Krahô: Mê pa inquêtjê kôt harên xà. Campinas: Pontes Editores, 2014b.

ALBUQUERQUE, F. E. (ED.). [xra]_008_Português Krahô. Campinas: Pontes Editores, 2014c.

ALBUQUERQUE, F. E. [xra]_009_Ciências Krahô. Campinas: Pontes Editores, 2016.

ALBUQUERQUE, F. E. (ED.). [xra]_005_Krahô jujarên xà kwý (Literatura krahô). Campinas: Pontes Editores, 2103.

ALBUQUERQUE, F. E.; KRAHÔ, R. Y. (EDS.). [xra]_004_Krahô jô ikhãhhôc kryjre mê cati. Campinas: Pontes Editores, 2013.

ALBUQUERQUE, F. E.; KRAHÔ, R. Y. (EDS.). [xra]_010_Gramática Pedagógica Krahô. Campinas: Pontes Editores, 2016.

ALBUQUERQUE, F. E.; PROFESSORES INDÍGENAS APINAYÉ. [apn]_006_Apinaje Grer ne Ujarenh ã Kagã: **livro de narrativas e cantigas Apinayé**. Fortaleza: Printcolor, 2007.

ALBUQUERQUE, F. E.; PROFESSORES INDÍGENAS APINAYÉ. [apn]_001_Panhî Kapêr - **Kupê Kapêr: Dicionário Escolar Apinayé – Português**. [s.l.] FALÉ-UFMG, 2012.

ALMEIDA, M. I. DE. [mbl]_005_O livro que conta histórias de antigamente (Mônãyxop 'ãgtux yôg tappet). Belo Horizonte: MEC / SEE-MG / PNUD, 1998.

ALUNOS DA ESCOLA ESTADUAL INDÍGENA KOGE EIARI. [bor]_003_Tugo Kuri Doge **Ewadarô Reo**. Cuiabá: [s.n.].

ALUNOS DO 3º MAGISTÉRIO XAVANTE. [xav]_003__Ai`utÉ ma rowatsu`u: **literatura infantil xavante**. Campo Grande: MEC / SEF / Universidade Católica Dom Bosco, 1999.

ALUNOS TAPIRAPÉ. [taf]_003_xeparama'ewa: cartilha tapirapé. São Félix do Araguaia: Prelazia de São Félix do Araguaia, 1987.

ALVES, W. R.; PROFESSORES MAXACALÍ (EDS.). [mbl]_001_Geografia da nossa aldeia: ũxuxet ax, hãmxek āgtux. Brasília: SEE-MG I MEC, 2000.

ASSOCIAÇÃO ESCOLA INDÍGENA TUKANO YUPURI (ED.). [tuo]_003_Ati umukhore suori nirã kihti. São Gabriel da Cachoeira; São Paulo: ISA / FOIRN / AEITY, 2011.

AWETI, W. et al. [awe]_001_Awytyza ti 'ingku: livro para alfabetização na língua Aweti. [s.l.] Instituto Socioambiental, 2002.

AZEVEDO, V. V.-B.; AZEVEDO, D. A.; OLIVEIRA, M. (EDS.). [tuo]_004_Mariye kihti añuse. São Gabriel da Cachoeira; Belo Horizonte; Brasília: AEITY / FALE-UFMG / SECAD-MEC, 2007.

BATISTA, C. N.; LADEIRA, M. E. (EDS.). [pak]_001_Parakanã Imopinimawa. [s.l.] Escola Parakanã, 2014.

CABALZAR, A. et al. (EDS.). [tue]_005_Bureko watotire, wametire 1. São Gabriel da Cachoeira; São Paulo: ISA / FOIRN / AEITY, 2009.

CABALZAR, F. D. (ED.). [tue]_001_Mariya dita - Iñanunuse masire/Escola Indígena. Brasília: ISA, 2001.

CABALZAR, F. D. [tue]_002_Kiti wederira tuohoarira. [s.l.] ISA, 2002.

COMUNIDADE INDÍGENA ITAPEREIRA. [yrl]_001_Yasu Yalery Nheengatu. Belo Horizonte: FALE/UFMG, 2007.

COMUNIDADE KURÂ-BAKAIRI. [bkq]_001_Iweniby Modo: Abecedário na Língua Bakairí. Anápolis: SIL, 2012a.

COMUNIDADE KURÂ-BAKAIRI. [bkq]_002_Kâsenomedâdyly 1 - Estou Aprendendo 1: Livro de Apoio 1 na Língua Bakairí. Anápolis: SIL, 2012b.

COMUNIDADE KURÂ-BAKAIRI. [bkq]_003_Kâsenomedâdyly 2 - Estou Aprendendo 2: Livro de Apoio 2 na Língua Bakairí. Anápolis: SIL, 2012c.

COMUNIDADE KURÂ-BAKAIRI. [bkq]_004_Kytanwem Tâdâsenomedâdo Kâintainhoem - Para Nós Lermos Na Nossa Língua: Livro de Transição: de Português para Bakairí. Anápolis: SIL, 2012d.

COMUNIDADE TAPIRAPÉ. [taf]_002_xanetawa parageta - histórias das nossas aldeias. São Paulo / Brasília: MEC / SEF / MARI, 1998.

D'ANGELIS, W. DA R. [kgp]_003_A língua KaingangPortal Kaingang, , [s.d.]. Disponível em: <http://www.portalkaingang.org/lgua_kaingang.pdf>

DE LIMA, T. B. (ED.). [kgk]_002 Amandaje nº 169 Ava Ka'aguygua rehegua Tetã Sãsoyguakuéruguáva ha Ñemobohechapyrã ojejapo va'erã pe OITre ñemoñe'epyre Ava ka'aguygua rehe. Tradução da Convenção nº 169 sobre Povos Indígenas e Tribais e Resolução referente à ação da OIT para a língua indígena brasileira Guarani Kaiowá. Brasília: OIT, 2012.

DE OLIVEIRA, G. M. (ED.). [xok]_001_Vãnhkala Jó Kabel. Florianópolis: Secretaria de Estado da Educação e do Desporto/SC, 1999.

- DIETZSCH, R. (ED.). **Catálogo ortográfica das línguas indígenas brasileiras: formulário Google Forms**, 2021. Disponível em: <https://docs.google.com/spreadsheets/d/1uFEAnNXNmV-5QBDbTcczdpK822u_kFCPmnq6HOWAkA/edit?resourcekey#gid=1025703120>
- DO NASCIMENTO, P. C. N. **[tuo]_002_Mari yeé kiti: nossas histórias**. São Gabriel da Cachoeira: MEC, 2001.
- DRUDE, S.; AWETI, W.; AWETI, A. **[awe]_002_A Ortografia da Língua Awetí**. v. 19, n. e019014, p. 1–23, 2019.
- EDUCADORES GUARANI DA CIDADE DE SÃO PAULO. **[gun]_002_Nhandereko nhemombé u tenonderã: Histórias para contar e sonhar**. São Paulo: SME / DOT, 2007.
- FARGETTI, C. M. (ED.). **[jur]_001_Yudja Kamena Dju’a Papera**. São Paulo: ISA, 1998.
- FAUSTO, C.; FRANCHETTO, B. **[kui]_008_Tisakisü : tradição e novas tecnologias da memória : Kuikuro do Alto Xingu = Tisakisü : tradition and the new technologies of memory : the Kuikuro of the Upper Xingu / Carlos Fausto, Bruna Franchetto**. [s.l: s.n.].
- FERNANDES, A. C. [DIAKURU]; FERNANDES, D. M. [KISIBI]. **[des]_001_Bueri kãdiri maririye: os ensinamentos que não se esquecem**. São Gabriel da Cachoeira: UNIRT / FOIRN, 2006.
- FERREIRA, H. P. **Projeto 914BRZ4019 Contrato SA-3993/2019**. Brasília: Museu do Índio - FUNAI, 2021.
- FONTES, A.; DINIZ, L. L.; TRUJILLO, T. P. (EDS.). **[kpc]_002_Ikadzekatakadapha**. São Paulo: ISA, 2005.
- FRANCESCHINI, D. (ED.). **[mav]_001_Hühü e’hap: IV etapa do curso de formação de professores indígenas Saterê-Mawe**. Manaus: Estado do Amazonas / Programa de Educação Escolar Indígena, 1997.
- FRANCESCHINI, D. **[mav]_002_Wantym sa’awy etiat: livro de leitura na língua sateré-mawé**. Brasília: OPISM / MEC, 2000a.
- FRANCESCHINI, D. **[mav]_003_Satere-mawe: mowe’eg hap**. Manaus: GOVERNO DO AMAZONAS - SEDUC - IER/AM, 2000b.
- FRANCESCHINI, D. (ED.). **[mav]_004_Warana sa’awy etiat: Livro de Leitura na Língua Saterê-Mawe**. Brasília: OPISM / MEC, 2000c.
- FRANCHETTO, B.; TRONCARELLI, M. C. (EDS.). **[kui]_003_Tisügühütu, KuKügühütu: livro de estudo da língua kalapalo/nahukuá/matipú**. São Paulo: ISA / ATIX, 2002.
- FRANCHETTO, B.; TRONCARELLI, MARIA CRISTINA; GAUDITANO, C. (EDS.). **[kui]_006_Kungatagohoha igei ngongoi- esta é a terra que nós plantamos**. Belo Horizonte: AIKAX FALE/UFMG SECAD/MEC, 2007a.
- FRANCHETTO, B.; TRONCARELLI, MARIA CRISTINA; GAUDITANO, C. (EDS.). **[kui]_007_Ngongoha igei kungatagoho - esta é a terra que nós plantamos**. Belo Horizonte: AIKAX FALE/UFMG SECAD/MEC, 2007b.
- GALLOIS, D. T. (ED.). **[pto]_001_Jo’e rekoha bokituteha ram: planejando como vamos continuar vivendo bem no futuro. Plano de Gestão Territorial e Ambiental TI Zo’é**. Santarém: Iepé / Funai, 2019.

- GALVÃO, W. S. [TÕRÂMUB.; GALVÃO, R. C. [GUAHARI Y. Ñ. [des]_002 **Livro dos antigos Desana: guahari diputiro porã**. São Gabriel da Cachoeira: ONIMRP / FOIRN, 2004.
- GAVAZZI, R. A.; ROCHA, M. C. (EDS.). [cbs]_003 **Mae shawãyã namaki hariri raya kaĩti xinã: Plano de Gestão territorial e ambiental da Terra Indígena Kaxinawá do Rio Humaitá**. Rio Branco: Comissão Pró-Índio do Acre, 2015.
- GONSALVES, A. et al. [kgk]_004 **upéicha rohai - ore kuatione ñe`ẽ peteĩha**. Dourados; Campo Grande: CIMI / CED MS, 1993.
- GRIFFITHS, G. [kbc]_001 **DICIONÁRIO da LÍNGUA KADIWÉU**. Cuiabá: SIL, 2012.
- HALL, J.; MCLEOD, R. A.; MITCHELL, V. (EDS.). [xav]_002 **Pequeno Dicionário Xavante-Português Português-Xavante**. Sociedade Internacional de Linguística: SIL, 2004.
- IKPENG, K.; IKPENG, I.; IKPENG, M. [txi]_001 **Ikpeng Orempantop**. São Paulo: ISA, 2001.
- ISA. [kui]_002 **Gekuilene: livro de saúde Kalapalo, Matipu e Nahukuá**. São Paulo: ISA, 2003.
- JECUPÉ, K. W. [gun]_001 **Tupã Tenondé : a criação do universo, da terra e do homem segundo a tradição oral Guarani / Kaka Werá Jecupé**. São Paulo: Peirópolis, 2001.
- JESUS, A. DE [KUI]; JESUS, P. DE [PUKUTHA]; AGUIAR, L. [KUENAKA]. [tae]_001 **Ennu ianáperi: história dos Tariano pelo clã Koivathe**. São Gabriel da Cachoeira: Coidi / FOIRN, 2018.
- KAIABI, TAKAP PI'YU TRUMAI; TRUMAI, MATSILAKE. [tpy]_002 **Ayey Kut'a: trumai wan kate daint'a**. São Paulo: ISA / ATIX, 2004.
- KATUKINA, B. A.; SENA, V. O. (EDS.). [knt]_001 **Noke shoviti: Mitos Katukina**. Rio Branco: Poronga, [s.d.].
- KAXINAWÁ, N. S. [cbs]_001 **NIXPU PIMA O ritual de passagem do povo Huni Kuĩ**. Rio Branco: Fale/UFMG, 2015.
- KEPPI, J.; GOMIDE, M. L. C. (EDS.). [arr]_001 **Alimentação Karo Arara: Saberes e Práticas**. São Leopoldo: Editora Oikos Ltda, 2016.
- KEPPI, J.; PRUIKSMA, N. (EDS.). [arr]_002 **Nossas vidas: histórias de mulheres Karo Arara. Iba'kât kanã: ma'pâyrap at kanã xet to'**. São Leopoldo: Editora Oikos Ltda, 2018.
- [kgk]_001 **Nembohoky Ne'e Tesai Rehehápe: Mbo'ehára kuéra Kaiowá ha Guarani**. Mato Grosso do Sul: Secretaria de Estado de Educação MS / MEC, 2002.
- KRAHÔ, R. Y.; ALBUQUERQUE, F. E. [xra]_001 **Livro de Alfabetização Krahô**. Fortaleza: Printcolor, 2009.
- KUIKURO, S.; KUIKURO, M.; KUIKURO, T. [kui]_001 **Tisügühütu, KuKügühütu: livro de leitura e atividades de escrita na língua kuikuro**. São Paulo: ISA, 2003.
- LADEIRA, M. I.; VITTI, V. T. (EDS.). [gun]_003 **Jaguatareí Nhemboé – caminhando e aprendendo com os Guarani Comunidade de Guarani-Mbyá de Aguapeú**. Brasília: MMA, 2006.
- LINKE, I. L. VAN V.; VAN VELTHEM, L. H. **O livro da argila: Ëliwë Pampila – Orino Papeh**. São Paulo: Iepé, 2017.

LUCKMANN, S.; FALCADE, N. T. (EDS.). **[kjp]_001_Gufã Ag Kajró: Ti êg kajrán-ja kâjatun ge tu êg nĩ**. São Leopoldo: Editora Oikos Ltda, 2008.

MAIA, M. **[tuo]_001_Isâ Yekisimia Masike = o conhecimento dos nossos antepassados : uma narrativa Oye / narradores, Moisés Maia (Akito), Tiago Maia (kimaro) ; colaboradores, Manuel Maia (Akito), Guilherme Maia (Doe), Laureano Maia (Akito) ; intérprete, Arlindo Maia (Yepara) .** [s.l.: s.n.].

MANA, J. P.; MONTE, N. L. **[cbs]_002_Shenipabu miyui : história dos antigos / Organização dos Professores Indígenas do Acre.** [s.l.: s.n.].

MANCHINERI, J.; MANCHINERI, G.; MANCHINERI, A. P. **[mpd]_001_Geografia Manchineri.** Rio Branco: [s.n.].

MATOS, B. DE A. (ED.). **[mcf]_001_Matses Dadauaid Mequiadte.** Brasília: CTI, 2008.

MAXAKALI, I. et al. **[mbi]_004_Hitupmã'ax: curar.** Belo Horizonte: FALE/Literaterras/UFMG, 2008.

NASCIMENTO, H. S. (ED.). **[mpq]_002_Matsesën Txu Darawakit - Palavras Matis.** Brasília: CTI, 2005.

NASCIMENTO, H. S.; MAKUANANTE, B.; LADEIRA, M. E. **[mpq]_001_Txapa.** Brasília: CTI, 2006.

NAVARRO, E. DE A. **[yrl]_003_Curso de Língua Geral (Nheengatu ou Tupi Moderno): A Língua das Origens da Civilização Amazônica.** São Paulo: Paym Gráfica e Editora, 2016.

ORGANIZAÇÃO GERAL DOS PROFESSORES TICUNA BILÍNGUES. **[tca]_001_Werigü arü ae.** Brasília: OGPTB / MEC, 2002a.

ORGANIZAÇÃO GERAL DOS PROFESSORES TICUNA BILÍNGUES. **[tca]_002_Cururugü Tchiga.** Brasília: OGPTB / MEC, 2002b.

OS BORORO DE MERURI-MT. **[bor]_001_Boe Eno Bakaru - Lendas Bororo.** Campo Grande/MS: Missão Salesiana de Mato Grosso, [s.d.].

PÃRÕKUMU, U.; KÊHÍRI, T. **[des]_003_Antes o mundo não existia: Mitologia dos antigos Desana-Kêhíripõrã.** São Gabriel da Cachoeira; São João Batista do Rio Tiquié: UNIRT / FOIRN, 1995.

[plu]_001 Waramwi: A Cobra Grande. [s.l.: s.n.].

PROFESSORES, ALUNOS E LIDERANÇAS WANANO. **[gvc]_001_Kootira Ya Me'Ne Buehina : Wa'Ikina Khiti Kootiria Yame'Ne.** São Gabriel da Cachoeira; São Paulo: ISA / FOIRN / Imprensa Oficial SP, 2004.

PROFESSORES BANIWA E CORIPACO. **[kpc]_001_Baniwa Coripaco Iemakaa.** São Paulo: ISA, 2001.

PROFESSORES BORORO DA ALDEIA HERURI E O ANCIÃO DA TRIBO, ANTÔNIO KANAJÓ. **[bor]_002_Mano: um ritual bororo e uma experiência didático-pedagógica.** Campo Grande/MS: MEC/SEDUC-MT/PNUD, [s.d.].

PROFESSORES INDÍGENAS APINAYÉ. [apn]_002_ **Āmnepēm Apinajejaja ujarēnh nē pika kôt mẽmoj pumunhã kagã. História e Geografia Apinayé.** [s.l.] Editora Curt Nimuendajú / UFT (Universidade Federal do Tocantins), 2007a.

PROFESSORES INDÍGENAS APINAYÉ. [apn]_004_ **Apinaje kamã nũm mẽ mry apỹnhã wa harēnh_ Matemática e Ciências Apinayé.** Campinas: Editora Curt Nimuendajú, 2007b.

PROFESSORES INDÍGENAS APINAYÉ. [apn]_007_ **Apinaje Kanexà Jarenh ã Kagã: receitas da medicina tradicional Apinayé.** Fortaleza: Printcolor, 2007c.

PROFESSORES MEHINAKU. [mmh]_001 **Imiehũnaku layaka : livro para alfabetização na língua mehinaku.** Brasília: ISA / ATIX, 2002.

PROFESSORES YANOMAMI. [wca]_001_ **Professores Yanomami.** [s.l: s.n.].

PROFESSORES YANOMAMI. [wca]_001_ **Të ã oni puhi mōyamiwei siki.** [s.l: s.n.].

PROFESSORES-ASSISTENTES DO PROJETO ÁRA VERÁ. [kgk]_003_ **Te'yi rembiapo / Mbo'ehára kuéra Kaiowá ha Guarani.** Mato Grosso do Sul: Secretaria de Estado de Educação MS / MEC, 2002.

RAMOS, J. B. [tue]_003_ **Utapiopona kuye poseminiã niromakaraye- pássaros-adornos dos filhos da cobra de pedra (Tuyuka).** Tradução: Higino Pimentel Tenório; Tradução: Guilherme Tenório; Tradução: José Fernandes Barbora. São Paulo, São Gabriel da Cachoeira: ISA / AEITU, 2012.

RURI'Õ, L.; DE BIASE, H. S. [xav]_001_ **Daró Idzô'uhu Watsuu - A História da Aldeia Abelhinha.** São Paulo: Master Book, 2000.

SASS, W. (ED.). [dny]_003_ **Ivimari Deni – Nossa língua deni: cartilha de alfabetização e textos de leitura.** São Leopoldo: Editora Oikos Ltda, 2003.

SASS, W. (ED.). [dny]_001_ **Ima Bute Denikha - Mitos Deni.** São Leopoldo: Editora Oikos Ltda, 2004.

SASS, W. (ED.). [knm]_001_ **Tâkuna Nawa Bũh Amteiyam Amkira: Mitos Kanamari.** São Leopoldo: Editora Oikos Ltda, 2007.

SASS, W. (ED.). [dny]_002_ **Niha Inavatuna Aru Vainade: Matemática Deni.** São Leopoldo: Editora Oikos Ltda, 2008.

SEKI, L.; PROFESSORES KAMAIURÁ (EDS.). [kay]_001_ **Kamajura Jemo'etap.** São Paulo: ISA, 1998.

SOCIEDADE INTERNACIONAL DE LINGÜÍSTICA. [mbl]_002_ **Dicionário Maxakalí-Português.** Cuiabá: Sociedade Internacional de Lingüística, 2005.

SOUZA, P. DE O. B. [kyz]_002_ **Algumas considerações sobre a escrita kaiabi.** Anais 15º COLE – Congresso de Leitura do Brasil. **Anais...** In: 15º COLE – CONGRESSO DE LEITURA DO BRASIL. Campinas: Associação de Leitura do Brasil, 2005. Disponível em: <http://alb.com.br/arquivo-morto/edicoes_anteriores/anais15/Sem05/patricia.htm>

TENÓRIO, H. P. [tue]_004_ **Keore- Utapiopona saiña hoa bauaneriputi.** São Paulo; São Gabriel da Cachoeira: ISA / FOIRN, 2004.

TISS, C. (ED.). [cul]_001_Zophinehedeni Kha Wima – Thothoho Tamine – Tuberculose Tamine. São Leopoldo: Editora Oikos Ltda, 2008.

TISS, F. (ED.). [cul]_002_Nehekomaneza ikha madie tobikani wahine kha wima: cartilha alcoolismo. São Leopoldo: Editora Oikos Ltda, 2007.

TORAL, A. A. DE (ED.). [kgp]_005_Ëg jamên kÿ mû: textos kanhgág. Brasília: APBKG/DkaÁustria/MECPNUD, 1997.

TRONCARELLI, M. C. [kyz]_003_Jane jemu'jawa ypyrungawa jane je'enga- livro para alfabetização na língua Kaiabi. São Paulo: ISA / MEC, 1999.

TRONCARELLI, M. C. (ED.). [wau]_001_Wauja Ogatakojatau: livro para alfabetização na língua Wauja. São Paulo: ISA / ATIX, 2004.

TRONCARELLI, M. C.; ZORTHEA, K.; DOS SANTOS, L. C. (EDS.). [suy]_001_Kisêdjê Kapêrê - livro para alfabetização na língua Suyá.pdf. São Paulo: ISA, 1999.

TRUMAI, T. et al. [tpy]_001_Trumai: livro para ensino da língua trumai. São Paulo: ISA / ATIX, 2002.

TUGNY, R. P. DE. [mbi]_003_Môgmôka yôg kutex xi âgtux : cantos e histórias do gavião-espírito / narradores, escritores e ilustradores tikmu'un da Terra Indígena de Água Boa ; estudo, organização e versão final, Rosângela Pereira de Tugny. [s.l: s.n.].

VAITI, V. C. [kgp]_002_Dicionário Bilingue Ilustrado: Língua Portuguesa e Língua Kaingang. São Paulo: FEUSP, 2003.

WAJÃPI, J. [oym]_001_Konomiokyry, Tanôgarã, Tanogarerã : Moma'e kô jane pota jôme. Tea pupe mã'eô ee mãekô, jane paje rupi korijô te mã'eô ee mãekô = Como funcionava as antigas práticas de resguardo dos rapazes e como funcionam os resguardos que fazemos até hoje / [autoria e desenhos de:] Jawaruwa Wajãpi. [s.l: s.n.].

WELPER, E. M. (ED.). [mzr]_001_Yorã Vana Wicha Ni Pei Rao: remédios do mato. [s.l.] CTI, 2005.

WIESEMANN, U. G. [kgp]_004_Dicionário Kaingang - Português | Português - Kaingang. 2ª ed. Curitiba: Editora Evangélica Esperança, 2011.

WÜRKER, E. [kui]_005_Gekuilene- livro de saúde Kuikuro. São Paulo: ISA / ATIX, 2003a.

WÜRKER, E. (ED.). [kyz]_001_Nana ae jeangui jera'au wi: livro de saúde Kaiabi. São Paulo: Instituto Socioambiental / ATIX, 2003b.

WÜRKER, E. (ED.). [suy]_002_Wa mbeto sujareni: livro de saúde Kisêdjê (Suiá). São Paulo: ISA / ATIX, 2003c.

WÜRKER, E. (ED.). [txi]_002_Ikpeng agëngpïnpe itowo: livro de saúde Ikpeng. São Paulo: ISA / ATIX, 2003d.

WÜRKER, E. (ED.). [jur]_002_Wãbiseha: livro de saúde na língua Yudja. São Paulo: ISA / ATIX, 2005a.

WÜRKER, E. [kay]_002_Jene katuwetap- livro de saúde na língua Kamaiurá. São Paulo: ISA / ATIX, 2005b.

WÜRKER, E. (ED.). [tpy]_003_ **Det'a yaw is- livro de saúde na língua Trumai**. São Paulo: ISA / ATIX, 2005c.

WÜRKER, E.; TRONCARELLI, M. C. (EDS.). [kui]_004_ **Tisakisü**. São Paulo: ISA, 1996.

WÜRKER, E.; TRONCARELLI, MARIA CRISTINA (EDS.). [txi]_003_ **Watpiro mīran = Livro sobre nutrição**. Belo Horizonte: FALE/UFMG, 2012.

WÜRKER, E.; WERNECK, A. (EDS.). [kre]_001_ **Priara jō Howkjya.pdf**. São Paulo: ISA, 2002.

[yrl]_002_ **Barekena Nheċga**. [s.l: s.n.].

Bibliografía consultada

- ACOSTA, N. **Michay (fonte tipográfica)**. Buenos Aires: [s.n.].
- ALCARAZ, J. **Typ09 - México DF: Tipografía nativa, entrevista con Marina Garone Gravier**, [s.d.]. Disponível em: <<https://atypimx09.blogspot.com/2009/08/tipografia-nativa-entrevista-con-marina.html>>. Acesso em: 25 ago. 2020
- ARTIGAS, C. **Inika (fonte tipográfica)**. Buenos Aires: [s.n.].
- CAMPOS, Y. **Mosonyi (fonte tipográfica)**. Buenos Aires: RNS Fonts, 2018a.
- CAMPOS, Y. **Programa tipográfico para el diseño de materiales bilingües yaruro-español**, 2018b.
- COSGAYA, P. La imprenta misionera. **tipoGráfica**, n. 59, p. 6–7, 2004a.
- COSGAYA, P. **Loreto: sin pecado concebida: Tipografía y lenguaje en Argentina, desde el siglo XVIII hasta hoy**. Tipos tópicos textos y contextos: primer congreso de tipografía, Valencia, 25-27 June, 2004. **Anais...** In: PRIMER CONGRESO DE TIPOGRAFÍA. Valencia: Asociación de Diseñadores de la Comunidad Valenciana, 2004b.
- CROFTS, M. Creating a Munduruku Orthography. **Visible Language**, v. 5, n. 1, p. 49–58, jan. 1971.
- DINIZ, K. C. **Tipografia e línguas indígenas: customização de fonte para uso por escritores e escritoras mebêngôkre (kayapó)**. Monografia de conclusão de curso de graduação – Brasília, Brasil: Universidade de Brasília, 2007a.
- DINIZ, K. C. Notas sobre tipografias para línguas indígenas do Brasil. **InfoDesign - Revista Brasileira de Design da Informação**, v. 4, n. 1, p. 36–46, 2007b.
- DINIZ, K. C. Marina Garone Gravier (Entrevista). **InfoDesign - Revista Brasileira de Design da Informação**, v. 6, n. 2, p. 50–54, 2009.
- DINIZ, K. C. **Desenho de letras em livros das Reduções Jesuíticas Guarani**. Dissertação de Mestrado – São Paulo: Universidade de São Paulo (USP). Faculdade de Arquitetura e Urbanismo, 2014.
- FONTANA, R. El idioma de las letras. **tipoGráfica**, n. 51, p. 18–29, 2002.
- GAULTNEY, V. J. **Gentium (fonte tipográfica)**. High Wycombe, UK: SIL International, 2003.
- GAULTNEY, V. J. **Multitudinous Alphabets: The challenges of extended Latin typeface design**. High Wycombe, UK, 2006.
- GIOVAGNOLI, C. **Andada Ht: tipografía para texto latinoamericanos**, 2011.
- GIOVAGNOLI, C. **Andada ht (fonte tipográfica)**. Argentina: Huerta Tipográfica, 2019.
- GIOVAGNOLI, C.; DEL PERAL, J. P. **Teclado Guaraní**. [s.l.] Huerta Tipográfica, 2011.

- GRAVIER, M. G. **Tipografía y diseño industrial: estudio teórico e histórico para la representación tipográfica de una lengua indígena.** Dissertação de Mestrado – México, DF: Universidad Nacional Autónoma de México, 2003.
- GRAVIER, M. G. **Historia de la tipografía colonial para lenguas indígenas.** Tipos tópicos textos y contextos: primer congreso de tipografía, Valencia, 25-27 June, 2004. **Anais...** In: PRIMER CONGRESO DE TIPOGRAFÍA. Valencia: Asociación de Diseñadores de la Comunidad Valenciana, 2004a.
- GRAVIER, M. G. Claves para una traducción sin traición. **tipoGráfica**, n. 60, p. 18–23, 2004b.
- GRAVIER, M. G. **La indianización del alfabeto. Recepción e apropiación escrita en América.** Las otras letras: segundo congreso de tipografía, Valencia, 23-25 June, 2006. **Anais...** In: SEGUNDO CONGRESO DE TIPOGRAFÍA, VALENCIA, 23-25 JUNE, 2006. Valencia: Asociación de Diseñadores de la Comunidad Valenciana, 2006.
- GRAVIER, M. G. Nuevos retratos para las viejas palabras: libros novohispanos en lenguas indígenas. **Istor: revista de historia internacional**, n. 31, p. 102–117, 2007.
- GRAVIER, M. G. **Historia de la tipografía colonial para lenguas indígenas.** Tese de doutorado – México, DF: Universidad Nacional Autónoma de México, 2009.
- GRAVIER, M. G. Kuati’á guarani: tres momentos de la edición tipográfica del guaraní (siglos XVII, XIX y XXI). In: LLUÍS I VIDAL-FOLCH, A.; JANER I COLL, P.; OSORNO MESA, C. (Eds.). **Literaturas amerindias.** Barcelona: Casa Amèrica Catalunya, 2010a. p. 133–140.
- GRAVIER, M. G. Reflexiones sobre la comunicación escrita de un grupo indígena mexicano. **Cultura Escrita Y Sociedad**, n. 11, p. 73–95, 2010b.
- GRAVIER, M. G. Cultura impresa colonial en lenguas indígenas: una visión histórica y regional. **Ensayos: Historia y Teoría del Arte**, v. 18, p. 99–145, 2010c.
- GRAVIER, M. G. La tipografía y las lenguas indígenas estrategias editoriales en La Nueva España. **Bibliofilia**, v. 113, n. 3, p. 355–373, 2011.
- GRAVIER, M. G. Tipografía e identidad lingüística. In: CONSOLO, C. (Ed.). **Tipografía en Latinoamérica.** São Paulo: Blucher, 2013a.
- GRAVIER, M. G. Letras para una lengua indomable. Tipografía y edición colonial en otomí. **Estudios de Historia Novohispana**, v. 49, p. 119–168, jul. 2013b.
- GRAVIER, M. G. El impresor Diego Fernández de León (1682-1710), pionero de la edición colonial poblana en lenguas indígenas. In: BARRIGA, R.; ZENDEJAS, E. H. (Eds.). **Lenguas, estructuras y hablantes. Estudios en homenaje a Thomas C. Smith Stark.** México, DF: El Colegio de México, 2014. v. 1p. 337–368.
- GRAVIER, M. G. Las cátedras universitarias de lenguas indígenas y la producción editorial en la Nueva España: una aproximación desde la historia del libro. In: RIVERA, M. S. (Ed.). **De eruditione americana. Prácticas de lectura y escritura en los ámbitos académicos novohispanos.** Ciudad de México: Universidad Nacional Autónoma de México, Instituto de Investigaciones Bibliográficas, Biblioteca Nacional / Hemeroteca Nacional, 2019. p. 127–174.
- GRAVIER, M. G.; WOLF, G.; CONNORS, J. **Sahagún’s Codex and Book Design in the Indigenous Context.** Florence: Villa I Tatti, 2011.

GÜRTLER, A. Los Romanos escribieron, pero no en alemán, ni tampoco en maya. **tipoGráfica**, n. 51, p. 20–27, 2003.

HEILBORN, J. **Kuarahy (fonte tipográfica)**. Assunção: [s.n.].

HEILBORN, J. **Jeroky (fonte tipográfica)**. Assunção: [s.n.].

HEILBORN, J. Hacia la lengua vista. **tipoGráfica**, n. 60, p. 25–27, 2004b.

HEILBORN, J. **Type and native languages in postcolonial America**. 2009 ATypI conference. **Anais...** In: TYP09: THE HEART OF THE LETTER. 2009 ATYPI CONFERENCE. México, DF: 26 out. 2009. Disponível em: <http://old.atypi.org/04_Mexico.html>

HEILBORN, J. **E'a (fonte tipográfica)**. Assunção: [s.n.].

HEILBORN, J. **Compilada (fonte tipográfica)**. Assunção: [s.n.].

HEILBORN, J. **Tipografia Para Guarani**. Disponível em: <<http://tipografiaparaguarani.org/>>. Acesso em: 21 set. 2020.

HEILBORN, J.; VILLAGRA, O. O. **From the jungle to the pixel**. 2009 ATypI conference. **Anais...** In: TYP09: THE HEART OF THE LETTER. 2009 ATYPI CONFERENCE. México, DF: 30 out. 2009. Disponível em: <http://old.atypi.org/04_Mexico.html>

HUDSON, J. **Sylfaen: Foundations of multiscript typography**Tiro Typeworks, , 2000. Disponível em: <http://www.tiro.com/Articles/sylfaen_article.pdf>. Acesso em: 25 ago. 2020

HUDSON, J. **Introducing the Brill Typeface to Scholarship: unrivalled completeness and typographical sophistication [Typeface Brochure]**Brill, , 2011. Disponível em: <https://brill.com/fileasset/downloads_static/static_typefacebrill_brilltypefacebrochure_2011.pdf>. Acesso em: 6 jan. 2021

HUDSON, J. **The Brill Types**, 16 maio 2014. Disponível em: <<http://www.tiro.com/John/Hudson-Brill-DECK.pdf>>. Acesso em: 25 ago. 2020

HUDSON, J.; SAVOIE, A. **Brill typeface**. Leiden: Brill, 2012.

HUDSON, J.; WADE, G. **Sylfaen typeface**. Seattle: Microsoft, 1998.

JAMRA, M.; PATEL, N. **Jamra Patel: Kigelia**. Disponível em: <<https://www.kigelia-font.com/>>. Acesso em: 25 ago. 2020.

LÓPEZ ROCHA, J. M. **Manuel López Rocha - Interview at ATypI 2017 Montréal**: ATypI 2017 Montréal.Montreal, 2017a. Disponível em: <<https://www.youtube.com/watch?v=BochY9xW4rI>>

LÓPEZ ROCHA, J. M. **Type design for american native languages: four case studies**: ATypI 2017 Montréal.Montreal, 16 set. 2017b. Disponível em: <<https://www.youtube.com/watch?v=IDFKz95E9Cs>>. Acesso em: 25 ago. 2020

LÓPEZ ROCHA, J. M.; GARCIA, S. **Gente**. Colômbia: [s.n.].

MILLS, R. **Huronia (fonte tipográfica)**. Brno: Rosetta, 2010a.

MILLS, R. **Huronia Type Specimen**Tiro Typeworks, , 2010b. Disponível em: <https://www.tiro.com/huronia/Huronia_Beta_16_OL.pdf>. Acesso em: 25 ago. 2020

- ROSS, S. **Yukatek (fonte tipográfica)**. Reading, UK: [s.n.].
- ROSS, S. **Yukatek typeface specimen**, 2010b. Disponível em: <<http://typefacedesign.net/typefaces/year/2010/yukatek-placeholder/>>
- ROSS, S. **From Glyphs to Letters: the Mayan Writing Reform**. Dissertation (Master of Arts in Typeface Design) – Reading, UK: University of Reading, 2010c.
- SCHWARZENBERG, J. N. S. **Poly (fonte tipográfica)**. [s.l.] Google, 2010.
- SHEN, J. **Lushootseed (fonte tipográfica)**. Seattle: [s.n.].
- SHEN, J. Aesthetic Innovation in Indigenous Typefaces: Designing a Lushootseed Font. **Glimpse**, n. 7, 2010.
- SHEN, J. **Lushootseed: Software and Fonts**Lushootseed, 31 jan. 2018a. Disponível em: <<https://tulaliplushootseed.com/software-and-fonts/>>. Acesso em: 25 ago. 2020
- SHEN, J. **Juliet Shen Lushootseed Indigenous Font Design at the San Francisco Public Library**San Francisco, 21 mar. 2018b. Disponível em: <<https://www.youtube.com/watch?v=l8hf1jk83HE>>. Acesso em: 25 ago. 2020
- SHEN, J. **Indigenous Font Design and the Ripple Effect with Juliet Shen**, 27 mar. 2018c. Disponível em: <<https://vimeo.com/262087396>>. Acesso em: 25 ago. 2020
- TERÁN, D. M. **A bridge to cultural diversity: A typographic exploration of Tuhun Savi, the Mixtec language. Diego Mier y Terán at the 2009 ATypI conference**. 2009 ATypI conference. **Anais...** In: TYP09: THE HEART OF THE LETTER. 2009 ATYPI CONFERENCE. México, DF: 2009.
- TERAN, D. M. Y. **Tuhun (fonte tipográfica)**. [s.l.: s.n.].
- UNNA, J. DE B. Impresiones de la tipografía mexicana. **tipoGráfica**, n. 59, p. 26–31, mar. 2004.
- UTSCH, A.; GRAVIER, M. G. (EDS.). **Red Latinoamericana de Cultura Gráfica. Bibliografía Latinoamericana de Cultura Gráfica - 2018**. Belo Horizonte: Faculdade de Letras, UFMG, 2018.
- UTSCH, A.; GRAVIER, M. G. (EDS.). **Red Latinoamericana de Cultura Gráfica. Bibliografía Latinoamericana de Cultura Gráfica - 2019**. Belo Horizonte: Faculdade de Letras, UFMG, 2019.

Apêndices

1. Formulário e respostas

Nome da instituição à qual você é afiliad(a)o

4. Escreva abaixo o nome da língua indígena *

5. Escreva abaixo a lista dos caracteres acentuados utilizados nesta língua. Se não há uso de acentos na língua, escreva somente "sem uso de acentos". *

6. Escreva aqui o código ISO 639 de três letras para esta língua. Exemplo: Pirahã [myp]. Os códigos ISO 639-3 são usados para representação de nomes de línguas. Você pode consultar os códigos no link a seguir: <https://www.ethnologue.com/browse>

7. No caso de dificuldade/problema técnico para escrever a lista dos caracteres acentuados, utilize o campo abaixo para explicar/descrever quais caracteres são utilizados. Exemplos de respostas: "não consegui digitar o g com til" ou "na língua xx são utilizados ainda os seguintes caracteres: i cortado com acento agudo, u com til e agudo".

Lembre-se: você pode fazer mais de uma contribuição. É só acessar o formulário novamente!

Este conteúdo não foi criado nem aprovado pela Google.

Google Formulários

Nome Completo	Instituição	Escreva abaixo o nome da língua indígena	Escreva abaixo a lista dos caracteres acentuados utilizados nesta língua. Se não há uso de acentos na língua, escreva somente "sem uso de acentos".	No caso de dificuldade/ problema técnico para escrever a lista dos caracteres acentuados, utilize o campo abaixo para explicar/ descrever quais caracteres são utilizados. Exemplos de respostas: "não consegui digitar o g com til" ou "na língua xx são utilizados ainda os seguintes caracteres: i cortado com acento agudo, u com til e agudo".	Escreva aqui o código ISO 639 de três letras para esta língua. Exemplo: Pira-hã [myp]. Os códigos ISO 639-3 são usados para representação de nomes de línguas. Você pode consultar os códigos no link a seguir: https://www.ethnologue.com/browse
teste	teste	teste			
DIONEY MOREIRA GOMES	Universidade de Brasília	Mundurukú	ã õ î û ë ÿ ģ		myu
Eduardo R. Ribeiro	pesquisador independente	Karajá	à è ò ù ã î õ ÿ		kpj
Áurea Cavalcante Santana	Universidade Federal de Mato Grosso	Chiquitano	ã, é, ê, ĩ, ó, ô, ù, û ' (p/ marcar a glotal)		[cax]
Áurea Cavalcante Santana	Universidade Federal de Mato Grosso - UFMT / PPCEL	Wakalitesu (Nambikwara)	ã, ä, ã, e, ê, ĩ, î, ĩ, o, u, û, ü		[nab]
Alan Vogel	Associação Internacional de Linguística - SIL Brasil	Jarawara	SEM USO DE ACENTOS		jaa
Kristine Sue Stenzel	Universidade Federal do Rio de Janeiro	Kotiria (Wanano) e Wa'ikhana (Piratapuyo) ambas da família Tukano Oriental	As ortografias de ambas as línguas são iguais e por isso estão aqui no mesmo formulário: os caracteres acentuados na ortografia são: < ñ ' ã ê ĩ õ û ü >>		Kotiria: gvc, Wa'ikhana: pir
VICENTE CARVALHO DE SOUSA	UNIVERSIDADE FEDERAL DO RIO DE JANEIRO	MĒBĒNGŌKRE/ KAYAPÓ	Ă Ă Ê Ē Ĭ Ō Ő Ū Ÿ Ŷ	NÃO CONSIGO INCLUIR O TIL NAS VOGAIS E, I, O, U. TAMBEM NÃO CONSIGO COLOCAR A GLOTTAL OCLUSIVA NAS VOGAIS DE ACORDO COM O IPHA	TXU 639-3
Sebastian Drude	Museu P. E. Goeldi	Aweti	Ă ä Ē ê Ĭ ĩ Ō ō Ū ū Ÿ ŷ	Em Aweti [awe], a consoante glotal é (idealmente) representada pelo caractere Unicode U+02bc, mas pode ser representada por outros símbolos mais problemáticos: por exemplo ' (U+2019) ou ' (U+0027). As vogais nasais podem ser representadas usando um trema em vez do til se este não for acessível em um certo sistema: Ă ä Ē ê Ĭ ĩ Ō ō Ū ū Ÿ ŷ	awe
Maria Filomena Spatti Sandalo	Unicamp	Kadiwéu	Sem uso de acentos	a oclusiva uvular sonora é representada por um g cortado	Kbc
Wesley Nascimento dos Santos	University of California, Berkeley	Kawahiva (variedades Juma, Karipuna, e Uru Eu Wau Wau)	(vogais): ä, ê, ĩ, ô, ū, ŷ, Ă, Ē, Ĭ, Ō, Ū, Ŷ, (consoante glotal representada com o apóstrofo) '		Juma [jua], Karipuna [kuq], Uru Eu Wau Wau [urz]

Nome Completo	Instituição	Escreva abaixo o nome da língua indígena	Escreva abaixo a lista dos caracteres acentuados utilizados nesta língua. Se não há uso de acentos na língua, escreva somente "sem uso de acentos".	No caso de dificuldade/ problema técnico para escrever a lista dos caracteres acentuados, utilize o campo abaixo para explicar/ descrever quais caracteres são utilizados. Exemplos de respostas: "não consegui digitar o g com til" ou "na língua xx são utilizados ainda os seguintes caracteres: i cortado com acento agudo, u com til e agudo".	Escreva aqui o código ISO 639 de três letras para esta língua. Exemplo: Pihrahã [myp]. Os códigos ISO 639-3 são usados para representação de nomes de línguas. Você pode consultar os códigos no link a seguir: https://www.ethnologue.com/browse
Wilmar da Rocha D'Angelis	UNICAMP	Kaingang / Kan-hgág (na ortografia própria)	á é ó ã î ü ë ÿ Á Ê Ó Ã Ī Ū Ē Ÿ	A ortografia do Kaingang (do Sul) também usa o apóstrofo para representar a consoante oclusiva glotal (em palavras como: pã'î, kó'y etc)	kgp
Wilmar da Rocha D'Angelis	UNICAMP	Kaingang Paulista / Kaingâk (na ortografia própria)	â ã é ê ï ó õ ü Â Ã Ē Ĕ Ī Ō Ū	Na ortografia do Kaingâk paulista se usa também o hífen (-) para representar a consoante oclusiva glotal (como por exemplo: pét-ó)	kgp (não há um código diferente para o dialeto paulista da língua kaingang)
Hein van der Voort	Museu Paraense Emílio Goeldi	Kwazá	ã, ê, êẽ, î, î, õ, õ, ü, ' ,	Não tive problema em Unicode.	xwa
Hein van der Voort	Museu Paraense Emílio Goeldi	Aikanã	ã, ê, î, î, ü, ü, ' ,	Não tem problema graças a Unicode.	tba
Hein van der Voort	Museu Paraense Emílio Goeldi	Arikapú	ã, ä, ä, ê, î, î, õ, õ, ü, ü, ' ,	Não teve problemas graças a Unicode.	ark
Idalina Maciel León	Consulado do Paraguái em Campo Grande MS/ BR	Língua Guarani de Paraguái	Apu'ã: levantar. Aikotevê: necessitar. Amoï: colocar. Aju: vengo. Ñani: correr.	Ĝ	ISO 639. Arrange: data, hoje, dia.
Idalina Maciel	Consulado do Paraguái em Campo Grande MS - BR	Guarani Paraguái	Pu'ã	Ĝuahẽ	ISO 639 aqt
Idalina Maciel León	Consulado Geral do Paraguái em Campo Grande MS/BR	Idioma Guarani Paraguái	Pu'ã	Pu'ã : são utilizado puso entre as vogais u é a, acento nasal na ã.	ISO 639 aqt
Idalina Maciel León	Consulado do Paraguái em Campo Grande MS BR	Idioma Guarani Paraguáio	Angaité	Angaité : ouve reforma ortográfica em 2018, e retiraram o acento da letra "e" na palavra Angaite.	ISO 630-3 aqt
Idalina Maciel León	Consulado Geral do Paraguái em Campo Grande MS/BR	Idioma Guarani Paraguáio	Â, á, Ã, ã, Ô, õ, Û, ü, Ý, ý, Ÿ, ÿ, Ī, î, Ē, ê, Ğ, ğ, Ñ, ñ		ISO 639-3 angaité. aqt. Ouve reforma ortográfica em 2018, retirou o acento agudo da letra "é", Angaite
Idalina Maciel León	Consulado Geral do Paraguái em Campo Grande / MS Brasil	Idioma Guarani Paraguáio	Pu'ã, leva puso é também til nasal na letra ã, são utilizados em todas as vogais til nasais, ã, ê, î, õ, ü, ý, ğ, ñ, e também o acento agudo á, é, í, ó, ú, ý, nunca no final da palavra, inclusive a letra consoante levam til nasais ex. Ğ, ğ	Não tive dificuldade.	ISO 639 - 3 Pu'ã aqt,
Alan Vogel	SIL-Brasil	Banawá	sem uso de acentos		jaa
Alan Vogel	SIL-Brasil	Jamamadi	sem uso de acentos		jaa
Alan Vogel	SIL-Brasil	Paumari	'b 'd		pad
Alan Vogel	SIL-Brasil	Deni			dny

Nome Completo	Instituição	Escreva abaixo o nome da língua indígena	Escreva abaixo a lista dos caracteres acentuados utilizados nesta língua. Se não há uso de acentos na língua, escreva somente "sem uso de acentos".	No caso de dificuldade/ problema técnico para escrever a lista dos caracteres acentuados, utilize o campo abaixo para explicar/ descrever quais caracteres são utilizados. Exemplos de respostas: "não consegui digitar o g com til" ou "na língua xx são utilizados ainda os seguintes caracteres: i cortado com acento agudo, u com til e agudo".	Escreva aqui o código ISO 639 de três letras para esta língua. Exemplo: Pírahã [myp]. Os códigos ISO 639-3 são usados para representação de nomes de línguas. Você pode consultar os códigos no link a seguir: https://www.ethnologue.com/browse
Mário André Coelho da Silva	Unicamp	Maxakalí	Ã, Ê, Ì, Ò, Û, ä, ë, ì, ò, ù, '.	Normalmente, Ê, Ì, Ò, Û, ä, ë, ì, ù são substituídos por Ê, Ì, Ò, Û, ê, í, û nas redes sociais, dada a dificuldade técnica de se digitar o til nessas letras. O uso do apóstrofo é pouco frequente e varia muito de acordo com a pessoa.	[mbl]
ana vilacy galucio	Museu Paraense Emilio Goeldi	Sakurabiat (também conhecida anteriormente como Mekens)	Ã Ö Ì Û Ê ä ö ì ù ë ''		skf
ana vilacy galucio	Museu Paraense Emílio Goeldi	Puruborá	ã õ ì ù ê ÿ â '.		pur
Wilmar da Rocha D'Angelis	UNICAMP	Kaingang	Á, á, Â, â, Ã, ã, Ê, ê, Ë, ë, Ì, ì, Ó, ó, Û, ù, Ȳ, ȳ		kgp
Wilmar da Rocha D'Angelis	UNICAMP	Nhandewa Guarani (ou Tupi-Guarani)	Á, á, Ê, ê, Ó, ó, Â, â, Ë, ë, Ì, ì, Ȳ, ȳ, Û, ù		nhd
Wilmar da Rocha D'Angelis	UNICAMP	Kaingang Paulista	Á, á, Â, â, Ã, ã, Ê, ê, Ë, ë, Ì, ì, Ó, ó, Ò, ò, Õ, õ, Û, ù		zgp
Wilmar da Rocha D'Angelis	UNICAMP	Nheengatu (Nheégatu)	Á, á, Ê, ê, Ë, ë, Ì, ì, Ú, ú, Ȳ, ȳ, Â, â, Ë, ë, Ì, ì, Û, ù		ysl
Wilmar da Rocha D'Angelis	UNICAMP	Ashaninka (Acre, BR)	Ã, ä, Ê, ê, Ì, ì, Ò, ò, Û, ù		ahs
Andrey Nikulin	UnB	Mëbêngôkre	à ÿ ä ÿ ê ë ì ó õ ù		txu
Andrey Nikulin	UnB	Kajkwakhrattxi (Tapayuna)	à ÿ ä ÿ ê ë ì ó õ ù		não há (Glottocode beic1238)
Andrey Nikulin	UnB	Kisêdjê	á ä ä ÿ ê ë ì ó õ ù		suy
Andrey Nikulin	UnB	Apinajé	à ÿ ä ÿ ê ë ì ó õ ù		apn
Andrey Nikulin	UnB	Parkatêjê	à ÿ ä ÿ ê ë ì ó õ ù		gvp
Andrey Nikulin	UnB	Canela & Krahô	à ÿ ä ÿ ê ë ì ó õ ù		ram & xra
Andrey Nikulin	UnB	Krikatí & Gavião-Pykobjê	ÿ ä ÿ ê ö '.		xri
Andrey Nikulin	UnB	Panará	ã ä ê ë ì ó õ ù		kre
Andrey Nikulin	UnB	Akwẽ-Xerente	â û ä ê ë ì ó õ		xer
Andrey Nikulin	UnB	Xavante	ö ä é ê ì ó ö '.	Talvez seja interessante alocar algumas teclas aos dígrafos comuns (ts, dz, nh).	xav
Andrey Nikulin	UnB	Kaingang	á ä ÿ é ê ì ó õ ù '.		kgp
Andrey Nikulin	UnB	Laklãnô (Xokleng)	á ä ÿ é ê ì ó õ ù '.		xok
Andrey Nikulin	UnB	Maxakalí	ã ê ì ó õ ù '.		mbl
Andrey Nikulin	UnB	Karajá (Iny rybè)	à è ò ù ä ÿ ì ò		kpj
Andrey Nikulin	UnB	Rikbaktsa	ã ÿ ê ì ó õ ù		rkb

Nome Completo	Instituição	Escreva abaixo o nome da língua indígena	Escreva abaixo a lista dos caracteres acentuados utilizados nesta língua. Se não há uso de acentos na língua, escreva somente "sem uso de acentos".	No caso de dificuldade/ problema técnico para escrever a lista dos caracteres acentuados, utilize o campo abaixo para explicar/ descrever quais caracteres são utilizados. Exemplos de respostas: "não consegui digitar o g com til" ou "na língua xx são utilizados ainda os seguintes caracteres: i cortado com acento agudo, u com til e agudo".	Escreva aqui o código ISO 639 de três letras para esta língua. Exemplo: Pírahã [myp]. Os códigos ISO 639-3 são usados para representação de nomes de línguas. Você pode consultar os códigos no link a seguir: https://www.ethnologue.com/browse
Andrey Nikulin	UnB	Djeoromitxi (Jabutí)	ã ê ì õ ä ü ps bz tx tx dj pf		jbt
Andrey Nikulin	UnB	Arikapú	ä ÿ ü ä ã ê ì õ ü ' dj tx		ark
Andrey Nikulin	UnB	Chiquitano (Mato Grosso)	ü ä ö ÿ ü é ó ' nh ch tch rch	A lista dos caracteres acentuados acima diz respeito a uma recente convenção adotada em algumas comunidades no Brasil (MT), ver Santana (2012). Na Bolívia, onde a maior parte dos falantes reside, é usado um alfabeto inteiramente diferente, com um conjunto distinto de caracteres especiais (i ñ ' aa ee ii oo uu ch ty xh).	cax
Andrey Nikulin	UnB	Mundurukú	ã õ ÿ ü ë ã'		myu
Andrey Nikulin	UnB	Tuparí	à ò ì ù è ã õ ÿ ü é ' tx		tpr
Andrey Nikulin	UnB	Baniwa-Koripako	ñ tt ph th tth tsh kh mh nh ñh wh rh lh yh		bwi
Mário Ramão Villalva Filho	UNILA - Universidade Federal da Integração Latino-Americana	GUARANI	ã - ë - ì - ò - ü - ÿ - ã - Ä - Ê - Ī - Ō - Ū - Ŷ - Ğ		ISO 639-3
Marcelo Silveira	Universidade Estadual de Londrina	Kaingang	Á, É, Ó, Æ, Ê, Ī, Ū, Ŷ, á, é, ó, æ, ê, ĩ, ū, ŷ, ' ,	um aluno usou o aplicativo da Microsoft para fazer teclados e, em cima do Português ABNT2 inseriu a possibilidade de acentuar ê, ĩ, ū, ŷ, que não são nativos do teclado em português	kgp
Luciana Raccanello Storto	Universidade de São Paulo	Karitiana	ã, ê, ĩ, õ, ŷ, ĵ, ' ,		ktn
Isabella Coutinho Costa	Universidade Estadual de Roraima	Ye'kwana	ä, ö, Ä, Ö, ' ,		Ye'kwana [mch] ISO 639-3
Ana Carla Dos Santos Bruno	INPA (Instituto Nacional de Pesquisas da Amazônia)	Waimiri Atroari	Sem Uso de acentos, mas utiliza o apóstrofo para representar a glotal	Tenho alunos indígenas que escrevem em Tukano que tem dificuldade com o /n/, /u/, /e/ e /i/ com o til e eles têm muita dificuldade. Além do U cortado.	ISO 639-3
Gustavo Godoy	nenhuma	ka'apor	ã, õ, ĩ, ū, ë, ' ' áí	til nas vogais	urb
Stella Telles	Universidade Federal de Pernambuco	Latundê	ã, á, ĩ, ū, ë	não consegui digitar o diacrítico de laringalização (til subscrito) em vogais.	[lnt]
Cristina Martins Fargetti	UNESP	juruna	uso de til em todas as vogais + ĩ ĩ		
Thiago Chacon	Universidade de Brasília	Kubeo	ĩ, í, ê, é, ĩ, í, ô, ó, ū, ú, ä, á		cub

Nome Completo	Instituição	Escreva abaixo o nome da língua indígena	Escreva abaixo a lista dos caracteres acentuados utilizados nesta língua. Se não há uso de acentos na língua, escreva somente "sem uso de acentos".	No caso de dificuldade/ problema técnico para escrever a lista dos caracteres acentuados, utilize o campo abaixo para explicar/ descrever quais caracteres são utilizados. Exemplos de respostas: "não consegui digitar o g com til" ou "na língua xx são utilizados ainda os seguintes caracteres: i cortado com acento agudo, u com til e agudo".	Escreva aqui o código ISO 639 de três letras para esta língua. Exemplo: Pírahã [myp]. Os códigos ISO 639-3 são usados para representação de nomes de línguas. Você pode consultar os códigos no link a seguir: https://www.ethnologue.com/browse
Thiago Chacon	Universidade de Brasília	Arutani	ë		atx
Thiago Chacon	Universidade de Brasília	Ninam	ë, ê, ï, ù, õ, ï, ã		shb
Thiago Chacon	UnB	Tukano	í, î, ï, ú, û, ô, õ, ö, á, â, ã, é, ê, ë, í, î, ï, ú, û, ô, õ, á, â, ã, é, ê, í, î, ï		tuo
Alan Vogel	SIL-Brasil	Aparai	ã e ï õ ü ÿ		apy
Alan Vogel	SIL-Brasil	Apinayé	ã à ê ë ï ó ô ü ÿ ÿ'		apn
Alan Vogel	SIL-Brasil	Apurinã	ã e ï ó		apu
Alan Vogel	SIL-Brasil	Bakairí	ã		bkq
Alan Vogel	SIL-Brasil	Canela	à á ê ë ï ó ô ü ÿ ÿ'		ram
Alan Vogel	SIL-Brasil	Cinta Larga	á á á é é é í í í ï ï ï ó ó ó'		cin
Alan Vogel	SIL-Brasil	Guajajara	à'		gub
Alan Vogel	SIL-Brasil	Borôro	č		bor
Alan Vogel	SIL-Brasil	Hixkaryána	à		hix
Alan Vogel	SIL-Brasil	Hupdë	ã á é ê ï ó ó ü ü'		jup
Alan Vogel	SIL-Brasil	Kadiwéu	g	g com corte horizontal	
Alan Vogel	SIL-Brasil	Kaingang	á á é ê ï ó ü ÿ'		kgp
Alan Vogel	SIL-Brasil	Kaiwá	á á ê ë ï ó ó ü ü ÿ ÿ'		kgk
Alan Vogel	SIL-Brasil	Asuriní do Tocantins	'		asu
Alan Vogel	SIL-Brasil	Guarani Mbyá	ã e ï ó ü ÿ'		gun
Alan Vogel	SIL-Brasil	Kamayurá	ã'		kay
Alan Vogel	SIL-Brasil	Karajá	ã ò õ ÿ		kpj
Alan Vogel	SIL-Brasil	Crioulo Karipúna	ã é é ó ó		kmv
Alan Vogel	SIL-Brasil	Karitiána	ã e ï ó ÿ'		ktn
Alan Vogel	SIL-Brasil	Kayabí	ã e ï ó ü ÿ'		kyz
Alan Vogel	SIL-Brasil	Kayapó	à á ê ë ï ó ó ü ü ÿ ÿ'		txu
Alan Vogel	SIL-Brasil	Maxakalí	ã e ï ó ü'		mbl
Alan Vogel	SIL-Brasil	Mundurukú	ã e ï ó ü ã'		myu
Alan Vogel	SIL-Brasil	Palikur	sem uso de acentos		plu
Alan Vogel	SIL-Brasil	Paresí	á é í ó		pab
Alan Vogel	SIL-Brasil	Rikbaktsa	ã e ï ó ü ÿ		rkb
Alan Vogel	SIL-Brasil	Sateré-Mawé	ã e ï ó ü ÿ'		mav
Alan Vogel	SIL-Brasil	Mamaindé	ã á á é é e í í ï ó ó ü ü ü'		wmd

Nome Completo	Instituição	Escreva abaixo o nome da língua indígena	Escreva abaixo a lista dos caracteres acentuados utilizados nesta língua. Se não há uso de acentos na língua, escreva somente "sem uso de acentos".	No caso de dificuldade/ problema técnico para escrever a lista dos caracteres acentuados, utilize o campo abaixo para explicar/ descrever quais caracteres são utilizados. Exemplos de respostas: "não consegui digitar o g com til" ou "na língua xx são utilizados ainda os seguintes caracteres: i cortado com acento agudo, u com til e agudo".	Escreva aqui o código ISO 639 de três letras para esta língua. Exemplo: Pirahã [myp]. Os códigos ISO 639-3 são usados para representação de nomes de línguas. Você pode consultar os códigos no link a seguir: https://www.ethnologue.com/browse
Alan Vogel	SIL-Brasil	Suruí	á â ã ä ē é ê ë í î ï ó ô õ ú û ü ÿ		sru
Alan Vogel	SIL-Brasil	Suruí do Pará	'		mdz
Alan Vogel	SIL-Brasil	Tenharim	ã ē ī ō ũ ÿ'		pah
Alan Vogel	SIL-Brasil	Terena	á â é ê í î ó ô ú '		ter
Alan Vogel	SIL-Brasil	Kaapor	ã ä ē é ī ō ō ũ ú ŷ '		urb
Alan Vogel	SIL-Brasil	Waurá	á â é ê í î ō ō ũ ŷ		wau
Alan Vogel	SIL-Brasil	Wayampí	ã ē ī ō ũ ŷ'		oym
Rodolfo Senn	SIL	Nadëb	ã é ê ī ó ō ũ ŷ ë ä ä ' (' antes das vogais indica laringalização)		mbj
Alan Vogel	SIL-Brasil	Nambikuára	ã ä ä ē é ê ë í î ō ũ ŷ ŷ 1 2 3	Os números representam tons, e são sobrescritos, e são colocados no fim da sílaba.	nab
Alan Vogel	SIL-Brasil	Xokleng	á â é ê ó ō ũ		xok
Alan Vogel	SIL-Brasil	Xavante	ã ä ä é ê ī ō ō ŷ'		xav

2. Tabela Combinada

Este anexo contém todo o levantamento de dados realizado, com as 239 tabelas combinadas em uma única planilha.

arquivo	Unicode	Glyphname	Character	bloco
[aoc]_001	0027	quotesingle	'	1. Basic Latin
[aoc]_001	00CF	Idieresis	Ï	2. Latin-1 Supplement
[aoc]_001	00D6	Odieresis	Ö	2. Latin-1 Supplement
[aoc]_001	00DC	Udieresis	Ü	2. Latin-1 Supplement
[aoc]_001	00EF	idieresis	ï	2. Latin-1 Supplement
[aoc]_001	00F6	odieresis	ö	2. Latin-1 Supplement
[aoc]_001	00FC	udieresis	ü	2. Latin-1 Supplement
[aoc]_001	2019	quoteright	'	11. General Punctuation
[apn]_001	00C0	Agrave	À	2. Latin-1 Supplement
[apn]_001	00C3	Atilde	Ã	2. Latin-1 Supplement
[apn]_001	00CA	Ecircumflex	Ê	2. Latin-1 Supplement
[apn]_001	00D4	Ocircumflex	Ô	2. Latin-1 Supplement
[apn]_001	00D5	Otilde	Õ	2. Latin-1 Supplement
[apn]_001	00E0	agrave	à	2. Latin-1 Supplement
[apn]_001	00E3	atilde	ã	2. Latin-1 Supplement
[apn]_001	00EA	ecircumflex	ê	2. Latin-1 Supplement
[apn]_001	00F4	ocircumflex	ô	2. Latin-1 Supplement
[apn]_001	00F5	otilde	õ	2. Latin-1 Supplement
[apn]_001	0128	ltilde	Ł	3. Latin Ext.-A
[apn]_001	0129	itilde	ł	3. Latin Ext.-A
[apn]_001	0168	Utilde	Ũ	3. Latin Ext.-A
[apn]_001	0169	utilde	ũ	3. Latin Ext.-A
[apn]_001	1EBC	Etilde	Ě	8. Latin Ext. Add.
[apn]_001	1EBD	etilde	ě	8. Latin Ext. Add.
[apn]_001	1EF2	Ygrave	Ỳ	8. Latin Ext. Add.
[apn]_001	1EF3	ygrave	ỳ	8. Latin Ext. Add.
[apn]_001	1EF8	Ytilde	Ỳ	8. Latin Ext. Add.
[apn]_001	1EF9	ytilde	ỳ	8. Latin Ext. Add.
[apn]_002	00C0	Agrave	À	2. Latin-1 Supplement
[apn]_002	00C3	Atilde	Ã	2. Latin-1 Supplement
[apn]_002	00CA	Ecircumflex	Ê	2. Latin-1 Supplement
[apn]_002	00D4	Ocircumflex	Ô	2. Latin-1 Supplement
[apn]_002	00D5	Otilde	Õ	2. Latin-1 Supplement
[apn]_002	00E0	agrave	à	2. Latin-1 Supplement
[apn]_002	00E3	atilde	ã	2. Latin-1 Supplement
[apn]_002	00EA	ecircumflex	ê	2. Latin-1 Supplement
[apn]_002	00F4	ocircumflex	ô	2. Latin-1 Supplement
[apn]_002	00F5	otilde	õ	2. Latin-1 Supplement
[apn]_002	0128	ltilde	Ł	3. Latin Ext.-A
[apn]_002	0129	itilde	ł	3. Latin Ext.-A
[apn]_002	0168	Utilde	Ũ	3. Latin Ext.-A
[apn]_002	0169	utilde	ũ	3. Latin Ext.-A
[apn]_002	1EBC	Etilde	Ě	8. Latin Ext. Add.
[apn]_002	1EBD	etilde	ě	8. Latin Ext. Add.
[apn]_002	1EF2	Ygrave	Ỳ	8. Latin Ext. Add.
[apn]_002	1EF3	ygrave	ỳ	8. Latin Ext. Add.
[apn]_002	1EF8	Ytilde	Ỳ	8. Latin Ext. Add.

arquivo	Unicode	Glyphname	Character	bloco
[apn]_002	1EF9	ytilde	ÿ	8. Latin Ext. Add.
[apn]_003	00C0	Agrave	À	2. Latin-1 Supplement
[apn]_003	00C3	Atilde	Ã	2. Latin-1 Supplement
[apn]_003	00CA	Ecircumflex	Ê	2. Latin-1 Supplement
[apn]_003	00D4	Ocircumflex	Ô	2. Latin-1 Supplement
[apn]_003	00D5	Otilde	Õ	2. Latin-1 Supplement
[apn]_003	00E0	agrave	à	2. Latin-1 Supplement
[apn]_003	00E3	atilde	ã	2. Latin-1 Supplement
[apn]_003	00EA	ecircumflex	ê	2. Latin-1 Supplement
[apn]_003	00F4	ocircumflex	ô	2. Latin-1 Supplement
[apn]_003	00F5	otilde	õ	2. Latin-1 Supplement
[apn]_003	0128	ltilde	ĺ	3. Latin Ext.-A
[apn]_003	0129	itilde	ı	3. Latin Ext.-A
[apn]_003	0168	Utilde	Ū	3. Latin Ext.-A
[apn]_003	0169	utilde	ū	3. Latin Ext.-A
[apn]_003	1EBC	Etilde	Ě	8. Latin Ext. Add.
[apn]_003	1EBD	etilde	ě	8. Latin Ext. Add.
[apn]_003	1EF2	Ygrave	Ỳ	8. Latin Ext. Add.
[apn]_003	1EF3	ygrave	ỳ	8. Latin Ext. Add.
[apn]_003	1EF8	Ytilde	Ỳ	8. Latin Ext. Add.
[apn]_003	1EF9	ytilde	ỳ	8. Latin Ext. Add.
[apn]_004	00C0	Agrave	À	2. Latin-1 Supplement
[apn]_004	00C3	Atilde	Ã	2. Latin-1 Supplement
[apn]_004	00CA	Ecircumflex	Ê	2. Latin-1 Supplement
[apn]_004	00D4	Ocircumflex	Ô	2. Latin-1 Supplement
[apn]_004	00D5	Otilde	Õ	2. Latin-1 Supplement
[apn]_004	00E0	agrave	à	2. Latin-1 Supplement
[apn]_004	00E3	atilde	ã	2. Latin-1 Supplement
[apn]_004	00EA	ecircumflex	ê	2. Latin-1 Supplement
[apn]_004	00F4	ocircumflex	ô	2. Latin-1 Supplement
[apn]_004	00F5	otilde	õ	2. Latin-1 Supplement
[apn]_004	0128	ltilde	ĺ	3. Latin Ext.-A
[apn]_004	0129	itilde	ı	3. Latin Ext.-A
[apn]_004	0168	Utilde	Ū	3. Latin Ext.-A
[apn]_004	0169	utilde	ū	3. Latin Ext.-A
[apn]_004	1EBC	Etilde	Ě	8. Latin Ext. Add.
[apn]_004	1EBD	etilde	ě	8. Latin Ext. Add.
[apn]_004	1EF2	Ygrave	Ỳ	8. Latin Ext. Add.
[apn]_004	1EF3	ygrave	ỳ	8. Latin Ext. Add.
[apn]_004	1EF8	Ytilde	Ỳ	8. Latin Ext. Add.
[apn]_004	1EF9	ytilde	ỳ	8. Latin Ext. Add.
[apn]_005	00C0	Agrave	À	2. Latin-1 Supplement
[apn]_005	00C3	Atilde	Ã	2. Latin-1 Supplement
[apn]_005	00CA	Ecircumflex	Ê	2. Latin-1 Supplement
[apn]_005	00D4	Ocircumflex	Ô	2. Latin-1 Supplement
[apn]_005	00D5	Otilde	Õ	2. Latin-1 Supplement
[apn]_005	00E0	agrave	à	2. Latin-1 Supplement

arquivo	Unicode	Glyphname	Character	bloco
[apn]_005	00E3	atilde	ã	2. Latin-1 Supplement
[apn]_005	00EA	ecircumflex	ê	2. Latin-1 Supplement
[apn]_005	00F4	ocircumflex	ô	2. Latin-1 Supplement
[apn]_005	00F5	otilde	õ	2. Latin-1 Supplement
[apn]_005	0128	ltilde	ĺ	3. Latin Ext.-A
[apn]_005	0129	itilde	ı̇	3. Latin Ext.-A
[apn]_005	0168	Utilde	Ū	3. Latin Ext.-A
[apn]_005	0169	utilde	ū	3. Latin Ext.-A
[apn]_005	1EBC	Etilde	Ě	8. Latin Ext. Add.
[apn]_005	1EBD	etilde	ě	8. Latin Ext. Add.
[apn]_005	1EF2	Ygrave	Ỳ	8. Latin Ext. Add.
[apn]_005	1EF3	ygrave	ỳ	8. Latin Ext. Add.
[apn]_005	1EF8	Ytilde	Ỳ̇	8. Latin Ext. Add.
[apn]_005	1EF9	ytilde	ỳ̇	8. Latin Ext. Add.
[apn]_006	00C0	Agrave	À	2. Latin-1 Supplement
[apn]_006	00C3	Atilde	Ã	2. Latin-1 Supplement
[apn]_006	00CA	Ecircumflex	Ê	2. Latin-1 Supplement
[apn]_006	00D4	Ocircumflex	Ô	2. Latin-1 Supplement
[apn]_006	00D5	Otilde	Õ	2. Latin-1 Supplement
[apn]_006	00E0	agrave	à	2. Latin-1 Supplement
[apn]_006	00E3	atilde	ã	2. Latin-1 Supplement
[apn]_006	00EA	ecircumflex	ê	2. Latin-1 Supplement
[apn]_006	00F4	ocircumflex	ô	2. Latin-1 Supplement
[apn]_006	00F5	otilde	õ	2. Latin-1 Supplement
[apn]_006	0128	ltilde	ĺ	3. Latin Ext.-A
[apn]_006	0129	itilde	ı̇	3. Latin Ext.-A
[apn]_006	0168	Utilde	Ū	3. Latin Ext.-A
[apn]_006	0169	utilde	ū	3. Latin Ext.-A
[apn]_006	1EBC	Etilde	Ě	8. Latin Ext. Add.
[apn]_006	1EBD	etilde	ě	8. Latin Ext. Add.
[apn]_006	1EF2	Ygrave	Ỳ	8. Latin Ext. Add.
[apn]_006	1EF3	ygrave	ỳ	8. Latin Ext. Add.
[apn]_006	1EF8	Ytilde	Ỳ̇	8. Latin Ext. Add.
[apn]_006	1EF9	ytilde	ỳ̇	8. Latin Ext. Add.
[apn]_007	00C0	Agrave	À	2. Latin-1 Supplement
[apn]_007	00C3	Atilde	Ã	2. Latin-1 Supplement
[apn]_007	00CA	Ecircumflex	Ê	2. Latin-1 Supplement
[apn]_007	00D4	Ocircumflex	Ô	2. Latin-1 Supplement
[apn]_007	00D5	Otilde	Õ	2. Latin-1 Supplement
[apn]_007	00E0	agrave	à	2. Latin-1 Supplement
[apn]_007	00E3	atilde	ã	2. Latin-1 Supplement
[apn]_007	00EA	ecircumflex	ê	2. Latin-1 Supplement
[apn]_007	00F4	ocircumflex	ô	2. Latin-1 Supplement
[apn]_007	00F5	otilde	õ	2. Latin-1 Supplement
[apn]_007	0128	ltilde	ĺ	3. Latin Ext.-A
[apn]_007	0129	itilde	ı̇	3. Latin Ext.-A
[apn]_007	0168	Utilde	Ū	3. Latin Ext.-A

arquivo	Unicode	Glyphname	Character	bloco
[apn]_007	0169	utilde	ũ	3. Latin Ext.-A
[apn]_007	1EBC	Etilde	Ě	8. Latin Ext. Add.
[apn]_007	1EBD	etilde	ě	8. Latin Ext. Add.
[apn]_007	1EF2	Ygrave	Ỳ	8. Latin Ext. Add.
[apn]_007	1EF3	ygrave	ỳ	8. Latin Ext. Add.
[apn]_007	1EF8	Ytilde	Ỳ	8. Latin Ext. Add.
[apn]_007	1EF9	ytilde	ỳ	8. Latin Ext. Add.
[apn]_008	00C0	Agrave	À	2. Latin-1 Supplement
[apn]_008	00C3	Atilde	Ã	2. Latin-1 Supplement
[apn]_008	00CA	Ecircumflex	Ê	2. Latin-1 Supplement
[apn]_008	00D4	Ocircumflex	Ô	2. Latin-1 Supplement
[apn]_008	00D5	Otilde	Õ	2. Latin-1 Supplement
[apn]_008	00E0	agrave	à	2. Latin-1 Supplement
[apn]_008	00E3	atilde	ã	2. Latin-1 Supplement
[apn]_008	00EA	ecircumflex	ê	2. Latin-1 Supplement
[apn]_008	00F4	ocircumflex	ô	2. Latin-1 Supplement
[apn]_008	00F5	otilde	õ	2. Latin-1 Supplement
[apn]_008	0128	Itilde	Ī	3. Latin Ext.-A
[apn]_008	0129	itilde	ī	3. Latin Ext.-A
[apn]_008	0168	Utilde	Ū	3. Latin Ext.-A
[apn]_008	0169	utilde	ũ	3. Latin Ext.-A
[apn]_008	1EBC	Etilde	Ě	8. Latin Ext. Add.
[apn]_008	1EBD	etilde	ě	8. Latin Ext. Add.
[apn]_008	1EF2	Ygrave	Ỳ	8. Latin Ext. Add.
[apn]_008	1EF3	ygrave	ỳ	8. Latin Ext. Add.
[apn]_008	1EF8	Ytilde	Ỳ	8. Latin Ext. Add.
[apn]_008	1EF9	ytilde	ỳ	8. Latin Ext. Add.
[apn]_009	00C0	Agrave	À	2. Latin-1 Supplement
[apn]_009	00C3	Atilde	Ã	2. Latin-1 Supplement
[apn]_009	00CA	Ecircumflex	Ê	2. Latin-1 Supplement
[apn]_009	00D4	Ocircumflex	Ô	2. Latin-1 Supplement
[apn]_009	00D5	Otilde	Õ	2. Latin-1 Supplement
[apn]_009	00E0	agrave	à	2. Latin-1 Supplement
[apn]_009	00E3	atilde	ã	2. Latin-1 Supplement
[apn]_009	00EA	ecircumflex	ê	2. Latin-1 Supplement
[apn]_009	00F4	ocircumflex	ô	2. Latin-1 Supplement
[apn]_009	00F5	otilde	õ	2. Latin-1 Supplement
[apn]_009	0128	Itilde	Ī	3. Latin Ext.-A
[apn]_009	0129	itilde	ī	3. Latin Ext.-A
[apn]_009	0168	Utilde	Ū	3. Latin Ext.-A
[apn]_009	0169	utilde	ũ	3. Latin Ext.-A
[apn]_009	1EBC	Etilde	Ě	8. Latin Ext. Add.
[apn]_009	1EBD	etilde	ě	8. Latin Ext. Add.
[apn]_009	1EF2	Ygrave	Ỳ	8. Latin Ext. Add.
[apn]_009	1EF3	ygrave	ỳ	8. Latin Ext. Add.
[apn]_009	1EF8	Ytilde	Ỳ	8. Latin Ext. Add.
[apn]_009	1EF9	ytilde	ỳ	8. Latin Ext. Add.

arquivo	Unicode	Glyphname	Character	bloco
[apn]_010	0027	quotesingle	'	1. Basic Latin
[apn]_010	00C0	Agrave	À	2. Latin-1 Supplement
[apn]_010	00C3	Atilde	Ã	2. Latin-1 Supplement
[apn]_010	00CA	Ecircumflex	Ê	2. Latin-1 Supplement
[apn]_010	00D4	Ocircumflex	Ô	2. Latin-1 Supplement
[apn]_010	00D5	Otilde	Õ	2. Latin-1 Supplement
[apn]_010	00E0	agrave	à	2. Latin-1 Supplement
[apn]_010	00E3	atilde	ã	2. Latin-1 Supplement
[apn]_010	00EA	ecircumflex	ê	2. Latin-1 Supplement
[apn]_010	00F4	ocircumflex	ô	2. Latin-1 Supplement
[apn]_010	00F5	otilde	õ	2. Latin-1 Supplement
[apn]_010	0128	ltilde	İ	3. Latin Ext.-A
[apn]_010	0129	itilde	ı	3. Latin Ext.-A
[apn]_010	0168	Utilde	Ū	3. Latin Ext.-A
[apn]_010	0169	utilde	ū	3. Latin Ext.-A
[apn]_010	1EBC	Etilde	Ě	8. Latin Ext. Add.
[apn]_010	1EBD	etilde	ě	8. Latin Ext. Add.
[apn]_010	1EF2	Ygrave	Ỳ	8. Latin Ext. Add.
[apn]_010	1EF3	ygrave	ỳ	8. Latin Ext. Add.
[apn]_010	1EF8	Ytilde	Ỳ	8. Latin Ext. Add.
[apn]_010	1EF9	ytilde	ỳ	8. Latin Ext. Add.
[apu]_001	00C3	Atilde	Ã	2. Latin-1 Supplement
[apu]_001	00D5	Otilde	Õ	2. Latin-1 Supplement
[apu]_001	00E3	atilde	ã	2. Latin-1 Supplement
[apu]_001	00F5	otilde	õ	2. Latin-1 Supplement
[apu]_001	0128	ltilde	İ	3. Latin Ext.-A
[apu]_001	0129	itilde	ı	3. Latin Ext.-A
[apu]_001	1EBC	Etilde	Ě	8. Latin Ext. Add.
[apu]_001	1EBD	etilde	ě	8. Latin Ext. Add.
[apy]_001	00C3	Atilde	Ã	2. Latin-1 Supplement
[apy]_001	00D5	Otilde	Õ	2. Latin-1 Supplement
[apy]_001	00E3	atilde	ã	2. Latin-1 Supplement
[apy]_001	00F5	otilde	õ	2. Latin-1 Supplement
[apy]_001	0128	ltilde	İ	3. Latin Ext.-A
[apy]_001	0129	itilde	ı	3. Latin Ext.-A
[apy]_001	0168	Utilde	Ū	3. Latin Ext.-A
[apy]_001	0169	utilde	ū	3. Latin Ext.-A
[apy]_001	1EBC	Etilde	Ě	8. Latin Ext. Add.
[apy]_001	1EBD	etilde	ě	8. Latin Ext. Add.
[apy]_001	1EF8	Ytilde	Ỳ	8. Latin Ext. Add.
[apy]_001	1EF9	ytilde	ỳ	8. Latin Ext. Add.
[ark]_001	0027	quotesingle	'	1. Basic Latin
[ark]_001	00C3	Atilde	Ã	2. Latin-1 Supplement
[ark]_001	00C4	Adieresis	Ä	2. Latin-1 Supplement
[ark]_001	00CF	Idieresis	İ	2. Latin-1 Supplement
[ark]_001	00D5	Otilde	Õ	2. Latin-1 Supplement
[ark]_001	00DC	Udieresis	Ü	2. Latin-1 Supplement

arquivo	Unicode	Glyphname	Character	bloco
[ark]_001	00E3	atilde	ã	2. Latin-1 Supplement
[ark]_001	00E4	adieresis	ä	2. Latin-1 Supplement
[ark]_001	00EF	idieresis	ï	2. Latin-1 Supplement
[ark]_001	00F5	otilde	õ	2. Latin-1 Supplement
[ark]_001	00FC	udieresis	ü	2. Latin-1 Supplement
[ark]_001	0128	ltilde	ĩ	3. Latin Ext.-A
[ark]_001	0129	itilde	ř	3. Latin Ext.-A
[ark]_001	0168	Utilde	Ū	3. Latin Ext.-A
[ark]_001	0169	utilde	ů	3. Latin Ext.-A
[ark]_001	0303	tildecomb	˜	7. Combining Marks
[ark]_001	1EBC	Etilde	Ě	8. Latin Ext. Add.
[ark]_001	1EBD	etilde	ě	8. Latin Ext. Add.
[ark]_001	00C4_0303	Adieresis_tildecomb	Ä̃	10. non-compliant
[ark]_001	00E4_0303	adieresis_tildecomb	ä̃	10. non-compliant
[ark]_002	0027	quotesingle	'	1. Basic Latin
[ark]_002	00C3	Atilde	Ā	2. Latin-1 Supplement
[ark]_002	00C4	Adieresis	Ă	2. Latin-1 Supplement
[ark]_002	00CF	Idieresis	İ	2. Latin-1 Supplement
[ark]_002	00D5	Otilde	Ō	2. Latin-1 Supplement
[ark]_002	00DC	Udieresis	Ū	2. Latin-1 Supplement
[ark]_002	00E3	atilde	ã	2. Latin-1 Supplement
[ark]_002	00E4	adieresis	ä	2. Latin-1 Supplement
[ark]_002	00EF	idieresis	ï	2. Latin-1 Supplement
[ark]_002	00F5	otilde	õ	2. Latin-1 Supplement
[ark]_002	00FC	udieresis	ü	2. Latin-1 Supplement
[ark]_002	0128	ltilde	ĩ	3. Latin Ext.-A
[ark]_002	0129	itilde	ř	3. Latin Ext.-A
[ark]_002	0168	Utilde	Ū	3. Latin Ext.-A
[ark]_002	0169	utilde	ů	3. Latin Ext.-A
[ark]_002	0303	tildecomb	˜	7. Combining Marks
[ark]_002	1EBC	Etilde	Ě	8. Latin Ext. Add.
[ark]_002	1EBD	etilde	ě	8. Latin Ext. Add.
[ark]_002	00C4_0303	Adieresis_tildecomb	Ä̃	10. non-compliant
[ark]_002	00E4_0303	adieresis_tildecomb	ä̃	10. non-compliant
[ark]_002	2019	quoteright	'	11. General Punctuation
[arr]_001	00C1	Aacute	Ā	2. Latin-1 Supplement
[arr]_001	00C2	Acircumflex	Ă	2. Latin-1 Supplement
[arr]_001	00C3	Atilde	Ā	2. Latin-1 Supplement
[arr]_001	00C9	Eacute	É	2. Latin-1 Supplement
[arr]_001	00CD	Iacute	Í	2. Latin-1 Supplement
[arr]_001	00D3	Oacute	Ó	2. Latin-1 Supplement
[arr]_001	00D5	Otilde	Ō	2. Latin-1 Supplement
[arr]_001	00DB	Ucircumflex	Ū	2. Latin-1 Supplement
[arr]_001	00E1	aacute	á	2. Latin-1 Supplement
[arr]_001	00E2	acircumflex	â	2. Latin-1 Supplement
[arr]_001	00E3	atilde	ã	2. Latin-1 Supplement
[arr]_001	00E9	eacute	é	2. Latin-1 Supplement

arquivo	Unicode	Glyphname	Character	bloco
[arr]_001	00ED	iacute	í	2. Latin-1 Supplement
[arr]_001	00F3	oacute	ó	2. Latin-1 Supplement
[arr]_001	00F5	otilde	õ	2. Latin-1 Supplement
[arr]_001	00FB	ucircumflex	û	2. Latin-1 Supplement
[arr]_001	0128	ltilde	ĩ	3. Latin Ext.-A
[arr]_001	0129	itilde	ı̃	3. Latin Ext.-A
[arr]_001	02BC	apostrophemod	'	6. Spc. Mod. Letters
[arr]_001	1EBC	Etilde	Ě	8. Latin Ext. Add.
[arr]_001	1EBD	etilde	ě	8. Latin Ext. Add.
[arr]_001	A78B	Saltillo	ı	9. Latin Ext.-D
[arr]_001	A78C	salttillo	ı	9. Latin Ext.-D
[arr]_001	0047_0303	G_tildecomb	Ĝ	10. non-compliant
[arr]_001	0067_0303	g_tildecomb	ĝ	10. non-compliant
[arr]_001	00C3_0301	Atilde_acutecomb	Ă	10. non-compliant
[arr]_001	00E3_0301	atilde_acutecomb	ă	10. non-compliant
[arr]_001	1EBC_0301	Etilde_acutecomb	Ě	10. non-compliant
[arr]_001	1EBD_0301	etilde_acutecomb	ě	10. non-compliant
[arr]_001	0128_0301	ltilde_acutecomb	ĩ	10. non-compliant
[arr]_001	0129_0301	itilde_acutecomb	ı̃	10. non-compliant
[arr]_001	00D5_0301	Otilde_acutecomb	Ŏ	10. non-compliant
[arr]_001	00F5_0301	otilde_acutecomb	õ	10. non-compliant
[arr]_002	00C1	Aacute	Á	2. Latin-1 Supplement
[arr]_002	00C2	Acircumflex	Â	2. Latin-1 Supplement
[arr]_002	00C3	Atilde	Ã	2. Latin-1 Supplement
[arr]_002	00C9	Eacute	É	2. Latin-1 Supplement
[arr]_002	00CD	Iacute	Í	2. Latin-1 Supplement
[arr]_002	00D3	Oacute	Ó	2. Latin-1 Supplement
[arr]_002	00D5	Otilde	Õ	2. Latin-1 Supplement
[arr]_002	00DA	Uacute	Ú	2. Latin-1 Supplement
[arr]_002	00DB	Ucircumflex	Û	2. Latin-1 Supplement
[arr]_002	00E1	aacute	á	2. Latin-1 Supplement
[arr]_002	00E2	acircumflex	â	2. Latin-1 Supplement
[arr]_002	00E3	atilde	ã	2. Latin-1 Supplement
[arr]_002	00E9	eacute	é	2. Latin-1 Supplement
[arr]_002	00ED	iacute	í	2. Latin-1 Supplement
[arr]_002	00F3	oacute	ó	2. Latin-1 Supplement
[arr]_002	00F5	otilde	õ	2. Latin-1 Supplement
[arr]_002	00FA	uacute	ú	2. Latin-1 Supplement
[arr]_002	00FB	ucircumflex	û	2. Latin-1 Supplement
[arr]_002	0128	ltilde	ĩ	3. Latin Ext.-A
[arr]_002	0129	itilde	ı̃	3. Latin Ext.-A
[arr]_002	02BC	apostrophemod	'	6. Spc. Mod. Letters
[arr]_002	1EBC	Etilde	Ě	8. Latin Ext. Add.
[arr]_002	1EBD	etilde	ě	8. Latin Ext. Add.
[arr]_002	A78B	Saltillo	ı	9. Latin Ext.-D
[arr]_002	A78C	salttillo	ı	9. Latin Ext.-D

arquivo	Unicode	Glyphname	Character	bloco
[arr]_002	0047_0303	G_tildecomb	Ĝ	10. non-compliant
[arr]_002	0067_0303	g_tildecomb	ĝ	10. non-compliant
[arr]_002	00C3_0301	Atilde_acutecomb	Ă	10. non-compliant
[arr]_002	00E3_0301	atilde_acutecomb	ă	10. non-compliant
[arr]_002	1EBC_0301	Etilde_acutecomb	Ě	10. non-compliant
[arr]_002	1EBD_0301	etilde_acutecomb	ě	10. non-compliant
[arr]_002	0128_0301	ltilde_acutecomb	Ĭ	10. non-compliant
[arr]_002	0129_0301	itilde_acutecomb	ĭ	10. non-compliant
[arr]_002	00D5_0301	Otilde_acutecomb	Ŏ	10. non-compliant
[arr]_002	00F5_0301	otilde_acutecomb	ŏ	10. non-compliant
[arr]_002	00C2_0301	Acircumflex_acutecomb	Ā	10. non-compliant
[arr]_002	00E2_0301	acircumflex_acutecomb	ā	10. non-compliant
[asu]_001	0027	quotesingle	'	1. Basic Latin
[atr]_001	02BC	apostrophemod	'	6. Spc. Mod. Letters
[atx]_001	00C1	Aacute	Á	2. Latin-1 Supplement
[atx]_001	00C3	Atilde	Ã	2. Latin-1 Supplement
[atx]_001	00D1	Ntilde	Ñ	2. Latin-1 Supplement
[atx]_001	00E1	aacute	á	2. Latin-1 Supplement
[atx]_001	00E3	atilde	ã	2. Latin-1 Supplement
[atx]_001	00F1	ntilde	ñ	2. Latin-1 Supplement
[atx]_001	0128	ltilde	Ĭ	3. Latin Ext.-A
[atx]_001	0129	itilde	ĭ	3. Latin Ext.-A
[atx]_002	00CB	Edieresis	Ë	2. Latin-1 Supplement
[atx]_002	00EB	edieresis	ë	2. Latin-1 Supplement
[awe]_001	00C1	Aacute	Á	2. Latin-1 Supplement
[awe]_001	00C3	Atilde	Ã	2. Latin-1 Supplement
[awe]_001	00D5	Otilde	Ŏ	2. Latin-1 Supplement
[awe]_001	00E1	aacute	á	2. Latin-1 Supplement
[awe]_001	00E3	atilde	ã	2. Latin-1 Supplement
[awe]_001	00F5	otilde	ŏ	2. Latin-1 Supplement
[awe]_001	0128	ltilde	Ĭ	3. Latin Ext.-A
[awe]_001	0129	itilde	ĭ	3. Latin Ext.-A
[awe]_001	0168	Utilde	Û	3. Latin Ext.-A
[awe]_001	0169	utilde	ü	3. Latin Ext.-A
[awe]_001	02BC	apostrophemod	'	6. Spc. Mod. Letters
[awe]_001	1EBC	Etilde	Ě	8. Latin Ext. Add.
[awe]_001	1EBD	etilde	ě	8. Latin Ext. Add.
[awe]_001	1EF8	Ytilde	Ÿ	8. Latin Ext. Add.
[awe]_001	1EF9	ytilde	ÿ	8. Latin Ext. Add.
[awe]_001	A78B	Saltillo	ı	9. Latin Ext.-D
[awe]_001	A78C	salttillo	ı	9. Latin Ext.-D
[awe]_002	00C3	Atilde	Ã	2. Latin-1 Supplement
[awe]_002	00D5	Otilde	Ŏ	2. Latin-1 Supplement
[awe]_002	00E3	atilde	ã	2. Latin-1 Supplement
[awe]_002	00F5	otilde	ŏ	2. Latin-1 Supplement
[awe]_002	0128	ltilde	Ĭ	3. Latin Ext.-A

arquivo	Unicode	Glyphname	Character	bloco
[awe]_002	0129	itilde	ĩ	3. Latin Ext.-A
[awe]_002	0168	Utilde	Ū	3. Latin Ext.-A
[awe]_002	0169	utilde	ū	3. Latin Ext.-A
[awe]_002	02BC	apostrophemod	'	6. Spc. Mod. Letters
[awe]_002	1EBC	Etilde	Ě	8. Latin Ext. Add.
[awe]_002	1EBD	etilde	ě	8. Latin Ext. Add.
[awe]_002	1EF8	Ytilde	Ÿ	8. Latin Ext. Add.
[awe]_002	1EF9	ytilde	ÿ	8. Latin Ext. Add.
[awe]_002	A78B	Saltillo	ı	9. Latin Ext.-D
[awe]_002	A78C	salttillo	ı	9. Latin Ext.-D
[awe]_003	00C3	Atilde	Ã	2. Latin-1 Supplement
[awe]_003	00D5	Otilde	Õ	2. Latin-1 Supplement
[awe]_003	00E3	atilde	ã	2. Latin-1 Supplement
[awe]_003	00F5	otilde	õ	2. Latin-1 Supplement
[awe]_003	0128	ltilde	Ĺ	3. Latin Ext.-A
[awe]_003	0129	itilde	ĩ	3. Latin Ext.-A
[awe]_003	0168	Utilde	Ū	3. Latin Ext.-A
[awe]_003	0169	utilde	ū	3. Latin Ext.-A
[awe]_003	02BC	apostrophemod	'	6. Spc. Mod. Letters
[awe]_003	1EBC	Etilde	Ě	8. Latin Ext. Add.
[awe]_003	1EBD	etilde	ě	8. Latin Ext. Add.
[awe]_003	1EF8	Ytilde	Ÿ	8. Latin Ext. Add.
[awe]_003	1EF9	ytilde	ÿ	8. Latin Ext. Add.
[beic1238]_001	00C0	Agrave	À	2. Latin-1 Supplement
[beic1238]_001	00C3	Atilde	Ã	2. Latin-1 Supplement
[beic1238]_001	00CA	Ecircumflex	Ê	2. Latin-1 Supplement
[beic1238]_001	00D4	Ocircumflex	Ô	2. Latin-1 Supplement
[beic1238]_001	00D5	Otilde	Õ	2. Latin-1 Supplement
[beic1238]_001	00E0	agrave	à	2. Latin-1 Supplement
[beic1238]_001	00E3	atilde	ã	2. Latin-1 Supplement
[beic1238]_001	00EA	ecircumflex	ê	2. Latin-1 Supplement
[beic1238]_001	00F4	ocircumflex	ô	2. Latin-1 Supplement
[beic1238]_001	00F5	otilde	õ	2. Latin-1 Supplement
[beic1238]_001	0128	ltilde	Ĺ	3. Latin Ext.-A
[beic1238]_001	0129	itilde	ĩ	3. Latin Ext.-A
[beic1238]_001	0168	Utilde	Ū	3. Latin Ext.-A
[beic1238]_001	0169	utilde	ū	3. Latin Ext.-A
[beic1238]_001	1EBC	Etilde	Ě	8. Latin Ext. Add.
[beic1238]_001	1EBD	etilde	ě	8. Latin Ext. Add.
[beic1238]_001	1EF2	Ygrave	Ỳ	8. Latin Ext. Add.
[beic1238]_001	1EF3	ygrave	ỳ	8. Latin Ext. Add.
[beic1238]_001	1EF8	Ytilde	Ÿ	8. Latin Ext. Add.
[beic1238]_001	1EF9	ytilde	ÿ	8. Latin Ext. Add.
[bkq]_001	00C2	Acircumflex	Â	2. Latin-1 Supplement
[bkq]_001	00E2	acircumflex	â	2. Latin-1 Supplement
[bkq]_002	00C2	Acircumflex	Â	2. Latin-1 Supplement

arquivo	Unicode	Glyphname	Character	bloco
[bkq]_002	00E2	acircumflex	â	2. Latin-1 Supplement
[bkq]_003	00C2	Acircumflex	Â	2. Latin-1 Supplement
[bkq]_003	00E2	acircumflex	â	2. Latin-1 Supplement
[bkq]_004	00C2	Acircumflex	Â	2. Latin-1 Supplement
[bkq]_004	00E2	acircumflex	â	2. Latin-1 Supplement
[bkq]_005	00C2	Acircumflex	Â	2. Latin-1 Supplement
[bkq]_005	00E2	acircumflex	â	2. Latin-1 Supplement
[bor]_001	010C	Ccaron	Č	3. Latin Ext.-A
[bor]_001	010D	ccaron	č	3. Latin Ext.-A
[bor]_002	0108	Ccircumflex	Ĉ	3. Latin Ext.-A
[bor]_002	0109	ccircumflex	ĉ	3. Latin Ext.-A
[bor]_004	010C	Ccaron	Č	3. Latin Ext.-A
[bor]_004	010D	ccaron	č	3. Latin Ext.-A
[bwi]_002	00D1	Ntilde	Ñ	2. Latin-1 Supplement
[bwi]_002	00F1	ntilde	ñ	2. Latin-1 Supplement
[cax]_001	0027	quotesingle	'	1. Basic Latin
[cax]_001	00C3	Atilde	Ã	2. Latin-1 Supplement
[cax]_001	00C9	Eacute	É	2. Latin-1 Supplement
[cax]_001	00D3	Oacute	Ó	2. Latin-1 Supplement
[cax]_001	00D5	Otilde	Õ	2. Latin-1 Supplement
[cax]_001	00DC	Udieresis	Ü	2. Latin-1 Supplement
[cax]_001	00E3	atilde	ã	2. Latin-1 Supplement
[cax]_001	00E9	eacute	é	2. Latin-1 Supplement
[cax]_001	00F3	oacute	ó	2. Latin-1 Supplement
[cax]_001	00F5	otilde	õ	2. Latin-1 Supplement
[cax]_001	00FC	udieresis	ü	2. Latin-1 Supplement
[cax]_001	0128	ltilde	İ	3. Latin Ext.-A
[cax]_001	0129	itilde	ı	3. Latin Ext.-A
[cax]_001	0168	Utilde	Ū	3. Latin Ext.-A
[cax]_001	0169	utilde	ū	3. Latin Ext.-A
[cax]_001	1EBC	Etilde	Ě	8. Latin Ext. Add.
[cax]_001	1EBD	etilde	ě	8. Latin Ext. Add.
[cax]_001	2018	quoteleft	‘	11. General Punctuation
[cax]_002	0027	quotesingle	'	1. Basic Latin
[cax]_002	00C3	Atilde	Ã	2. Latin-1 Supplement
[cax]_002	00C9	Eacute	É	2. Latin-1 Supplement
[cax]_002	00D3	Oacute	Ó	2. Latin-1 Supplement
[cax]_002	00D5	Otilde	Õ	2. Latin-1 Supplement
[cax]_002	00DC	Udieresis	Ü	2. Latin-1 Supplement
[cax]_002	00E3	atilde	ã	2. Latin-1 Supplement
[cax]_002	00E9	eacute	é	2. Latin-1 Supplement
[cax]_002	00F3	oacute	ó	2. Latin-1 Supplement
[cax]_002	00F5	otilde	õ	2. Latin-1 Supplement
[cax]_002	00FC	udieresis	ü	2. Latin-1 Supplement
[cax]_002	0128	ltilde	İ	3. Latin Ext.-A
[cax]_002	0129	itilde	ı	3. Latin Ext.-A
[cax]_002	0168	Utilde	Ū	3. Latin Ext.-A

arquivo	Unicode	Glyphname	Character	bloco
[cax]_002	0169	utilde	ũ	3. Latin Ext.-A
[cax]_002	2019	quoteright	'	11. General Punctuation
[cax]_003	0027	quotesingle	'	1. Basic Latin
[cax]_003	00D1	Ntilde	Ñ	2. Latin-1 Supplement
[cax]_003	00F1	ntilde	ñ	2. Latin-1 Supplement
[cax]_003	0197	lstroke	‡	4. Latin Ext.-B
[cax]_003	0268	istroke	‡	5. IPA Extensions
[cax]_003	2019	quoteright	'	11. General Punctuation
[cbs]_001	00C1	Aacute	Á	2. Latin-1 Supplement
[cbs]_001	00C3	Atilde	Ã	2. Latin-1 Supplement
[cbs]_001	00E1	aacute	á	2. Latin-1 Supplement
[cbs]_001	00E3	atilde	ã	2. Latin-1 Supplement
[cbs]_001	0128	ltilde	İ	3. Latin Ext.-A
[cbs]_001	0129	itilde	ı	3. Latin Ext.-A
[cbs]_001	0168	Utilde	Û	3. Latin Ext.-A
[cbs]_001	0169	utilde	ü	3. Latin Ext.-A
[cbs]_001	1EBC	Etilde	Ě	8. Latin Ext. Add.
[cbs]_001	1EBD	etilde	ě	8. Latin Ext. Add.
[cbs]_002	00C3	Atilde	Ã	2. Latin-1 Supplement
[cbs]_002	00E3	atilde	ã	2. Latin-1 Supplement
[cbs]_002	0128	ltilde	İ	3. Latin Ext.-A
[cbs]_002	0129	itilde	ı	3. Latin Ext.-A
[cbs]_002	0168	Utilde	Û	3. Latin Ext.-A
[cbs]_002	0169	utilde	ü	3. Latin Ext.-A
[cbs]_002	1EBC	Etilde	Ě	8. Latin Ext. Add.
[cbs]_002	1EBD	etilde	ě	8. Latin Ext. Add.
[cbs]_003	00C1	Aacute	Á	2. Latin-1 Supplement
[cbs]_003	00C3	Atilde	Ã	2. Latin-1 Supplement
[cbs]_003	00E1	aacute	á	2. Latin-1 Supplement
[cbs]_003	00E3	atilde	ã	2. Latin-1 Supplement
[cbs]_003	0128	ltilde	İ	3. Latin Ext.-A
[cbs]_003	0129	itilde	ı	3. Latin Ext.-A
[cbs]_003	0168	Utilde	Û	3. Latin Ext.-A
[cbs]_003	0169	utilde	ü	3. Latin Ext.-A
[cbs]_003	1EBC	Etilde	Ě	8. Latin Ext. Add.
[cbs]_003	1EBD	etilde	ě	8. Latin Ext. Add.
[cin]_001	0027	quotesingle	'	1. Basic Latin
[cin]_001	00C1	Aacute	Á	2. Latin-1 Supplement
[cin]_001	00C2	Acircumflex	Â	2. Latin-1 Supplement
[cin]_001	00C3	Atilde	Ã	2. Latin-1 Supplement
[cin]_001	00C9	Eacute	É	2. Latin-1 Supplement
[cin]_001	00CA	Ecircumflex	Ê	2. Latin-1 Supplement
[cin]_001	00CD	Iacute	Í	2. Latin-1 Supplement
[cin]_001	00CE	Icircumflex	Î	2. Latin-1 Supplement
[cin]_001	00D3	Oacute	Ó	2. Latin-1 Supplement
[cin]_001	00D4	Ocircumflex	Ô	2. Latin-1 Supplement
[cin]_001	00D5	Otilde	Õ	2. Latin-1 Supplement

arquivo	Unicode	Glyphname	Character	bloco
[cin]_001	00E1	aacute	á	2. Latin-1 Supplement
[cin]_001	00E2	acircumflex	â	2. Latin-1 Supplement
[cin]_001	00E3	atilde	ã	2. Latin-1 Supplement
[cin]_001	00E9	eacute	é	2. Latin-1 Supplement
[cin]_001	00EA	ecircumflex	ê	2. Latin-1 Supplement
[cin]_001	00ED	iacute	í	2. Latin-1 Supplement
[cin]_001	00EE	icircumflex	î	2. Latin-1 Supplement
[cin]_001	00F3	oacute	ó	2. Latin-1 Supplement
[cin]_001	00F4	ocircumflex	ô	2. Latin-1 Supplement
[cin]_001	00F5	otilde	õ	2. Latin-1 Supplement
[cin]_001	0128	ltilde	ĩ	3. Latin Ext.-A
[cin]_001	0129	itilde	ï	3. Latin Ext.-A
[cin]_001	0197	lstroke	†	4. Latin Ext.-B
[cin]_001	0268	istroke	‡	5. IPA Extensions
[cin]_001	0301	acutecomb	´	7. Combining Marks
[cin]_001	0303	tildecomb	˜	7. Combining Marks
[cin]_001	0341	acutetonecomb	ˆ	7. Combining Marks
[cin]_001	1EBC	Etilde	Ě	8. Latin Ext. Add.
[cin]_001	1EBD	etilde	ě	8. Latin Ext. Add.
[cin]_001	0197_0301	lstroke_acutecomb	†´	10. non-compliant
[cin]_001	0268_0301	istroke_acutecomb	‡´	10. non-compliant
[cin]_001	0197_0303	lstroke_tildecomb	†˜	10. non-compliant
[cin]_001	0268_0303	istroke_tildecomb	‡˜	10. non-compliant
[cni]_001	00C3	Atilde	Ã	2. Latin-1 Supplement
[cni]_001	00E3	atilde	ã	2. Latin-1 Supplement
[cni]_001	0128	ltilde	ĩ	3. Latin Ext.-A
[cni]_001	0129	itilde	ï	3. Latin Ext.-A
[cni]_001	0168	Utilde	Ū	3. Latin Ext.-A
[cni]_001	0169	utilde	ū	3. Latin Ext.-A
[cni]_001	1EBC	Etilde	Ě	8. Latin Ext. Add.
[cni]_001	1EBD	etilde	ě	8. Latin Ext. Add.
[cub]_001	00C3	Atilde	Ã	2. Latin-1 Supplement
[cub]_001	00D5	Otilde	Õ	2. Latin-1 Supplement
[cub]_001	00E3	atilde	ã	2. Latin-1 Supplement
[cub]_001	00F5	otilde	õ	2. Latin-1 Supplement
[cub]_001	0128	ltilde	ĩ	3. Latin Ext.-A
[cub]_001	0129	itilde	ï	3. Latin Ext.-A
[cub]_001	0168	Utilde	Ū	3. Latin Ext.-A
[cub]_001	0169	utilde	ū	3. Latin Ext.-A
[cub]_001	0197	lstroke	†	4. Latin Ext.-B
[cub]_001	0268	istroke	‡	5. IPA Extensions
[cub]_001	0301	acutecomb	´	7. Combining Marks
[cub]_001	0303	tildecomb	˜	7. Combining Marks
[cub]_001	0341	acutetonecomb	ˆ	7. Combining Marks
[cub]_001	1E4C	Otildeacute	Ó	8. Latin Ext. Add.
[cub]_001	1E4D	otildeacute	ó	8. Latin Ext. Add.
[cub]_001	1E78	Utildeacute	Ú	8. Latin Ext. Add.

arquivo	Unicode	Glyphname	Character	bloco
[cub]_001	1E79	utildeacute	ú	8. Latin Ext. Add.
[cub]_001	1EBC	Etilde	Ë	8. Latin Ext. Add.
[cub]_001	1EBD	etilde	ë	8. Latin Ext. Add.
[cub]_001	00C3_0301	Atilde_acutecomb	Á	10. non-compliant
[cub]_001	00E3_0301	atilde_acutecomb	á	10. non-compliant
[cub]_001	1EBC_0301	Etilde_acutecomb	Ë	10. non-compliant
[cub]_001	1EBD_0301	etilde_acutecomb	ë	10. non-compliant
[cub]_001	0128_0301	ltilde_acutecomb	í	10. non-compliant
[cub]_001	0129_0301	itilde_acutecomb	ï	10. non-compliant
[cub]_001	00D5_0301	Otilde_acutecomb	Õ	10. non-compliant
[cub]_001	00F5_0301	otilde_acutecomb	õ	10. non-compliant
[cub]_001	0168_0301	Utilde_acutecomb	Û	10. non-compliant
[cub]_001	0169_0301	utilde_acutecomb	ü	10. non-compliant
[cub]_001	0197_0303	lstroke_tildecomb	ř	10. non-compliant
[cub]_001	0268_0303	istroke_tildecomb	ř	10. non-compliant
[cub]_001	0197_0303_0301	lstroke_tildecomb_acutecomb	ř	10. non-compliant
[cub]_001	0268_0303_0301	istroke_tildecomb_acutecomb	ř	10. non-compliant
[des]_001	00C1	Aacute	Á	2. Latin-1 Supplement
[des]_001	00C3	Atilde	Ã	2. Latin-1 Supplement
[des]_001	00C9	Eacute	É	2. Latin-1 Supplement
[des]_001	00CD	Iacute	Í	2. Latin-1 Supplement
[des]_001	00D1	Ntilde	Ñ	2. Latin-1 Supplement
[des]_001	00D3	Oacute	Ó	2. Latin-1 Supplement
[des]_001	00D5	Otilde	Õ	2. Latin-1 Supplement
[des]_001	00DA	Uacute	Ú	2. Latin-1 Supplement
[des]_001	00E1	aacute	á	2. Latin-1 Supplement
[des]_001	00E3	atilde	ã	2. Latin-1 Supplement
[des]_001	00E9	eacute	é	2. Latin-1 Supplement
[des]_001	00ED	iacute	í	2. Latin-1 Supplement
[des]_001	00F1	ntilde	ñ	2. Latin-1 Supplement
[des]_001	00F3	oacute	ó	2. Latin-1 Supplement
[des]_001	00F5	otilde	õ	2. Latin-1 Supplement
[des]_001	00FA	uacute	ú	2. Latin-1 Supplement
[des]_001	0128	ltilde	í	3. Latin Ext.-A
[des]_001	0129	itilde	ï	3. Latin Ext.-A
[des]_001	0168	Utilde	Û	3. Latin Ext.-A
[des]_001	0169	utilde	ü	3. Latin Ext.-A
[des]_001	244.0	Ubar	Ů	4. Latin Ext.-B
[des]_001	289.0	ubar	ů	5. IPA Extensions
[des]_001	02BC	apostrophemod	'	6. Spc. Mod. Letters
[des]_001	1EBC	Etilde	Ë	8. Latin Ext. Add.
[des]_001	1EBD	etilde	ë	8. Latin Ext. Add.
[des]_001	A78B	Saltillo	ı	9. Latin Ext.-D
[des]_001	A78C	salttillo	ı	9. Latin Ext.-D
[des]_001	0244_0301	Ubar_acutecomb	Ů	10. non-compliant
[des]_001	0289_0301	ubar_acutecomb	ů	10. non-compliant

archivo	Unicode	Glyphname	Character	bloco
[des]_001	0244_0303	Ubar_tildecomb	Ů	10. non-compliant
[des]_001	0289_0303	ubar_tildecomb	ů	10. non-compliant
[des]_002	00C1	Aacute	Á	2. Latin-1 Supplement
[des]_002	00C3	Atilde	Ã	2. Latin-1 Supplement
[des]_002	00C9	Eacute	É	2. Latin-1 Supplement
[des]_002	00CD	Iacute	Í	2. Latin-1 Supplement
[des]_002	00D1	Ntilde	Ñ	2. Latin-1 Supplement
[des]_002	00D3	Oacute	Ó	2. Latin-1 Supplement
[des]_002	00D5	Otilde	Õ	2. Latin-1 Supplement
[des]_002	00DA	Uacute	Ú	2. Latin-1 Supplement
[des]_002	00E1	aacute	á	2. Latin-1 Supplement
[des]_002	00E3	atilde	ã	2. Latin-1 Supplement
[des]_002	00E9	eacute	é	2. Latin-1 Supplement
[des]_002	00ED	iacute	í	2. Latin-1 Supplement
[des]_002	00F1	ntilde	ñ	2. Latin-1 Supplement
[des]_002	00F3	oacute	ó	2. Latin-1 Supplement
[des]_002	00F5	otilde	õ	2. Latin-1 Supplement
[des]_002	00FA	uacute	ú	2. Latin-1 Supplement
[des]_002	0128	Itilde	İ	3. Latin Ext.-A
[des]_002	0129	itilde	ı	3. Latin Ext.-A
[des]_002	0168	Utilde	Ū	3. Latin Ext.-A
[des]_002	0169	utilde	ū	3. Latin Ext.-A
[des]_002	244.0	Ubar	Ů	4. Latin Ext.-B
[des]_002	289.0	ubar	ů	5. IPA Extensions
[des]_002	02BC	apostrophemod	’	6. Spc. Mod. Letters
[des]_002	1EBC	Etilde	Ě	8. Latin Ext. Add.
[des]_002	1EBD	etilde	ě	8. Latin Ext. Add.
[des]_002	A78B	Saltillo	ı	9. Latin Ext.-D
[des]_002	A78C	salttillo	ı	9. Latin Ext.-D
[des]_002	0244_0301	Ubar_acutecomb	Ů̇	10. non-compliant
[des]_002	0289_0301	ubar_acutecomb	ů̇	10. non-compliant
[des]_002	0244_0303	Ubar_tildecomb	Ů	10. non-compliant
[des]_002	0289_0303	ubar_tildecomb	ů	10. non-compliant
[des]_003	00C1	Aacute	Á	2. Latin-1 Supplement
[des]_003	00C3	Atilde	Ã	2. Latin-1 Supplement
[des]_003	00C9	Eacute	É	2. Latin-1 Supplement
[des]_003	00CD	Iacute	Í	2. Latin-1 Supplement
[des]_003	00D1	Ntilde	Ñ	2. Latin-1 Supplement
[des]_003	00D3	Oacute	Ó	2. Latin-1 Supplement
[des]_003	00D5	Otilde	Õ	2. Latin-1 Supplement
[des]_003	00DA	Uacute	Ú	2. Latin-1 Supplement
[des]_003	00E1	aacute	á	2. Latin-1 Supplement
[des]_003	00E3	atilde	ã	2. Latin-1 Supplement
[des]_003	00E9	eacute	é	2. Latin-1 Supplement
[des]_003	00ED	iacute	í	2. Latin-1 Supplement
[des]_003	00F1	ntilde	ñ	2. Latin-1 Supplement

arquivo	Unicode	Glyphname	Character	bloco
[des]_003	00F3	oacute	ó	2. Latin-1 Supplement
[des]_003	00F5	otilde	õ	2. Latin-1 Supplement
[des]_003	00FA	uacute	ú	2. Latin-1 Supplement
[des]_003	0128	ltilde	ĺ	3. Latin Ext.-A
[des]_003	0129	itilde	ı̇	3. Latin Ext.-A
[des]_003	0168	Utilde	Ū	3. Latin Ext.-A
[des]_003	0169	utilde	ū	3. Latin Ext.-A
[des]_003	244.0	Ubar	Ʊ	4. Latin Ext.-B
[des]_003	289.0	ubar	ƶ	5. IPA Extensions
[des]_003	02BC	apostrophemod	'	6. Spc. Mod. Letters
[des]_003	1EBC	Etilde	Ě	8. Latin Ext. Add.
[des]_003	1EBD	etilde	ě	8. Latin Ext. Add.
[des]_003	A78B	Saltillo	ı̂	9. Latin Ext.-D
[des]_003	A78C	salttillo	ı̇	9. Latin Ext.-D
[des]_003	0244_0301	Ubar_acutecomb	Ʊ̇	10. non-compliant
[des]_003	0289_0301	ubar_acutecomb	ƶ̇	10. non-compliant
[des]_003	0244_0303	Ubar_tildecomb	Ʊ̂	10. non-compliant
[des]_003	0289_0303	ubar_tildecomb	ƶ̂	10. non-compliant
[dny]_001	02BC	apostrophemod	'	6. Spc. Mod. Letters
[dny]_001	A78B	Saltillo	ı̂	9. Latin Ext.-D
[dny]_001	A78C	salttillo	ı̇	9. Latin Ext.-D
[dny]_002	02BC	apostrophemod	'	6. Spc. Mod. Letters
[dny]_002	A78B	Saltillo	ı̂	9. Latin Ext.-D
[dny]_002	A78C	salttillo	ı̇	9. Latin Ext.-D
[dny]_003	02BC	apostrophemod	'	6. Spc. Mod. Letters
[dny]_003	A78B	Saltillo	ı̂	9. Latin Ext.-D
[dny]_003	A78C	salttillo	ı̇	9. Latin Ext.-D
[gae]_001	00C3	Atilde	Ã	2. Latin-1 Supplement
[gae]_001	00E3	atilde	ã	2. Latin-1 Supplement
[gae]_001	0128	ltilde	ĺ	3. Latin Ext.-A
[gae]_001	0129	itilde	ı̇	3. Latin Ext.-A
[gae]_001	0168	Utilde	Ū	3. Latin Ext.-A
[gae]_001	0169	utilde	ū	3. Latin Ext.-A
[gae]_001	1EBC	Etilde	Ě	8. Latin Ext. Add.
[gae]_001	1EBD	etilde	ě	8. Latin Ext. Add.
[gta]_001	00C3	Atilde	Ã	2. Latin-1 Supplement
[gta]_001	00E3	atilde	ã	2. Latin-1 Supplement
[gta]_001	0128	ltilde	ĺ	3. Latin Ext.-A
[gta]_001	0129	itilde	ı̇	3. Latin Ext.-A
[gta]_001	0168	Utilde	Ū	3. Latin Ext.-A
[gta]_001	0169	utilde	ū	3. Latin Ext.-A
[gta]_001	1EBC	Etilde	Ě	8. Latin Ext. Add.
[gta]_001	1EBD	etilde	ě	8. Latin Ext. Add.
[gta]_001	1EF8	Ytilde	Ÿ	8. Latin Ext. Add.
[gta]_001	1EF9	ytilde	ÿ	8. Latin Ext. Add.

arquivo	Unicode	Glyphname	Character	bloco
[gub]_001	0027	quotesingle	'	1. Basic Latin
[gub]_001	00C0	Agrave	À	2. Latin-1 Supplement
[gub]_001	00E0	agrave	à	2. Latin-1 Supplement
[gug]_001	0027	quotesingle	'	1. Basic Latin
[gug]_001	00C1	Aacute	Á	2. Latin-1 Supplement
[gug]_001	00C3	Atilde	Ã	2. Latin-1 Supplement
[gug]_001	00C9	Eacute	É	2. Latin-1 Supplement
[gug]_001	00CD	Iacute	Í	2. Latin-1 Supplement
[gug]_001	00D1	Ntilde	Ñ	2. Latin-1 Supplement
[gug]_001	00D3	Oacute	Ó	2. Latin-1 Supplement
[gug]_001	00D5	Otilde	Õ	2. Latin-1 Supplement
[gug]_001	00DA	Uacute	Ú	2. Latin-1 Supplement
[gug]_001	00DD	Yacute	Ý	2. Latin-1 Supplement
[gug]_001	00E1	aacute	á	2. Latin-1 Supplement
[gug]_001	00E3	atilde	ã	2. Latin-1 Supplement
[gug]_001	00E9	eacute	é	2. Latin-1 Supplement
[gug]_001	00ED	iacute	í	2. Latin-1 Supplement
[gug]_001	00F1	ntilde	ñ	2. Latin-1 Supplement
[gug]_001	00F3	oacute	ó	2. Latin-1 Supplement
[gug]_001	00F5	otilde	õ	2. Latin-1 Supplement
[gug]_001	00FA	uacute	ú	2. Latin-1 Supplement
[gug]_001	00FD	yacute	ý	2. Latin-1 Supplement
[gug]_001	0128	ltilde	Ĺ	3. Latin Ext.-A
[gug]_001	0129	itilde	İ	3. Latin Ext.-A
[gug]_001	0168	Utilde	Ū	3. Latin Ext.-A
[gug]_001	0169	utilde	ū	3. Latin Ext.-A
[gug]_001	0303	tildecomb	̃	7. Combining Marks
[gug]_001	1EBC	Etilde	Ě	8. Latin Ext. Add.
[gug]_001	1EBD	etilde	ě	8. Latin Ext. Add.
[gug]_001	1EF8	Ytilde	Ÿ	8. Latin Ext. Add.
[gug]_001	1EF9	ytilde	ÿ	8. Latin Ext. Add.
[gug]_001	0047_0303	G_tildecomb	Ĝ	10. non-compliant
[gug]_001	0067_0303	g_tildecomb	ğ	10. non-compliant
[gug]_002	00C3	Atilde	Ã	2. Latin-1 Supplement
[gug]_002	00D5	Otilde	Õ	2. Latin-1 Supplement
[gug]_002	00E3	atilde	ã	2. Latin-1 Supplement
[gug]_002	00F5	otilde	õ	2. Latin-1 Supplement
[gug]_002	0128	ltilde	Ĺ	3. Latin Ext.-A
[gug]_002	0129	itilde	İ	3. Latin Ext.-A
[gug]_002	0168	Utilde	Ū	3. Latin Ext.-A
[gug]_002	0169	utilde	ū	3. Latin Ext.-A
[gug]_002	0303	tildecomb	̃	7. Combining Marks
[gug]_002	1EBC	Etilde	Ě	8. Latin Ext. Add.
[gug]_002	1EBD	etilde	ě	8. Latin Ext. Add.
[gug]_002	1EF8	Ytilde	Ÿ	8. Latin Ext. Add.
[gug]_002	1EF9	ytilde	ÿ	8. Latin Ext. Add.
[gug]_002	0047_0303	G_tildecomb	Ĝ	10. non-compliant

archivo	Unicode	Glyphname	Character	bloco
[gug]_002	0067_0303	g_tildecomb	ġ	10. non-compliant
[gun]_001	00C1	Aacute	Á	2. Latin-1 Supplement
[gun]_001	00C3	Atilde	Ã	2. Latin-1 Supplement
[gun]_001	00C9	Eacute	É	2. Latin-1 Supplement
[gun]_001	00D1	Ntilde	Ñ	2. Latin-1 Supplement
[gun]_001	00D3	Oacute	Ó	2. Latin-1 Supplement
[gun]_001	00D5	Otilde	Õ	2. Latin-1 Supplement
[gun]_001	00DA	Uacute	Ú	2. Latin-1 Supplement
[gun]_001	00E1	aacute	á	2. Latin-1 Supplement
[gun]_001	00E3	atilde	ã	2. Latin-1 Supplement
[gun]_001	00E9	eacute	é	2. Latin-1 Supplement
[gun]_001	00F1	ntilde	ñ	2. Latin-1 Supplement
[gun]_001	00F3	oacute	ó	2. Latin-1 Supplement
[gun]_001	00F5	otilde	õ	2. Latin-1 Supplement
[gun]_001	00FA	uacute	ú	2. Latin-1 Supplement
[gun]_001	0168	Utilde	Û	3. Latin Ext.-A
[gun]_001	0169	utilde	ü	3. Latin Ext.-A
[gun]_001	02BC	apostrophemod	'	6. Spc. Mod. Letters
[gun]_001	1EF8	Ytilde	Ÿ	8. Latin Ext. Add.
[gun]_001	1EF9	ytilde	ÿ	8. Latin Ext. Add.
[gun]_001	A78B	Saltillo	ı	9. Latin Ext.-D
[gun]_001	A78C	salttillo	ı	9. Latin Ext.-D
[gun]_002	00C3	Atilde	Ã	2. Latin-1 Supplement
[gun]_002	00D5	Otilde	Õ	2. Latin-1 Supplement
[gun]_002	00E3	atilde	ã	2. Latin-1 Supplement
[gun]_002	00F5	otilde	õ	2. Latin-1 Supplement
[gun]_002	0128	ltilde	Ĺ	3. Latin Ext.-A
[gun]_002	0129	itilde	ĺ	3. Latin Ext.-A
[gun]_002	0168	Utilde	Û	3. Latin Ext.-A
[gun]_002	0169	utilde	ü	3. Latin Ext.-A
[gun]_002	02BC	apostrophemod	'	6. Spc. Mod. Letters
[gun]_002	1EBC	Etilde	Ě	8. Latin Ext. Add.
[gun]_002	1EBD	etilde	ě	8. Latin Ext. Add.
[gun]_002	1EF8	Ytilde	Ÿ	8. Latin Ext. Add.
[gun]_002	1EF9	ytilde	ÿ	8. Latin Ext. Add.
[gun]_002	A78B	Saltillo	ı	9. Latin Ext.-D
[gun]_002	A78C	salttillo	ı	9. Latin Ext.-D
[gun]_004	0027	quotesingle	'	1. Basic Latin
[gun]_004	00C3	Atilde	Ã	2. Latin-1 Supplement
[gun]_004	00D5	Otilde	Õ	2. Latin-1 Supplement
[gun]_004	00E3	atilde	ã	2. Latin-1 Supplement
[gun]_004	00F5	otilde	õ	2. Latin-1 Supplement
[gun]_004	0128	ltilde	Ĺ	3. Latin Ext.-A
[gun]_004	0129	itilde	ĺ	3. Latin Ext.-A
[gun]_004	0168	Utilde	Û	3. Latin Ext.-A
[gun]_004	0169	utilde	ü	3. Latin Ext.-A

arquivo	Unicode	Glyphname	Character	bloco
[gun]_004	1EBC	Etilde	Ě	8. Latin Ext. Add.
[gun]_004	1EBD	etilde	ě	8. Latin Ext. Add.
[gun]_004	1EF8	Ytilde	Ÿ	8. Latin Ext. Add.
[gun]_004	1EF9	ytilde	ÿ	8. Latin Ext. Add.
[gvc]_001	00C1	Aacute	Á	2. Latin-1 Supplement
[gvc]_001	00C3	Atilde	Ã	2. Latin-1 Supplement
[gvc]_001	00C9	Eacute	É	2. Latin-1 Supplement
[gvc]_001	00D1	Ntilde	Ñ	2. Latin-1 Supplement
[gvc]_001	00D5	Otilde	Õ	2. Latin-1 Supplement
[gvc]_001	00DA	Uacute	Ú	2. Latin-1 Supplement
[gvc]_001	00E1	aacute	á	2. Latin-1 Supplement
[gvc]_001	00E3	atilde	ã	2. Latin-1 Supplement
[gvc]_001	00E9	eacute	é	2. Latin-1 Supplement
[gvc]_001	00F1	ntilde	ñ	2. Latin-1 Supplement
[gvc]_001	00F5	otilde	õ	2. Latin-1 Supplement
[gvc]_001	00FA	uacute	ú	2. Latin-1 Supplement
[gvc]_001	0128	ltilde	Ĺ	3. Latin Ext.-A
[gvc]_001	0129	itilde	ĺ	3. Latin Ext.-A
[gvc]_001	0168	Utilde	Ū	3. Latin Ext.-A
[gvc]_001	0169	utilde	ū	3. Latin Ext.-A
[gvc]_001	0244	Ubar	Ɔ	4. Latin Ext.-B
[gvc]_001	0289	ubar	ɸ	5. IPA Extensions
[gvc]_001	02BC	apostrophemod	'	6. Spc. Mod. Letters
[gvc]_001	1EBC	Etilde	Ě	8. Latin Ext. Add.
[gvc]_001	1EBD	etilde	ě	8. Latin Ext. Add.
[gvc]_001	A78B	Saltillo	ı	9. Latin Ext.-D
[gvc]_001	A78C	salttillo	ı	9. Latin Ext.-D
[gvc]_001	0244_0303	Ubar_tildecomb	Ɔ̃	10. non-compliant
[gvc]_001	0289_0303	ubar_tildecomb	ɸ̃	10. non-compliant
[gvc]_002	0027	quotesingle	'	1. Basic Latin
[gvc]_002	00C3	Atilde	Ã	2. Latin-1 Supplement
[gvc]_002	00D1	Ntilde	Ñ	2. Latin-1 Supplement
[gvc]_002	00D5	Otilde	Õ	2. Latin-1 Supplement
[gvc]_002	00E3	atilde	ã	2. Latin-1 Supplement
[gvc]_002	00F1	ntilde	ñ	2. Latin-1 Supplement
[gvc]_002	00F5	otilde	õ	2. Latin-1 Supplement
[gvc]_002	0128	ltilde	Ĺ	3. Latin Ext.-A
[gvc]_002	0129	itilde	ĺ	3. Latin Ext.-A
[gvc]_002	0168	Utilde	Ū	3. Latin Ext.-A
[gvc]_002	0169	utilde	ū	3. Latin Ext.-A
[gvc]_002	0244	Ubar	Ɔ	4. Latin Ext.-B
[gvc]_002	0289	ubar	ɸ	5. IPA Extensions
[gvc]_002	0303	tildecomb	̃	7. Combining Marks
[gvc]_002	1EBC	Etilde	Ě	8. Latin Ext. Add.
[gvc]_002	1EBD	etilde	ě	8. Latin Ext. Add.
[gvc]_002	0244_0303	Ubar_tildecomb	Ɔ̃	10. non-compliant

arquivo	Unicode	Glyphname	Character	bloco
[gvc]_002	0289_0303	ubar_tildecomb	ũ	10. non-compliant
[gvp]_001	00C0	Agrave	À	2. Latin-1 Supplement
[gvp]_001	00C3	Atilde	Ã	2. Latin-1 Supplement
[gvp]_001	00CA	Ecircumflex	Ê	2. Latin-1 Supplement
[gvp]_001	00D4	Ocircumflex	Ô	2. Latin-1 Supplement
[gvp]_001	00D5	Otilde	Õ	2. Latin-1 Supplement
[gvp]_001	00E0	agrave	à	2. Latin-1 Supplement
[gvp]_001	00E3	atilde	ã	2. Latin-1 Supplement
[gvp]_001	00EA	ecircumflex	ê	2. Latin-1 Supplement
[gvp]_001	00F4	ocircumflex	ô	2. Latin-1 Supplement
[gvp]_001	00F5	otilde	õ	2. Latin-1 Supplement
[gvp]_001	0128	ltilde	Ĺ	3. Latin Ext.-A
[gvp]_001	0129	itilde	Ī	3. Latin Ext.-A
[gvp]_001	0168	Utilde	Ū	3. Latin Ext.-A
[gvp]_001	0169	utilde	ū	3. Latin Ext.-A
[gvp]_001	1EBC	Etilde	Ě	8. Latin Ext. Add.
[gvp]_001	1EBD	etilde	ě	8. Latin Ext. Add.
[gvp]_001	1EF2	Ygrave	Ỳ	8. Latin Ext. Add.
[gvp]_001	1EF3	ygrave	ỳ	8. Latin Ext. Add.
[gvp]_001	1EF8	Ytilde	Ÿ	8. Latin Ext. Add.
[gvp]_001	1EF9	ytilde	ÿ	8. Latin Ext. Add.
[hix]_001	00C0	Agrave	À	2. Latin-1 Supplement
[hix]_001	00E0	agrave	à	2. Latin-1 Supplement
[int]_001	00C1	Aacute	Á	2. Latin-1 Supplement
[int]_001	00C3	Atilde	Ã	2. Latin-1 Supplement
[int]_001	00E1	aacute	á	2. Latin-1 Supplement
[int]_001	00E3	atilde	ã	2. Latin-1 Supplement
[int]_001	0128	ltilde	Ĺ	3. Latin Ext.-A
[int]_001	0129	itilde	Ī	3. Latin Ext.-A
[int]_001	0168	Utilde	Ū	3. Latin Ext.-A
[int]_001	0169	utilde	ū	3. Latin Ext.-A
[int]_001	0303	tildecomb	˜	7. Combining Marks
[int]_001	0330	tildebelowcomb	˘	7. Combining Marks
[int]_001	1E1A	Etildebelow	Ẹ̌	8. Latin Ext. Add.
[int]_001	1E1B	etildebelow	ẹ̌	8. Latin Ext. Add.
[int]_001	1E2C	ltildebelow	Ḷ́	8. Latin Ext. Add.
[int]_001	1E2D	itildebelow	Ị̄	8. Latin Ext. Add.
[int]_001	1E74	Utildebelow	Ụ̄	8. Latin Ext. Add.
[int]_001	1E75	utildebelow	ụ̄	8. Latin Ext. Add.
[int]_001	1EBC	Etilde	Ě	8. Latin Ext. Add.
[int]_001	1EBD	etilde	ě	8. Latin Ext. Add.
[int]_001	0041_0330	A_tildebelowcomb	Ǻ	10. non-compliant
[int]_001	0061_0330	a_tildebelowcomb	ǻ	10. non-compliant
[int]_001	004F_0330	O_tildebelowcomb	Ǫ	10. non-compliant
[int]_001	006F_0330	o_tildebelowcomb	ǫ	10. non-compliant
[int]_001	00C3_0330	Atilde_tildebelowcomb	Ạ̊́	10. non-compliant
[int]_001	00E3_0330	atilde_tildebelowcomb	ạ̊́	10. non-compliant

arquivo	Unicode	Glyphname	Character	bloco
[int]_001	1EBC_0330	Etilde_tildebelowcomb	Ě	10. non-compliant
[int]_001	1EBD_0330	etilde_tildebelowcomb	ě	10. non-compliant
[int]_001	0128_0330	ltilde_tildebelowcomb	Ī	10. non-compliant
[int]_001	0129_0330	itilde_tildebelowcomb	ī	10. non-compliant
[int]_001	0168_0330	Utilde_tildebelowcomb	Ū	10. non-compliant
[int]_001	0169_0330	utilde_tildebelowcomb	ū	10. non-compliant
[ite]_001	00D1	Ntilde	Ñ	2. Latin-1 Supplement
[ite]_001	00F1	ntilde	ñ	2. Latin-1 Supplement
[jbt]_001	00C3	Atilde	Ã	2. Latin-1 Supplement
[jbt]_001	00C4	Adieresis	Ä	2. Latin-1 Supplement
[jbt]_001	00D5	Otilde	Õ	2. Latin-1 Supplement
[jbt]_001	00DC	Udieresis	Ü	2. Latin-1 Supplement
[jbt]_001	00E3	atilde	ã	2. Latin-1 Supplement
[jbt]_001	00E4	adieresis	ä	2. Latin-1 Supplement
[jbt]_001	00F5	otilde	õ	2. Latin-1 Supplement
[jbt]_001	00FC	udieresis	ü	2. Latin-1 Supplement
[jbt]_001	0128	ltilde	Ī	3. Latin Ext.-A
[jbt]_001	0129	itilde	ī	3. Latin Ext.-A
[jbt]_001	1EBC	Etilde	Ě	8. Latin Ext. Add.
[jbt]_001	1EBD	etilde	ě	8. Latin Ext. Add.
[jua]_001	0027	quotesingle	'	1. Basic Latin
[jua]_001	00C3	Atilde	Ã	2. Latin-1 Supplement
[jua]_001	00D5	Otilde	Õ	2. Latin-1 Supplement
[jua]_001	00E3	atilde	ã	2. Latin-1 Supplement
[jua]_001	00F5	otilde	õ	2. Latin-1 Supplement
[jua]_001	0128	ltilde	Ī	3. Latin Ext.-A
[jua]_001	0129	itilde	ī	3. Latin Ext.-A
[jua]_001	0168	Utilde	Ū	3. Latin Ext.-A
[jua]_001	0169	utilde	ū	3. Latin Ext.-A
[jua]_001	1EBC	Etilde	Ě	8. Latin Ext. Add.
[jua]_001	1EBD	etilde	ě	8. Latin Ext. Add.
[jua]_001	1EF8	Ytilde	Ÿ	8. Latin Ext. Add.
[jua]_001	1EF9	ytilde	ÿ	8. Latin Ext. Add.
[jup]_001	0027	quotesingle	'	1. Basic Latin
[jup]_001	00C1	Aacute	Á	2. Latin-1 Supplement
[jup]_001	00C3	Atilde	Ã	2. Latin-1 Supplement
[jup]_001	00C9	Eacute	É	2. Latin-1 Supplement
[jup]_001	00D3	Oacute	Ó	2. Latin-1 Supplement
[jup]_001	00D5	Otilde	Õ	2. Latin-1 Supplement
[jup]_001	00E1	aacute	á	2. Latin-1 Supplement
[jup]_001	00E3	atilde	ã	2. Latin-1 Supplement
[jup]_001	00E9	eacute	é	2. Latin-1 Supplement
[jup]_001	00F3	oacute	ó	2. Latin-1 Supplement
[jup]_001	00F5	otilde	õ	2. Latin-1 Supplement
[jup]_001	0168	Utilde	Ū	3. Latin Ext.-A
[jup]_001	0169	utilde	ū	3. Latin Ext.-A
[jup]_001	0197	Istroke	†	4. Latin Ext.-B

archivo	Unicode	Glyphname	Character	bloco
[jup]_001	0268	istroke	ı	5. IPA Extensions
[jup]_001	0303	tildecomb	˜	7. Combining Marks
[jup]_001	1EBC	Etilde	Ě	8. Latin Ext. Add.
[jup]_001	1EBD	etilde	ě	8. Latin Ext. Add.
[jup]_001	0197_0303	Istroke_tildecomb	İ̃	10. non-compliant
[jup]_001	0268_0303	istroke_tildecomb	ı̃	10. non-compliant
[jur]_001	00C3	Atilde	Ā	2. Latin-1 Supplement
[jur]_001	00CF	Idieresis	Ï	2. Latin-1 Supplement
[jur]_001	00E3	atilde	ā	2. Latin-1 Supplement
[jur]_001	00EF	idieresis	ï	2. Latin-1 Supplement
[jur]_001	0128	ltilde	Ĺ	3. Latin Ext.-A
[jur]_001	0129	itilde	ĺ	3. Latin Ext.-A
[jur]_001	0168	Utilde	Ū	3. Latin Ext.-A
[jur]_001	0169	utilde	ū	3. Latin Ext.-A
[jur]_001	02BC	apostrophemod	'	6. Spc. Mod. Letters
[jur]_001	A78B	Saltillo	ı	9. Latin Ext.-D
[jur]_001	A78C	salttillo	ı̂	9. Latin Ext.-D
[jur]_001	00CF_0303	Idieresis_tildecomb	Ï̃	10. non-compliant
[jur]_001	00EF_0303	idieresis_tildecomb	ï̃	10. non-compliant
[jur]_002	00C3	Atilde	Ā	2. Latin-1 Supplement
[jur]_002	00CF	Idieresis	Ï	2. Latin-1 Supplement
[jur]_002	00E3	atilde	ā	2. Latin-1 Supplement
[jur]_002	00EF	idieresis	ï	2. Latin-1 Supplement
[jur]_002	0128	ltilde	Ĺ	3. Latin Ext.-A
[jur]_002	0129	itilde	ĺ	3. Latin Ext.-A
[jur]_002	0168	Utilde	Ū	3. Latin Ext.-A
[jur]_002	0169	utilde	ū	3. Latin Ext.-A
[jur]_002	02BC	apostrophemod	'	6. Spc. Mod. Letters
[jur]_002	1EBC	Etilde	Ě	8. Latin Ext. Add.
[jur]_002	1EBD	etilde	ě	8. Latin Ext. Add.
[jur]_002	A78B	Saltillo	ı	9. Latin Ext.-D
[jur]_002	A78C	salttillo	ı̂	9. Latin Ext.-D
[jur]_002	00CF_0303	Idieresis_tildecomb	Ï̃	10. non-compliant
[jur]_002	00EF_0303	idieresis_tildecomb	ï̃	10. non-compliant
[jur]_003	00C3	Atilde	Ā	2. Latin-1 Supplement
[jur]_003	00CF	Idieresis	Ï	2. Latin-1 Supplement
[jur]_003	00D5	Otilde	Ŏ	2. Latin-1 Supplement
[jur]_003	00E3	atilde	ā	2. Latin-1 Supplement
[jur]_003	00EF	idieresis	ï	2. Latin-1 Supplement
[jur]_003	00F5	otilde	õ	2. Latin-1 Supplement
[jur]_003	0128	ltilde	Ĺ	3. Latin Ext.-A
[jur]_003	0129	itilde	ĺ	3. Latin Ext.-A
[jur]_003	0168	Utilde	Ū	3. Latin Ext.-A
[jur]_003	0169	utilde	ū	3. Latin Ext.-A
[jur]_003	0303	tildecomb	˜	7. Combining Marks
[jur]_003	1EBC	Etilde	Ě	8. Latin Ext. Add.

arquivo	Unicode	Glyphname	Character	bloco
[jur]_003	1EBD	etilde	ě	8. Latin Ext. Add.
[jur]_003	00CF_0303	Idieresis_tildecomb	ǐ	10. non-compliant
[jur]_003	00EF_0303	idieresis_tildecomb	ǐ	10. non-compliant
[kay]_001	00C3	Atilde	Ā	2. Latin-1 Supplement
[kay]_001	00D5	Otilde	Ō	2. Latin-1 Supplement
[kay]_001	00E3	atilde	ā	2. Latin-1 Supplement
[kay]_001	00F5	otilde	ō	2. Latin-1 Supplement
[kay]_001	0128	ltilde	Ĺ	3. Latin Ext.-A
[kay]_001	0129	itilde	ĺ	3. Latin Ext.-A
[kay]_001	0168	Utilde	Ū	3. Latin Ext.-A
[kay]_001	0169	utilde	ū	3. Latin Ext.-A
[kay]_001	1EBC	Etilde	Ě	8. Latin Ext. Add.
[kay]_001	1EBD	etilde	ě	8. Latin Ext. Add.
[kay]_001	1EF8	Ytilde	Ÿ	8. Latin Ext. Add.
[kay]_001	1EF9	ytilde	ÿ	8. Latin Ext. Add.
[kay]_002	00C3	Atilde	Ā	2. Latin-1 Supplement
[kay]_002	00D5	Otilde	Ō	2. Latin-1 Supplement
[kay]_002	00E3	atilde	ā	2. Latin-1 Supplement
[kay]_002	00F5	otilde	ō	2. Latin-1 Supplement
[kay]_002	0128	ltilde	Ĺ	3. Latin Ext.-A
[kay]_002	0129	itilde	ĺ	3. Latin Ext.-A
[kay]_002	0168	Utilde	Ū	3. Latin Ext.-A
[kay]_002	0169	utilde	ū	3. Latin Ext.-A
[kay]_002	1EBC	Etilde	Ě	8. Latin Ext. Add.
[kay]_002	1EBD	etilde	ě	8. Latin Ext. Add.
[kay]_002	1EF8	Ytilde	Ÿ	8. Latin Ext. Add.
[kay]_002	1EF9	ytilde	ÿ	8. Latin Ext. Add.
[kay]_003	0027	quotesingle	'	1. Basic Latin
[kay]_003	00C3	Atilde	Ā	2. Latin-1 Supplement
[kay]_003	00E3	atilde	ā	2. Latin-1 Supplement
[kbc]_001	01E4	Gstroke	Ɔ	4. Latin Ext.-B
[kbc]_001	01E5	gstroke	ɔ	4. Latin Ext.-B
[kbc]_002	01E4	Gstroke	Ɔ	4. Latin Ext.-B
[kbc]_002	01E5	gstroke	ɔ	4. Latin Ext.-B
[kbc]_003	01E4	Gstroke	Ɔ	4. Latin Ext.-B
[kbc]_003	01E5	gstroke	ɔ	4. Latin Ext.-B
[kgk]_001	00C1	Aacute	Ā	2. Latin-1 Supplement
[kgk]_001	00C3	Atilde	Ā	2. Latin-1 Supplement
[kgk]_001	00C9	Eacute	Ē	2. Latin-1 Supplement
[kgk]_001	00CD	Iacute	Ī	2. Latin-1 Supplement
[kgk]_001	00D1	Ntilde	Ñ	2. Latin-1 Supplement
[kgk]_001	00D3	Oacute	Ō	2. Latin-1 Supplement
[kgk]_001	00D5	Otilde	Ō	2. Latin-1 Supplement
[kgk]_001	00DA	Uacute	Ū	2. Latin-1 Supplement
[kgk]_001	00DD	Yacute	Ÿ	2. Latin-1 Supplement
[kgk]_001	00E1	aacute	ā	2. Latin-1 Supplement
[kgk]_001	00E3	atilde	ā	2. Latin-1 Supplement

archivo	Unicode	Glyphname	Character	bloco
[kgk]_001	00E9	eacute	é	2. Latin-1 Supplement
[kgk]_001	00ED	iacute	í	2. Latin-1 Supplement
[kgk]_001	00F1	ntilde	ñ	2. Latin-1 Supplement
[kgk]_001	00F3	oacute	ó	2. Latin-1 Supplement
[kgk]_001	00F5	otilde	õ	2. Latin-1 Supplement
[kgk]_001	00FA	uacute	ú	2. Latin-1 Supplement
[kgk]_001	00FD	yacute	ý	2. Latin-1 Supplement
[kgk]_001	0128	ltilde	ĩ	3. Latin Ext.-A
[kgk]_001	0129	itilde	ï	3. Latin Ext.-A
[kgk]_001	02BC	apostrophemod	'	6. Spc. Mod. Letters
[kgk]_001	1EBC	Etilde	Ě	8. Latin Ext. Add.
[kgk]_001	1EBD	etilde	ě	8. Latin Ext. Add.
[kgk]_001	1EF8	Ytilde	Ÿ	8. Latin Ext. Add.
[kgk]_001	1EF9	ytilde	ÿ	8. Latin Ext. Add.
[kgk]_001	A78B	Saltillo	ı	9. Latin Ext.-D
[kgk]_001	A78C	saltillo	ı	9. Latin Ext.-D
[kgk]_001	0047_0303	G_tildecomb	Ĝ	10. non-compliant
[kgk]_001	0067_0303	g_tildecomb	ĝ	10. non-compliant
[kgk]_002	00C1	Aacute	Á	2. Latin-1 Supplement
[kgk]_002	00C3	Atilde	Ã	2. Latin-1 Supplement
[kgk]_002	00C9	Eacute	É	2. Latin-1 Supplement
[kgk]_002	00D1	Ntilde	Ñ	2. Latin-1 Supplement
[kgk]_002	00D3	Oacute	Ó	2. Latin-1 Supplement
[kgk]_002	00D5	Otilde	Õ	2. Latin-1 Supplement
[kgk]_002	00DD	Yacute	Ý	2. Latin-1 Supplement
[kgk]_002	00E1	aacute	á	2. Latin-1 Supplement
[kgk]_002	00E3	atilde	ã	2. Latin-1 Supplement
[kgk]_002	00E9	eacute	é	2. Latin-1 Supplement
[kgk]_002	00F1	ntilde	ñ	2. Latin-1 Supplement
[kgk]_002	00F3	oacute	ó	2. Latin-1 Supplement
[kgk]_002	00F5	otilde	õ	2. Latin-1 Supplement
[kgk]_002	00FD	yacute	ý	2. Latin-1 Supplement
[kgk]_002	0128	ltilde	ĩ	3. Latin Ext.-A
[kgk]_002	0129	itilde	ï	3. Latin Ext.-A
[kgk]_002	0168	Utilde	Ū	3. Latin Ext.-A
[kgk]_002	0169	utilde	ū	3. Latin Ext.-A
[kgk]_002	02BC	apostrophemod	'	6. Spc. Mod. Letters
[kgk]_002	1EBC	Etilde	Ě	8. Latin Ext. Add.
[kgk]_002	1EBD	etilde	ě	8. Latin Ext. Add.
[kgk]_002	1EF8	Ytilde	Ÿ	8. Latin Ext. Add.
[kgk]_002	1EF9	ytilde	ÿ	8. Latin Ext. Add.
[kgk]_002	A78B	Saltillo	ı	9. Latin Ext.-D
[kgk]_002	A78C	saltillo	ı	9. Latin Ext.-D
[kgk]_002	0047_0303	G_tildecomb	Ĝ	10. non-compliant
[kgk]_002	0067_0303	g_tildecomb	ĝ	10. non-compliant
[kgk]_003_	00C1	Aacute	Á	2. Latin-1 Supplement

arquivo	Unicode	Glyphname	Character	bloco
[kgk]_003_	00C3	Atilde	Ã	2. Latin-1 Supplement
[kgk]_003_	00C9	Eacute	É	2. Latin-1 Supplement
[kgk]_003_	00D1	Ntilde	Ñ	2. Latin-1 Supplement
[kgk]_003_	00D3	Oacute	Ó	2. Latin-1 Supplement
[kgk]_003_	00D5	Otilde	Õ	2. Latin-1 Supplement
[kgk]_003_	00DD	Yacute	Ý	2. Latin-1 Supplement
[kgk]_003_	00E1	aacute	á	2. Latin-1 Supplement
[kgk]_003_	00E3	atilde	ã	2. Latin-1 Supplement
[kgk]_003_	00E9	eacute	é	2. Latin-1 Supplement
[kgk]_003_	00F1	ntilde	ñ	2. Latin-1 Supplement
[kgk]_003_	00F3	oacute	ó	2. Latin-1 Supplement
[kgk]_003_	00F5	otilde	õ	2. Latin-1 Supplement
[kgk]_003_	00FD	yacute	ý	2. Latin-1 Supplement
[kgk]_003_	0128	ltilde	ĩ	3. Latin Ext.-A
[kgk]_003_	0129	itilde	ı̇	3. Latin Ext.-A
[kgk]_003_	0168	Utilde	Û	3. Latin Ext.-A
[kgk]_003_	0169	utilde	ü	3. Latin Ext.-A
[kgk]_003_	02BC	apostrophemod	'	6. Spc. Mod. Letters
[kgk]_003_	1EBC	Etilde	Ě	8. Latin Ext. Add.
[kgk]_003_	1EBD	etilde	ě	8. Latin Ext. Add.
[kgk]_003_	1EF8	Ytilde	Ÿ	8. Latin Ext. Add.
[kgk]_003_	1EF9	ytilde	ÿ	8. Latin Ext. Add.
[kgk]_003_	A78B	Saltillo	ı̂	9. Latin Ext.-D
[kgk]_003_	A78C	salttillo	ı̃	9. Latin Ext.-D
[kgk]_003_	0047_0303	G_tildecomb	Ĝ	10. non-compliant
[kgk]_003_	0067_0303	g_tildecomb	ğ	10. non-compliant
[kgk]_004	00C1	Aacute	Á	2. Latin-1 Supplement
[kgk]_004	00C3	Atilde	Ã	2. Latin-1 Supplement
[kgk]_004	00C9	Eacute	É	2. Latin-1 Supplement
[kgk]_004	00CD	Iacute	Í	2. Latin-1 Supplement
[kgk]_004	00D1	Ntilde	Ñ	2. Latin-1 Supplement
[kgk]_004	00D3	Oacute	Ó	2. Latin-1 Supplement
[kgk]_004	00D5	Otilde	Õ	2. Latin-1 Supplement
[kgk]_004	00DA	Uacute	Ú	2. Latin-1 Supplement
[kgk]_004	00DD	Yacute	Ý	2. Latin-1 Supplement
[kgk]_004	00E1	aacute	á	2. Latin-1 Supplement
[kgk]_004	00E3	atilde	ã	2. Latin-1 Supplement
[kgk]_004	00E9	eacute	é	2. Latin-1 Supplement
[kgk]_004	00ED	Iacute	í	2. Latin-1 Supplement
[kgk]_004	00F1	ntilde	ñ	2. Latin-1 Supplement
[kgk]_004	00F3	oacute	ó	2. Latin-1 Supplement
[kgk]_004	00F5	otilde	õ	2. Latin-1 Supplement
[kgk]_004	00FA	uacute	ú	2. Latin-1 Supplement
[kgk]_004	00FD	yacute	ý	2. Latin-1 Supplement
[kgk]_004	0128	ltilde	ĩ	3. Latin Ext.-A
[kgk]_004	0129	itilde	ı̇	3. Latin Ext.-A

arquivo	Unicode	Glyphname	Character	bloco
[kgk]_004	0168	Utilde	Ū	3. Latin Ext.-A
[kgk]_004	0169	utilde	ū	3. Latin Ext.-A
[kgk]_004	02BC	apostrophemod	'	6. Spc. Mod. Letters
[kgk]_004	1EBC	Etilde	Ě	8. Latin Ext. Add.
[kgk]_004	1EBD	etilde	ě	8. Latin Ext. Add.
[kgk]_004	1EF8	Ytilde	Ÿ	8. Latin Ext. Add.
[kgk]_004	1EF9	ytilde	ÿ	8. Latin Ext. Add.
[kgk]_004	A78B	Saltillo	ı	9. Latin Ext.-D
[kgk]_004	A78C	salttillo	ı	9. Latin Ext.-D
[kgk]_004	0047_0303	G_tildecomb	Ĝ	10. non-compliant
[kgk]_004	0067_0303	g_tildecomb	ĝ	10. non-compliant
[kgk]_005	0027	quotesingle	'	1. Basic Latin
[kgk]_005	00C1	Aacute	Á	2. Latin-1 Supplement
[kgk]_005	00C3	Atilde	Ã	2. Latin-1 Supplement
[kgk]_005	00C9	Eacute	É	2. Latin-1 Supplement
[kgk]_005	00D3	Oacute	Ó	2. Latin-1 Supplement
[kgk]_005	00D5	Otilde	Õ	2. Latin-1 Supplement
[kgk]_005	00DA	Uacute	Ú	2. Latin-1 Supplement
[kgk]_005	00DD	Yacute	Ý	2. Latin-1 Supplement
[kgk]_005	00E1	aacute	á	2. Latin-1 Supplement
[kgk]_005	00E3	atilde	ã	2. Latin-1 Supplement
[kgk]_005	00E9	eacute	é	2. Latin-1 Supplement
[kgk]_005	00F3	oacute	ó	2. Latin-1 Supplement
[kgk]_005	00F5	otilde	õ	2. Latin-1 Supplement
[kgk]_005	00FA	uacute	ú	2. Latin-1 Supplement
[kgk]_005	00FD	yacute	ý	2. Latin-1 Supplement
[kgk]_005	0128	Itilde	Ī	3. Latin Ext.-A
[kgk]_005	0129	itilde	ī	3. Latin Ext.-A
[kgk]_005	0168	Utilde	Ū	3. Latin Ext.-A
[kgk]_005	0169	utilde	ū	3. Latin Ext.-A
[kgk]_005	1EBC	Etilde	Ě	8. Latin Ext. Add.
[kgk]_005	1EBD	etilde	ě	8. Latin Ext. Add.
[kgk]_005	1EF8	Ytilde	Ÿ	8. Latin Ext. Add.
[kgk]_005	1EF9	ytilde	ÿ	8. Latin Ext. Add.
[kgrp]_001	00C1	Aacute	Á	2. Latin-1 Supplement
[kgrp]_001	00C3	Atilde	Ã	2. Latin-1 Supplement
[kgrp]_001	00C9	Eacute	É	2. Latin-1 Supplement
[kgrp]_001	00D3	Oacute	Ó	2. Latin-1 Supplement
[kgrp]_001	00E1	aacute	á	2. Latin-1 Supplement
[kgrp]_001	00E3	atilde	ã	2. Latin-1 Supplement
[kgrp]_001	00E9	eacute	é	2. Latin-1 Supplement
[kgrp]_001	00F3	oacute	ó	2. Latin-1 Supplement
[kgrp]_001	0128	Itilde	Ī	3. Latin Ext.-A
[kgrp]_001	0129	itilde	ī	3. Latin Ext.-A
[kgrp]_001	0168	Utilde	Ū	3. Latin Ext.-A
[kgrp]_001	0169	utilde	ū	3. Latin Ext.-A

arquivo	Unicode	Glyphname	Character	bloco
[kgrp]_001	1EBC	Etilde	Ě	8. Latin Ext. Add.
[kgrp]_001	1EBD	etilde	ě	8. Latin Ext. Add.
[kgrp]_001	1EF8	Ytilde	Ÿ	8. Latin Ext. Add.
[kgrp]_001	1EF9	ytilde	ÿ	8. Latin Ext. Add.
[kgrp]_0010	0027	quotesingle	'	1. Basic Latin
[kgrp]_0010	00C1	Aacute	Á	2. Latin-1 Supplement
[kgrp]_0010	00C3	Atilde	Ã	2. Latin-1 Supplement
[kgrp]_0010	00C9	Eacute	É	2. Latin-1 Supplement
[kgrp]_0010	00D3	Oacute	Ó	2. Latin-1 Supplement
[kgrp]_0010	00E1	aacute	á	2. Latin-1 Supplement
[kgrp]_0010	00E3	atilde	ã	2. Latin-1 Supplement
[kgrp]_0010	00E9	eacute	é	2. Latin-1 Supplement
[kgrp]_0010	00F3	oacute	ó	2. Latin-1 Supplement
[kgrp]_0010	0128	Itilde	İ	3. Latin Ext.-A
[kgrp]_0010	0129	itilde	ı	3. Latin Ext.-A
[kgrp]_0010	0168	Utilde	Ū	3. Latin Ext.-A
[kgrp]_0010	0169	utilde	ū	3. Latin Ext.-A
[kgrp]_0010	1EBC	Etilde	Ě	8. Latin Ext. Add.
[kgrp]_0010	1EBD	etilde	ě	8. Latin Ext. Add.
[kgrp]_0010	1EF8	Ytilde	Ÿ	8. Latin Ext. Add.
[kgrp]_0010	1EF9	ytilde	ÿ	8. Latin Ext. Add.
[kgrp]_0011	0027	quotesingle	'	1. Basic Latin
[kgrp]_0011	00C1	Aacute	Á	2. Latin-1 Supplement
[kgrp]_0011	00C3	Atilde	Ã	2. Latin-1 Supplement
[kgrp]_0011	00C9	Eacute	É	2. Latin-1 Supplement
[kgrp]_0011	00D3	Oacute	Ó	2. Latin-1 Supplement
[kgrp]_0011	00E1	aacute	á	2. Latin-1 Supplement
[kgrp]_0011	00E3	atilde	ã	2. Latin-1 Supplement
[kgrp]_0011	00E9	eacute	é	2. Latin-1 Supplement
[kgrp]_0011	00F3	oacute	ó	2. Latin-1 Supplement
[kgrp]_0011	0128	Itilde	İ	3. Latin Ext.-A
[kgrp]_0011	0129	itilde	ı	3. Latin Ext.-A
[kgrp]_0011	0168	Utilde	Ū	3. Latin Ext.-A
[kgrp]_0011	0169	utilde	ū	3. Latin Ext.-A
[kgrp]_0011	1EBC	Etilde	Ě	8. Latin Ext. Add.
[kgrp]_0011	1EBD	etilde	ě	8. Latin Ext. Add.
[kgrp]_0011	1EF8	Ytilde	Ÿ	8. Latin Ext. Add.
[kgrp]_0011	1EF9	ytilde	ÿ	8. Latin Ext. Add.
[kgrp]_002	00C2	Acircumflex	Â	2. Latin-1 Supplement
[kgrp]_002	00C3	Atilde	Ã	2. Latin-1 Supplement
[kgrp]_002	00C9	Eacute	É	2. Latin-1 Supplement
[kgrp]_002	00D3	Oacute	Ó	2. Latin-1 Supplement
[kgrp]_002	00D5	Otilde	Õ	2. Latin-1 Supplement
[kgrp]_002	00E2	acircumflex	â	2. Latin-1 Supplement
[kgrp]_002	00E3	atilde	ã	2. Latin-1 Supplement
[kgrp]_002	00E9	eacute	é	2. Latin-1 Supplement
[kgrp]_002	00F3	oacute	ó	2. Latin-1 Supplement

arquivo	Unicode	Glyphname	Character	bloco
[kgrp]_002	00F5	otilde	õ	2. Latin-1 Supplement
[kgrp]_002	0128	ltilde	ĩ	3. Latin Ext.-A
[kgrp]_002	0129	itilde	ı̇	3. Latin Ext.-A
[kgrp]_002	1EBC	Etilde	Ě	8. Latin Ext. Add.
[kgrp]_002	1EBD	etilde	ě	8. Latin Ext. Add.
[kgrp]_003	00C1	Aacute	Á	2. Latin-1 Supplement
[kgrp]_003	00C3	Atilde	Ã	2. Latin-1 Supplement
[kgrp]_003	00C9	Eacute	É	2. Latin-1 Supplement
[kgrp]_003	00D3	Oacute	Ó	2. Latin-1 Supplement
[kgrp]_003	00E1	aacute	á	2. Latin-1 Supplement
[kgrp]_003	00E3	atilde	ã	2. Latin-1 Supplement
[kgrp]_003	00E9	eacute	é	2. Latin-1 Supplement
[kgrp]_003	00F3	oacute	ó	2. Latin-1 Supplement
[kgrp]_003	0128	ltilde	ĩ	3. Latin Ext.-A
[kgrp]_003	0129	itilde	ı̇	3. Latin Ext.-A
[kgrp]_003	0168	Utilde	Û	3. Latin Ext.-A
[kgrp]_003	0169	utilde	ü	3. Latin Ext.-A
[kgrp]_003	1EBC	Etilde	Ě	8. Latin Ext. Add.
[kgrp]_003	1EBD	etilde	ě	8. Latin Ext. Add.
[kgrp]_003	1EF8	Ytilde	Ÿ	8. Latin Ext. Add.
[kgrp]_003	1EF9	ytilde	ÿ	8. Latin Ext. Add.
[kgrp]_004	00C1	Aacute	Á	2. Latin-1 Supplement
[kgrp]_004	00C3	Atilde	Ã	2. Latin-1 Supplement
[kgrp]_004	00C9	Eacute	É	2. Latin-1 Supplement
[kgrp]_004	00D3	Oacute	Ó	2. Latin-1 Supplement
[kgrp]_004	00E1	aacute	á	2. Latin-1 Supplement
[kgrp]_004	00E3	atilde	ã	2. Latin-1 Supplement
[kgrp]_004	00E9	eacute	é	2. Latin-1 Supplement
[kgrp]_004	00F3	oacute	ó	2. Latin-1 Supplement
[kgrp]_004	0128	ltilde	ĩ	3. Latin Ext.-A
[kgrp]_004	0129	itilde	ı̇	3. Latin Ext.-A
[kgrp]_004	0168	Utilde	Û	3. Latin Ext.-A
[kgrp]_004	0169	utilde	ü	3. Latin Ext.-A
[kgrp]_004	02BC	apostrophemod	'	6. Spc. Mod. Letters
[kgrp]_004	1EBC	Etilde	Ě	8. Latin Ext. Add.
[kgrp]_004	1EBD	etilde	ě	8. Latin Ext. Add.
[kgrp]_004	1EF8	Ytilde	Ÿ	8. Latin Ext. Add.
[kgrp]_004	1EF9	ytilde	ÿ	8. Latin Ext. Add.
[kgrp]_004	A78B	Saltillo	ı̂	9. Latin Ext.-D
[kgrp]_004	A78C	salttillo	ı̃	9. Latin Ext.-D
[kgrp]_005	00C1	Aacute	Á	2. Latin-1 Supplement
[kgrp]_005	00C3	Atilde	Ã	2. Latin-1 Supplement
[kgrp]_005	00C9	Eacute	É	2. Latin-1 Supplement
[kgrp]_005	00D3	Oacute	Ó	2. Latin-1 Supplement
[kgrp]_005	00D5	Otilde	Õ	2. Latin-1 Supplement
[kgrp]_005	00E1	aacute	á	2. Latin-1 Supplement

arquivo	Unicode	Glyphname	Character	bloco
[kgp]_005	00E3	atilde	ã	2. Latin-1 Supplement
[kgp]_005	00E9	eacute	é	2. Latin-1 Supplement
[kgp]_005	00F3	oacute	ó	2. Latin-1 Supplement
[kgp]_005	00F5	otilde	õ	2. Latin-1 Supplement
[kgp]_005	0128	ltilde	ĩ	3. Latin Ext.-A
[kgp]_005	0129	itilde	ı̃	3. Latin Ext.-A
[kgp]_005	0168	Utilde	Ū	3. Latin Ext.-A
[kgp]_005	0169	utilde	ū	3. Latin Ext.-A
[kgp]_005	1EBC	Etilde	Ě	8. Latin Ext. Add.
[kgp]_005	1EBD	etilde	ě	8. Latin Ext. Add.
[kgp]_005	1EF8	Ytilde	Ÿ	8. Latin Ext. Add.
[kgp]_005	1EF9	ytilde	ÿ	8. Latin Ext. Add.
[kgp]_006	0027	quotesingle	'	1. Basic Latin
[kgp]_006	00C1	Aacute	Á	2. Latin-1 Supplement
[kgp]_006	00C3	Atilde	Ã	2. Latin-1 Supplement
[kgp]_006	00C9	Eacute	É	2. Latin-1 Supplement
[kgp]_006	00D3	Oacute	Ó	2. Latin-1 Supplement
[kgp]_006	00E1	aacute	á	2. Latin-1 Supplement
[kgp]_006	00E3	atilde	ã	2. Latin-1 Supplement
[kgp]_006	00E9	eacute	é	2. Latin-1 Supplement
[kgp]_006	00F3	oacute	ó	2. Latin-1 Supplement
[kgp]_006	0128	ltilde	ĩ	3. Latin Ext.-A
[kgp]_006	0129	itilde	ı̃	3. Latin Ext.-A
[kgp]_006	0168	Utilde	Ū	3. Latin Ext.-A
[kgp]_006	0169	utilde	ū	3. Latin Ext.-A
[kgp]_006	1EBC	Etilde	Ě	8. Latin Ext. Add.
[kgp]_006	1EBD	etilde	ě	8. Latin Ext. Add.
[kgp]_006	1EF8	Ytilde	Ÿ	8. Latin Ext. Add.
[kgp]_006	1EF9	ytilde	ÿ	8. Latin Ext. Add.
[kgp]_007	00C2	Acircumflex	Â	2. Latin-1 Supplement
[kgp]_007	00C3	Atilde	Ã	2. Latin-1 Supplement
[kgp]_007	00C9	Eacute	É	2. Latin-1 Supplement
[kgp]_007	00D3	Oacute	Ó	2. Latin-1 Supplement
[kgp]_007	00D5	Otilde	Õ	2. Latin-1 Supplement
[kgp]_007	00E2	acircumflex	â	2. Latin-1 Supplement
[kgp]_007	00E3	atilde	ã	2. Latin-1 Supplement
[kgp]_007	00E9	eacute	é	2. Latin-1 Supplement
[kgp]_007	00F3	oacute	ó	2. Latin-1 Supplement
[kgp]_007	00F5	otilde	õ	2. Latin-1 Supplement
[kgp]_007	0128	ltilde	ĩ	3. Latin Ext.-A
[kgp]_007	0129	itilde	ı̃	3. Latin Ext.-A
[kgp]_007	0168	Utilde	Ū	3. Latin Ext.-A
[kgp]_007	0169	utilde	ū	3. Latin Ext.-A
[kgp]_007	1EBC	Etilde	Ě	8. Latin Ext. Add.
[kgp]_007	1EBD	etilde	ě	8. Latin Ext. Add.
[kgp]_008	00C1	Aacute	Á	2. Latin-1 Supplement
[kgp]_008	00C3	Atilde	Ã	2. Latin-1 Supplement

arquivo	Unicode	Glyphname	Character	bloco
[kgp]_008	00C9	Eacute	É	2. Latin-1 Supplement
[kgp]_008	00D3	Oacute	Ó	2. Latin-1 Supplement
[kgp]_008	00E1	aacute	á	2. Latin-1 Supplement
[kgp]_008	00E3	atilde	ã	2. Latin-1 Supplement
[kgp]_008	00E9	eacute	é	2. Latin-1 Supplement
[kgp]_008	00F3	oacute	ó	2. Latin-1 Supplement
[kgp]_008	0128	Itilde	Ï	3. Latin Ext.-A
[kgp]_008	0129	itilde	ï	3. Latin Ext.-A
[kgp]_008	0168	Utilde	Û	3. Latin Ext.-A
[kgp]_008	0169	utilde	ü	3. Latin Ext.-A
[kgp]_008	1EBC	Etilde	Ě	8. Latin Ext. Add.
[kgp]_008	1EBD	etilde	ě	8. Latin Ext. Add.
[kgp]_008	1EF8	Ytilde	ÿ	8. Latin Ext. Add.
[kgp]_008	1EF9	ytilde	ÿ	8. Latin Ext. Add.
[kgp]_009	0027	quotesingle	'	1. Basic Latin
[kgp]_009	00C1	Aacute	Á	2. Latin-1 Supplement
[kgp]_009	00C3	Atilde	Ã	2. Latin-1 Supplement
[kgp]_009	00C9	Eacute	É	2. Latin-1 Supplement
[kgp]_009	00D3	Oacute	Ó	2. Latin-1 Supplement
[kgp]_009	00E1	aacute	á	2. Latin-1 Supplement
[kgp]_009	00E3	atilde	ã	2. Latin-1 Supplement
[kgp]_009	00E9	eacute	é	2. Latin-1 Supplement
[kgp]_009	00F3	oacute	ó	2. Latin-1 Supplement
[kgp]_009	0128	Itilde	Ï	3. Latin Ext.-A
[kgp]_009	0129	itilde	ï	3. Latin Ext.-A
[kgp]_009	0168	Utilde	Û	3. Latin Ext.-A
[kgp]_009	0169	utilde	ü	3. Latin Ext.-A
[kgp]_009	1EBC	Etilde	Ě	8. Latin Ext. Add.
[kgp]_009	1EBD	etilde	ě	8. Latin Ext. Add.
[kgp]_009	1EF8	Ytilde	ÿ	8. Latin Ext. Add.
[kgp]_009	1EF9	ytilde	ÿ	8. Latin Ext. Add.
[kgp]_009	2019	quoteright	'	11. General Punctuation
[kmv]_001	00C1	Aacute	Á	2. Latin-1 Supplement
[kmv]_001	00C3	Atilde	Ã	2. Latin-1 Supplement
[kmv]_001	00C9	Eacute	É	2. Latin-1 Supplement
[kmv]_001	00CD	Iacute	Í	2. Latin-1 Supplement
[kmv]_001	00D3	Oacute	Ó	2. Latin-1 Supplement
[kmv]_001	00D5	Otilde	Õ	2. Latin-1 Supplement
[kmv]_001	00DA	Uacute	Ú	2. Latin-1 Supplement
[kmv]_001	00E1	aacute	á	2. Latin-1 Supplement
[kmv]_001	00E3	atilde	ã	2. Latin-1 Supplement
[kmv]_001	00E9	eacute	é	2. Latin-1 Supplement
[kmv]_001	00ED	Iacute	í	2. Latin-1 Supplement
[kmv]_001	00F3	oacute	ó	2. Latin-1 Supplement
[kmv]_001	00F5	otilde	õ	2. Latin-1 Supplement
[kmv]_001	00FA	uacute	ú	2. Latin-1 Supplement
[kmv]_001	0128	Itilde	Ï	3. Latin Ext.-A

arquivo	Unicode	Glyphname	Character	bloco
[kmv]_001	0129	itilde	ĩ	3. Latin Ext.-A
[kmv]_001	0168	Utilde	Ū	3. Latin Ext.-A
[kmv]_001	0169	utilde	ū	3. Latin Ext.-A
[kmv]_001	1EBC	Etilde	Ě	8. Latin Ext. Add.
[kmv]_001	1EBD	etilde	ě	8. Latin Ext. Add.
[kmv]_003	00C3	Atilde	Ā	2. Latin-1 Supplement
[kmv]_003	00C9	Eacute	É	2. Latin-1 Supplement
[kmv]_003	00D3	Oacute	Ó	2. Latin-1 Supplement
[kmv]_003	00D5	Otilde	Õ	2. Latin-1 Supplement
[kmv]_003	00E3	atilde	ā	2. Latin-1 Supplement
[kmv]_003	00E9	eacute	é	2. Latin-1 Supplement
[kmv]_003	00F3	oacute	ó	2. Latin-1 Supplement
[kmv]_003	00F5	otilde	õ	2. Latin-1 Supplement
[kmv]_003	1EBC	Etilde	Ě	8. Latin Ext. Add.
[kmv]_003	1EBD	etilde	ě	8. Latin Ext. Add.
[knm]_001	00C1	Aacute	Á	2. Latin-1 Supplement
[knm]_001	00C2	Acircumflex	Â	2. Latin-1 Supplement
[knm]_001	00C9	Eacute	É	2. Latin-1 Supplement
[knm]_001	00CA	Ecircumflex	Ê	2. Latin-1 Supplement
[knm]_001	00D4	Ocircumflex	Ô	2. Latin-1 Supplement
[knm]_001	00DA	Uacute	Ú	2. Latin-1 Supplement
[knm]_001	00E1	aacute	á	2. Latin-1 Supplement
[knm]_001	00E2	acircumflex	â	2. Latin-1 Supplement
[knm]_001	00E9	eacute	é	2. Latin-1 Supplement
[knm]_001	00EA	ecircumflex	ê	2. Latin-1 Supplement
[knm]_001	00F4	ocircumflex	ô	2. Latin-1 Supplement
[knm]_001	00FA	uacute	ú	2. Latin-1 Supplement
[knt]_001	00C3	Atilde	Ā	2. Latin-1 Supplement
[knt]_001	00C4	Adieresis	Ä	2. Latin-1 Supplement
[knt]_001	00D5	Otilde	Õ	2. Latin-1 Supplement
[knt]_001	00E1	aacute	á	2. Latin-1 Supplement
[knt]_001	00E2	acircumflex	â	2. Latin-1 Supplement
[knt]_001	00E3	atilde	ā	2. Latin-1 Supplement
[knt]_001	00E4	adieresis	ä	2. Latin-1 Supplement
[knt]_001	00E9	eacute	é	2. Latin-1 Supplement
[knt]_001	00EA	ecircumflex	ê	2. Latin-1 Supplement
[knt]_001	00ED	iacute	í	2. Latin-1 Supplement
[knt]_001	00EE	icircumflex	î	2. Latin-1 Supplement
[knt]_001	00F3	oacute	ó	2. Latin-1 Supplement
[knt]_001	00F4	ocircumflex	ô	2. Latin-1 Supplement
[knt]_001	00F5	otilde	õ	2. Latin-1 Supplement
[knt]_001	00F6	odieresis	ö	2. Latin-1 Supplement
[knt]_001	0128	Itilde	Ī	3. Latin Ext.-A
[knt]_001	0129	itilde	ĩ	3. Latin Ext.-A
[knt]_001	02BC	apostrophemod	'	6. Spc. Mod. Letters
[knt]_001	1EBC	Etilde	Ě	8. Latin Ext. Add.
[knt]_001	1EBD	etilde	ě	8. Latin Ext. Add.

arquivo	Unicode	Glyphname	Character	bloco
[kpc]_001	00C3	Atilde	Ã	2. Latin-1 Supplement
[kpc]_001	00D1	Ntilde	Ñ	2. Latin-1 Supplement
[kpc]_001	00E3	atilde	ã	2. Latin-1 Supplement
[kpc]_001	00F1	ntilde	ñ	2. Latin-1 Supplement
[kpc]_002	00C1	Aacute	Á	2. Latin-1 Supplement
[kpc]_002	00C3	Atilde	Ã	2. Latin-1 Supplement
[kpc]_002	00C9	Eacute	É	2. Latin-1 Supplement
[kpc]_002	00CD	iacute	Í	2. Latin-1 Supplement
[kpc]_002	00D1	Ntilde	Ñ	2. Latin-1 Supplement
[kpc]_002	00D3	Oacute	Ó	2. Latin-1 Supplement
[kpc]_002	00E1	aacute	á	2. Latin-1 Supplement
[kpc]_002	00E3	atilde	ã	2. Latin-1 Supplement
[kpc]_002	00E9	eacute	é	2. Latin-1 Supplement
[kpc]_002	00ED	iacute	í	2. Latin-1 Supplement
[kpc]_002	00F1	ntilde	ñ	2. Latin-1 Supplement
[kpc]_002	00F3	oacute	ó	2. Latin-1 Supplement
[kpij]_001	00C0	Agrave	À	2. Latin-1 Supplement
[kpij]_001	00C3	Atilde	Ã	2. Latin-1 Supplement
[kpij]_001	00C8	Egrave	È	2. Latin-1 Supplement
[kpij]_001	00D2	Ograve	Ò	2. Latin-1 Supplement
[kpij]_001	00D5	Otilde	Õ	2. Latin-1 Supplement
[kpij]_001	00D9	Ugrave	Ù	2. Latin-1 Supplement
[kpij]_001	00E0	agrave	à	2. Latin-1 Supplement
[kpij]_001	00E3	atilde	ã	2. Latin-1 Supplement
[kpij]_001	00E8	egrave	è	2. Latin-1 Supplement
[kpij]_001	00F2	ograve	ò	2. Latin-1 Supplement
[kpij]_001	00F5	otilde	õ	2. Latin-1 Supplement
[kpij]_001	00F9	ugrave	ù	2. Latin-1 Supplement
[kpij]_001	0128	ltilde	Ĺ	3. Latin Ext.-A
[kpij]_001	0129	itilde	ĺ	3. Latin Ext.-A
[kpij]_001	1EF8	Ytilde	Ÿ	8. Latin Ext. Add.
[kpij]_001	1EF9	ytilde	ÿ	8. Latin Ext. Add.
[kpij]_002	00C0	Agrave	À	2. Latin-1 Supplement
[kpij]_002	00C3	Atilde	Ã	2. Latin-1 Supplement
[kpij]_002	00C8	Egrave	È	2. Latin-1 Supplement
[kpij]_002	00D2	Ograve	Ò	2. Latin-1 Supplement
[kpij]_002	00D5	Otilde	Õ	2. Latin-1 Supplement
[kpij]_002	00D9	Ugrave	Ù	2. Latin-1 Supplement
[kpij]_002	00E0	agrave	à	2. Latin-1 Supplement
[kpij]_002	00E3	atilde	ã	2. Latin-1 Supplement
[kpij]_002	00E8	egrave	è	2. Latin-1 Supplement
[kpij]_002	00F2	ograve	ò	2. Latin-1 Supplement
[kpij]_002	00F5	otilde	õ	2. Latin-1 Supplement
[kpij]_002	00F9	ugrave	ù	2. Latin-1 Supplement
[kpij]_002	0128	ltilde	Ĺ	3. Latin Ext.-A
[kpij]_002	0129	itilde	ĺ	3. Latin Ext.-A
[kpij]_002	1EF8	Ytilde	Ÿ	8. Latin Ext. Add.

arquivo	Unicode	Glyphname	Character	bloco
[kpj]_002	1EF9	ytilde	ÿ	8. Latin Ext. Add.
[kpj]_003	00C0	Agrave	À	2. Latin-1 Supplement
[kpj]_003	00C3	Atilde	Ã	2. Latin-1 Supplement
[kpj]_003	00C8	Egrave	È	2. Latin-1 Supplement
[kpj]_003	00D2	Ograve	Ò	2. Latin-1 Supplement
[kpj]_003	00D5	Otilde	Õ	2. Latin-1 Supplement
[kpj]_003	00D9	Ugrave	Ù	2. Latin-1 Supplement
[kpj]_003	00E0	agrave	à	2. Latin-1 Supplement
[kpj]_003	00E3	atilde	ã	2. Latin-1 Supplement
[kpj]_003	00E8	egrave	è	2. Latin-1 Supplement
[kpj]_003	00F2	ograve	ò	2. Latin-1 Supplement
[kpj]_003	00F5	otilde	õ	2. Latin-1 Supplement
[kpj]_003	00F9	ugrave	ù	2. Latin-1 Supplement
[kpj]_003	0128	ltilde	ĺ	3. Latin Ext.-A
[kpj]_003	0129	itilde	ı	3. Latin Ext.-A
[kpj]_003	1EF8	Ytilde	ÿ	8. Latin Ext. Add.
[kpj]_003	1EF9	ytilde	ÿ	8. Latin Ext. Add.
[kre]_001	00C2	Acircumflex	Â	2. Latin-1 Supplement
[kre]_001	00C3	Atilde	Ã	2. Latin-1 Supplement
[kre]_001	00CA	Ecircumflex	Ê	2. Latin-1 Supplement
[kre]_001	00D4	Ocircumflex	Ô	2. Latin-1 Supplement
[kre]_001	00D5	Otilde	Õ	2. Latin-1 Supplement
[kre]_001	00E2	acircumflex	â	2. Latin-1 Supplement
[kre]_001	00E3	atilde	ã	2. Latin-1 Supplement
[kre]_001	00EA	ecircumflex	ê	2. Latin-1 Supplement
[kre]_001	00F4	ocircumflex	ô	2. Latin-1 Supplement
[kre]_001	00F5	otilde	õ	2. Latin-1 Supplement
[kre]_001	0128	ltilde	ĺ	3. Latin Ext.-A
[kre]_001	0129	itilde	ı	3. Latin Ext.-A
[kre]_001	0168	Utilde	Ů	3. Latin Ext.-A
[kre]_001	0169	utilde	ů	3. Latin Ext.-A
[kre]_001	1EBC	Etilde	Ě	8. Latin Ext. Add.
[kre]_001	1EBD	etilde	ě	8. Latin Ext. Add.
[kre]_002	00C2	Acircumflex	Â	2. Latin-1 Supplement
[kre]_002	00C3	Atilde	Ã	2. Latin-1 Supplement
[kre]_002	00CA	Ecircumflex	Ê	2. Latin-1 Supplement
[kre]_002	00D4	Ocircumflex	Ô	2. Latin-1 Supplement
[kre]_002	00D5	Otilde	Õ	2. Latin-1 Supplement
[kre]_002	00E2	acircumflex	â	2. Latin-1 Supplement
[kre]_002	00E3	atilde	ã	2. Latin-1 Supplement
[kre]_002	00EA	ecircumflex	ê	2. Latin-1 Supplement
[kre]_002	00F4	ocircumflex	ô	2. Latin-1 Supplement
[kre]_002	00F5	otilde	õ	2. Latin-1 Supplement
[kre]_002	0128	ltilde	ĺ	3. Latin Ext.-A
[kre]_002	0129	itilde	ı	3. Latin Ext.-A
[kre]_002	0168	Utilde	Ů	3. Latin Ext.-A
[kre]_002	0169	utilde	ů	3. Latin Ext.-A

arquivo	Unicode	Glyphname	Character	bloco
[kre]_002	1EBC	Etilde	Ě	8. Latin Ext. Add.
[kre]_002	1EBD	etilde	ě	8. Latin Ext. Add.
[ktn]_001	0027	quotesingle	'	1. Basic Latin
[ktn]_001	00C3	Atilde	Ã	2. Latin-1 Supplement
[ktn]_001	00D5	Otilde	Õ	2. Latin-1 Supplement
[ktn]_001	00E3	atilde	ã	2. Latin-1 Supplement
[ktn]_001	00F5	otilde	õ	2. Latin-1 Supplement
[ktn]_001	0128	ltilde	Ĺ	3. Latin Ext.-A
[ktn]_001	0129	itilde	ĺ	3. Latin Ext.-A
[ktn]_001	0303	tildecomb	˜	7. Combining Marks
[ktn]_001	1EBC	Etilde	Ě	8. Latin Ext. Add.
[ktn]_001	1EBD	etilde	ě	8. Latin Ext. Add.
[ktn]_001	1EF8	Ytilde	Ÿ	8. Latin Ext. Add.
[ktn]_001	1EF9	ytilde	ÿ	8. Latin Ext. Add.
[ktn]_001	004A_0303	J_tildecomb	Ĵ	10. non-compliant
[ktn]_001	006A_0303	j_tildecomb	ĵ	10. non-compliant
[ktn]_002	0027	quotesingle	'	1. Basic Latin
[ktn]_002	00C3	Atilde	Ã	2. Latin-1 Supplement
[ktn]_002	00D5	Otilde	Õ	2. Latin-1 Supplement
[ktn]_002	00E3	atilde	ã	2. Latin-1 Supplement
[ktn]_002	00F5	otilde	õ	2. Latin-1 Supplement
[ktn]_002	0128	ltilde	Ĺ	3. Latin Ext.-A
[ktn]_002	0129	itilde	ĺ	3. Latin Ext.-A
[ktn]_002	1EBC	Etilde	Ě	8. Latin Ext. Add.
[ktn]_002	1EBD	etilde	ě	8. Latin Ext. Add.
[ktn]_002	1EF8	Ytilde	Ÿ	8. Latin Ext. Add.
[ktn]_002	1EF9	ytilde	ÿ	8. Latin Ext. Add.
[kui]_001	00C1	Aacute	Á	2. Latin-1 Supplement
[kui]_001	00C3	Atilde	Ã	2. Latin-1 Supplement
[kui]_001	00C9	Eacute	É	2. Latin-1 Supplement
[kui]_001	00CD	Iacute	Í	2. Latin-1 Supplement
[kui]_001	00D3	Oacute	Ó	2. Latin-1 Supplement
[kui]_001	00D5	Otilde	Õ	2. Latin-1 Supplement
[kui]_001	00DA	Uacute	Ú	2. Latin-1 Supplement
[kui]_001	00DC	Udieresis	Ü	2. Latin-1 Supplement
[kui]_001	00E1	aacute	á	2. Latin-1 Supplement
[kui]_001	00E3	atilde	ã	2. Latin-1 Supplement
[kui]_001	00E9	eacute	é	2. Latin-1 Supplement
[kui]_001	00ED	iacute	í	2. Latin-1 Supplement
[kui]_001	00F3	oacute	ó	2. Latin-1 Supplement
[kui]_001	00F5	otilde	õ	2. Latin-1 Supplement
[kui]_001	00FA	uacute	ú	2. Latin-1 Supplement
[kui]_001	00FC	udieresis	ü	2. Latin-1 Supplement
[kui]_001	0128	ltilde	Ĺ	3. Latin Ext.-A
[kui]_001	0129	itilde	ĺ	3. Latin Ext.-A
[kui]_001	0168	Utilde	Û	3. Latin Ext.-A
[kui]_001	0169	utilde	ü	3. Latin Ext.-A

arquivo	Unicode	Glyphname	Character	bloco
[kui]_001	1EBC	Etilde	Ě	8. Latin Ext. Add.
[kui]_001	1EBD	etilde	ě	8. Latin Ext. Add.
[kui]_002	00C1	Aacute	Á	2. Latin-1 Supplement
[kui]_002	00C3	Atilde	Ã	2. Latin-1 Supplement
[kui]_002	00C9	Eacute	É	2. Latin-1 Supplement
[kui]_002	00CD	iacute	Í	2. Latin-1 Supplement
[kui]_002	00D3	Oacute	Ó	2. Latin-1 Supplement
[kui]_002	00D5	Otilde	Õ	2. Latin-1 Supplement
[kui]_002	00DA	Uacute	Ú	2. Latin-1 Supplement
[kui]_002	00DC	Udieresis	Ü	2. Latin-1 Supplement
[kui]_002	00E1	aacute	á	2. Latin-1 Supplement
[kui]_002	00E3	atilde	ã	2. Latin-1 Supplement
[kui]_002	00E9	eacute	é	2. Latin-1 Supplement
[kui]_002	00ED	iacute	í	2. Latin-1 Supplement
[kui]_002	00F3	oacute	ó	2. Latin-1 Supplement
[kui]_002	00F5	otilde	õ	2. Latin-1 Supplement
[kui]_002	00FA	uacute	ú	2. Latin-1 Supplement
[kui]_002	00FC	udieresis	ü	2. Latin-1 Supplement
[kui]_002	0128	ltilde	Ĺ	3. Latin Ext.-A
[kui]_002	0129	itilde	ĺ	3. Latin Ext.-A
[kui]_002	0168	Utilde	Ů	3. Latin Ext.-A
[kui]_002	0169	utilde	ů	3. Latin Ext.-A
[kui]_002	1EBC	Etilde	Ě	8. Latin Ext. Add.
[kui]_002	1EBD	etilde	ě	8. Latin Ext. Add.
[kui]_003	00C1	Aacute	Á	2. Latin-1 Supplement
[kui]_003	00C3	Atilde	Ã	2. Latin-1 Supplement
[kui]_003	00C9	Eacute	É	2. Latin-1 Supplement
[kui]_003	00CD	iacute	Í	2. Latin-1 Supplement
[kui]_003	00D3	Oacute	Ó	2. Latin-1 Supplement
[kui]_003	00D5	Otilde	Õ	2. Latin-1 Supplement
[kui]_003	00DA	Uacute	Ú	2. Latin-1 Supplement
[kui]_003	00DC	Udieresis	Ü	2. Latin-1 Supplement
[kui]_003	00E1	aacute	á	2. Latin-1 Supplement
[kui]_003	00E3	atilde	ã	2. Latin-1 Supplement
[kui]_003	00E9	eacute	é	2. Latin-1 Supplement
[kui]_003	00ED	iacute	í	2. Latin-1 Supplement
[kui]_003	00F3	oacute	ó	2. Latin-1 Supplement
[kui]_003	00F5	otilde	õ	2. Latin-1 Supplement
[kui]_003	00FA	uacute	ú	2. Latin-1 Supplement
[kui]_003	00FC	udieresis	ü	2. Latin-1 Supplement
[kui]_003	0128	ltilde	Ĺ	3. Latin Ext.-A
[kui]_003	0129	itilde	ĺ	3. Latin Ext.-A
[kui]_003	0168	Utilde	Ů	3. Latin Ext.-A
[kui]_003	0169	utilde	ů	3. Latin Ext.-A
[kui]_003	1EBC	Etilde	Ě	8. Latin Ext. Add.
[kui]_003	1EBD	etilde	ě	8. Latin Ext. Add.
[kui]_004	00C1	Aacute	Á	2. Latin-1 Supplement

arquivo	Unicode	Glyphname	Character	bloco
[kui]_004	00C2	Acircumflex	Â	2. Latin-1 Supplement
[kui]_004	00C3	Atilde	Ã	2. Latin-1 Supplement
[kui]_004	00C9	Eacute	É	2. Latin-1 Supplement
[kui]_004	00CD	Iacute	Í	2. Latin-1 Supplement
[kui]_004	00D1	Ntilde	Ñ	2. Latin-1 Supplement
[kui]_004	00D3	Oacute	Ó	2. Latin-1 Supplement
[kui]_004	00DA	Uacute	Ú	2. Latin-1 Supplement
[kui]_004	00DC	Udieresis	Ü	2. Latin-1 Supplement
[kui]_004	00E1	aacute	á	2. Latin-1 Supplement
[kui]_004	00E2	acircumflex	â	2. Latin-1 Supplement
[kui]_004	00E3	atilde	ã	2. Latin-1 Supplement
[kui]_004	00E9	eacute	é	2. Latin-1 Supplement
[kui]_004	00ED	iacute	í	2. Latin-1 Supplement
[kui]_004	00F1	ntilde	ñ	2. Latin-1 Supplement
[kui]_004	00F3	oacute	ó	2. Latin-1 Supplement
[kui]_004	00FA	uacute	ú	2. Latin-1 Supplement
[kui]_004	00FC	udieresis	ü	2. Latin-1 Supplement
[kui]_005	00C3	Atilde	Ã	2. Latin-1 Supplement
[kui]_005	00C9	Eacute	É	2. Latin-1 Supplement
[kui]_005	00D5	Otilde	Õ	2. Latin-1 Supplement
[kui]_005	00DC	Udieresis	Ü	2. Latin-1 Supplement
[kui]_005	00E3	atilde	ã	2. Latin-1 Supplement
[kui]_005	00E9	eacute	é	2. Latin-1 Supplement
[kui]_005	00F5	otilde	õ	2. Latin-1 Supplement
[kui]_005	00FC	udieresis	ü	2. Latin-1 Supplement
[kui]_005	0128	Itilde	İ	3. Latin Ext.-A
[kui]_005	0129	itilde	ı	3. Latin Ext.-A
[kui]_005	0168	Utilde	Ū	3. Latin Ext.-A
[kui]_005	0169	utilde	ū	3. Latin Ext.-A
[kui]_005	1EBC	Etilde	Ě	8. Latin Ext. Add.
[kui]_005	1EBD	etilde	ě	8. Latin Ext. Add.
[kui]_006	00C1	Aacute	Á	2. Latin-1 Supplement
[kui]_006	00C3	Atilde	Ã	2. Latin-1 Supplement
[kui]_006	00C9	Eacute	É	2. Latin-1 Supplement
[kui]_006	00D5	Otilde	Õ	2. Latin-1 Supplement
[kui]_006	00DC	Udieresis	Ü	2. Latin-1 Supplement
[kui]_006	00E1	aacute	á	2. Latin-1 Supplement
[kui]_006	00E3	atilde	ã	2. Latin-1 Supplement
[kui]_006	00E9	eacute	é	2. Latin-1 Supplement
[kui]_006	00F5	otilde	õ	2. Latin-1 Supplement
[kui]_006	00FC	udieresis	ü	2. Latin-1 Supplement
[kui]_006	0128	Itilde	İ	3. Latin Ext.-A
[kui]_006	0129	itilde	ı	3. Latin Ext.-A
[kui]_006	0168	Utilde	Ū	3. Latin Ext.-A
[kui]_006	0169	utilde	ū	3. Latin Ext.-A
[kui]_006	1EBC	Etilde	Ě	8. Latin Ext. Add.
[kui]_006	1EBD	etilde	ě	8. Latin Ext. Add.

arquivo	Unicode	Glyphname	Character	bloco
[kui]_007	00C1	Aacute	Á	2. Latin-1 Supplement
[kui]_007	00C3	Atilde	Ã	2. Latin-1 Supplement
[kui]_007	00C9	Eacute	É	2. Latin-1 Supplement
[kui]_007	00D3	Oacute	Ó	2. Latin-1 Supplement
[kui]_007	00D5	Otilde	Õ	2. Latin-1 Supplement
[kui]_007	00DC	Udieresis	Ü	2. Latin-1 Supplement
[kui]_007	00E1	aacute	á	2. Latin-1 Supplement
[kui]_007	00E3	atilde	ã	2. Latin-1 Supplement
[kui]_007	00E9	eacute	é	2. Latin-1 Supplement
[kui]_007	00F3	oacute	ó	2. Latin-1 Supplement
[kui]_007	00F5	otilde	õ	2. Latin-1 Supplement
[kui]_007	00FC	udieresis	ü	2. Latin-1 Supplement
[kui]_007	0128	ltilde	İ	3. Latin Ext.-A
[kui]_007	0129	itilde	ı	3. Latin Ext.-A
[kui]_007	0168	Utilde	Ū	3. Latin Ext.-A
[kui]_007	0169	utilde	ū	3. Latin Ext.-A
[kui]_007	1EBC	Etilde	Ě	8. Latin Ext. Add.
[kui]_007	1EBD	etilde	ě	8. Latin Ext. Add.
[kui]_008	00C3	Atilde	Ã	2. Latin-1 Supplement
[kui]_008	00DC	Udieresis	Ü	2. Latin-1 Supplement
[kui]_008	00E3	atilde	ã	2. Latin-1 Supplement
[kui]_008	00FC	udieresis	ü	2. Latin-1 Supplement
[kuq]_001	0027	quotesingle	'	1. Basic Latin
[kuq]_001	00C3	Atilde	Ã	2. Latin-1 Supplement
[kuq]_001	00D5	Otilde	Õ	2. Latin-1 Supplement
[kuq]_001	00E3	atilde	ã	2. Latin-1 Supplement
[kuq]_001	00F5	otilde	õ	2. Latin-1 Supplement
[kuq]_001	0128	ltilde	İ	3. Latin Ext.-A
[kuq]_001	0129	itilde	ı	3. Latin Ext.-A
[kuq]_001	0168	Utilde	Ū	3. Latin Ext.-A
[kuq]_001	0169	utilde	ū	3. Latin Ext.-A
[kuq]_001	1EBC	Etilde	Ě	8. Latin Ext. Add.
[kuq]_001	1EBD	etilde	ě	8. Latin Ext. Add.
[kuq]_001	1EF8	Ytilde	Ÿ	8. Latin Ext. Add.
[kuq]_001	1EF9	ytilde	ÿ	8. Latin Ext. Add.
[kyz]_001	00C3	Atilde	Ã	2. Latin-1 Supplement
[kyz]_001	00D5	Otilde	Õ	2. Latin-1 Supplement
[kyz]_001	00E3	atilde	ã	2. Latin-1 Supplement
[kyz]_001	00F5	otilde	õ	2. Latin-1 Supplement
[kyz]_001	0128	ltilde	İ	3. Latin Ext.-A
[kyz]_001	0129	itilde	ı	3. Latin Ext.-A
[kyz]_001	0168	Utilde	Ū	3. Latin Ext.-A
[kyz]_001	0169	utilde	ū	3. Latin Ext.-A
[kyz]_001	02BC	apostrophemod	'	6. Spc. Mod. Letters
[kyz]_001	1EBC	Etilde	Ě	8. Latin Ext. Add.
[kyz]_001	1EBD	etilde	ě	8. Latin Ext. Add.
[kyz]_001	1EF8	Ytilde	Ÿ	8. Latin Ext. Add.

arquivo	Unicode	Glyphname	Character	bloco
[kyz]_001	1EF9	ytilde	ÿ	8. Latin Ext. Add.
[kyz]_001	A78B	Saltillo	ı	9. Latin Ext.-D
[kyz]_001	A78C	salttillo	ı	9. Latin Ext.-D
[kyz]_002	00C2	Acircumflex	Â	2. Latin-1 Supplement
[kyz]_002	00C3	Atilde	Ã	2. Latin-1 Supplement
[kyz]_002	00CC	Igrave	ì	2. Latin-1 Supplement
[kyz]_002	00D5	Otilde	Õ	2. Latin-1 Supplement
[kyz]_002	00D9	Ugrave	Ù	2. Latin-1 Supplement
[kyz]_002	00E2	acircumflex	â	2. Latin-1 Supplement
[kyz]_002	00E3	atilde	ã	2. Latin-1 Supplement
[kyz]_002	00EC	igrave	ì	2. Latin-1 Supplement
[kyz]_002	00F5	otilde	õ	2. Latin-1 Supplement
[kyz]_002	00F9	ugrave	ù	2. Latin-1 Supplement
[kyz]_002	0128	ltilde	ĺ	3. Latin Ext.-A
[kyz]_002	0129	itilde	ı	3. Latin Ext.-A
[kyz]_002	0168	Utilde	Ū	3. Latin Ext.-A
[kyz]_002	0169	utilde	ū	3. Latin Ext.-A
[kyz]_002	02BC	apostrophemod	'	6. Spc. Mod. Letters
[kyz]_002	1EBC	Etilde	Ě	8. Latin Ext. Add.
[kyz]_002	1EBD	etilde	ě	8. Latin Ext. Add.
[kyz]_002	A78B	Saltillo	ı	9. Latin Ext.-D
[kyz]_002	A78C	salttillo	ı	9. Latin Ext.-D
[kyz]_003	00C3	Atilde	Ã	2. Latin-1 Supplement
[kyz]_003	00D5	Otilde	Õ	2. Latin-1 Supplement
[kyz]_003	00E3	atilde	ã	2. Latin-1 Supplement
[kyz]_003	00F5	otilde	õ	2. Latin-1 Supplement
[kyz]_003	0128	ltilde	ĺ	3. Latin Ext.-A
[kyz]_003	0129	itilde	ı	3. Latin Ext.-A
[kyz]_003	0168	Utilde	Ū	3. Latin Ext.-A
[kyz]_003	0169	utilde	ū	3. Latin Ext.-A
[kyz]_003	02BC	apostrophemod	'	6. Spc. Mod. Letters
[kyz]_003	1EBC	Etilde	Ě	8. Latin Ext. Add.
[kyz]_003	1EBD	etilde	ě	8. Latin Ext. Add.
[kyz]_003	A78B	Saltillo	ı	9. Latin Ext.-D
[kyz]_003	A78C	salttillo	ı	9. Latin Ext.-D
[kyz]_004	0027	quotesingle	'	1. Basic Latin
[kyz]_004	00C3	Atilde	Ã	2. Latin-1 Supplement
[kyz]_004	00D5	Otilde	Õ	2. Latin-1 Supplement
[kyz]_004	00E3	atilde	ã	2. Latin-1 Supplement
[kyz]_004	00F5	otilde	õ	2. Latin-1 Supplement
[kyz]_004	0128	ltilde	ĺ	3. Latin Ext.-A
[kyz]_004	0129	itilde	ı	3. Latin Ext.-A
[kyz]_004	0168	Utilde	Ū	3. Latin Ext.-A
[kyz]_004	0169	utilde	ū	3. Latin Ext.-A
[kyz]_004	1EBC	Etilde	Ě	8. Latin Ext. Add.

arquivo	Unicode	Glyphname	Character	bloco
[kyz]_004	1EBD	etilde	ě	8. Latin Ext. Add.
[kyz]_004	1EF8	Ytilde	Ÿ	8. Latin Ext. Add.
[kyz]_004	1EF9	ytilde	ÿ	8. Latin Ext. Add.
[mav]_001	00C3	Atilde	Ã	2. Latin-1 Supplement
[mav]_001	00E3	atilde	ã	2. Latin-1 Supplement
[mav]_001	0100	Amacron	Ā	3. Latin Ext.-A
[mav]_001	0101	amacron	ā	3. Latin Ext.-A
[mav]_001	0112	Emacron	Ē	3. Latin Ext.-A
[mav]_001	0113	emacron	ē	3. Latin Ext.-A
[mav]_001	0128	Itilde	Ī	3. Latin Ext.-A
[mav]_001	0129	itilde	ī	3. Latin Ext.-A
[mav]_001	012A	Imacron	Ī	3. Latin Ext.-A
[mav]_001	012B	imacron	ī	3. Latin Ext.-A
[mav]_001	014C	Omacron	Ō	3. Latin Ext.-A
[mav]_001	014D	omacron	ō	3. Latin Ext.-A
[mav]_001	016A	Umacron	Ū	3. Latin Ext.-A
[mav]_001	016B	umacron	ū	3. Latin Ext.-A
[mav]_001	0232	Ymacron	Ŷ	4. Latin Ext.-B
[mav]_001	0233	ymacron	ÿ	4. Latin Ext.-B
[mav]_001	02BC	apostrophemod	'	6. Spc. Mod. Letters
[mav]_001	1EBC	Etilde	Ě	8. Latin Ext. Add.
[mav]_001	1EBD	etilde	ě	8. Latin Ext. Add.
[mav]_001	A78B	Saltillo	ı	9. Latin Ext.-D
[mav]_001	A78C	salttillo	ı	9. Latin Ext.-D
[mav]_001	0047_0303	G_tildecomb	Ĝ	10. non-compliant
[mav]_001	0067_0303	g_tildecomb	ĝ	10. non-compliant
[mav]_002	0100	Amacron	Ā	3. Latin Ext.-A
[mav]_002	0101	amacron	ā	3. Latin Ext.-A
[mav]_002	0112	Emacron	Ē	3. Latin Ext.-A
[mav]_002	0113	emacron	ē	3. Latin Ext.-A
[mav]_002	012A	Imacron	Ī	3. Latin Ext.-A
[mav]_002	012B	imacron	ī	3. Latin Ext.-A
[mav]_002	014C	Omacron	Ō	3. Latin Ext.-A
[mav]_002	014D	omacron	ō	3. Latin Ext.-A
[mav]_002	016A	Umacron	Ū	3. Latin Ext.-A
[mav]_002	016B	umacron	ū	3. Latin Ext.-A
[mav]_002	0232	Ymacron	Ŷ	4. Latin Ext.-B
[mav]_002	0233	ymacron	ÿ	4. Latin Ext.-B
[mav]_002	02BC	apostrophemod	'	6. Spc. Mod. Letters
[mav]_002	A78B	Saltillo	ı	9. Latin Ext.-D
[mav]_002	A78C	salttillo	ı	9. Latin Ext.-D
[mav]_002	0047_0303	G_tildecomb	Ĝ	10. non-compliant
[mav]_002	0067_0303	g_tildecomb	ĝ	10. non-compliant
[mav]_003	00C3	Atilde	Ã	2. Latin-1 Supplement
[mav]_003	00E3	atilde	ã	2. Latin-1 Supplement
[mav]_003	0100	Amacron	Ā	3. Latin Ext.-A

arquivo	Unicode	Glyphname	Character	bloco
[mav]_003	0101	amacron	ā	3. Latin Ext.-A
[mav]_003	0112	Emacron	Ē	3. Latin Ext.-A
[mav]_003	0113	emacron	ē	3. Latin Ext.-A
[mav]_003	0128	ltilde	ĩ	3. Latin Ext.-A
[mav]_003	0129	itilde	ī	3. Latin Ext.-A
[mav]_003	012A	lmacron	Ī	3. Latin Ext.-A
[mav]_003	012B	imacron	ī	3. Latin Ext.-A
[mav]_003	014C	Omacron	Ō	3. Latin Ext.-A
[mav]_003	014D	omacron	ō	3. Latin Ext.-A
[mav]_003	0168	Utilde	Ū	3. Latin Ext.-A
[mav]_003	0169	utilde	ū	3. Latin Ext.-A
[mav]_003	016A	Umacron	Ū	3. Latin Ext.-A
[mav]_003	016B	umacron	ū	3. Latin Ext.-A
[mav]_003	02BC	apostrophemod	'	6. Spc. Mod. Letters
[mav]_003	1EBC	Etilde	Ě	8. Latin Ext. Add.
[mav]_003	1EBD	etilde	ě	8. Latin Ext. Add.
[mav]_003	A78B	Saltillo	ı	9. Latin Ext.-D
[mav]_003	A78C	salttillo	ı	9. Latin Ext.-D
[mav]_003	0047_0303	G_tildecomb	Ĝ	10. non-compliant
[mav]_003	0067_0303	g_tildecomb	ĝ	10. non-compliant
[mav]_004	0027	quotesingle	'	1. Basic Latin
[mav]_004	00C3	Atilde	Ã	2. Latin-1 Supplement
[mav]_004	00D5	Otilde	Õ	2. Latin-1 Supplement
[mav]_004	00E3	atilde	ã	2. Latin-1 Supplement
[mav]_004	00F5	otilde	õ	2. Latin-1 Supplement
[mav]_004	0168	Utilde	Ū	3. Latin Ext.-A
[mav]_004	0169	utilde	ū	3. Latin Ext.-A
[mav]_004	0303	tildecomb	˜	7. Combining Marks
[mav]_004	1EBC	Etilde	Ě	8. Latin Ext. Add.
[mav]_004	1EBD	etilde	ě	8. Latin Ext. Add.
[mav]_004	1EF8	Ytilde	Ÿ	8. Latin Ext. Add.
[mav]_004	1EF9	ytilde	ÿ	8. Latin Ext. Add.
[mbj]_001	0027	quotesingle	'	1. Basic Latin
[mbj]_001	00C3	Atilde	Ã	2. Latin-1 Supplement
[mbj]_001	00C4	Adieresis	Ä	2. Latin-1 Supplement
[mbj]_001	00C9	Eacute	É	2. Latin-1 Supplement
[mbj]_001	00CB	Edieresis	Ë	2. Latin-1 Supplement
[mbj]_001	00D3	Oacute	Ó	2. Latin-1 Supplement
[mbj]_001	00D5	Otilde	Õ	2. Latin-1 Supplement
[mbj]_001	00E3	atilde	ã	2. Latin-1 Supplement
[mbj]_001	00E4	adieresis	ä	2. Latin-1 Supplement
[mbj]_001	00E9	eacute	é	2. Latin-1 Supplement
[mbj]_001	00EB	edieresis	ë	2. Latin-1 Supplement
[mbj]_001	00F3	oacute	ó	2. Latin-1 Supplement
[mbj]_001	00F5	otilde	õ	2. Latin-1 Supplement
[mbj]_001	0128	ltilde	ĩ	3. Latin Ext.-A

arquivo	Unicode	Glyphname	Character	bloco
[mbj]_001	0129	itilde	ĩ	3. Latin Ext.-A
[mbj]_001	0168	Utilde	Ū	3. Latin Ext.-A
[mbj]_001	0169	utilde	ū	3. Latin Ext.-A
[mbj]_001	0303	tildecomb	̃	7. Combining Marks
[mbj]_001	1EBC	Etilde	Ě	8. Latin Ext. Add.
[mbj]_001	1EBD	etilde	ě	8. Latin Ext. Add.
[mbj]_001	1EF8	Ytilde	Ÿ	8. Latin Ext. Add.
[mbj]_001	1EF9	ytilde	ÿ	8. Latin Ext. Add.
[mbj]_001	00C4_0303	Adieresis_tildecomb	Ä̃	10. non-compliant
[mbj]_001	00E4_0303	adieresis_tildecomb	ä̃	10. non-compliant
[mbl]_001	00C3	Atilde	Ā	2. Latin-1 Supplement
[mbl]_001	00D5	Otilde	Ō	2. Latin-1 Supplement
[mbl]_001	00E3	atilde	ā	2. Latin-1 Supplement
[mbl]_001	00F5	otilde	ō	2. Latin-1 Supplement
[mbl]_001	0128	ltilde	Ĭ	3. Latin Ext.-A
[mbl]_001	0129	itilde	ĩ	3. Latin Ext.-A
[mbl]_001	0168	Utilde	Ū	3. Latin Ext.-A
[mbl]_001	0169	utilde	ū	3. Latin Ext.-A
[mbl]_001	02BC	apostrophemod	'	6. Spc. Mod. Letters
[mbl]_001	A78B	Saltillo	ı	9. Latin Ext.-D
[mbl]_001	A78C	salttillo	ı	9. Latin Ext.-D
[mbl]_002	00C3	Atilde	Ā	2. Latin-1 Supplement
[mbl]_002	00D5	Otilde	Ō	2. Latin-1 Supplement
[mbl]_002	00E3	atilde	ā	2. Latin-1 Supplement
[mbl]_002	00F5	otilde	ō	2. Latin-1 Supplement
[mbl]_002	0128	ltilde	Ĭ	3. Latin Ext.-A
[mbl]_002	0129	itilde	ĩ	3. Latin Ext.-A
[mbl]_002	0168	Utilde	Ū	3. Latin Ext.-A
[mbl]_002	0169	utilde	ū	3. Latin Ext.-A
[mbl]_002	1EBC	Etilde	Ě	8. Latin Ext. Add.
[mbl]_002	1EBD	etilde	ě	8. Latin Ext. Add.
[mbl]_003	00C3	Atilde	Ā	2. Latin-1 Supplement
[mbl]_003	00D5	Otilde	Ō	2. Latin-1 Supplement
[mbl]_003	00E3	atilde	ā	2. Latin-1 Supplement
[mbl]_003	00F5	otilde	ō	2. Latin-1 Supplement
[mbl]_003	0128	ltilde	Ĭ	3. Latin Ext.-A
[mbl]_003	0129	itilde	ĩ	3. Latin Ext.-A
[mbl]_003	0168	Utilde	Ū	3. Latin Ext.-A
[mbl]_003	0169	utilde	ū	3. Latin Ext.-A
[mbl]_003	1EBC	Etilde	Ě	8. Latin Ext. Add.
[mbl]_003	1EBD	etilde	ě	8. Latin Ext. Add.
[mbl]_004	00C3	Atilde	Ā	2. Latin-1 Supplement
[mbl]_004	00D5	Otilde	Ō	2. Latin-1 Supplement
[mbl]_004	00E3	atilde	ā	2. Latin-1 Supplement
[mbl]_004	00F5	otilde	ō	2. Latin-1 Supplement
[mbl]_004	0128	ltilde	Ĭ	3. Latin Ext.-A

arquivo	Unicode	Glyphname	Character	bloco
[mb1]_004	0129	itilde	ĩ	3. Latin Ext.-A
[mb1]_004	0168	Utilde	Û	3. Latin Ext.-A
[mb1]_004	0169	utilde	ü	3. Latin Ext.-A
[mb1]_004	1EBC	Etilde	Ě	8. Latin Ext. Add.
[mb1]_004	1EBD	etilde	ě	8. Latin Ext. Add.
[mb1]_006	0027	quotesingle	'	1. Basic Latin
[mb1]_006	00C3	Atilde	Ã	2. Latin-1 Supplement
[mb1]_006	00D5	Otilde	Õ	2. Latin-1 Supplement
[mb1]_006	00E3	atilde	ã	2. Latin-1 Supplement
[mb1]_006	00F5	otilde	õ	2. Latin-1 Supplement
[mb1]_006	0128	ltilde	Ĺ	3. Latin Ext.-A
[mb1]_006	0129	itilde	ĩ	3. Latin Ext.-A
[mb1]_006	0168	Utilde	Û	3. Latin Ext.-A
[mb1]_006	0169	utilde	ü	3. Latin Ext.-A
[mb1]_006	1EBC	Etilde	Ě	8. Latin Ext. Add.
[mb1]_006	1EBD	etilde	ě	8. Latin Ext. Add.
[mb1]_007	0027	quotesingle	'	1. Basic Latin
[mb1]_007	00C3	Atilde	Ã	2. Latin-1 Supplement
[mb1]_007	00D5	Otilde	Õ	2. Latin-1 Supplement
[mb1]_007	00E3	atilde	ã	2. Latin-1 Supplement
[mb1]_007	00F5	otilde	õ	2. Latin-1 Supplement
[mb1]_007	0128	ltilde	Ĺ	3. Latin Ext.-A
[mb1]_007	0129	itilde	ĩ	3. Latin Ext.-A
[mb1]_007	0168	Utilde	Û	3. Latin Ext.-A
[mb1]_007	0169	utilde	ü	3. Latin Ext.-A
[mb1]_007	1EBC	Etilde	Ě	8. Latin Ext. Add.
[mb1]_007	1EBD	etilde	ě	8. Latin Ext. Add.
[mb1]_007	2019	quoteright	'	11. General Punctuation
[mb1]_008	0027	quotesingle	'	1. Basic Latin
[mb1]_008	00C3	Atilde	Ã	2. Latin-1 Supplement
[mb1]_008	00D5	Otilde	Õ	2. Latin-1 Supplement
[mb1]_008	00E3	atilde	ã	2. Latin-1 Supplement
[mb1]_008	00F5	otilde	õ	2. Latin-1 Supplement
[mb1]_008	0128	ltilde	Ĺ	3. Latin Ext.-A
[mb1]_008	0129	itilde	ĩ	3. Latin Ext.-A
[mb1]_008	0168	Utilde	Û	3. Latin Ext.-A
[mb1]_008	0169	utilde	ü	3. Latin Ext.-A
[mb1]_008	1EBC	Etilde	Ě	8. Latin Ext. Add.
[mb1]_008	1EBD	etilde	ě	8. Latin Ext. Add.
[mcf]_001	00CB	Edieresis	Ë	2. Latin-1 Supplement
[mcf]_001	00EB	edieresis	ë	2. Latin-1 Supplement
[mch]_001	0027	quotesingle	'	1. Basic Latin
[mch]_001	00C4	Adieresis	Ä	2. Latin-1 Supplement
[mch]_001	00D6	Odieresis	Ö	2. Latin-1 Supplement
[mch]_001	00E4	adieresis	ä	2. Latin-1 Supplement
[mch]_001	00F6	odieresis	ö	2. Latin-1 Supplement
[mch]_002	00C4	Adieresis	Ä	2. Latin-1 Supplement

arquivo	Unicode	Glyphname	Character	bloco
[mch]_002	00D6	Odieresis	Ö	2. Latin-1 Supplement
[mch]_002	00E4	adieresis	ä	2. Latin-1 Supplement
[mch]_002	00F6	odieresis	ö	2. Latin-1 Supplement
[mch]_002	A78B	Saltillo	ı	9. Latin Ext.-D
[mch]_002	A78C	salttillo	ı	9. Latin Ext.-D
[mdz]_001	0027	quotesingle	'	1. Basic Latin
[mmh]_001	00C1	Aacute	Á	2. Latin-1 Supplement
[mmh]_001	00C3	Atilde	Ã	2. Latin-1 Supplement
[mmh]_001	00C9	Eacute	É	2. Latin-1 Supplement
[mmh]_001	00CD	Iacute	Í	2. Latin-1 Supplement
[mmh]_001	00DC	Udieresis	Ü	2. Latin-1 Supplement
[mmh]_001	00E1	aacute	á	2. Latin-1 Supplement
[mmh]_001	00E3	atilde	ã	2. Latin-1 Supplement
[mmh]_001	00E9	eacute	é	2. Latin-1 Supplement
[mmh]_001	00ED	iacute	í	2. Latin-1 Supplement
[mmh]_001	00FC	udieresis	ü	2. Latin-1 Supplement
[mmh]_001	0128	Itilde	İ	3. Latin Ext.-A
[mmh]_001	0129	itilde	ı	3. Latin Ext.-A
[mmh]_001	0168	Utilde	Ū	3. Latin Ext.-A
[mmh]_001	0169	utilde	ū	3. Latin Ext.-A
[mmh]_001	1EBC	Etilde	Ě	8. Latin Ext. Add.
[mmh]_001	1EBD	etilde	ě	8. Latin Ext. Add.
[mmh]_001	00DC_0303	Udieresis_tildecomb	Û	10. non-compliant
[mmh]_001	00FC_0303	udieresis_tildecomb	ü	10. non-compliant
[mpq]_001	00CB	Edieresis	Ě	2. Latin-1 Supplement
[mpq]_001	00EB	edieresis	ě	2. Latin-1 Supplement
[mpq]_002	00CB	Edieresis	Ě	2. Latin-1 Supplement
[mpq]_002	00EB	edieresis	ě	2. Latin-1 Supplement
[myu]_001	00C3	Atilde	Ã	2. Latin-1 Supplement
[myu]_001	00D5	Otilde	Õ	2. Latin-1 Supplement
[myu]_001	00E3	atilde	ã	2. Latin-1 Supplement
[myu]_001	00F5	otilde	õ	2. Latin-1 Supplement
[myu]_001	0128	Itilde	İ	3. Latin Ext.-A
[myu]_001	0129	itilde	ı	3. Latin Ext.-A
[myu]_001	0168	Utilde	Ū	3. Latin Ext.-A
[myu]_001	0169	utilde	ū	3. Latin Ext.-A
[myu]_001	0303	tildecomb	˜	7. Combining Marks
[myu]_001	1EBC	Etilde	Ě	8. Latin Ext. Add.
[myu]_001	1EBD	etilde	ě	8. Latin Ext. Add.
[myu]_001	1EF8	Ytilde	Ÿ	8. Latin Ext. Add.
[myu]_001	1EF9	ytilde	ÿ	8. Latin Ext. Add.
[myu]_001	0047_0303	G_tildecomb	Ĝ	10. non-compliant
[myu]_001	0067_0303	g_tildecomb	ĝ	10. non-compliant
[myu]_002	0027	quotesingle	'	1. Basic Latin
[myu]_002	00C3	Atilde	Ã	2. Latin-1 Supplement
[myu]_002	00D5	Otilde	Õ	2. Latin-1 Supplement

arquivo	Unicode	Glyphname	Character	bloco
[myu]_002	00E3	atilde	ã	2. Latin-1 Supplement
[myu]_002	00F5	otilde	õ	2. Latin-1 Supplement
[myu]_002	0128	ltilde	Ĺ	3. Latin Ext.-A
[myu]_002	0129	itilde	ĩ	3. Latin Ext.-A
[myu]_002	0168	Utilde	Ū	3. Latin Ext.-A
[myu]_002	0169	utilde	ū	3. Latin Ext.-A
[myu]_002	0303	tildecomb	˜	7. Combining Marks
[myu]_002	1EBC	Etilde	Ě	8. Latin Ext. Add.
[myu]_002	1EBD	etilde	ě	8. Latin Ext. Add.
[myu]_002	0047_0303	G_tildecomb	Ĝ	10. non-compliant
[myu]_002	0067_0303	g_tildecomb	ĝ	10. non-compliant
[myu]_002	2019	quoteright	'	11. General Punctuation
[mzr]_001	00C1	Aacute	Á	2. Latin-1 Supplement
[mzr]_001	00C3	Atilde	Ã	2. Latin-1 Supplement
[mzr]_001	00C9	Eacute	É	2. Latin-1 Supplement
[mzr]_001	00CD	Iacute	Í	2. Latin-1 Supplement
[mzr]_001	00D3	Oacute	Ó	2. Latin-1 Supplement
[mzr]_001	00D5	Otilde	Õ	2. Latin-1 Supplement
[mzr]_001	00E1	aacute	á	2. Latin-1 Supplement
[mzr]_001	00E3	atilde	ã	2. Latin-1 Supplement
[mzr]_001	00E9	eacute	é	2. Latin-1 Supplement
[mzr]_001	00ED	iacute	í	2. Latin-1 Supplement
[mzr]_001	00F3	oacute	ó	2. Latin-1 Supplement
[mzr]_001	00F5	otilde	õ	2. Latin-1 Supplement
[mzr]_001	0128	ltilde	Ĺ	3. Latin Ext.-A
[mzr]_001	0129	itilde	ĩ	3. Latin Ext.-A
[mzr]_001	1EBC	Etilde	Ě	8. Latin Ext. Add.
[mzr]_001	1EBD	etilde	ě	8. Latin Ext. Add.
[nab]_001	00C3	Atilde	Ã	2. Latin-1 Supplement
[nab]_001	00E3	atilde	ã	2. Latin-1 Supplement
[nab]_001	0128	ltilde	Ĺ	3. Latin Ext.-A
[nab]_001	0129	itilde	ĩ	3. Latin Ext.-A
[nab]_001	0168	Utilde	Ū	3. Latin Ext.-A
[nab]_001	0169	utilde	ū	3. Latin Ext.-A
[nab]_001	0303	tildecomb	˜	7. Combining Marks
[nab]_001	0330	tildebelowcomb	˘	7. Combining Marks
[nab]_001	1E1A	Etildebelow	Ě̃	8. Latin Ext. Add.
[nab]_001	1E1B	etildebelow	ě̃	8. Latin Ext. Add.
[nab]_001	1E2C	ltildebelow	Ĺ̃	8. Latin Ext. Add.
[nab]_001	1E2D	itildebelow	ĩ̃	8. Latin Ext. Add.
[nab]_001	1E74	Utildebelow	Ū̃	8. Latin Ext. Add.
[nab]_001	1E75	utildebelow	ū̃	8. Latin Ext. Add.
[nab]_001	1EBC	Etilde	Ě	8. Latin Ext. Add.
[nab]_001	1EBD	etilde	ě	8. Latin Ext. Add.
[nab]_001	0041_0330	A_tildebelowcomb	Ă	10. non-compliant
[nab]_001	0061_0330	a_tildebelowcomb	ă	10. non-compliant
[nab]_001	004F_0330	O_tildebelowcomb	Ȫ	10. non-compliant

arquivo	Unicode	Glyphname	Character	bloco
[nab]_001	006F_0330	o_tildebelowcomb	o	10. non-compliant
[nab]_001	00C3_0330	Atilde_tildebelowcomb	Ã	10. non-compliant
[nab]_001	00E3_0330	atilde_tildebelowcomb	ã	10. non-compliant
[nab]_001	1EBC_0330	Etilde_tildebelowcomb	Ě	10. non-compliant
[nab]_001	1EBD_0330	etilde_tildebelowcomb	ě	10. non-compliant
[nab]_001	0128_0330	ltilde_tildebelowcomb	Ī	10. non-compliant
[nab]_001	0129_0330	itilde_tildebelowcomb	ī	10. non-compliant
[nab]_001	0168_0330	Utilde_tildebelowcomb	Ū	10. non-compliant
[nab]_001	0169_0330	utilde_tildebelowcomb	ū	10. non-compliant
[nab]_002	00C3	Atilde	Ã	2. Latin-1 Supplement
[nab]_002	00E3	atilde	ã	2. Latin-1 Supplement
[nab]_002	0128	ltilde	Ī	3. Latin Ext.-A
[nab]_002	0129	itilde	ī	3. Latin Ext.-A
[nab]_002	0168	Utilde	Ū	3. Latin Ext.-A
[nab]_002	0169	utilde	ū	3. Latin Ext.-A
[nab]_002	0331	macronbelowcomb	_	7. Combining Marks
[nab]_002	1EBC	Etilde	Ě	8. Latin Ext. Add.
[nab]_002	1EBD	etilde	ě	8. Latin Ext. Add.
[nab]_002	0041_0331	A_macronbelowcomb	Ǻ	10. non-compliant
[nab]_002	0061_0331	a_macronbelowcomb	ǻ	10. non-compliant
[nab]_002	0045_0331	E_macronbelowcomb	Ǽ	10. non-compliant
[nab]_002	0065_0331	e_macronbelowcomb	ǽ	10. non-compliant
[nab]_002	0049_0331	I_macronbelowcomb	Ǫ	10. non-compliant
[nab]_002	0069_0331	i_macronbelowcomb	ǫ	10. non-compliant
[nab]_002	004F_0331	O_macronbelowcomb	Ǿ	10. non-compliant
[nab]_002	006F_0331	o_macronbelowcomb	ǿ	10. non-compliant
[nab]_002	0055_0331	U_macronbelowcomb	ǘ	10. non-compliant
[nab]_002	0075_0331	u_macronbelowcomb	Ǚ	10. non-compliant
[nab]_002	00C3_0331	Atilde_macronbelowcomb	Ǻ	10. non-compliant
[nab]_002	00E3_0331	atilde_macronbelowcomb	ǻ	10. non-compliant
[nab]_002	1EBC_0331	Etilde_macronbelowcomb	Ě	10. non-compliant
[nab]_002	1EBD_0331	etilde_macronbelowcomb	ě	10. non-compliant
[nab]_002	0128_0331	ltilde_macronbelowcomb	Ī	10. non-compliant
[nab]_002	0129_0331	itilde_macronbelowcomb	ī	10. non-compliant
[nab]_002	0168_0331	Utilde_macronbelowcomb	Ū	10. non-compliant
[nab]_002	0169_0331	utilde_macronbelowcomb	ū	10. non-compliant
[nhd]_001	00C1	Aacute	Á	2. Latin-1 Supplement
[nhd]_001	00C3	Atilde	Ã	2. Latin-1 Supplement
[nhd]_001	00C9	Eacute	É	2. Latin-1 Supplement
[nhd]_001	00D3	Oacute	Ó	2. Latin-1 Supplement
[nhd]_001	00E1	aacute	á	2. Latin-1 Supplement
[nhd]_001	00E3	atilde	ã	2. Latin-1 Supplement
[nhd]_001	00E9	eacute	é	2. Latin-1 Supplement
[nhd]_001	00F3	oacute	ó	2. Latin-1 Supplement
[nhd]_001	0128	ltilde	Ī	3. Latin Ext.-A
[nhd]_001	0129	itilde	ī	3. Latin Ext.-A
[nhd]_001	0168	Utilde	Ū	3. Latin Ext.-A

arquivo	Unicode	Glyphname	Character	bloco
[nhd]_001	0169	utilde	ũ	3. Latin Ext.-A
[nhd]_001	1EBC	Etilde	Ě	8. Latin Ext. Add.
[nhd]_001	1EBD	etilde	ě	8. Latin Ext. Add.
[nhd]_001	1EF8	Ytilde	Ÿ	8. Latin Ext. Add.
[nhd]_001	1EF9	ytilde	ÿ	8. Latin Ext. Add.
[oym]_001	00C3	Atilde	Ă	2. Latin-1 Supplement
[oym]_001	00D5	Otilde	Ŏ	2. Latin-1 Supplement
[oym]_001	00E3	atilde	ă	2. Latin-1 Supplement
[oym]_001	00F5	otilde	õ	2. Latin-1 Supplement
[oym]_001	0128	ltilde	Ĺ	3. Latin Ext.-A
[oym]_001	0129	itilde	ĺ	3. Latin Ext.-A
[oym]_001	02BC	apostrophemod	'	6. Spc. Mod. Letters
[oym]_001	1EBC	Etilde	Ě	8. Latin Ext. Add.
[oym]_001	1EBD	etilde	ě	8. Latin Ext. Add.
[oym]_001	A78B	Saltillo	ı	9. Latin Ext.-D
[oym]_001	A78C	salttillo	ı	9. Latin Ext.-D
[oym]_002	0027	quotesingle	'	1. Basic Latin
[oym]_002	00C3	Atilde	Ă	2. Latin-1 Supplement
[oym]_002	00D5	Otilde	Ŏ	2. Latin-1 Supplement
[oym]_002	00E3	atilde	ă	2. Latin-1 Supplement
[oym]_002	00F5	otilde	õ	2. Latin-1 Supplement
[oym]_002	0128	ltilde	Ĺ	3. Latin Ext.-A
[oym]_002	0129	itilde	ĺ	3. Latin Ext.-A
[oym]_002	0168	Utilde	Ŭ	3. Latin Ext.-A
[oym]_002	0169	utilde	ũ	3. Latin Ext.-A
[oym]_002	1EBC	Etilde	Ě	8. Latin Ext. Add.
[oym]_002	1EBD	etilde	ě	8. Latin Ext. Add.
[oym]_002	1EF8	Ytilde	Ÿ	8. Latin Ext. Add.
[oym]_002	1EF9	ytilde	ÿ	8. Latin Ext. Add.
[pab]_001	00C1	Aacute	Á	2. Latin-1 Supplement
[pab]_001	00C9	Eacute	É	2. Latin-1 Supplement
[pab]_001	00CD	Iacute	Í	2. Latin-1 Supplement
[pab]_001	00D3	Oacute	Ó	2. Latin-1 Supplement
[pab]_001	00E1	aacute	á	2. Latin-1 Supplement
[pab]_001	00E9	eacute	é	2. Latin-1 Supplement
[pab]_001	00ED	iacute	í	2. Latin-1 Supplement
[pab]_001	00F3	oacute	ó	2. Latin-1 Supplement
[pad]_001	0027	quotesingle	'	1. Basic Latin
[pah]_001	0027	quotesingle	'	1. Basic Latin
[pah]_001	00C3	Atilde	Ă	2. Latin-1 Supplement
[pah]_001	00D5	Otilde	Ŏ	2. Latin-1 Supplement
[pah]_001	00E3	atilde	ă	2. Latin-1 Supplement
[pah]_001	00F5	otilde	õ	2. Latin-1 Supplement
[pah]_001	0128	ltilde	Ĺ	3. Latin Ext.-A
[pah]_001	0129	itilde	ĺ	3. Latin Ext.-A
[pah]_001	0168	Utilde	Ŭ	3. Latin Ext.-A

arquivo	Unicode	Glyphname	Character	bloco
[pah]_001	0169	utilde	ũ	3. Latin Ext.-A
[pah]_001	1EBC	Etilde	Ě	8. Latin Ext. Add.
[pah]_001	1EBD	etilde	ě	8. Latin Ext. Add.
[pah]_001	1EF8	Ytilde	Ÿ	8. Latin Ext. Add.
[pah]_001	1EF9	ytilde	ÿ	8. Latin Ext. Add.
[pah]_002	0027	quotesingle	'	1. Basic Latin
[pah]_002	00C3	Atilde	Ā	2. Latin-1 Supplement
[pah]_002	00D5	Otilde	Ō	2. Latin-1 Supplement
[pah]_002	00E3	atilde	ā	2. Latin-1 Supplement
[pah]_002	00F5	otilde	ō	2. Latin-1 Supplement
[pah]_002	0128	ltilde	Ĺ	3. Latin Ext.-A
[pah]_002	0129	itilde	ĺ	3. Latin Ext.-A
[pah]_002	0168	Utilde	Ū	3. Latin Ext.-A
[pah]_002	0169	utilde	ū	3. Latin Ext.-A
[pah]_002	0303	tildecomb	˜	7. Combining Marks
[pah]_002	1EBC	Etilde	Ě	8. Latin Ext. Add.
[pah]_002	1EBD	etilde	ě	8. Latin Ext. Add.
[pah]_002	0047_0303	G_tildecomb	Ĝ	10. non-compliant
[pah]_002	0067_0303	g_tildecomb	ĝ	10. non-compliant
[pak]_001	02BC	apostrophemod	'	6. Spc. Mod. Letters
[pak]_001	A78B	Saltillo	ı	9. Latin Ext.-D
[pak]_001	A78C	salttillo	ı	9. Latin Ext.-D
[pir]_001	0027	quotesingle	'	1. Basic Latin
[pir]_001	00C3	Atilde	Ā	2. Latin-1 Supplement
[pir]_001	00D1	Ntilde	Ñ	2. Latin-1 Supplement
[pir]_001	00D5	Otilde	Ō	2. Latin-1 Supplement
[pir]_001	00E3	atilde	ā	2. Latin-1 Supplement
[pir]_001	00F1	ntilde	ñ	2. Latin-1 Supplement
[pir]_001	00F5	otilde	ō	2. Latin-1 Supplement
[pir]_001	0128	ltilde	Ĺ	3. Latin Ext.-A
[pir]_001	0129	itilde	ĺ	3. Latin Ext.-A
[pir]_001	0168	Utilde	Ū	3. Latin Ext.-A
[pir]_001	0169	utilde	ū	3. Latin Ext.-A
[pir]_001	0244	Ubar	Ů	4. Latin Ext.-B
[pir]_001	0289	ubar	ů	5. IPA Extensions
[pir]_001	0303	tildecomb	˜	7. Combining Marks
[pir]_001	1EBC	Etilde	Ě	8. Latin Ext. Add.
[pir]_001	1EBD	etilde	ě	8. Latin Ext. Add.
[pir]_001	0244_0303	Ubar_tildecomb	Ů	10. non-compliant
[pir]_001	0289_0303	ubar_tildecomb	ů	10. non-compliant
[pto]_001	00C3	Atilde	Ā	2. Latin-1 Supplement
[pto]_001	00D5	Otilde	Ō	2. Latin-1 Supplement
[pto]_001	00E3	atilde	ā	2. Latin-1 Supplement
[pto]_001	00F5	otilde	ō	2. Latin-1 Supplement
[pto]_001	0128	ltilde	Ĺ	3. Latin Ext.-A
[pto]_001	0129	itilde	ĺ	3. Latin Ext.-A

arquivo	Unicode	Glyphname	Character	bloco
[pto]_001	0168	Utilde	Û	3. Latin Ext.-A
[pto]_001	0169	utilde	ü	3. Latin Ext.-A
[pto]_001	02BC	apostrophemod	'	6. Spc. Mod. Letters
[pto]_001	1EBC	Etilde	Ě	8. Latin Ext. Add.
[pto]_001	1EBD	etilde	ě	8. Latin Ext. Add.
[pto]_001	A78B	Saltillo	ı	9. Latin Ext.-D
[pto]_001	A78C	salttillo	ı	9. Latin Ext.-D
[pur]_001	0027	quotesingle	'	1. Basic Latin
[pur]_001	00C2	Acircumflex	Â	2. Latin-1 Supplement
[pur]_001	00C3	Atilde	Ã	2. Latin-1 Supplement
[pur]_001	00D5	Otilde	Õ	2. Latin-1 Supplement
[pur]_001	00E2	acircumflex	â	2. Latin-1 Supplement
[pur]_001	00E3	atilde	ã	2. Latin-1 Supplement
[pur]_001	00F5	otilde	õ	2. Latin-1 Supplement
[pur]_001	0128	ltilde	Ĭ	3. Latin Ext.-A
[pur]_001	0129	itilde	ĭ	3. Latin Ext.-A
[pur]_001	0168	Utilde	Û	3. Latin Ext.-A
[pur]_001	0169	utilde	ü	3. Latin Ext.-A
[pur]_001	1EBC	Etilde	Ě	8. Latin Ext. Add.
[pur]_001	1EBD	etilde	ě	8. Latin Ext. Add.
[pur]_001	1EF8	Ytilde	Ÿ	8. Latin Ext. Add.
[pur]_001	1EF9	ytilde	ÿ	8. Latin Ext. Add.
[ram]_001	00C0	Agrave	À	2. Latin-1 Supplement
[ram]_001	00C3	Atilde	Ã	2. Latin-1 Supplement
[ram]_001	00CA	Ecircumflex	Ê	2. Latin-1 Supplement
[ram]_001	00D4	Ocircumflex	Ô	2. Latin-1 Supplement
[ram]_001	00D5	Otilde	Õ	2. Latin-1 Supplement
[ram]_001	00E0	agrave	à	2. Latin-1 Supplement
[ram]_001	00E3	atilde	ã	2. Latin-1 Supplement
[ram]_001	00EA	ecircumflex	ê	2. Latin-1 Supplement
[ram]_001	00F4	ocircumflex	ô	2. Latin-1 Supplement
[ram]_001	00F5	otilde	õ	2. Latin-1 Supplement
[ram]_001	0128	ltilde	Ĭ	3. Latin Ext.-A
[ram]_001	0129	itilde	ĭ	3. Latin Ext.-A
[ram]_001	0168	Utilde	Û	3. Latin Ext.-A
[ram]_001	0169	utilde	ü	3. Latin Ext.-A
[ram]_001	1EBC	Etilde	Ě	8. Latin Ext. Add.
[ram]_001	1EBD	etilde	ě	8. Latin Ext. Add.
[ram]_001	1EF2	Ygrave	Ỳ	8. Latin Ext. Add.
[ram]_001	1EF3	ygrave	ỳ	8. Latin Ext. Add.
[ram]_001	1EF8	Ytilde	Ÿ	8. Latin Ext. Add.
[ram]_001	1EF9	ytilde	ÿ	8. Latin Ext. Add.
[ram]_002	00C0	Agrave	À	2. Latin-1 Supplement
[ram]_002	00C3	Atilde	Ã	2. Latin-1 Supplement
[ram]_002	00CA	Ecircumflex	Ê	2. Latin-1 Supplement
[ram]_002	00D4	Ocircumflex	Ô	2. Latin-1 Supplement

arquivo	Unicode	Glyphname	Character	bloco
[ram]_002	00D5	Otilde	Õ	2. Latin-1 Supplement
[ram]_002	00E0	agrave	à	2. Latin-1 Supplement
[ram]_002	00E3	atilde	ã	2. Latin-1 Supplement
[ram]_002	00EA	ecircumflex	ê	2. Latin-1 Supplement
[ram]_002	00F4	ocircumflex	ô	2. Latin-1 Supplement
[ram]_002	00F5	otilde	õ	2. Latin-1 Supplement
[ram]_002	0128	ltilde	ĺ	3. Latin Ext.-A
[ram]_002	0129	itilde	ı	3. Latin Ext.-A
[ram]_002	0168	Utilde	Ū	3. Latin Ext.-A
[ram]_002	0169	utilde	ū	3. Latin Ext.-A
[ram]_002	1EBC	Etilde	Ě	8. Latin Ext. Add.
[ram]_002	1EBD	etilde	ě	8. Latin Ext. Add.
[ram]_002	1EF2	Ygrave	ÿ	8. Latin Ext. Add.
[ram]_002	1EF3	ygrave	ÿ	8. Latin Ext. Add.
[ram]_002	1EF8	Ytilde	Ÿ	8. Latin Ext. Add.
[ram]_002	1EF9	ytilde	ÿ	8. Latin Ext. Add.
[rkb]_001	00C3	Atilde	Ă	2. Latin-1 Supplement
[rkb]_001	00D5	Otilde	Õ	2. Latin-1 Supplement
[rkb]_001	00E3	atilde	ã	2. Latin-1 Supplement
[rkb]_001	00F5	otilde	õ	2. Latin-1 Supplement
[rkb]_001	0128	ltilde	ĺ	3. Latin Ext.-A
[rkb]_001	0129	itilde	ı	3. Latin Ext.-A
[rkb]_001	0168	Utilde	Ū	3. Latin Ext.-A
[rkb]_001	0169	utilde	ū	3. Latin Ext.-A
[rkb]_001	1EBC	Etilde	Ě	8. Latin Ext. Add.
[rkb]_001	1EBD	etilde	ě	8. Latin Ext. Add.
[rkb]_001	1EF8	Ytilde	Ÿ	8. Latin Ext. Add.
[rkb]_001	1EF9	ytilde	ÿ	8. Latin Ext. Add.
[rkb]_002	00C3	Atilde	Ă	2. Latin-1 Supplement
[rkb]_002	00D5	Otilde	Õ	2. Latin-1 Supplement
[rkb]_002	00E3	atilde	ã	2. Latin-1 Supplement
[rkb]_002	00F5	otilde	õ	2. Latin-1 Supplement
[rkb]_002	0128	ltilde	ĺ	3. Latin Ext.-A
[rkb]_002	0129	itilde	ı	3. Latin Ext.-A
[rkb]_002	0168	Utilde	Ū	3. Latin Ext.-A
[rkb]_002	0169	utilde	ū	3. Latin Ext.-A
[rkb]_002	1EBC	Etilde	Ě	8. Latin Ext. Add.
[rkb]_002	1EBD	etilde	ě	8. Latin Ext. Add.
[rkb]_002	1EF8	Ytilde	Ÿ	8. Latin Ext. Add.
[rkb]_002	1EF9	ytilde	ÿ	8. Latin Ext. Add.
[shb]_001	00C3	Atilde	Ă	2. Latin-1 Supplement
[shb]_001	00CB	Edieresis	Ë	2. Latin-1 Supplement
[shb]_001	00D5	Otilde	Õ	2. Latin-1 Supplement
[shb]_001	00E3	atilde	ã	2. Latin-1 Supplement
[shb]_001	00EB	edieresis	ë	2. Latin-1 Supplement
[shb]_001	00F5	otilde	õ	2. Latin-1 Supplement
[shb]_001	0128	ltilde	ĺ	3. Latin Ext.-A

arquivo	Unicode	Glyphname	Character	bloco
[shb]_001	0129	itilde	ĩ	3. Latin Ext.-A
[shb]_001	0168	Utilde	Û	3. Latin Ext.-A
[shb]_001	0169	utilde	ü	3. Latin Ext.-A
[shb]_001	0197	lstroke	ł	4. Latin Ext.-B
[shb]_001	0268	istroke	ı	5. IPA Extensions
[shb]_001	0303	tildecomb	˜	7. Combining Marks
[shb]_001	00CB_0303	Edieresis_tildecomb	Ë	10. non-compliant
[shb]_001	00EB_0303	edieresis_tildecomb	ë	10. non-compliant
[shb]_001	0197_0303	lstroke_tildecomb	ł̃	10. non-compliant
[shb]_001	0268_0303	istroke_tildecomb	ı̃	10. non-compliant
[shb]_002	00C3	Atilde	Ã	2. Latin-1 Supplement
[shb]_002	00CB	Edieresis	Ë	2. Latin-1 Supplement
[shb]_002	00D5	Otilde	Õ	2. Latin-1 Supplement
[shb]_002	00E3	atilde	ã	2. Latin-1 Supplement
[shb]_002	00EB	edieresis	ë	2. Latin-1 Supplement
[shb]_002	00F5	otilde	ö	2. Latin-1 Supplement
[shb]_002	0128	ltilde	Ł	3. Latin Ext.-A
[shb]_002	0129	itilde	ĩ	3. Latin Ext.-A
[shb]_002	0168	Utilde	Û	3. Latin Ext.-A
[shb]_002	0169	utilde	ü	3. Latin Ext.-A
[shb]_002	0197	lstroke	ł	4. Latin Ext.-B
[shb]_002	0268	istroke	ı	5. IPA Extensions
[shb]_002	0303	tildecomb	˜	7. Combining Marks
[shb]_002	00CB_0303	Edieresis_tildecomb	Ë	10. non-compliant
[shb]_002	00EB_0303	edieresis_tildecomb	ë	10. non-compliant
[shb]_002	0197_0303	lstroke_tildecomb	ł̃	10. non-compliant
[shb]_002	0268_0303	istroke_tildecomb	ı̃	10. non-compliant
[skf]_001	0027	quotesingle	'	1. Basic Latin
[skf]_001	00C3	Atilde	Ã	2. Latin-1 Supplement
[skf]_001	00D5	Otilde	Õ	2. Latin-1 Supplement
[skf]_001	00E3	atilde	ã	2. Latin-1 Supplement
[skf]_001	00F5	otilde	ö	2. Latin-1 Supplement
[skf]_001	0128	ltilde	Ł	3. Latin Ext.-A
[skf]_001	0129	itilde	ĩ	3. Latin Ext.-A
[skf]_001	0168	Utilde	Û	3. Latin Ext.-A
[skf]_001	0169	utilde	ü	3. Latin Ext.-A
[skf]_001	1EBC	Etilda	Ě	8. Latin Ext. Add.
[skf]_001	1EBD	etilde	ě	8. Latin Ext. Add.
[sru]_001	00C1	Aacute	Á	2. Latin-1 Supplement
[sru]_001	00C3	Atilde	Ã	2. Latin-1 Supplement
[sru]_001	00C9	Eacute	É	2. Latin-1 Supplement
[sru]_001	00CD	Iacute	Í	2. Latin-1 Supplement
[sru]_001	00D3	Oacute	Ó	2. Latin-1 Supplement
[sru]_001	00D5	Otilde	Õ	2. Latin-1 Supplement
[sru]_001	00DA	Uacute	Ú	2. Latin-1 Supplement
[sru]_001	00E1	aacute	á	2. Latin-1 Supplement
[sru]_001	00E3	atilde	ã	2. Latin-1 Supplement

arquivo	Unicode	Glyphname	Character	bloco
[sru]_001	00E9	eacute	é	2. Latin-1 Supplement
[sru]_001	00ED	iacute	í	2. Latin-1 Supplement
[sru]_001	00F3	oacute	ó	2. Latin-1 Supplement
[sru]_001	00F5	otilde	õ	2. Latin-1 Supplement
[sru]_001	00FA	uacute	ú	2. Latin-1 Supplement
[sru]_001	0128	ltilde	ĩ	3. Latin Ext.-A
[sru]_001	0129	itilde	ı̃	3. Latin Ext.-A
[sru]_001	0168	Utilde	Ū	3. Latin Ext.-A
[sru]_001	0169	utilde	ũ	3. Latin Ext.-A
[sru]_001	0301	acutecomb	´	7. Combining Marks
[sru]_001	0303	tildecomb	˘	7. Combining Marks
[sru]_001	0341	acutetonecomb	ˆ	7. Combining Marks
[sru]_001	1E4C	Otildeacute	Ô	8. Latin Ext. Add.
[sru]_001	1E4D	otildeacute	ô	8. Latin Ext. Add.
[sru]_001	1E78	Utildeacute	Ū	8. Latin Ext. Add.
[sru]_001	1E79	utildeacute	ũ	8. Latin Ext. Add.
[sru]_001	1EBC	Etilde	Ě	8. Latin Ext. Add.
[sru]_001	1EBD	etilde	ě	8. Latin Ext. Add.
[sru]_001	0047_0303	G_tildecomb	Ĝ	10. non-compliant
[sru]_001	0067_0303	g_tildecomb	ĝ	10. non-compliant
[sru]_001	00C3_0301	Atilde_acutecomb	Ă	10. non-compliant
[sru]_001	00E3_0301	atilde_acutecomb	ă	10. non-compliant
[sru]_001	1EBC_0301	Etilde_acutecomb	Ě	10. non-compliant
[sru]_001	1EBD_0301	etilde_acutecomb	ě	10. non-compliant
[sru]_001	0128_0301	ltilde_acutecomb	ĩ	10. non-compliant
[sru]_001	0129_0301	itilde_acutecomb	ı̃	10. non-compliant
[sru]_001	00D5_0301	Otilde_acutecomb	Ô	10. non-compliant
[sru]_001	00F5_0301	otilde_acutecomb	ô	10. non-compliant
[sru]_001	0168_0301	Utilde_acutecomb	Ū	10. non-compliant
[sru]_001	0169_0301	utilde_acutecomb	ũ	10. non-compliant
[suy]_001	00C1	Aacute	Á	2. Latin-1 Supplement
[suy]_001	00C2	Acircumflex	Â	2. Latin-1 Supplement
[suy]_001	00C3	Atilde	Ã	2. Latin-1 Supplement
[suy]_001	00CA	Ecircumflex	Ê	2. Latin-1 Supplement
[suy]_001	00D4	Ocircumflex	Ô	2. Latin-1 Supplement
[suy]_001	00D5	Otilde	Õ	2. Latin-1 Supplement
[suy]_001	00E1	aacute	á	2. Latin-1 Supplement
[suy]_001	00E2	acircumflex	â	2. Latin-1 Supplement
[suy]_001	00E3	atilde	ã	2. Latin-1 Supplement
[suy]_001	00EA	ecircumflex	ê	2. Latin-1 Supplement
[suy]_001	00F4	ocircumflex	ô	2. Latin-1 Supplement
[suy]_001	00F5	otilde	õ	2. Latin-1 Supplement
[suy]_001	0128	ltilde	ĩ	3. Latin Ext.-A
[suy]_001	0129	itilde	ı̃	3. Latin Ext.-A
[suy]_001	0168	Utilde	Ū	3. Latin Ext.-A
[suy]_001	0169	utilde	ũ	3. Latin Ext.-A
[suy]_001	1EBC	Etilde	Ě	8. Latin Ext. Add.

arquivo	Unicode	Glyphname	Character	bloco
[suy]_001	1EBD	etilde	ě	8. Latin Ext. Add.
[suy]_001	1EF8	Ytilde	Ÿ	8. Latin Ext. Add.
[suy]_001	1EF9	ytilde	ÿ	8. Latin Ext. Add.
[suy]_002	00C1	Aacute	Á	2. Latin-1 Supplement
[suy]_002	00C2	Acircumflex	Â	2. Latin-1 Supplement
[suy]_002	00C3	Atilde	Ã	2. Latin-1 Supplement
[suy]_002	00CA	Ecircumflex	Ê	2. Latin-1 Supplement
[suy]_002	00D4	Ocircumflex	Ô	2. Latin-1 Supplement
[suy]_002	00D5	Otilde	Õ	2. Latin-1 Supplement
[suy]_002	00E1	aacute	á	2. Latin-1 Supplement
[suy]_002	00E2	acircumflex	â	2. Latin-1 Supplement
[suy]_002	00E3	atilde	ã	2. Latin-1 Supplement
[suy]_002	00EA	ecircumflex	ê	2. Latin-1 Supplement
[suy]_002	00F4	ocircumflex	ô	2. Latin-1 Supplement
[suy]_002	00F5	otilde	õ	2. Latin-1 Supplement
[suy]_002	0128	ltilde	Ĺ	3. Latin Ext.-A
[suy]_002	0129	itilde	ĺ	3. Latin Ext.-A
[suy]_002	0168	Utilde	Ū	3. Latin Ext.-A
[suy]_002	0169	utilde	ū	3. Latin Ext.-A
[suy]_002	1EBC	Etilde	Ě	8. Latin Ext. Add.
[suy]_002	1EBD	etilde	ě	8. Latin Ext. Add.
[suy]_002	1EF8	Ytilde	Ÿ	8. Latin Ext. Add.
[suy]_002	1EF9	ytilde	ÿ	8. Latin Ext. Add.
[suy]_003	00C1	Aacute	Á	2. Latin-1 Supplement
[suy]_003	00C2	Acircumflex	Â	2. Latin-1 Supplement
[suy]_003	00C3	Atilde	Ã	2. Latin-1 Supplement
[suy]_003	00CA	Ecircumflex	Ê	2. Latin-1 Supplement
[suy]_003	00D4	Ocircumflex	Ô	2. Latin-1 Supplement
[suy]_003	00D5	Otilde	Õ	2. Latin-1 Supplement
[suy]_003	00E1	aacute	á	2. Latin-1 Supplement
[suy]_003	00E2	acircumflex	â	2. Latin-1 Supplement
[suy]_003	00E3	atilde	ã	2. Latin-1 Supplement
[suy]_003	00EA	ecircumflex	ê	2. Latin-1 Supplement
[suy]_003	00F4	ocircumflex	ô	2. Latin-1 Supplement
[suy]_003	00F5	otilde	õ	2. Latin-1 Supplement
[suy]_003	0128	ltilde	Ĺ	3. Latin Ext.-A
[suy]_003	0129	itilde	ĺ	3. Latin Ext.-A
[suy]_003	0168	Utilde	Ū	3. Latin Ext.-A
[suy]_003	0169	utilde	ū	3. Latin Ext.-A
[suy]_003	1EBC	Etilde	Ě	8. Latin Ext. Add.
[suy]_003	1EBD	etilde	ě	8. Latin Ext. Add.
[suy]_003	1EF8	Ytilde	Ÿ	8. Latin Ext. Add.
[suy]_003	1EF9	ytilde	ÿ	8. Latin Ext. Add.
[swo]_001	00C3	Atilde	Ã	2. Latin-1 Supplement
[swo]_001	00D5	Otilde	Õ	2. Latin-1 Supplement
[swo]_001	00E3	atilde	ã	2. Latin-1 Supplement
[swo]_001	00F5	otilde	õ	2. Latin-1 Supplement

arquivo	Unicode	Glyphname	Character	bloco
[swo]_001	0128	ltilde	ĩ	3. Latin Ext.-A
[swo]_001	0129	itilde	ı̇	3. Latin Ext.-A
[swo]_001	0168	Utilde	Û	3. Latin Ext.-A
[swo]_001	0169	utilde	ü	3. Latin Ext.-A
[swo]_001	1EBC	Etilde	Ě	8. Latin Ext. Add.
[swo]_001	1EBD	etilde	ě	8. Latin Ext. Add.
[swo]_001	1EF8	Ytilde	Ÿ	8. Latin Ext. Add.
[swo]_001	1EF9	ytilde	ÿ	8. Latin Ext. Add.
[taf]_002	02BC	apostrophemod	'	6. Spc. Mod. Letters
[taf]_002	A78B	Saltillo	ı̂	9. Latin Ext.-D
[taf]_002	A78C	salttillo	ı̃	9. Latin Ext.-D
[taf]_003	02BC	apostrophemod	'	6. Spc. Mod. Letters
[taf]_003	A78B	Saltillo	ı̂	9. Latin Ext.-D
[taf]_003	A78C	salttillo	ı̃	9. Latin Ext.-D
[tba]_001	0027	quotesingle	'	1. Basic Latin
[tba]_001	00C3	Atilde	Ã	2. Latin-1 Supplement
[tba]_001	00DC	Udieresis	Ü	2. Latin-1 Supplement
[tba]_001	00E3	atilde	ã	2. Latin-1 Supplement
[tba]_001	00FC	udieresis	ü	2. Latin-1 Supplement
[tba]_001	0128	ltilde	ĩ	3. Latin Ext.-A
[tba]_001	0129	itilde	ı̇	3. Latin Ext.-A
[tba]_001	0168	Utilde	Û	3. Latin Ext.-A
[tba]_001	0169	utilde	ü	3. Latin Ext.-A
[tba]_001	0197	lstroke	ł	4. Latin Ext.-B
[tba]_001	0268	istroke	ı̇	5. IPA Extensions
[tba]_001	0303	tildecomb	˜	7. Combining Marks
[tba]_001	1EBC	Etilde	Ě	8. Latin Ext. Add.
[tba]_001	1EBD	etilde	ě	8. Latin Ext. Add.
[tba]_001	00DC_0303	Udieresis_tildecomb	Ü̃	10. non-compliant
[tba]_001	00FC_0303	udieresis_tildecomb	ü̃	10. non-compliant
[tba]_001	0197_0303	lstroke_tildecomb	ł̃	10. non-compliant
[tba]_001	0268_0303	istroke_tildecomb	ı̇̃	10. non-compliant
[tca]_001	00C3	Atilde	Ã	2. Latin-1 Supplement
[tca]_001	00CA	Ecircumflex	Ê	2. Latin-1 Supplement
[tca]_001	00CD	iacute	Í	2. Latin-1 Supplement
[tca]_001	00D5	Otilde	Õ	2. Latin-1 Supplement
[tca]_001	00DC	Udieresis	Ü	2. Latin-1 Supplement
[tca]_001	00E3	atilde	ã	2. Latin-1 Supplement
[tca]_001	00EA	ecircumflex	ê	2. Latin-1 Supplement
[tca]_001	00ED	iacute	í	2. Latin-1 Supplement
[tca]_001	00F5	otilde	õ	2. Latin-1 Supplement
[tca]_001	00FC	udieresis	ü	2. Latin-1 Supplement
[tca]_001	02BC	apostrophemod	'	6. Spc. Mod. Letters
[tca]_001	A78B	Saltillo	ı̂	9. Latin Ext.-D
[tca]_001	A78C	salttillo	ı̃	9. Latin Ext.-D

arquivo	Unicode	Glyphname	Character	bloco
[tca]_002	00C3	Atilde	Ã	2. Latin-1 Supplement
[tca]_002	00CD	lacute	Í	2. Latin-1 Supplement
[tca]_002	00DC	Udieresis	Û	2. Latin-1 Supplement
[tca]_002	00E3	atilde	ã	2. Latin-1 Supplement
[tca]_002	00ED	iacute	í	2. Latin-1 Supplement
[tca]_002	00FC	udieresis	ü	2. Latin-1 Supplement
[tca]_002	0128	ltilde	ĺ	3. Latin Ext.-A
[tca]_002	0129	itilde	ı	3. Latin Ext.-A
[tca]_002	0168	Utilde	Û	3. Latin Ext.-A
[tca]_002	0169	utilde	ü	3. Latin Ext.-A
[tca]_002	02BC	apostrophemod	'	6. Spc. Mod. Letters
[tca]_002	1EBC	Etilde	Ë	8. Latin Ext. Add.
[tca]_002	1EBD	etilde	ë	8. Latin Ext. Add.
[tca]_002	A78B	Saltillo	ı	9. Latin Ext.-D
[tca]_002	A78C	saltillo	ı	9. Latin Ext.-D
[tca]_002	00DC_0303	Udieresis_tildecomb	Û	10. non-compliant
[tca]_002	00FC_0303	udieresis_tildecomb	ü	10. non-compliant
[ter]_001	00C1	Aacute	Á	2. Latin-1 Supplement
[ter]_001	00C2	Acircumflex	Â	2. Latin-1 Supplement
[ter]_001	00C9	Eacute	É	2. Latin-1 Supplement
[ter]_001	00CA	Ecircumflex	Ê	2. Latin-1 Supplement
[ter]_001	00CD	lacute	Í	2. Latin-1 Supplement
[ter]_001	00CE	Icircumflex	Î	2. Latin-1 Supplement
[ter]_001	00D3	Oacute	Ó	2. Latin-1 Supplement
[ter]_001	00D4	Ocircumflex	Ô	2. Latin-1 Supplement
[ter]_001	00DA	Uacute	Ú	2. Latin-1 Supplement
[ter]_001	00DB	Ucircumflex	Û	2. Latin-1 Supplement
[ter]_001	00E1	aacute	á	2. Latin-1 Supplement
[ter]_001	00E2	acircumflex	â	2. Latin-1 Supplement
[ter]_001	00E9	eacute	é	2. Latin-1 Supplement
[ter]_001	00EA	ecircumflex	ê	2. Latin-1 Supplement
[ter]_001	00ED	iacute	í	2. Latin-1 Supplement
[ter]_001	00EE	icircumflex	î	2. Latin-1 Supplement
[ter]_001	00F3	oacute	ó	2. Latin-1 Supplement
[ter]_001	00F4	ocircumflex	ô	2. Latin-1 Supplement
[ter]_001	00FA	uacute	ú	2. Latin-1 Supplement
[ter]_001	00FB	ucircumflex	û	2. Latin-1 Supplement
[ter]_002	00C1	Aacute	Á	2. Latin-1 Supplement
[ter]_002	00C2	Acircumflex	Â	2. Latin-1 Supplement
[ter]_002	00C9	Eacute	É	2. Latin-1 Supplement
[ter]_002	00CA	Ecircumflex	Ê	2. Latin-1 Supplement
[ter]_002	00CD	lacute	Í	2. Latin-1 Supplement
[ter]_002	00CE	Icircumflex	Î	2. Latin-1 Supplement
[ter]_002	00D3	Oacute	Ó	2. Latin-1 Supplement
[ter]_002	00D4	Ocircumflex	Ô	2. Latin-1 Supplement
[ter]_002	00DA	Uacute	Ú	2. Latin-1 Supplement

arquivo	Unicode	Glyphname	Character	bloco
[ter]_002	00DB	Ucircumflex	Û	2. Latin-1 Supplement
[ter]_002	00E1	aacute	á	2. Latin-1 Supplement
[ter]_002	00E2	acircumflex	â	2. Latin-1 Supplement
[ter]_002	00E9	eacute	é	2. Latin-1 Supplement
[ter]_002	00EA	ecircumflex	ê	2. Latin-1 Supplement
[ter]_002	00ED	iacute	í	2. Latin-1 Supplement
[ter]_002	00EE	icircumflex	î	2. Latin-1 Supplement
[ter]_002	00F3	oacute	ó	2. Latin-1 Supplement
[ter]_002	00F4	ocircumflex	ô	2. Latin-1 Supplement
[ter]_002	00FA	uacute	ú	2. Latin-1 Supplement
[ter]_002	00FB	ucircumflex	û	2. Latin-1 Supplement
[ter]_003	00C1	Aacute	Á	2. Latin-1 Supplement
[ter]_003	00C2	Acircumflex	Â	2. Latin-1 Supplement
[ter]_003	00C9	Eacute	É	2. Latin-1 Supplement
[ter]_003	00CA	Ecircumflex	Ê	2. Latin-1 Supplement
[ter]_003	00CD	Iacute	Í	2. Latin-1 Supplement
[ter]_003	00CE	Icircumflex	Î	2. Latin-1 Supplement
[ter]_003	00D3	Oacute	Ó	2. Latin-1 Supplement
[ter]_003	00D4	Ocircumflex	Ô	2. Latin-1 Supplement
[ter]_003	00DA	Uacute	Ú	2. Latin-1 Supplement
[ter]_003	00DB	Ucircumflex	Û	2. Latin-1 Supplement
[ter]_003	00E1	aacute	á	2. Latin-1 Supplement
[ter]_003	00E2	acircumflex	â	2. Latin-1 Supplement
[ter]_003	00E9	eacute	é	2. Latin-1 Supplement
[ter]_003	00EA	ecircumflex	ê	2. Latin-1 Supplement
[ter]_003	00ED	iacute	í	2. Latin-1 Supplement
[ter]_003	00EE	icircumflex	î	2. Latin-1 Supplement
[ter]_003	00F3	oacute	ó	2. Latin-1 Supplement
[ter]_003	00F4	ocircumflex	ô	2. Latin-1 Supplement
[ter]_003	00FA	uacute	ú	2. Latin-1 Supplement
[ter]_003	00FB	ucircumflex	û	2. Latin-1 Supplement
[ter]_004	00C1	Aacute	Á	2. Latin-1 Supplement
[ter]_004	00C2	Acircumflex	Â	2. Latin-1 Supplement
[ter]_004	00C9	Eacute	É	2. Latin-1 Supplement
[ter]_004	00CA	Ecircumflex	Ê	2. Latin-1 Supplement
[ter]_004	00CD	Iacute	Í	2. Latin-1 Supplement
[ter]_004	00CE	Icircumflex	Î	2. Latin-1 Supplement
[ter]_004	00D3	Oacute	Ó	2. Latin-1 Supplement
[ter]_004	00D4	Ocircumflex	Ô	2. Latin-1 Supplement
[ter]_004	00DA	Uacute	Ú	2. Latin-1 Supplement
[ter]_004	00DB	Ucircumflex	Û	2. Latin-1 Supplement
[ter]_004	00E1	aacute	á	2. Latin-1 Supplement
[ter]_004	00E2	acircumflex	â	2. Latin-1 Supplement
[ter]_004	00E9	eacute	é	2. Latin-1 Supplement
[ter]_004	00EA	ecircumflex	ê	2. Latin-1 Supplement
[ter]_004	00ED	iacute	í	2. Latin-1 Supplement
[ter]_004	00EE	icircumflex	î	2. Latin-1 Supplement

arquivo	Unicode	Glyphname	Character	bloco
[ter]_004	00F3	oacute	ó	2. Latin-1 Supplement
[ter]_004	00F4	ocircumflex	ô	2. Latin-1 Supplement
[ter]_004	00FA	uacute	ú	2. Latin-1 Supplement
[ter]_004	00FB	ucircumflex	û	2. Latin-1 Supplement
[ter]_005	0027	quotesingle	'	1. Basic Latin
[ter]_005	00C1	Aacute	Á	2. Latin-1 Supplement
[ter]_005	00C2	Acircumflex	Â	2. Latin-1 Supplement
[ter]_005	00C9	Eacute	É	2. Latin-1 Supplement
[ter]_005	00CA	Ecircumflex	Ê	2. Latin-1 Supplement
[ter]_005	00CD	Iacute	Í	2. Latin-1 Supplement
[ter]_005	00CE	Icircumflex	Î	2. Latin-1 Supplement
[ter]_005	00D3	Oacute	Ó	2. Latin-1 Supplement
[ter]_005	00D4	Ocircumflex	Ô	2. Latin-1 Supplement
[ter]_005	00DA	Uacute	Ú	2. Latin-1 Supplement
[ter]_005	00DB	Ucircumflex	Û	2. Latin-1 Supplement
[ter]_005	00E1	aacute	á	2. Latin-1 Supplement
[ter]_005	00E2	acircumflex	â	2. Latin-1 Supplement
[ter]_005	00E9	eacute	é	2. Latin-1 Supplement
[ter]_005	00EA	ecircumflex	ê	2. Latin-1 Supplement
[ter]_005	00ED	iacute	í	2. Latin-1 Supplement
[ter]_005	00EE	icircumflex	î	2. Latin-1 Supplement
[ter]_005	00F3	oacute	ó	2. Latin-1 Supplement
[ter]_005	00F4	ocircumflex	ô	2. Latin-1 Supplement
[ter]_005	00FA	uacute	ú	2. Latin-1 Supplement
[ter]_005	00FB	ucircumflex	û	2. Latin-1 Supplement
[tpr]_001	0027	quotesingle	'	1. Basic Latin
[tpr]_001	00C0	Agrave	À	2. Latin-1 Supplement
[tpr]_001	00C3	Atilde	Ã	2. Latin-1 Supplement
[tpr]_001	00C8	Egrave	È	2. Latin-1 Supplement
[tpr]_001	00CC	Igrave	Ì	2. Latin-1 Supplement
[tpr]_001	00D2	Ograve	Ò	2. Latin-1 Supplement
[tpr]_001	00D5	Otilde	Õ	2. Latin-1 Supplement
[tpr]_001	00D9	Ugrave	Ù	2. Latin-1 Supplement
[tpr]_001	00E0	agrave	à	2. Latin-1 Supplement
[tpr]_001	00E3	atilde	ã	2. Latin-1 Supplement
[tpr]_001	00E8	egrave	è	2. Latin-1 Supplement
[tpr]_001	00EC	igrave	ì	2. Latin-1 Supplement
[tpr]_001	00F2	ograve	ò	2. Latin-1 Supplement
[tpr]_001	00F5	otilde	õ	2. Latin-1 Supplement
[tpr]_001	00F9	ugrave	ù	2. Latin-1 Supplement
[tpr]_001	0128	Itilde	Ï	3. Latin Ext.-A
[tpr]_001	0129	itilde	ï	3. Latin Ext.-A
[tpr]_001	0168	Utilde	Û	3. Latin Ext.-A
[tpr]_001	0169	utilde	ü	3. Latin Ext.-A
[tpr]_001	1EBC	Etilde	Ě	8. Latin Ext. Add.
[tpr]_001	1EBD	etilde	ě	8. Latin Ext. Add.
[tpr]_001	2019	quoteright	'	11. General Punctuation

arquivo	Unicode	Glyphname	Character	bloco
[tpy]_001	00CF	Idieresis	ÿ	2. Latin-1 Supplement
[tpy]_001	00EF	idieresis	ÿ	2. Latin-1 Supplement
[tpy]_001	021A	Tcommaaccent	Ŧ	4. Latin Ext.-B
[tpy]_001	021B	tcommaaccent	ŧ	4. Latin Ext.-B
[tpy]_001	02BC	apostrophemod	'	6. Spc. Mod. Letters
[tpy]_001	A78B	Saltillo	ı	9. Latin Ext.-D
[tpy]_001	A78C	salttillo	ı	9. Latin Ext.-D
[tpy]_002	00CF	Idieresis	ÿ	2. Latin-1 Supplement
[tpy]_002	00EF	idieresis	ÿ	2. Latin-1 Supplement
[tpy]_002	021A	Tcommaaccent	Ŧ	4. Latin Ext.-B
[tpy]_002	021B	tcommaaccent	ŧ	4. Latin Ext.-B
[tpy]_002	02BC	apostrophemod	'	6. Spc. Mod. Letters
[tpy]_002	A78B	Saltillo	ı	9. Latin Ext.-D
[tpy]_002	A78C	salttillo	ı	9. Latin Ext.-D
[tpy]_003	00CF	Idieresis	ÿ	2. Latin-1 Supplement
[tpy]_003	00EF	idieresis	ÿ	2. Latin-1 Supplement
[tpy]_003	021A	Tcommaaccent	Ŧ	4. Latin Ext.-B
[tpy]_003	021B	tcommaaccent	ŧ	4. Latin Ext.-B
[tpy]_003	02BC	apostrophemod	'	6. Spc. Mod. Letters
[tpy]_003	A78B	Saltillo	ı	9. Latin Ext.-D
[tpy]_003	A78C	salttillo	ı	9. Latin Ext.-D
[tue]_001	00C3	Atilde	Ã	2. Latin-1 Supplement
[tue]_001	00D1	Ntilde	Ñ	2. Latin-1 Supplement
[tue]_001	00D5	Otilde	Õ	2. Latin-1 Supplement
[tue]_001	00E3	atilde	ã	2. Latin-1 Supplement
[tue]_001	00F1	ntilde	ñ	2. Latin-1 Supplement
[tue]_001	00F5	otilde	õ	2. Latin-1 Supplement
[tue]_001	0128	ltilde	Ł	3. Latin Ext.-A
[tue]_001	0129	itilde	ł	3. Latin Ext.-A
[tue]_001	0168	Utilde	Ū	3. Latin Ext.-A
[tue]_001	0169	utilde	ū	3. Latin Ext.-A
[tue]_001	0244	Ubar	Ʊ	4. Latin Ext.-B
[tue]_001	0289	ubar	Ʊ	5. IPA Extensions
[tue]_001	1EBC	Etilde	Ě	8. Latin Ext. Add.
[tue]_001	1EBD	etilde	ě	8. Latin Ext. Add.
[tue]_001	0244_0303	Ubar_tildecomb	Ʊ̃	10. non-compliant
[tue]_001	0289_0303	ubar_tildecomb	Ʊ̃	10. non-compliant
[tue]_002	00C3	Atilde	Ã	2. Latin-1 Supplement
[tue]_002	00D1	Ntilde	Ñ	2. Latin-1 Supplement
[tue]_002	00D5	Otilde	Õ	2. Latin-1 Supplement
[tue]_002	00E3	atilde	ã	2. Latin-1 Supplement
[tue]_002	00F1	ntilde	ñ	2. Latin-1 Supplement
[tue]_002	00F5	otilde	õ	2. Latin-1 Supplement
[tue]_002	0128	ltilde	Ł	3. Latin Ext.-A
[tue]_002	0129	itilde	ł	3. Latin Ext.-A

arquivo	Unicode	Glyphname	Character	bloco
[tue]_002	0168	Utilde	Ū	3. Latin Ext.-A
[tue]_002	0169	utilde	ū	3. Latin Ext.-A
[tue]_002	0244	Ubar	Ū̄	4. Latin Ext.-B
[tue]_002	0289	ubar	ū̄	5. IPA Extensions
[tue]_002	1EBC	Etilde	Ě	8. Latin Ext. Add.
[tue]_002	1EBD	etilde	ě	8. Latin Ext. Add.
[tue]_002	0244_0303	Ubar_tildecomb	Ū̄̃	10. non-compliant
[tue]_002	0289_0303	ubar_tildecomb	ū̄̃	10. non-compliant
[tue]_003	00C3	Atilde	Ā	2. Latin-1 Supplement
[tue]_003	00D1	Ntilde	Ñ	2. Latin-1 Supplement
[tue]_003	00D5	Otilde	Ō	2. Latin-1 Supplement
[tue]_003	00E3	atilde	ā	2. Latin-1 Supplement
[tue]_003	00F1	ntilde	ñ	2. Latin-1 Supplement
[tue]_003	00F5	otilde	ō	2. Latin-1 Supplement
[tue]_003	0128	ltilde	Ĺ	3. Latin Ext.-A
[tue]_003	0129	itilde	ĺ	3. Latin Ext.-A
[tue]_003	0168	Utilde	Ū	3. Latin Ext.-A
[tue]_003	0169	utilde	ū	3. Latin Ext.-A
[tue]_003	0244	Ubar	Ū̄	4. Latin Ext.-B
[tue]_003	0289	ubar	ū̄	5. IPA Extensions
[tue]_003	1EBC	Etilde	Ě	8. Latin Ext. Add.
[tue]_003	1EBD	etilde	ě	8. Latin Ext. Add.
[tue]_003	0244_0303	Ubar_tildecomb	Ū̄̃	10. non-compliant
[tue]_003	0289_0303	ubar_tildecomb	ū̄̃	10. non-compliant
[tue]_004	00C3	Atilde	Ā	2. Latin-1 Supplement
[tue]_004	00D1	Ntilde	Ñ	2. Latin-1 Supplement
[tue]_004	00D5	Otilde	Ō	2. Latin-1 Supplement
[tue]_004	00E3	atilde	ā	2. Latin-1 Supplement
[tue]_004	00F1	ntilde	ñ	2. Latin-1 Supplement
[tue]_004	00F5	otilde	ō	2. Latin-1 Supplement
[tue]_004	0128	ltilde	Ĺ	3. Latin Ext.-A
[tue]_004	0129	itilde	ĺ	3. Latin Ext.-A
[tue]_004	0168	Utilde	Ū	3. Latin Ext.-A
[tue]_004	0169	utilde	ū	3. Latin Ext.-A
[tue]_004	0244	Ubar	Ū̄	4. Latin Ext.-B
[tue]_004	0289	ubar	ū̄	5. IPA Extensions
[tue]_004	1EBC	Etilde	Ě	8. Latin Ext. Add.
[tue]_004	1EBD	etilde	ě	8. Latin Ext. Add.
[tue]_004	0244_0303	Ubar_tildecomb	Ū̄̃	10. non-compliant
[tue]_004	0289_0303	ubar_tildecomb	ū̄̃	10. non-compliant
[tue]_005	00C3	Atilde	Ā	2. Latin-1 Supplement
[tue]_005	00D1	Ntilde	Ñ	2. Latin-1 Supplement
[tue]_005	00D5	Otilde	Ō	2. Latin-1 Supplement
[tue]_005	00E3	atilde	ā	2. Latin-1 Supplement
[tue]_005	00F1	ntilde	ñ	2. Latin-1 Supplement
[tue]_005	00F5	otilde	ō	2. Latin-1 Supplement
[tue]_005	0128	ltilde	Ĺ	3. Latin Ext.-A

arquivo	Unicode	Glyphname	Character	bloco
[tue]_005	0129	itilde	ĩ	3. Latin Ext.-A
[tue]_005	0168	Utilde	Û	3. Latin Ext.-A
[tue]_005	0169	utilde	ü	3. Latin Ext.-A
[tue]_005	0244	Ubar	Ů	4. Latin Ext.-B
[tue]_005	0289	ubar	ɯ	5. IPA Extensions
[tue]_005	1EBC	Etilde	Ě	8. Latin Ext. Add.
[tue]_005	1EBD	etilde	ě	8. Latin Ext. Add.
[tue]_005	0244_0303	Ubar_tildecomb	Ů̃	10. non-compliant
[tue]_005	0289_0303	ubar_tildecomb	ɯ̃	10. non-compliant
[tuo]_001	00C1	Aacute	Á	2. Latin-1 Supplement
[tuo]_001	00C2	Acircumflex	Â	2. Latin-1 Supplement
[tuo]_001	00C3	Atilde	Ã	2. Latin-1 Supplement
[tuo]_001	00C9	Eacute	É	2. Latin-1 Supplement
[tuo]_001	00CA	Ecircumflex	Ê	2. Latin-1 Supplement
[tuo]_001	00CD	Iacute	Í	2. Latin-1 Supplement
[tuo]_001	00CE	Icircumflex	Î	2. Latin-1 Supplement
[tuo]_001	00D3	Oacute	Ó	2. Latin-1 Supplement
[tuo]_001	00D4	Ocircumflex	Ô	2. Latin-1 Supplement
[tuo]_001	00D5	Otilde	Õ	2. Latin-1 Supplement
[tuo]_001	00DA	Uacute	Ú	2. Latin-1 Supplement
[tuo]_001	00DB	Ucircumflex	Û	2. Latin-1 Supplement
[tuo]_001	00E1	aacute	á	2. Latin-1 Supplement
[tuo]_001	00E2	acircumflex	â	2. Latin-1 Supplement
[tuo]_001	00E3	atilde	ã	2. Latin-1 Supplement
[tuo]_001	00E9	eacute	é	2. Latin-1 Supplement
[tuo]_001	00EA	ecircumflex	ê	2. Latin-1 Supplement
[tuo]_001	00ED	iacute	í	2. Latin-1 Supplement
[tuo]_001	00EE	icircumflex	î	2. Latin-1 Supplement
[tuo]_001	00F3	oacute	ó	2. Latin-1 Supplement
[tuo]_001	00F4	ocircumflex	ô	2. Latin-1 Supplement
[tuo]_001	00F5	otilde	õ	2. Latin-1 Supplement
[tuo]_001	00FA	uacute	ú	2. Latin-1 Supplement
[tuo]_001	00FB	ucircumflex	û	2. Latin-1 Supplement
[tuo]_001	0128	Itilde	İ	3. Latin Ext.-A
[tuo]_001	0129	itilde	ı	3. Latin Ext.-A
[tuo]_001	0168	Utilde	Û	3. Latin Ext.-A
[tuo]_001	0169	utilde	ü	3. Latin Ext.-A
[tuo]_001	0197	Istroke	†	4. Latin Ext.-B
[tuo]_001	0268	istroke	‡	5. IPA Extensions
[tuo]_001	02BC	apostrophemod	’	6. Spc. Mod. Letters
[tuo]_001	1EBC	Etilde	Ě	8. Latin Ext. Add.
[tuo]_001	1EBD	etilde	ě	8. Latin Ext. Add.
[tuo]_001	A78B	Saltillo	ı	9. Latin Ext.-D
[tuo]_001	A78C	saltillo	ı	9. Latin Ext.-D
[tuo]_001	0197_0301	Istroke_acutecomb	†̃	10. non-compliant
[tuo]_001	0268_0301	istroke_acutecomb	‡̃	10. non-compliant

arquivo	Unicode	Glyphname	Character	bloco
[tuo]_001	0197_0302	lstroke_circumflexcomb	ĥ	10. non-compliant
[tuo]_001	0268_0302	istroke_circumflexcomb	ĥ	10. non-compliant
[tuo]_001	0197_0303	lstroke_tildecomb	ĥ̃	10. non-compliant
[tuo]_001	0268_0303	istroke_tildecomb	ĥ̃	10. non-compliant
[tuo]_002	00C1	Aacute	Á	2. Latin-1 Supplement
[tuo]_002	00C2	Acircumflex	Â	2. Latin-1 Supplement
[tuo]_002	00C3	Atilde	Ã	2. Latin-1 Supplement
[tuo]_002	00C9	Eacute	É	2. Latin-1 Supplement
[tuo]_002	00CA	Ecircumflex	Ê	2. Latin-1 Supplement
[tuo]_002	00CD	Iacute	Í	2. Latin-1 Supplement
[tuo]_002	00CE	Icircumflex	Î	2. Latin-1 Supplement
[tuo]_002	00D3	Oacute	Ó	2. Latin-1 Supplement
[tuo]_002	00D4	Ocircumflex	Ô	2. Latin-1 Supplement
[tuo]_002	00D5	Otilde	Õ	2. Latin-1 Supplement
[tuo]_002	00DA	Uacute	Ú	2. Latin-1 Supplement
[tuo]_002	00DB	Ucircumflex	Û	2. Latin-1 Supplement
[tuo]_002	00E1	aacute	á	2. Latin-1 Supplement
[tuo]_002	00E2	acircumflex	â	2. Latin-1 Supplement
[tuo]_002	00E3	atilde	ã	2. Latin-1 Supplement
[tuo]_002	00E9	eacute	é	2. Latin-1 Supplement
[tuo]_002	00EA	ecircumflex	ê	2. Latin-1 Supplement
[tuo]_002	00ED	iacute	í	2. Latin-1 Supplement
[tuo]_002	00EE	icircumflex	î	2. Latin-1 Supplement
[tuo]_002	00F3	oacute	ó	2. Latin-1 Supplement
[tuo]_002	00F4	ocircumflex	ô	2. Latin-1 Supplement
[tuo]_002	00F5	otilde	õ	2. Latin-1 Supplement
[tuo]_002	00FA	uacute	ú	2. Latin-1 Supplement
[tuo]_002	00FB	ucircumflex	û	2. Latin-1 Supplement
[tuo]_002	0128	ltilde	Ĺ	3. Latin Ext.-A
[tuo]_002	0129	itilde	Ĩ	3. Latin Ext.-A
[tuo]_002	0168	Utilde	Ū	3. Latin Ext.-A
[tuo]_002	0169	utilde	ŭ	3. Latin Ext.-A
[tuo]_002	0197	lstroke	ł	4. Latin Ext.-B
[tuo]_002	0268	istroke	ł	5. IPA Extensions
[tuo]_002	02BC	apostrophemod	’	6. Spc. Mod. Letters
[tuo]_002	A78B	Saltillo	ı	9. Latin Ext.-D
[tuo]_002	A78C	salttillo	ı	9. Latin Ext.-D
[tuo]_002	00C3_0301	Atilde_acutecomb	Ă	10. non-compliant
[tuo]_002	00E3_0301	atilde_acutecomb	ă	10. non-compliant
[tuo]_002	0197_0301	lstroke_acutecomb	Ł	10. non-compliant
[tuo]_002	0268_0301	istroke_acutecomb	ł	10. non-compliant
[tuo]_002	0197_0302	lstroke_circumflexcomb	ł̂	10. non-compliant
[tuo]_002	0268_0302	istroke_circumflexcomb	ł̂	10. non-compliant
[tuo]_002	0197_0303	lstroke_tildecomb	ł̃	10. non-compliant
[tuo]_002	0268_0303	istroke_tildecomb	ł̃	10. non-compliant
[tuo]_003	00C3	Atilde	Ã	2. Latin-1 Supplement

arquivo	Unicode	Glyphname	Character	bloco
[tuo]_003	00D1	Ntilde	Ñ	2. Latin-1 Supplement
[tuo]_003	00D5	Otilde	Õ	2. Latin-1 Supplement
[tuo]_003	00E3	atilde	ã	2. Latin-1 Supplement
[tuo]_003	00F1	ntilde	ñ	2. Latin-1 Supplement
[tuo]_003	00F5	otilde	õ	2. Latin-1 Supplement
[tuo]_003	0128	ltilde	Ĺ	3. Latin Ext.-A
[tuo]_003	0129	itilde	ĺ	3. Latin Ext.-A
[tuo]_003	0168	Utilde	Ū	3. Latin Ext.-A
[tuo]_003	0169	utilde	ū	3. Latin Ext.-A
[tuo]_003	0244	Ubar	Ʊ	4. Latin Ext.-B
[tuo]_003	0289	ubar	ƶ	5. IPA Extensions
[tuo]_003	1EBC	Etilde	Ě	8. Latin Ext. Add.
[tuo]_003	1EBD	etilde	ě	8. Latin Ext. Add.
[tuo]_003	0244_0303	Ubar_tildecomb	Ʊ̃	10. non-compliant
[tuo]_003	0289_0303	ubar_tildecomb	ƶ̃	10. non-compliant
[tuo]_004	00C1	Aacute	Á	2. Latin-1 Supplement
[tuo]_004	00C3	Atilde	Ã	2. Latin-1 Supplement
[tuo]_004	00CD	iacute	Í	2. Latin-1 Supplement
[tuo]_004	00D1	Ntilde	Ñ	2. Latin-1 Supplement
[tuo]_004	00D3	Oacute	Ó	2. Latin-1 Supplement
[tuo]_004	00D5	Otilde	Õ	2. Latin-1 Supplement
[tuo]_004	00E1	aacute	á	2. Latin-1 Supplement
[tuo]_004	00E3	atilde	ã	2. Latin-1 Supplement
[tuo]_004	00ED	iacute	í	2. Latin-1 Supplement
[tuo]_004	00F1	ntilde	ñ	2. Latin-1 Supplement
[tuo]_004	00F3	oacute	ó	2. Latin-1 Supplement
[tuo]_004	00F5	otilde	õ	2. Latin-1 Supplement
[tuo]_004	0128	ltilde	Ĺ	3. Latin Ext.-A
[tuo]_004	0129	itilde	ĺ	3. Latin Ext.-A
[tuo]_004	0168	Utilde	Ū	3. Latin Ext.-A
[tuo]_004	0169	utilde	ū	3. Latin Ext.-A
[tuo]_004	0244	Ubar	Ʊ	4. Latin Ext.-B
[tuo]_004	0289	ubar	ƶ	5. IPA Extensions
[tuo]_004	1EBC	Etilde	Ě	8. Latin Ext. Add.
[tuo]_004	1EBD	etilde	ě	8. Latin Ext. Add.
[tuo]_004	0244_0303	Ubar_tildecomb	Ʊ̃	10. non-compliant
[tuo]_004	0289_0303	ubar_tildecomb	ƶ̃	10. non-compliant
[tuo]_005	00C1	Aacute	Á	2. Latin-1 Supplement
[tuo]_005	00C2	Acircumflex	Â	2. Latin-1 Supplement
[tuo]_005	00C3	Atilde	Ã	2. Latin-1 Supplement
[tuo]_005	00C9	Eacute	É	2. Latin-1 Supplement
[tuo]_005	00CA	Ecircumflex	Ê	2. Latin-1 Supplement
[tuo]_005	00CD	iacute	Í	2. Latin-1 Supplement
[tuo]_005	00CE	Icircumflex	Î	2. Latin-1 Supplement
[tuo]_005	00D1	Ntilde	Ñ	2. Latin-1 Supplement
[tuo]_005	00D3	Oacute	Ó	2. Latin-1 Supplement
[tuo]_005	00D4	Ocircumflex	Ô	2. Latin-1 Supplement

arquivo	Unicode	Glyphname	Character	bloco
[tuo]_005	00D5	Otilde	Õ	2. Latin-1 Supplement
[tuo]_005	00E1	aacute	á	2. Latin-1 Supplement
[tuo]_005	00E2	acircumflex	â	2. Latin-1 Supplement
[tuo]_005	00E3	atilde	ã	2. Latin-1 Supplement
[tuo]_005	00E9	eacute	é	2. Latin-1 Supplement
[tuo]_005	00EA	ecircumflex	ê	2. Latin-1 Supplement
[tuo]_005	00ED	iacute	í	2. Latin-1 Supplement
[tuo]_005	00EE	icircumflex	î	2. Latin-1 Supplement
[tuo]_005	00F1	ntilde	ñ	2. Latin-1 Supplement
[tuo]_005	00F3	oacute	ó	2. Latin-1 Supplement
[tuo]_005	00F4	ocircumflex	ô	2. Latin-1 Supplement
[tuo]_005	00F5	otilde	õ	2. Latin-1 Supplement
[tuo]_005	0128	ltilde	ĩ	3. Latin Ext.-A
[tuo]_005	0129	itilde	ï	3. Latin Ext.-A
[tuo]_005	0197	lstroke	†	4. Latin Ext.-B
[tuo]_005	0268	istroke	ı	5. IPA Extensions
[tuo]_006	00C1	Aacute	Á	2. Latin-1 Supplement
[tuo]_006	00C2	Acircumflex	Â	2. Latin-1 Supplement
[tuo]_006	00C3	Atilde	Ã	2. Latin-1 Supplement
[tuo]_006	00C9	Eacute	É	2. Latin-1 Supplement
[tuo]_006	00CA	Ecircumflex	Ê	2. Latin-1 Supplement
[tuo]_006	00CD	Iacute	Í	2. Latin-1 Supplement
[tuo]_006	00CE	Icircumflex	Î	2. Latin-1 Supplement
[tuo]_006	00D3	Oacute	Ó	2. Latin-1 Supplement
[tuo]_006	00D4	Ocircumflex	Ô	2. Latin-1 Supplement
[tuo]_006	00D5	Otilde	Õ	2. Latin-1 Supplement
[tuo]_006	00DA	Uacute	Ú	2. Latin-1 Supplement
[tuo]_006	00DB	Ucircumflex	Û	2. Latin-1 Supplement
[tuo]_006	00E1	aacute	á	2. Latin-1 Supplement
[tuo]_006	00E2	acircumflex	â	2. Latin-1 Supplement
[tuo]_006	00E3	atilde	ã	2. Latin-1 Supplement
[tuo]_006	00E9	eacute	é	2. Latin-1 Supplement
[tuo]_006	00EA	ecircumflex	ê	2. Latin-1 Supplement
[tuo]_006	00ED	iacute	í	2. Latin-1 Supplement
[tuo]_006	00EE	icircumflex	î	2. Latin-1 Supplement
[tuo]_006	00F3	oacute	ó	2. Latin-1 Supplement
[tuo]_006	00F4	ocircumflex	ô	2. Latin-1 Supplement
[tuo]_006	00F5	otilde	õ	2. Latin-1 Supplement
[tuo]_006	00FA	uacute	ú	2. Latin-1 Supplement
[tuo]_006	00FB	ucircumflex	û	2. Latin-1 Supplement
[tuo]_006	0128	ltilde	ĩ	3. Latin Ext.-A
[tuo]_006	0129	itilde	ï	3. Latin Ext.-A
[tuo]_006	0168	Utilde	Û	3. Latin Ext.-A
[tuo]_006	0169	utilde	ü	3. Latin Ext.-A
[tuo]_006	0197	lstroke	†	4. Latin Ext.-B
[tuo]_006	0268	istroke	ı	5. IPA Extensions
[tuo]_006	0301	acutecomb	´	7. Combining Marks

arquivo	Unicode	Glyphname	Character	bloco
[tuo]_006	0302	circumflexcomb	ˆ	7. Combining Marks
[tuo]_006	0303	tildecomb	˜	7. Combining Marks
[tuo]_006	0341	acutetonecomb	´	7. Combining Marks
[tuo]_006	1E4C	Otildeacute	Ó	8. Latin Ext. Add.
[tuo]_006	1E4D	otildeacute	ó	8. Latin Ext. Add.
[tuo]_006	1E78	Utildeacute	Ú	8. Latin Ext. Add.
[tuo]_006	1E79	utildeacute	ú	8. Latin Ext. Add.
[tuo]_006	1EBC	Etilde	Ě	8. Latin Ext. Add.
[tuo]_006	1EBD	etilde	ě	8. Latin Ext. Add.
[tuo]_006	00C3_0301	Atilde_acutecomb	Á	10. non-compliant
[tuo]_006	00E3_0301	atilde_acutecomb	á	10. non-compliant
[tuo]_006	1EBC_0301	Etilde_acutecomb	Ě	10. non-compliant
[tuo]_006	1EBD_0301	etilde_acutecomb	ě	10. non-compliant
[tuo]_006	0128_0301	ltilde_acutecomb	Í	10. non-compliant
[tuo]_006	0129_0301	itilde_acutecomb	í	10. non-compliant
[tuo]_006	00D5_0301	Otilde_acutecomb	Ó	10. non-compliant
[tuo]_006	00F5_0301	otilde_acutecomb	ó	10. non-compliant
[tuo]_006	0168_0301	Utilde_acutecomb	Ú	10. non-compliant
[tuo]_006	0169_0301	utilde_acutecomb	ú	10. non-compliant
[tuo]_006	00C3_0302	Atilde_circumflexcomb	Ă	10. non-compliant
[tuo]_006	00E3_0302	atilde_circumflexcomb	ă	10. non-compliant
[tuo]_006	1EBC_0302	Etilde_circumflexcomb	Ě	10. non-compliant
[tuo]_006	1EBD_0302	etilde_circumflexcomb	ě	10. non-compliant
[tuo]_006	0128_0302	ltilde_circumflexcomb	Î	10. non-compliant
[tuo]_006	0129_0302	itilde_circumflexcomb	î	10. non-compliant
[tuo]_006	00D5_0302	Otilde_circumflexcomb	Ŏ	10. non-compliant
[tuo]_006	00F5_0302	otilde_circumflexcomb	õ	10. non-compliant
[tuo]_006	0168_0302	Utilde_circumflexcomb	Û	10. non-compliant
[tuo]_006	0169_0302	utilde_circumflexcomb	û	10. non-compliant
[tuo]_006	0197_0301	lstroke_acutecomb	ƒ	10. non-compliant
[tuo]_006	0268_0301	istroke_acutecomb	f	10. non-compliant
[tuo]_006	0197_0302	lstroke_circumflexcomb	ĥ	10. non-compliant
[tuo]_006	0268_0302	istroke_circumflexcomb	h	10. non-compliant
[tuo]_006	0197_0303	lstroke_tildecomb	ƒ̃	10. non-compliant
[tuo]_006	0268_0303	istroke_tildecomb	f̃	10. non-compliant
[tuo]_006	0197_0303_0301	lstroke_tildecomb_acutecomb	ƒ̃́	10. non-compliant
[tuo]_006	0268_0303_0301	istroke_tildecomb_acutecomb	f̃́	10. non-compliant
[tuo]_006	0197_0303_0302	lstroke_tildecomb_circumflexcomb	ĥ̃	10. non-compliant
[tuo]_006	0268_0303_0302	istroke_tildecomb_circumflexcomb	h̃	10. non-compliant
[txi]_001	00CF	Idieresis	ÿ	2. Latin-1 Supplement
[txi]_001	00EF	idieresis	ÿ	2. Latin-1 Supplement
[txi]_002	00CF	Idieresis	ÿ	2. Latin-1 Supplement
[txi]_002	00EF	idieresis	ÿ	2. Latin-1 Supplement
[txi]_003	00CF	Idieresis	ÿ	2. Latin-1 Supplement
[txi]_003	00EF	idieresis	ÿ	2. Latin-1 Supplement

arquivo	Unicode	Glyphname	Character	bloco
[txu]_001	00C0	Agrave	À	2. Latin-1 Supplement
[txu]_001	00C3	Atilde	Ã	2. Latin-1 Supplement
[txu]_001	00CA	Ecircumflex	Ê	2. Latin-1 Supplement
[txu]_001	00D4	Ocircumflex	Ô	2. Latin-1 Supplement
[txu]_001	00D5	Otilde	Õ	2. Latin-1 Supplement
[txu]_001	00E0	agrave	à	2. Latin-1 Supplement
[txu]_001	00E3	atilde	ã	2. Latin-1 Supplement
[txu]_001	00EA	ecircumflex	ê	2. Latin-1 Supplement
[txu]_001	00F4	ocircumflex	ô	2. Latin-1 Supplement
[txu]_001	00F5	otilde	õ	2. Latin-1 Supplement
[txu]_001	0128	Itilde	Ï	3. Latin Ext.-A
[txu]_001	0129	itilde	ï	3. Latin Ext.-A
[txu]_001	0168	Utilde	Û	3. Latin Ext.-A
[txu]_001	0169	utilde	ü	3. Latin Ext.-A
[txu]_001	1EBC	Etilde	Ë	8. Latin Ext. Add.
[txu]_001	1EBD	etilde	ë	8. Latin Ext. Add.
[txu]_001	1EF2	Ygrave	ÿ	8. Latin Ext. Add.
[txu]_001	1EF3	ygrave	ÿ	8. Latin Ext. Add.
[txu]_001	1EF8	Ytilde	ÿ	8. Latin Ext. Add.
[txu]_001	1EF9	ytilde	ÿ	8. Latin Ext. Add.
[txu]_002	00C0	Agrave	À	2. Latin-1 Supplement
[txu]_002	00C3	Atilde	Ã	2. Latin-1 Supplement
[txu]_002	00CA	Ecircumflex	Ê	2. Latin-1 Supplement
[txu]_002	00D4	Ocircumflex	Ô	2. Latin-1 Supplement
[txu]_002	00D5	Otilde	Õ	2. Latin-1 Supplement
[txu]_002	00E0	agrave	à	2. Latin-1 Supplement
[txu]_002	00E3	atilde	ã	2. Latin-1 Supplement
[txu]_002	00EA	ecircumflex	ê	2. Latin-1 Supplement
[txu]_002	00F4	ocircumflex	ô	2. Latin-1 Supplement
[txu]_002	00F5	otilde	õ	2. Latin-1 Supplement
[txu]_002	0128	Itilde	Ï	3. Latin Ext.-A
[txu]_002	0129	itilde	ï	3. Latin Ext.-A
[txu]_002	0168	Utilde	Û	3. Latin Ext.-A
[txu]_002	0169	utilde	ü	3. Latin Ext.-A
[txu]_002	1EBC	Etilde	Ë	8. Latin Ext. Add.
[txu]_002	1EBD	etilde	ë	8. Latin Ext. Add.
[txu]_002	1EF2	Ygrave	ÿ	8. Latin Ext. Add.
[txu]_002	1EF3	ygrave	ÿ	8. Latin Ext. Add.
[txu]_002	1EF8	Ytilde	ÿ	8. Latin Ext. Add.
[txu]_002	1EF9	ytilde	ÿ	8. Latin Ext. Add.
[txu]_003	0027	quotesingle	'	1. Basic Latin
[txu]_003	00C0	Agrave	À	2. Latin-1 Supplement
[txu]_003	00C3	Atilde	Ã	2. Latin-1 Supplement
[txu]_003	00CA	Ecircumflex	Ê	2. Latin-1 Supplement
[txu]_003	00D4	Ocircumflex	Ô	2. Latin-1 Supplement
[txu]_003	00D5	Otilde	Õ	2. Latin-1 Supplement
[txu]_003	00E0	agrave	à	2. Latin-1 Supplement

arquivo	Unicode	Glyphname	Character	bloco
[txu]_003	00E3	atilde	ã	2. Latin-1 Supplement
[txu]_003	00EA	ecircumflex	ê	2. Latin-1 Supplement
[txu]_003	00F4	ocircumflex	ô	2. Latin-1 Supplement
[txu]_003	00F5	otilde	õ	2. Latin-1 Supplement
[txu]_003	0128	ltilde	ĩ	3. Latin Ext.-A
[txu]_003	0129	itilde	ı̃	3. Latin Ext.-A
[txu]_003	0168	Utilde	Û	3. Latin Ext.-A
[txu]_003	0169	utilde	ü	3. Latin Ext.-A
[txu]_003	1EBC	Etilde	Ě	8. Latin Ext. Add.
[txu]_003	1EBD	etilde	ě	8. Latin Ext. Add.
[txu]_003	1EF2	Ygrave	Ỳ	8. Latin Ext. Add.
[txu]_003	1EF3	ygrave	ỳ	8. Latin Ext. Add.
[txu]_003	1EF8	Ytilde	Ỳ̃	8. Latin Ext. Add.
[txu]_003	1EF9	ytilde	ỳ̃	8. Latin Ext. Add.
[urb]_001	00C1	Aacute	Á	2. Latin-1 Supplement
[urb]_001	00C3	Atilde	Ã	2. Latin-1 Supplement
[urb]_001	00D5	Otilde	Õ	2. Latin-1 Supplement
[urb]_001	00E1	aacute	á	2. Latin-1 Supplement
[urb]_001	00E3	atilde	ã	2. Latin-1 Supplement
[urb]_001	00F5	otilde	õ	2. Latin-1 Supplement
[urb]_001	0128	ltilde	ĩ	3. Latin Ext.-A
[urb]_001	0129	itilde	ı̃	3. Latin Ext.-A
[urb]_001	0168	Utilde	Û	3. Latin Ext.-A
[urb]_001	0169	utilde	ü	3. Latin Ext.-A
[urb]_001	1EBC	Etilde	Ě	8. Latin Ext. Add.
[urb]_001	1EBD	etilde	ě	8. Latin Ext. Add.
[urb]_001	A78B	Saltillo	ı̇	9. Latin Ext.-D
[urb]_001	A78C	saltillo	ı̇	9. Latin Ext.-D
[urb]_002	0027	quotesingle	'	1. Basic Latin
[urb]_002	00C1	Aacute	Á	2. Latin-1 Supplement
[urb]_002	00C3	Atilde	Ã	2. Latin-1 Supplement
[urb]_002	00C9	Eacute	É	2. Latin-1 Supplement
[urb]_002	00D3	Oacute	Ó	2. Latin-1 Supplement
[urb]_002	00D5	Otilde	Õ	2. Latin-1 Supplement
[urb]_002	00DA	Uacute	Ú	2. Latin-1 Supplement
[urb]_002	00DD	Yacute	Ý	2. Latin-1 Supplement
[urb]_002	00E1	aacute	á	2. Latin-1 Supplement
[urb]_002	00E3	atilde	ã	2. Latin-1 Supplement
[urb]_002	00E9	eacute	é	2. Latin-1 Supplement
[urb]_002	00F3	oacute	ó	2. Latin-1 Supplement
[urb]_002	00F5	otilde	õ	2. Latin-1 Supplement
[urb]_002	00FA	uacute	ú	2. Latin-1 Supplement
[urb]_002	00FD	yacute	ý	2. Latin-1 Supplement
[urb]_002	0128	ltilde	ĩ	3. Latin Ext.-A
[urb]_002	0129	itilde	ı̃	3. Latin Ext.-A
[urb]_002	0168	Utilde	Û	3. Latin Ext.-A

arquivo	Unicode	Glyphname	Character	bloco
[urb]_002	0169	utilde	ũ	3. Latin Ext.-A
[urb]_002	1EBC	Etilde	Ě	8. Latin Ext. Add.
[urb]_002	1EBD	etilde	ě	8. Latin Ext. Add.
[urb]_002	1EF8	Ytilde	Ÿ	8. Latin Ext. Add.
[urb]_002	1EF9	ytilde	ÿ	8. Latin Ext. Add.
[urz]_001	0027	quotesingle	'	1. Basic Latin
[urz]_001	00C3	Atilde	Ã	2. Latin-1 Supplement
[urz]_001	00D5	Otilde	Õ	2. Latin-1 Supplement
[urz]_001	00E3	atilde	ã	2. Latin-1 Supplement
[urz]_001	00F5	otilde	õ	2. Latin-1 Supplement
[urz]_001	0128	ltilde	Ĺ	3. Latin Ext.-A
[urz]_001	0129	itilde	ĺ	3. Latin Ext.-A
[urz]_001	0168	Utilde	Ŭ	3. Latin Ext.-A
[urz]_001	0169	utilde	ũ	3. Latin Ext.-A
[urz]_001	1EBC	Etilde	Ě	8. Latin Ext. Add.
[urz]_001	1EBD	etilde	ě	8. Latin Ext. Add.
[urz]_001	1EF8	Ytilde	Ÿ	8. Latin Ext. Add.
[urz]_001	1EF9	ytilde	ÿ	8. Latin Ext. Add.
[wau]_001	00C1	Aacute	Á	2. Latin-1 Supplement
[wau]_001	00C3	Atilde	Ã	2. Latin-1 Supplement
[wau]_001	00C9	Eacute	É	2. Latin-1 Supplement
[wau]_001	00CD	Iacute	Í	2. Latin-1 Supplement
[wau]_001	00DA	Uacute	Ú	2. Latin-1 Supplement
[wau]_001	00E1	aacute	á	2. Latin-1 Supplement
[wau]_001	00E3	atilde	ã	2. Latin-1 Supplement
[wau]_001	00E9	eacute	é	2. Latin-1 Supplement
[wau]_001	00ED	iacute	í	2. Latin-1 Supplement
[wau]_001	00FA	uacute	ú	2. Latin-1 Supplement
[wau]_001	0128	ltilde	Ĺ	3. Latin Ext.-A
[wau]_001	0129	itilde	ĺ	3. Latin Ext.-A
[wau]_001	0168	Utilde	Ŭ	3. Latin Ext.-A
[wau]_001	0169	utilde	ũ	3. Latin Ext.-A
[wau]_001	1EBC	Etilde	Ě	8. Latin Ext. Add.
[wau]_001	1EBD	etilde	ě	8. Latin Ext. Add.
[wau]_001	004F_0331	O_macronbelowcomb	Ō	10. non-compliant
[wau]_001	006F_0331	o_macronbelowcomb	ō	10. non-compliant
[wau]_001	00D3_0331	Oacute_macronbelowcomb	Ŏ	10. non-compliant
[wau]_001	00F3_0331	oacute_macronbelowcomb	ŏ	10. non-compliant
[wau]_001	00D5_0331	Otilde_macronbelowcomb	Ȫ	10. non-compliant
[wau]_001	00F5_0331	otilde_macronbelowcomb	ȫ	10. non-compliant
[wau]_002	00C1	Aacute	Á	2. Latin-1 Supplement
[wau]_002	00C3	Atilde	Ã	2. Latin-1 Supplement
[wau]_002	00C9	Eacute	É	2. Latin-1 Supplement
[wau]_002	00CD	Iacute	Í	2. Latin-1 Supplement
[wau]_002	00D3	Oacute	Ó	2. Latin-1 Supplement
[wau]_002	00D5	Otilde	Õ	2. Latin-1 Supplement
[wau]_002	00DA	Uacute	Ú	2. Latin-1 Supplement

arquivo	Unicode	Glyphname	Character	bloco
[wau]_002	00E1	aacute	á	2. Latin-1 Supplement
[wau]_002	00E3	atilde	ã	2. Latin-1 Supplement
[wau]_002	00E9	eacute	é	2. Latin-1 Supplement
[wau]_002	00ED	iacute	í	2. Latin-1 Supplement
[wau]_002	00F3	oacute	ó	2. Latin-1 Supplement
[wau]_002	00F5	otilde	õ	2. Latin-1 Supplement
[wau]_002	00FA	uacute	ú	2. Latin-1 Supplement
[wau]_002	0128	ltilde	ĩ	3. Latin Ext.-A
[wau]_002	0129	itilde	ı̃	3. Latin Ext.-A
[wau]_002	0168	Utilde	Ū	3. Latin Ext.-A
[wau]_002	0169	utilde	ū	3. Latin Ext.-A
[wau]_002	0331	macronbelowcomb	_	7. Combining Marks
[wau]_002	1EBC	Etilde	Ě	8. Latin Ext. Add.
[wau]_002	1EBD	etilde	ě	8. Latin Ext. Add.
[wau]_002	004F_0331	O_macronbelowcomb	Ō	10. non-compliant
[wau]_002	006F_0331	o_macronbelowcomb	o	10. non-compliant
[wau]_002	00D5_0331	Otilde_macronbelowcomb	Ŏ	10. non-compliant
[wau]_002	00F5_0331	otilde_macronbelowcomb	õ	10. non-compliant
[wau]_002	00D3_0331	Oacute_macronbelowcomb	Ō	10. non-compliant
[wau]_002	00F3_0331	oacute_macronbelowcomb	ó	10. non-compliant
[wca]_001	00C3	Atilde	Ã	2. Latin-1 Supplement
[wca]_001	00CB	Edieresis	Ë	2. Latin-1 Supplement
[wca]_001	00D5	Otilde	Õ	2. Latin-1 Supplement
[wca]_001	00E3	atilde	ã	2. Latin-1 Supplement
[wca]_001	00EB	edieresis	ë	2. Latin-1 Supplement
[wca]_001	00F5	otilde	õ	2. Latin-1 Supplement
[wca]_001	0197	Istroke	†	4. Latin Ext.-B
[wca]_001	0268	istroke	‡	5. IPA Extensions
[wca]_001	00CB_0303	Edieresis_tildecomb	Ë	10. non-compliant
[wca]_001	00EB_0303	edieresis_tildecomb	ë	10. non-compliant
[wmd]_004	0027	quotesingle	'	1. Basic Latin
[wmd]_004	00C3	Atilde	Ã	2. Latin-1 Supplement
[wmd]_004	00D5	Otilde	Õ	2. Latin-1 Supplement
[wmd]_004	00E3	atilde	ã	2. Latin-1 Supplement
[wmd]_004	00F5	otilde	õ	2. Latin-1 Supplement
[wmd]_004	0128	ltilde	ĩ	3. Latin Ext.-A
[wmd]_004	0129	itilde	ı̃	3. Latin Ext.-A
[wmd]_004	0168	Utilde	Ū	3. Latin Ext.-A
[wmd]_004	0169	utilde	ū	3. Latin Ext.-A
[wmd]_004	0331	macronbelowcomb	_	7. Combining Marks
[wmd]_004	1EBC	Etilde	Ě	8. Latin Ext. Add.
[wmd]_004	1EBD	etilde	ě	8. Latin Ext. Add.
[wmd]_004	0041_0331	A_macronbelowcomb	Ā	10. non-compliant
[wmd]_004	0061_0331	a_macronbelowcomb	a	10. non-compliant
[wmd]_004	0045_0331	E_macronbelowcomb	Ē	10. non-compliant
[wmd]_004	0065_0331	e_macronbelowcomb	e	10. non-compliant
[wmd]_004	0049_0331	I_macronbelowcomb	Ī	10. non-compliant

arquivo	Unicode	Glyphname	Character	bloco
[wmd]_004	0069_0331	i_macronbelowcomb	ï	10. non-compliant
[wmd]_004	004F_0331	O_macronbelowcomb	Ō	10. non-compliant
[wmd]_004	006F_0331	o_macronbelowcomb	ō	10. non-compliant
[wmd]_004	0055_0331	U_macronbelowcomb	Ū	10. non-compliant
[wmd]_004	0075_0331	u_macronbelowcomb	ū	10. non-compliant
[wmd]_004	00C3_0331	Atilde_macronbelowcomb	Ã	10. non-compliant
[wmd]_004	00E3_0331	atilde_macronbelowcomb	ã	10. non-compliant
[wmd]_004	0128_0331	ltilde_macronbelowcomb	Ĭ	10. non-compliant
[wmd]_004	0129_0331	itilde_macronbelowcomb	ĩ	10. non-compliant
[wmd]_004	0168_0331	Utilde_macronbelowcomb	Ū	10. non-compliant
[wmd]_004	0169_0331	utilde_macronbelowcomb	ū	10. non-compliant
[xav]_001	00C3	Atilde	Ã	2. Latin-1 Supplement
[xav]_001	00C9	Eacute	É	2. Latin-1 Supplement
[xav]_001	00CA	Ecircumflex	Ê	2. Latin-1 Supplement
[xav]_001	00D3	Oacute	Ó	2. Latin-1 Supplement
[xav]_001	00D4	Ocircumflex	Ô	2. Latin-1 Supplement
[xav]_001	00D5	Otilde	Õ	2. Latin-1 Supplement
[xav]_001	00D6	Odieresis	Ö	2. Latin-1 Supplement
[xav]_001	00E3	atilde	ã	2. Latin-1 Supplement
[xav]_001	00E9	eacute	é	2. Latin-1 Supplement
[xav]_001	00EA	ecircumflex	ê	2. Latin-1 Supplement
[xav]_001	00F3	oacute	ó	2. Latin-1 Supplement
[xav]_001	00F4	ocircumflex	ô	2. Latin-1 Supplement
[xav]_001	00F5	otilde	õ	2. Latin-1 Supplement
[xav]_001	00F6	odieresis	ö	2. Latin-1 Supplement
[xav]_001	0128	ltilde	Ĭ	3. Latin Ext.-A
[xav]_001	0129	itilde	ĩ	3. Latin Ext.-A
[xav]_001	02BC	apostrophemod	'	6. Spc. Mod. Letters
[xav]_001	1EBC	Etilde	Ě	8. Latin Ext. Add.
[xav]_001	1EBD	etilde	ě	8. Latin Ext. Add.
[xav]_001	A78B	Saltillo	ı	9. Latin Ext.-D
[xav]_001	A78C	salttillo	ı	9. Latin Ext.-D
[xav]_002	00C0	Agrave	À	2. Latin-1 Supplement
[xav]_002	00C2	Acircumflex	Â	2. Latin-1 Supplement
[xav]_002	00C3	Atilde	Ã	2. Latin-1 Supplement
[xav]_002	00C9	Eacute	É	2. Latin-1 Supplement
[xav]_002	00CC	Igrave	Ì	2. Latin-1 Supplement
[xav]_002	00D4	Ocircumflex	Ô	2. Latin-1 Supplement
[xav]_002	00D5	Otilde	Õ	2. Latin-1 Supplement
[xav]_002	00D9	Ugrave	Ù	2. Latin-1 Supplement
[xav]_002	00E0	agrave	à	2. Latin-1 Supplement
[xav]_002	00E2	acircumflex	â	2. Latin-1 Supplement
[xav]_002	00E3	atilde	ã	2. Latin-1 Supplement
[xav]_002	00E9	eacute	é	2. Latin-1 Supplement
[xav]_002	00EC	igrave	ì	2. Latin-1 Supplement
[xav]_002	00F4	ocircumflex	ô	2. Latin-1 Supplement

arquivo	Unicode	Glyphname	Character	bloco
[xav]_002	00F5	otilde	õ	2. Latin-1 Supplement
[xav]_002	00F9	ugrave	ù	2. Latin-1 Supplement
[xav]_002	0128	ltilde	ĩ	3. Latin Ext.-A
[xav]_002	0129	itilde	ı̃	3. Latin Ext.-A
[xav]_002	02BC	apostrophemod	'	6. Spc. Mod. Letters
[xav]_002	1EBC	Etilde	Ě	8. Latin Ext. Add.
[xav]_002	1EBD	etilde	ě	8. Latin Ext. Add.
[xav]_002	A78B	Saltillo	ı	9. Latin Ext.-D
[xav]_002	A78C	salttillo	ı̇	9. Latin Ext.-D
[xav]_003	00C3	Atilde	Ã	2. Latin-1 Supplement
[xav]_003	00C9	Eacute	É	2. Latin-1 Supplement
[xav]_003	00D4	Ocircumflex	Ô	2. Latin-1 Supplement
[xav]_003	00D5	Otilde	Õ	2. Latin-1 Supplement
[xav]_003	00D6	Odieresis	Ö	2. Latin-1 Supplement
[xav]_003	00E3	atilde	ã	2. Latin-1 Supplement
[xav]_003	00E9	eacute	é	2. Latin-1 Supplement
[xav]_003	00F4	ocircumflex	ô	2. Latin-1 Supplement
[xav]_003	00F5	otilde	õ	2. Latin-1 Supplement
[xav]_003	00F6	odieresis	ö	2. Latin-1 Supplement
[xav]_003	0128	ltilde	ĩ	3. Latin Ext.-A
[xav]_003	0129	itilde	ı̃	3. Latin Ext.-A
[xav]_003	02BC	apostrophemod	'	6. Spc. Mod. Letters
[xav]_003	1EBC	Etilde	Ě	8. Latin Ext. Add.
[xav]_003	1EBD	etilde	ě	8. Latin Ext. Add.
[xav]_003	A78B	Saltillo	ı	9. Latin Ext.-D
[xav]_003	A78C	salttillo	ı̇	9. Latin Ext.-D
[xav]_004	0027	quotesingle	'	1. Basic Latin
[xav]_004	00C3	Atilde	Ã	2. Latin-1 Supplement
[xav]_004	00C9	Eacute	É	2. Latin-1 Supplement
[xav]_004	00D4	Ocircumflex	Ô	2. Latin-1 Supplement
[xav]_004	00D5	Otilde	Õ	2. Latin-1 Supplement
[xav]_004	00D6	Odieresis	Ö	2. Latin-1 Supplement
[xav]_004	00E3	atilde	ã	2. Latin-1 Supplement
[xav]_004	00E9	eacute	é	2. Latin-1 Supplement
[xav]_004	00F4	ocircumflex	ô	2. Latin-1 Supplement
[xav]_004	00F5	otilde	õ	2. Latin-1 Supplement
[xav]_004	00F6	odieresis	ö	2. Latin-1 Supplement
[xav]_004	0128	ltilde	ĩ	3. Latin Ext.-A
[xav]_004	0129	itilde	ı̃	3. Latin Ext.-A
[xav]_004	1EBC	Etilde	Ě	8. Latin Ext. Add.
[xav]_004	1EBD	etilde	ě	8. Latin Ext. Add.
[xav]_004	2019	quoteright	'	11. General Punctuation
[xav]_005	0027	quotesingle	'	1. Basic Latin
[xav]_005	00C0	Agrave	À	2. Latin-1 Supplement
[xav]_005	00C2	Acircumflex	Â	2. Latin-1 Supplement
[xav]_005	00C3	Atilde	Ã	2. Latin-1 Supplement

arquivo	Unicode	Glyphname	Character	bloco
[xav]_005	00C9	Eacute	É	2. Latin-1 Supplement
[xav]_005	00D4	Ocircumflex	Ô	2. Latin-1 Supplement
[xav]_005	00D5	Otilde	Õ	2. Latin-1 Supplement
[xav]_005	00D9	Ugrave	Ù	2. Latin-1 Supplement
[xav]_005	00E0	agrave	à	2. Latin-1 Supplement
[xav]_005	00E2	acircumflex	â	2. Latin-1 Supplement
[xav]_005	00E3	atilde	ã	2. Latin-1 Supplement
[xav]_005	00E9	eacute	é	2. Latin-1 Supplement
[xav]_005	00F4	ocircumflex	ô	2. Latin-1 Supplement
[xav]_005	00F5	otilde	õ	2. Latin-1 Supplement
[xav]_005	00F9	ugrave	ù	2. Latin-1 Supplement
[xav]_005	0128	ltilde	ĩ	3. Latin Ext.-A
[xav]_005	0129	itilde	ı̃	3. Latin Ext.-A
[xav]_005	1EBC	Etilde	Ě	8. Latin Ext. Add.
[xav]_005	1EBD	etilde	ě	8. Latin Ext. Add.
[xer]_001	00C2	Acircumflex	Â	2. Latin-1 Supplement
[xer]_001	00C3	Atilde	Ã	2. Latin-1 Supplement
[xer]_001	00CA	Ecircumflex	Ê	2. Latin-1 Supplement
[xer]_001	00D4	Ocircumflex	Ô	2. Latin-1 Supplement
[xer]_001	00D5	Otilde	Õ	2. Latin-1 Supplement
[xer]_001	00DB	Ucircumflex	Û	2. Latin-1 Supplement
[xer]_001	00E2	acircumflex	â	2. Latin-1 Supplement
[xer]_001	00E3	atilde	ã	2. Latin-1 Supplement
[xer]_001	00EA	ecircumflex	ê	2. Latin-1 Supplement
[xer]_001	00F4	ocircumflex	ô	2. Latin-1 Supplement
[xer]_001	00F5	otilde	õ	2. Latin-1 Supplement
[xer]_001	00FB	ucircumflex	û	2. Latin-1 Supplement
[xer]_001	0128	ltilde	ĩ	3. Latin Ext.-A
[xer]_001	0129	itilde	ı̃	3. Latin Ext.-A
[xer]_001	1EBC	Etilde	Ě	8. Latin Ext. Add.
[xer]_001	1EBD	etilde	ě	8. Latin Ext. Add.
[xok]_001	00C1	Aacute	Á	2. Latin-1 Supplement
[xok]_001	00C3	Atilde	Ã	2. Latin-1 Supplement
[xok]_001	00C9	Eacute	É	2. Latin-1 Supplement
[xok]_001	00D3	Oacute	Ó	2. Latin-1 Supplement
[xok]_001	00D5	Otilde	Õ	2. Latin-1 Supplement
[xok]_001	00E1	aacute	á	2. Latin-1 Supplement
[xok]_001	00E3	atilde	ã	2. Latin-1 Supplement
[xok]_001	00E9	eacute	é	2. Latin-1 Supplement
[xok]_001	00F3	oacute	ó	2. Latin-1 Supplement
[xok]_001	00F5	otilde	õ	2. Latin-1 Supplement
[xok]_001	0168	Utilde	Ũ	3. Latin Ext.-A
[xok]_001	0169	utilde	ũ	3. Latin Ext.-A
[xok]_001	1EBC	Etilde	Ě	8. Latin Ext. Add.
[xok]_001	1EBD	etilde	ě	8. Latin Ext. Add.
[xok]_001	1EF8	Ytilde	Ÿ	8. Latin Ext. Add.
[xok]_001	1EF9	ytilde	ÿ	8. Latin Ext. Add.

arquivo	Unicode	Glyphname	Character	bloco
[xok]_002	0027	quotesingle	'	1. Basic Latin
[xok]_002	00C1	Aacute	Á	2. Latin-1 Supplement
[xok]_002	00C3	Atilde	Ã	2. Latin-1 Supplement
[xok]_002	00C9	Eacute	É	2. Latin-1 Supplement
[xok]_002	00D3	Oacute	Ó	2. Latin-1 Supplement
[xok]_002	00D5	Otilde	Õ	2. Latin-1 Supplement
[xok]_002	00E1	aacute	á	2. Latin-1 Supplement
[xok]_002	00E3	atilde	ã	2. Latin-1 Supplement
[xok]_002	00E9	eacute	é	2. Latin-1 Supplement
[xok]_002	00F3	oacute	ó	2. Latin-1 Supplement
[xok]_002	00F5	otilde	õ	2. Latin-1 Supplement
[xok]_002	0128	ltilde	İ	3. Latin Ext.-A
[xok]_002	0129	itilde	ı	3. Latin Ext.-A
[xok]_002	0168	Utilde	Ū	3. Latin Ext.-A
[xok]_002	0169	utilde	ū	3. Latin Ext.-A
[xok]_002	1EBC	Etilde	Ě	8. Latin Ext. Add.
[xok]_002	1EBD	etilde	ě	8. Latin Ext. Add.
[xok]_002	1EF8	Ytilde	Ÿ	8. Latin Ext. Add.
[xok]_002	1EF9	ytilde	ÿ	8. Latin Ext. Add.
[xok]_002	2019	quoteright	'	11. General Punctuation
[xok]_003	00C1	Aacute	Á	2. Latin-1 Supplement
[xok]_003	00C3	Atilde	Ã	2. Latin-1 Supplement
[xok]_003	00C9	Eacute	É	2. Latin-1 Supplement
[xok]_003	00D3	Oacute	Ó	2. Latin-1 Supplement
[xok]_003	00D5	Otilde	Õ	2. Latin-1 Supplement
[xok]_003	00E1	aacute	á	2. Latin-1 Supplement
[xok]_003	00E3	atilde	ã	2. Latin-1 Supplement
[xok]_003	00E9	eacute	é	2. Latin-1 Supplement
[xok]_003	00F3	oacute	ó	2. Latin-1 Supplement
[xok]_003	00F5	otilde	õ	2. Latin-1 Supplement
[xok]_003	0168	Utilde	Ū	3. Latin Ext.-A
[xok]_003	0169	utilde	ū	3. Latin Ext.-A
[xok]_003	1EBC	Etilde	Ě	8. Latin Ext. Add.
[xok]_003	1EBD	etilde	ě	8. Latin Ext. Add.
[xor]_001	00CB	Edieresis	Ë	2. Latin-1 Supplement
[xor]_001	00EB	edieresis	ë	2. Latin-1 Supplement
[xra]_001	00C0	Agrave	À	2. Latin-1 Supplement
[xra]_001	00C3	Atilde	Ã	2. Latin-1 Supplement
[xra]_001	00CA	Ecircumflex	Ê	2. Latin-1 Supplement
[xra]_001	00D4	Ocircumflex	Ô	2. Latin-1 Supplement
[xra]_001	00D5	Otilde	Õ	2. Latin-1 Supplement
[xra]_001	00E0	agrave	à	2. Latin-1 Supplement
[xra]_001	00E3	atilde	ã	2. Latin-1 Supplement
[xra]_001	00EA	ecircumflex	ê	2. Latin-1 Supplement
[xra]_001	00F4	ocircumflex	ô	2. Latin-1 Supplement
[xra]_001	00F5	otilde	õ	2. Latin-1 Supplement
[xra]_001	0128	ltilde	İ	3. Latin Ext.-A

arquivo	Unicode	Glyphname	Character	bloco
[xra]_001	0129	itilde	ĩ	3. Latin Ext.-A
[xra]_001	0168	Utilde	Ū	3. Latin Ext.-A
[xra]_001	0169	utilde	ū	3. Latin Ext.-A
[xra]_001	1EBC	Etilde	Ě	8. Latin Ext. Add.
[xra]_001	1EBD	etilde	ě	8. Latin Ext. Add.
[xra]_001	1EF2	Ygrave	Ỳ	8. Latin Ext. Add.
[xra]_001	1EF3	ygrave	ỳ	8. Latin Ext. Add.
[xra]_001	1EF8	Ytilde	Ỳ	8. Latin Ext. Add.
[xra]_001	1EF9	ytilde	ỳ	8. Latin Ext. Add.
[xra]_002	00C0	Agrave	À	2. Latin-1 Supplement
[xra]_002	00C1	Aacute	Á	2. Latin-1 Supplement
[xra]_002	00C3	Atilde	Ã	2. Latin-1 Supplement
[xra]_002	00CA	Ecircumflex	Ê	2. Latin-1 Supplement
[xra]_002	00D4	Ocircumflex	Ô	2. Latin-1 Supplement
[xra]_002	00D5	Otilde	Õ	2. Latin-1 Supplement
[xra]_002	00E0	agrave	à	2. Latin-1 Supplement
[xra]_002	00E1	aacute	á	2. Latin-1 Supplement
[xra]_002	00E3	atilde	ã	2. Latin-1 Supplement
[xra]_002	00EA	ecircumflex	ê	2. Latin-1 Supplement
[xra]_002	00F4	ocircumflex	ô	2. Latin-1 Supplement
[xra]_002	00F5	otilde	õ	2. Latin-1 Supplement
[xra]_002	0128	Itilde	Ĩ	3. Latin Ext.-A
[xra]_002	0129	itilde	ĩ	3. Latin Ext.-A
[xra]_002	0168	Utilde	Ū	3. Latin Ext.-A
[xra]_002	0169	utilde	ū	3. Latin Ext.-A
[xra]_002	1EBC	Etilde	Ě	8. Latin Ext. Add.
[xra]_002	1EBD	etilde	ě	8. Latin Ext. Add.
[xra]_002	1EF2	Ygrave	Ỳ	8. Latin Ext. Add.
[xra]_002	1EF3	ygrave	ỳ	8. Latin Ext. Add.
[xra]_002	1EF8	Ytilde	Ỳ	8. Latin Ext. Add.
[xra]_002	1EF9	ytilde	ỳ	8. Latin Ext. Add.
[xra]_003	00C0	Agrave	À	2. Latin-1 Supplement
[xra]_003	00C3	Atilde	Ã	2. Latin-1 Supplement
[xra]_003	00CA	Ecircumflex	Ê	2. Latin-1 Supplement
[xra]_003	00D4	Ocircumflex	Ô	2. Latin-1 Supplement
[xra]_003	00D5	Otilde	Õ	2. Latin-1 Supplement
[xra]_003	00E0	agrave	à	2. Latin-1 Supplement
[xra]_003	00E3	atilde	ã	2. Latin-1 Supplement
[xra]_003	00EA	ecircumflex	ê	2. Latin-1 Supplement
[xra]_003	00F4	ocircumflex	ô	2. Latin-1 Supplement
[xra]_003	00F5	otilde	õ	2. Latin-1 Supplement
[xra]_003	0128	Itilde	Ĩ	3. Latin Ext.-A
[xra]_003	0129	itilde	ĩ	3. Latin Ext.-A
[xra]_003	0168	Utilde	Ū	3. Latin Ext.-A
[xra]_003	0169	utilde	ū	3. Latin Ext.-A
[xra]_003	1EBC	Etilde	Ě	8. Latin Ext. Add.
[xra]_003	1EBD	etilde	ě	8. Latin Ext. Add.

arquivo	Unicode	Glyphname	Character	bloco
[xra]_003	1EF2	Ygrave	Ỳ	8. Latin Ext. Add.
[xra]_003	1EF3	ygrave	ỳ	8. Latin Ext. Add.
[xra]_003	1EF8	Ytilde	Ỳ	8. Latin Ext. Add.
[xra]_003	1EF9	ytilde	ỳ	8. Latin Ext. Add.
[xra]_004	00C0	Agrave	À	2. Latin-1 Supplement
[xra]_004	00C3	Atilde	Ã	2. Latin-1 Supplement
[xra]_004	00CA	Ecircumflex	Ê	2. Latin-1 Supplement
[xra]_004	00D4	Ocircumflex	Ô	2. Latin-1 Supplement
[xra]_004	00D5	Otilde	Õ	2. Latin-1 Supplement
[xra]_004	00E0	agrave	à	2. Latin-1 Supplement
[xra]_004	00E3	atilde	ã	2. Latin-1 Supplement
[xra]_004	00EA	ecircumflex	ê	2. Latin-1 Supplement
[xra]_004	00F4	ocircumflex	ô	2. Latin-1 Supplement
[xra]_004	00F5	otilde	õ	2. Latin-1 Supplement
[xra]_004	0128	ltilde	Ĺ	3. Latin Ext.-A
[xra]_004	0129	itilde	ĩ	3. Latin Ext.-A
[xra]_004	0168	Utilde	Ū	3. Latin Ext.-A
[xra]_004	0169	utilde	ū	3. Latin Ext.-A
[xra]_004	1EBC	Etilde	Ē	8. Latin Ext. Add.
[xra]_004	1EBD	etilde	ē	8. Latin Ext. Add.
[xra]_004	1EF2	Ygrave	Ỳ	8. Latin Ext. Add.
[xra]_004	1EF3	ygrave	ỳ	8. Latin Ext. Add.
[xra]_004	1EF8	Ytilde	Ỳ	8. Latin Ext. Add.
[xra]_004	1EF9	ytilde	ỳ	8. Latin Ext. Add.
[xra]_005	00C0	Agrave	À	2. Latin-1 Supplement
[xra]_005	00C3	Atilde	Ã	2. Latin-1 Supplement
[xra]_005	00CA	Ecircumflex	Ê	2. Latin-1 Supplement
[xra]_005	00D4	Ocircumflex	Ô	2. Latin-1 Supplement
[xra]_005	00D5	Otilde	Õ	2. Latin-1 Supplement
[xra]_005	00E0	agrave	à	2. Latin-1 Supplement
[xra]_005	00E3	atilde	ã	2. Latin-1 Supplement
[xra]_005	00EA	ecircumflex	ê	2. Latin-1 Supplement
[xra]_005	00F4	ocircumflex	ô	2. Latin-1 Supplement
[xra]_005	00F5	otilde	õ	2. Latin-1 Supplement
[xra]_005	0128	ltilde	Ĺ	3. Latin Ext.-A
[xra]_005	0129	itilde	ĩ	3. Latin Ext.-A
[xra]_005	0168	Utilde	Ū	3. Latin Ext.-A
[xra]_005	0169	utilde	ū	3. Latin Ext.-A
[xra]_005	1EBC	Etilde	Ē	8. Latin Ext. Add.
[xra]_005	1EBD	etilde	ē	8. Latin Ext. Add.
[xra]_005	1EF2	Ygrave	Ỳ	8. Latin Ext. Add.
[xra]_005	1EF3	ygrave	ỳ	8. Latin Ext. Add.
[xra]_005	1EF8	Ytilde	Ỳ	8. Latin Ext. Add.
[xra]_005	1EF9	ytilde	ỳ	8. Latin Ext. Add.
[xra]_006	00C0	Agrave	À	2. Latin-1 Supplement
[xra]_006	00C3	Atilde	Ã	2. Latin-1 Supplement
[xra]_006	00CA	Ecircumflex	Ê	2. Latin-1 Supplement

arquivo	Unicode	Glyphname	Character	bloco
[xra]_006	00D4	Ocircumflex	Ô	2. Latin-1 Supplement
[xra]_006	00D5	Otilde	Õ	2. Latin-1 Supplement
[xra]_006	00E0	agrave	à	2. Latin-1 Supplement
[xra]_006	00E3	atilde	ã	2. Latin-1 Supplement
[xra]_006	00EA	ecircumflex	ê	2. Latin-1 Supplement
[xra]_006	00F4	ocircumflex	ô	2. Latin-1 Supplement
[xra]_006	00F5	otilde	õ	2. Latin-1 Supplement
[xra]_006	0128	ltilde	Ī	3. Latin Ext.-A
[xra]_006	0129	itilde	ī	3. Latin Ext.-A
[xra]_006	0168	Utilde	Ū	3. Latin Ext.-A
[xra]_006	0169	utilde	ū	3. Latin Ext.-A
[xra]_006	1EBC	Etilde	Ě	8. Latin Ext. Add.
[xra]_006	1EBD	etilde	ě	8. Latin Ext. Add.
[xra]_006	1EF2	Ygrave	ÿ	8. Latin Ext. Add.
[xra]_006	1EF3	ygrave	ÿ	8. Latin Ext. Add.
[xra]_006	1EF8	Ytilde	Ÿ	8. Latin Ext. Add.
[xra]_006	1EF9	ytilde	ÿ	8. Latin Ext. Add.
[xra]_007	00C0	Agrave	À	2. Latin-1 Supplement
[xra]_007	00C3	Atilde	Ã	2. Latin-1 Supplement
[xra]_007	00CA	Ecircumflex	Ê	2. Latin-1 Supplement
[xra]_007	00D4	Ocircumflex	Ô	2. Latin-1 Supplement
[xra]_007	00D5	Otilde	Õ	2. Latin-1 Supplement
[xra]_007	00E0	agrave	à	2. Latin-1 Supplement
[xra]_007	00E3	atilde	ã	2. Latin-1 Supplement
[xra]_007	00EA	ecircumflex	ê	2. Latin-1 Supplement
[xra]_007	00F4	ocircumflex	ô	2. Latin-1 Supplement
[xra]_007	00F5	otilde	õ	2. Latin-1 Supplement
[xra]_007	0128	ltilde	Ī	3. Latin Ext.-A
[xra]_007	0129	itilde	ī	3. Latin Ext.-A
[xra]_007	0168	Utilde	Ū	3. Latin Ext.-A
[xra]_007	0169	utilde	ū	3. Latin Ext.-A
[xra]_007	1EBC	Etilde	Ě	8. Latin Ext. Add.
[xra]_007	1EBD	etilde	ě	8. Latin Ext. Add.
[xra]_007	1EF2	Ygrave	ÿ	8. Latin Ext. Add.
[xra]_007	1EF3	ygrave	ÿ	8. Latin Ext. Add.
[xra]_007	1EF8	Ytilde	Ÿ	8. Latin Ext. Add.
[xra]_007	1EF9	ytilde	ÿ	8. Latin Ext. Add.
[xra]_008	00C0	Agrave	À	2. Latin-1 Supplement
[xra]_008	00C3	Atilde	Ã	2. Latin-1 Supplement
[xra]_008	00CA	Ecircumflex	Ê	2. Latin-1 Supplement
[xra]_008	00D4	Ocircumflex	Ô	2. Latin-1 Supplement
[xra]_008	00D5	Otilde	Õ	2. Latin-1 Supplement
[xra]_008	00E0	agrave	à	2. Latin-1 Supplement
[xra]_008	00E3	atilde	ã	2. Latin-1 Supplement
[xra]_008	00EA	ecircumflex	ê	2. Latin-1 Supplement
[xra]_008	00F4	ocircumflex	ô	2. Latin-1 Supplement
[xra]_008	00F5	otilde	õ	2. Latin-1 Supplement

arquivo	Unicode	Glyphname	Character	bloco
[xra]_008	0128	Itilde	İ	3. Latin Ext.-A
[xra]_008	0129	itilde	ı	3. Latin Ext.-A
[xra]_008	0168	Utilde	Ū	3. Latin Ext.-A
[xra]_008	0169	utilde	ū	3. Latin Ext.-A
[xra]_008	1EBC	Etilde	Ě	8. Latin Ext. Add.
[xra]_008	1EBD	etilde	ě	8. Latin Ext. Add.
[xra]_008	1EF2	Ygrave	Ỳ	8. Latin Ext. Add.
[xra]_008	1EF3	ygrave	ỳ	8. Latin Ext. Add.
[xra]_008	1EF8	Ytilde	Ỳ	8. Latin Ext. Add.
[xra]_008	1EF9	ytilde	ỳ	8. Latin Ext. Add.
[xra]_009	00C0	Agrave	À	2. Latin-1 Supplement
[xra]_009	00C3	Atilde	Ã	2. Latin-1 Supplement
[xra]_009	00CA	Ecircumflex	Ê	2. Latin-1 Supplement
[xra]_009	00D4	Ocircumflex	Ô	2. Latin-1 Supplement
[xra]_009	00D5	Otilde	Õ	2. Latin-1 Supplement
[xra]_009	00E0	agrave	à	2. Latin-1 Supplement
[xra]_009	00E3	atilde	ã	2. Latin-1 Supplement
[xra]_009	00EA	ecircumflex	ê	2. Latin-1 Supplement
[xra]_009	00F4	ocircumflex	ô	2. Latin-1 Supplement
[xra]_009	00F5	otilde	õ	2. Latin-1 Supplement
[xra]_009	0128	Itilde	İ	3. Latin Ext.-A
[xra]_009	0129	itilde	ı	3. Latin Ext.-A
[xra]_009	0168	Utilde	Ū	3. Latin Ext.-A
[xra]_009	0169	utilde	ū	3. Latin Ext.-A
[xra]_009	1EBC	Etilde	Ě	8. Latin Ext. Add.
[xra]_009	1EBD	etilde	ě	8. Latin Ext. Add.
[xra]_009	1EF2	Ygrave	Ỳ	8. Latin Ext. Add.
[xra]_009	1EF3	ygrave	ỳ	8. Latin Ext. Add.
[xra]_009	1EF8	Ytilde	Ỳ	8. Latin Ext. Add.
[xra]_009	1EF9	ytilde	ỳ	8. Latin Ext. Add.
[xra]_010	00C0	Agrave	À	2. Latin-1 Supplement
[xra]_010	00C3	Atilde	Ã	2. Latin-1 Supplement
[xra]_010	00CA	Ecircumflex	Ê	2. Latin-1 Supplement
[xra]_010	00D4	Ocircumflex	Ô	2. Latin-1 Supplement
[xra]_010	00D5	Otilde	Õ	2. Latin-1 Supplement
[xra]_010	00E0	agrave	à	2. Latin-1 Supplement
[xra]_010	00E3	atilde	ã	2. Latin-1 Supplement
[xra]_010	00EA	ecircumflex	ê	2. Latin-1 Supplement
[xra]_010	00F4	ocircumflex	ô	2. Latin-1 Supplement
[xra]_010	00F5	otilde	õ	2. Latin-1 Supplement
[xra]_010	0128	Itilde	İ	3. Latin Ext.-A
[xra]_010	0129	itilde	ı	3. Latin Ext.-A
[xra]_010	0168	Utilde	Ū	3. Latin Ext.-A
[xra]_010	0169	utilde	ū	3. Latin Ext.-A
[xra]_010	1EBC	Etilde	Ě	8. Latin Ext. Add.
[xra]_010	1EBD	etilde	ě	8. Latin Ext. Add.
[xra]_010	1EF2	Ygrave	Ỳ	8. Latin Ext. Add.

arquivo	Unicode	Glyphname	Character	bloco
[xra]_010	1EF3	ygrave	ỳ	8. Latin Ext. Add.
[xra]_010	1EF8	Ytilde	Ỳ	8. Latin Ext. Add.
[xra]_010	1EF9	ytilde	ỳ	8. Latin Ext. Add.
[xra]_012	00C0	Agrave	À	2. Latin-1 Supplement
[xra]_012	00C3	Atilde	Ã	2. Latin-1 Supplement
[xra]_012	00CA	Ecircumflex	Ê	2. Latin-1 Supplement
[xra]_012	00D4	Ocircumflex	Ô	2. Latin-1 Supplement
[xra]_012	00D5	Otilde	Õ	2. Latin-1 Supplement
[xra]_012	00E0	agrave	à	2. Latin-1 Supplement
[xra]_012	00E3	atilde	ã	2. Latin-1 Supplement
[xra]_012	00EA	ecircumflex	ê	2. Latin-1 Supplement
[xra]_012	00F4	ocircumflex	ô	2. Latin-1 Supplement
[xra]_012	00F5	otilde	õ	2. Latin-1 Supplement
[xra]_012	0128	Itilde	İ	3. Latin Ext.-A
[xra]_012	0129	itilde	ı	3. Latin Ext.-A
[xra]_012	0168	Utilde	Ū	3. Latin Ext.-A
[xra]_012	0169	utilde	ū	3. Latin Ext.-A
[xra]_012	1EBC	Etilde	Ě	8. Latin Ext. Add.
[xra]_012	1EBD	etilde	ě	8. Latin Ext. Add.
[xra]_012	1EF2	Ygrave	Ỳ	8. Latin Ext. Add.
[xra]_012	1EF3	ygrave	ỳ	8. Latin Ext. Add.
[xra]_012	1EF8	Ytilde	Ỳ	8. Latin Ext. Add.
[xra]_012	1EF9	ytilde	ỳ	8. Latin Ext. Add.
[xri]_001	0027	quotesingle	'	1. Basic Latin
[xri]_001	00C3	Atilde	Ã	2. Latin-1 Supplement
[xri]_001	00D5	Otilde	Õ	2. Latin-1 Supplement
[xri]_001	00E3	atilde	ã	2. Latin-1 Supplement
[xri]_001	00F5	otilde	õ	2. Latin-1 Supplement
[xri]_001	1EBC	Etilde	Ě	8. Latin Ext. Add.
[xri]_001	1EBD	etilde	ě	8. Latin Ext. Add.
[xri]_001	1EF2	Ygrave	Ỳ	8. Latin Ext. Add.
[xri]_001	1EF3	ygrave	ỳ	8. Latin Ext. Add.
[xri]_001	1EF8	Ytilde	Ỳ	8. Latin Ext. Add.
[xri]_001	1EF9	ytilde	ỳ	8. Latin Ext. Add.
[xri]_001	2019	quoteright	'	11. General Punctuation
[xsu]_001	00C3	Atilde	Ã	2. Latin-1 Supplement
[xsu]_001	00C4	Adieresis	Ä	2. Latin-1 Supplement
[xsu]_001	00D5	Otilde	Õ	2. Latin-1 Supplement
[xsu]_001	00D6	Odieresis	Ö	2. Latin-1 Supplement
[xsu]_001	00E3	atilde	ã	2. Latin-1 Supplement
[xsu]_001	00E4	adieresis	ä	2. Latin-1 Supplement
[xsu]_001	00F5	otilde	õ	2. Latin-1 Supplement
[xsu]_001	00F6	odieresis	ö	2. Latin-1 Supplement
[xsu]_001	0128	Itilde	İ	3. Latin Ext.-A
[xsu]_001	0129	itilde	ı	3. Latin Ext.-A
[xsu]_001	0168	Utilde	Ū	3. Latin Ext.-A
[xsu]_001	0169	utilde	ū	3. Latin Ext.-A

arquivo	Unicode	Glyphname	Character	bloco
[xsu]_001	0303	tildecomb	̃	7. Combining Marks
[xsu]_001	1EBC	Etilde	Ě	8. Latin Ext. Add.
[xsu]_001	1EBD	etilde	ě	8. Latin Ext. Add.
[xsu]_001	00C4_0303	Adieresis_tildecomb	Ä̃	10. non-compliant
[xsu]_001	00E4_0303	adieresis_tildecomb	ä̃	10. non-compliant
[xsu]_001	00D6_0303	Odieresis_tildecomb	Ö̃	10. non-compliant
[xsu]_001	00F6_0303	odieresis_tildecomb	ö̃	10. non-compliant
[xwa]_001	0027	quotesingle	'	1. Basic Latin
[xwa]_001	00C3	Atilde	Ã	2. Latin-1 Supplement
[xwa]_001	00D5	Otilde	Õ	2. Latin-1 Supplement
[xwa]_001	00D6	Odieresis	Ö	2. Latin-1 Supplement
[xwa]_001	00E3	atilde	ã	2. Latin-1 Supplement
[xwa]_001	00F5	otilde	õ	2. Latin-1 Supplement
[xwa]_001	00F6	odieresis	ö	2. Latin-1 Supplement
[xwa]_001	0128	ltilde	ĩ	3. Latin Ext.-A
[xwa]_001	0129	itilde	ı̃	3. Latin Ext.-A
[xwa]_001	0168	Utilde	Û	3. Latin Ext.-A
[xwa]_001	0169	utilde	ü	3. Latin Ext.-A
[xwa]_001	0197	lstroke	ł	4. Latin Ext.-B
[xwa]_001	0268	istroke	ı̇	5. IPA Extensions
[xwa]_001	0303	tildecomb	̃	7. Combining Marks
[xwa]_001	1EBC	Etilde	Ě	8. Latin Ext. Add.
[xwa]_001	1EBD	etilde	ě	8. Latin Ext. Add.
[xwa]_001	0197_0303	lstroke_tildecomb	ł̃	10. non-compliant
[xwa]_001	0268_0303	istroke_tildecomb	ı̇̃	10. non-compliant
[yrl]_003	00C1	Aacute	Á	2. Latin-1 Supplement
[yrl]_003	00C3	Atilde	Ã	2. Latin-1 Supplement
[yrl]_003	00C9	Eacute	É	2. Latin-1 Supplement
[yrl]_003	00CD	Iacute	Í	2. Latin-1 Supplement
[yrl]_003	00DA	Uacute	Ú	2. Latin-1 Supplement
[yrl]_003	00E1	aacute	á	2. Latin-1 Supplement
[yrl]_003	00E3	atilde	ã	2. Latin-1 Supplement
[yrl]_003	00E9	eacute	é	2. Latin-1 Supplement
[yrl]_003	00ED	iacute	í	2. Latin-1 Supplement
[yrl]_003	00FA	uacute	ú	2. Latin-1 Supplement
[yrl]_003	0128	ltilde	ĩ	3. Latin Ext.-A
[yrl]_003	0129	itilde	ı̃	3. Latin Ext.-A
[yrl]_003	0168	Utilde	Û	3. Latin Ext.-A
[yrl]_003	0169	utilde	ü	3. Latin Ext.-A
[yrl]_003	1EBC	Etilde	Ě	8. Latin Ext. Add.
[yrl]_003	1EBD	etilde	ě	8. Latin Ext. Add.
[yrl]_004	00C1	Aacute	Á	2. Latin-1 Supplement
[yrl]_004	00C3	Atilde	Ã	2. Latin-1 Supplement
[yrl]_004	00C9	Eacute	É	2. Latin-1 Supplement
[yrl]_004	00CA	Ecircumflex	Ê	2. Latin-1 Supplement
[yrl]_004	00CD	Iacute	Í	2. Latin-1 Supplement
[yrl]_004	00DA	Uacute	Ú	2. Latin-1 Supplement

arquivo	Unicode	Glyphname	Character	bloco
[yrl]_004	00DD	Yacute	Ý	2. Latin-1 Supplement
[yrl]_004	00E1	aacute	á	2. Latin-1 Supplement
[yrl]_004	00E3	atilde	ã	2. Latin-1 Supplement
[yrl]_004	00E9	eacute	é	2. Latin-1 Supplement
[yrl]_004	00EA	ecircumflex	ê	2. Latin-1 Supplement
[yrl]_004	00ED	iacute	í	2. Latin-1 Supplement
[yrl]_004	00FA	uacute	ú	2. Latin-1 Supplement
[yrl]_004	00FD	yacute	ý	2. Latin-1 Supplement
[yrl]_004	0128	ltilde	Ĺ	3. Latin Ext.-A
[yrl]_004	0129	itilde	ĩ	3. Latin Ext.-A
[yrl]_004	0168	Utilde	Ū	3. Latin Ext.-A
[yrl]_004	0169	utilde	ū	3. Latin Ext.-A
[yrl]_004	1EBC	Etilde	Ě	8. Latin Ext. Add.
[yrl]_004	1EBD	etilde	ě	8. Latin Ext. Add.
[zkp]_001	00C1	Aacute	Á	2. Latin-1 Supplement
[zkp]_001	00C2	Acircumflex	Â	2. Latin-1 Supplement
[zkp]_001	00C3	Atilde	Ã	2. Latin-1 Supplement
[zkp]_001	00C9	Eacute	É	2. Latin-1 Supplement
[zkp]_001	00CA	Ecircumflex	Ê	2. Latin-1 Supplement
[zkp]_001	00D3	Oacute	Ó	2. Latin-1 Supplement
[zkp]_001	00D4	Ocircumflex	Ô	2. Latin-1 Supplement
[zkp]_001	00D5	Otilde	Õ	2. Latin-1 Supplement
[zkp]_001	00E1	aacute	á	2. Latin-1 Supplement
[zkp]_001	00E2	acircumflex	â	2. Latin-1 Supplement
[zkp]_001	00E3	atilde	ã	2. Latin-1 Supplement
[zkp]_001	00E9	eacute	é	2. Latin-1 Supplement
[zkp]_001	00EA	ecircumflex	ê	2. Latin-1 Supplement
[zkp]_001	00F3	oacute	ó	2. Latin-1 Supplement
[zkp]_001	00F4	ocircumflex	ô	2. Latin-1 Supplement
[zkp]_001	00F5	otilde	õ	2. Latin-1 Supplement
[zkp]_001	0168	Utilde	Ū	3. Latin Ext.-A
[zkp]_001	0169	utilde	ū	3. Latin Ext.-A
[zkp]_001	1EBC	Etilde	Ě	8. Latin Ext. Add.
[zkp]_001	1EBD	etilde	ě	8. Latin Ext. Add.

3. E-mails

From: Rafael Dietzsch rafael.dietzsch@gmail.com
Subject: Re: The Evolution of Character Codes, 1874-1968
Date: 7 October 2020 16:31
To: Eric Fischer enf@pobox.com



Dear Eric

How are you?

I have a question, regarding the method for inputting accented letters with ASCII.

Your paper (as well as Haralambous book, Fonts & Encodings), describe that "accented letters were to be transmitted as a sequence of three characters: the letter, backspace, and the accent."

I am curious about the outputs using this methods. Can you point me any printed examples?

My point is: is it possible to render accented letters properly with 7-bit ASCII? Today we can using combining marks and anchors, but I am not sure this methods were available those days.

In this regard, even if the spacing accents (or even non-spacing) were available in ASCII, what was the method to align them properly, on top of the base letters? The only way I see this happening is with monospaced characters. Is that correct?

Looking forward for you opinion on this.

Best
Rafael

From: Eric Fischer enf@pobox.com
Subject: Re: The Evolution of Character Codes, 1874-1968
Date: 8 October 2020 01:43
To: Rafael Dietzsch rafael.dietzsch@gmail.com



Thanks for asking. Yes, the intent was that this would be used with monospaced characters. I am not aware of any current software that supports this kind of overstriking, other than some Unix programs that support underlined and boldface text through backspacing, since that is what nroff expects to work in its output. Unfortunately I can't point to any actual printer output showing the backspace convention in use for accented letters, only the frequent use of backspace for underlining in Multics and early Unix documentation. I will keep an eye open for examples and if I see any will forward them on to you.

Eric

From: Eric Fischer enf@pobox.com
Subject: Re: The Evolution of Character Codes, 1874-1968
Date: 8 October 2020 02:00
To: Rafael Dietzsch rafael.dietzsch@gmail.com



OK, I found an example of overstruck diacritical marks. Still a contrived example, but showing what it looked like from an actual printer.

<https://archive.org/details/compleclearx820dunc/page/22/mode/2up>

From: Yannis Haralambous yannis.haralambous@imt-atlantique.fr
Subject: Question
Date: 7 October 2020 15:51
To: rafael.dietzsch@gmail.com



dear Rafael,

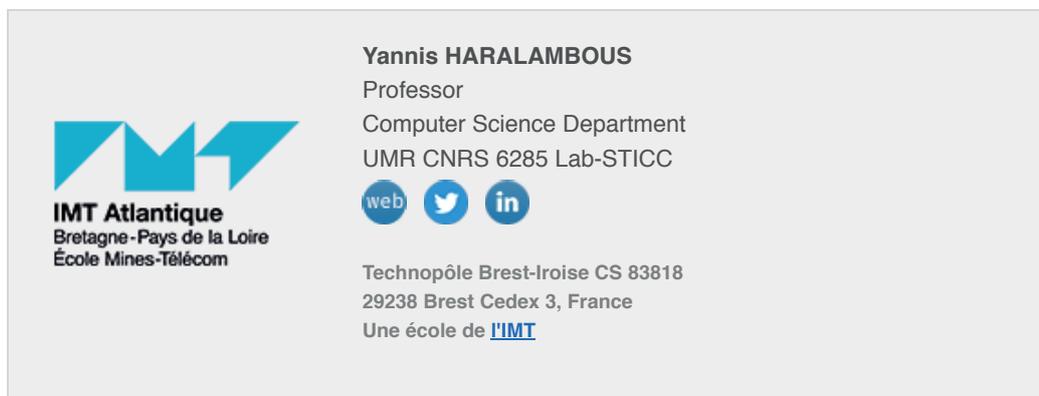
yes Gerry is a very dear friend!

Your topic sounds extremely interesting! A pity you didn't submit a paper for the /grafematik/ 2020 conference last June.

Let me know what I can do for you.

Cheers

Yannis



Yannis HARALAMBOUS
Professor
Computer Science Department
UMR CNRS 6285 Lab-STICC

Technopôle Brest-Iroise CS 83818
29238 Brest Cedex 3, France
Une école de [l'IMT](#)


IMT Atlantique
Bretagne - Pays de la Loire
École Mines-Télécom

From: Rafael Dietzsch rafael.dietzsch@gmail.com
Subject: Re: Question
Date: 7 October 2020 17:01
To: Yannis Haralambous yannis.haralambous@imt-atlantique.fr



Dear Yannis

Thank you for your quick reply.

I am sorry for not submitting a paper for the grafematik conference; I was not aware of it. I will sure do it next time.

There goes my question, regarding the method for inputting accented letters with ASCII.

Your book (as well as Eric Fischer's paper), describe that accented letters were to be transmitted as a sequence of three characters: the letter, backspace, and the accent.

I am curious about the outputs using this methods. Can you point me any printed examples?

My point is: is it possible to render accented letters properly with 7-bit ASCII? Today we can do this using combining marks and anchors, but I am not sure this methods were available those days.

In this regard, even if the spacing accents (or even non-spacing) were available in ASCII, what was the method to align them properly, on top of the base letters? The only way I see this happening is with monospaced characters. Is that correct?

Looking forward for you opinion on this.

Best
Rafael

From: **Yannis Haralambous** yannis.haralambous@imt-atlantique.fr
Subject: Re: Question
Date: 7 October 2020 17:37
To: Rafael Dietzsch rafael.dietzsch@gmail.com



Your book (as well as Eric Fischer's paper), describe that accented letters were to be transmitted as a sequence of three characters: the letter, backspace, and the accent.

In those days one of the main virtues of designing encoding was "how to obtain the most possible outputs with the least possible character slots?"
This why the backslash was invented: with the slash and the backslash you got the or \vee and the and \wedge characters for free: one additional position \backslash gives you two outputs \wedge and \vee (not to mention the parallel sign // :-)

So they said: by adding isolated diacritics and the backspace we can obtain "all" European characters, no need for 8-bit encodings, 7 bits are sufficient.

I am curious about the outputs using this methods. Can you point me any printed examples?

For that you need to find a computer listing from the 60s or 70s, not something easy to find.

My point is: is it possible to render accented letters properly with 7-bit ASCII? Today we can do this using combining marks and anchors, but I am not sure this methods were available those days.

You are reasoning in 21st century terms. At that time "rendering" meant different things depending on the device you were using. An ink-dot printer would indeed render 7-bit ASCII accented letters as badly as the rest. And phototypesetters were using their own codes, not related to ASCII at all. Donald Knuth was the first to build a device-independent file format and in TeX accents are placed through the \backslash accent primitive, which is quite complex (and mostly works fine).

So you are asking the wrong question. Nobody would ever use 7-bit accented ASCII directly on a device where you would see the difference between good positioning of diacritics and bad positioning. For that they had expensive machines, each one with its own binary format. That was ages before PostScript (which was heavily influenced by TeX).

In this regard, even if the spacing accents (or even non-spacing) were available in ASCII, what was the method to align them properly, on top of the base letters?

As I said: on low resolution devices (and I mean REALLY low, 144 dpi was already a great luxury) the question was not asked and monospaced characters were the norm. On high-resolution devices they had special codes.

The only way I see this happening is with monospaced characters. Is that correct?

Looking forward for you opinion on this.

When I first touched a computer, in 1979, I had to punch cards. On the (ridiculously small) keyboard there were no diacritics in the first place and no backspace key (I think), so in those days, even though I was in France, we didn't type accented letters, it was perfectly normal to write French without diacritics. Papers and theses were written on the typewriter (changing balls every few words), and then TeX arrived in 1986 (in my school) and it was the Messiah!

In 1991 I was commissioned by the École polytechnique in Palaiseau to teach TeX to the ladies of the internal printing house. They were using Alphatype phototypesetters with enormous keyboards and many codes to type, TeX was a relief for them. Take a look here, page 76ff:

<https://www.govinfo.gov/content/pkg/GOVPUB-C13-49f6b3c3ddb96d84cc5b060de5ae76/pdf/GOVPUB-C13-49f6b3c3ddb96d84cc5b060de5ae76.pdf>

I'm sorry I don't have any additional information.

Cheers,

Yannis

From: Rafael Dietzsch rafael.dietzsch@gmail.com
Subject: Re: Question
Date: 8 October 2020 22:00
To: Yannis Haralambous yannis.haralambous@imt-atlantique.fr



Dear Yannis

Thank you for your explanations, very helpful. Very generous of you.

I was, indeed, asking the wrong questions.

Regarding the link you send, I am afraid I did not understand what you want me to look at.

I got in touch with Eric Fischer, that shared some ideas, and some samples:

OK, I found an example of overstruck diacritical marks. Still a contrived example, but showing what it looked like from an actual printer.

<https://archive.org/details/completeclearartex820dunc/page/22/mode/2up>

Best
Rafael

From: Rafael Dietzsch rafael.dietzsch@gmail.com
Subject: Sylfaen article
Date: 27 December 2018 11:58
To: John Hudson john@tiro.ca



Hi John

How are you? Hope well.

I was searching for the article you wrote about Sylfaen, but could not find it in your website. Is it possible to have access to this document?

best
Rafa

From: John Hudson john@tiro.ca
Subject: Re: Sylfaen article
Date: 27 December 2018 16:51
To: Rafael Dietzsch rafael.dietzsch@gmail.com



Dear Rafa,

The old Sylfaen article is here:
http://www.tiro.com/Articles/sylfaen_article.pdf

It is strange to re-read this now, and realise the degree to which a lot of what I wrote about then is now common knowledge or practice. The whole WRIT project, of which Sylfaen was a part, came about in the wake of the 'OpenType Jamboree' in 1997, at which the new font format was first presented to a significant group of font developers. It became clear from talking to colleagues at that event that most had very little knowledge or experience beyond Western European characters.

Regards, John

--

John Hudson
Tiro Typeworks Ltd www.tiro.com
Salish Sea, BC tiro@tiro.com

NOTE: In the interests of productivity, I am currently dealing with email on only two days per week, usually Monday and Thursday unless this schedule is disrupted by travel. If you need to contact me urgently, please use some other method of communication. Thank you.

From: Rafael Dietzsch rafael.dietzsch@gmail.com
Subject: Re: Sylfaen article
Date: 10 July 2021 12:30
To: John Hudson john@tiro.ca



Hi John

All good?

As I finally got to the methodology chapter of my thesis, I re-read some of your works. It is nice to see that the way I structured the orthographic cataloging of Brazilian indigenous languages follows your methods (spreadsheets first, design later), and the results are consistent. The data collection was really painful, but in the end, I have information on more than 100 languages.

The cataloging is structured in spreadsheets like the one below:
<https://docs.google.com/spreadsheets/d/1xSS6U0t5T1S9sdXv6Sm9RU-n-VToBieRLSLzeDZ3Ymfw/edit?usp=sharing>

I was also re-reading your last e-mail, about Sylfaen. I was wondering about what happened with WRIT. Can you share more information about this? I have also read in your presentation about African Alphabets, a little statement about WRIT:

The WRIT approach
Research and document language- and region-specific orthographic requirements and glyph preferences.
Pros: Fascinating
Cons: Massively time consuming; virtually inexhaustible; non- generative

Is this tool still used?

After finishing my thesis, I think it would be interesting to invest in a tool similar to WRIT, with a proper user interface, to help linguists. I think there are good chances of finding funding from the Brazilian government.

Best
Rafa

From: John Hudson john@tiro.ca
Subject: Re: Sylfaen article
Date: 11 July 2021 12:10
To: Rafael Dietzsch rafael.dietzsch@gmail.com



Dear Rafael,

As far as I know, WRIT was completely abandoned in the late 1990s when we stopped work on it. It was written to work on the Windows NT platform, so I am not even sure it would run on recent versions of Windows, although the database could still be useful.

I have been approached recently about the possibility of making an open source repository of information on North American indigenous languages, but for various reasons I am hesitant to take it on. As you say: the data collection is painful.

Are you familiar with Rosetta's Hyperglot? I think it is the best of recent efforts to catalogue character requirements for languages, has a good size database, and both a web interface and a version that can run from the command line (the latter is particularly useful for generating a list of language supported by a font).
<https://hyperglot.rosettatype.com/>

It would be great to cross-reference your data with any languages already listed on Hyperglot, and perhaps to add your data?

J.

Anexos

1. Glossário, Terminologia tipográfica

Este Glossário foi publicado originalmente no livro *Como criar tipos: do esboço à tela* (2019), de Cristóbal Henestrosa, Laura Meseguer e José Scaglione. O mesmo texto foi utilizado como base para este glossário, mediante autorização dos autores.

Alfabeto

Conjunto de signos (letras) usados na grafia de uma língua. Num alfabeto, cada letra ou grafema corresponde a um som ou fonema. Logo, pode-se considerar que os sistemas de escrita que utilizam o alfabeto são escritas fonéticas.

Altura-x

(altura-de-x, altura de olho médio, *x-height*) Distância da linha de base à linha mediana de um determinado tipo, que normalmente corresponde à altura da letra x minúscula.

Âncoras

(*anchors*, attachment points) Pontos de referência para o posicionamento dos diacríticos em relação aos caracteres tomados como base.

Ascendente

(*ascender*) Projeção de uma haste vertical acima da altura-x, como no b, d, f, h, k e h.

Barra

(haste transversal) Traço reto horizontal de um caractere, como na letra E.

Cabeçalho de arquivo

(*file header*) É a informação no código da fonte que permite seu funcionamento e vínculo com o sistema operacional e o aplicativo de editoração eletrônica. Por exemplo criação das tabelas para nomes das fontes (*naming*), identificação (*font identification*), suporte e licenciamento.

Caixa alta

(maiúscula, versal, *uppercase, capitals*) Conjunto das letras maiúsculas de um alfabeto. Nos tipos de metal, refere-se à parte inferior da caixa tipográfica na qual esses caracteres são colocados.

Caixa baixa

(minúscula, *lowercase*) Conjunto das letras minúsculas de um alfabeto. Nos tipos de metal, refere-se à parte superior da caixa tipográfica na qual esses caracteres são colocados.

Caligrafia

Forma de escrita manual executada com cuidado especial, que costuma ter finalidade ornamental ou estética.

Caractere

Segundo o Unicode Consortium, um caractere é a menor componente da linguagem escrita com valor semântico; refere-se ao significado abstrato e/ou forma, e não à somente uma forma específica. Em linhas gerais, um caractere é uma unidade de informação que corresponde a uma letra ou grafema.

Claro

(*light*) Versão de uma família tipográfica com espessura menor do que a regular.

Condensada

(estreita, compactada, *condensed*) Variante de uma família tipográfica que ocupa menos espaço horizontal do que a versão regular.

Conexão

(ligação, *connection, joint*) Ponto de união entre traços, como na intersecção entre uma curva e uma reta ou entre duas diagonais.

Contraforma

(olho, *counter*) Espaço em branco interno de uma letra. Pode ser fechada ou aberta.

Contraste

É a diferença entre a espessura dos traços mais grossos e mais finos das letras.

Cor tipográfica

(cinza, cor) Tom de cinza gerado visualmente pela mancha do texto.

Corpo

(*body, body size*) Medida vertical de um caractere tendo como referência o retângulo no qual ele se inscreve, cuja dimensão se mede, em linhas gerais, a partir da base das descendentes até o topo do acento das maiúsculas, independentemente de sua aparência impressa. Normalmente medido em pontos tipográficos.

Curva

Traço curvo de um caractere. As curvas circulares por vezes são chamadas de bojos, e as semicirculares, de laços.

Curvas de Bézier

Método de descrição matemática de curvas desenvolvido por Pierre Bézier por volta de 1960 para o traçado de desenhos técnicos no design automotivo. Posteriormente, esse

método foi adotado pelos inventores do PostScript (linguagem que permitiu o desenvolvimento de sistemas de impressão de alta qualidade a partir do computador) para a geração de desenhos vetoriais e contornos de fontes.

Descendente

(*descender*) Projeção de um traço abaixo da linha de base, como no g, j e p.

Ductus

Movimento ou caminho que faz a ferramenta ao escrever.

Egípcia

(mecânica, serifa quadrada, *egyptian, slab serif*) Modelo tipográfico caracterizado por traços pesados, baixo contraste e serifas quadrangulares.

Entrelinha

(*leading, interlinear spacing*) Espaço entre duas linhas de um texto composto. Na tipografia digital, é a distância entre as linhas de base de duas linhas consecutivas de texto. Na composição com tipos de chumbo, era o espaço introduzido entre duas linhas de caracteres por uma tira de metal.

Escrita

(*writing*) Ato ou efeito de escrever; aquilo que se escreve. Forma de produzir letras ou sinais por gestos manuais utilizando uma ferramenta (lápiz, caneta, pincel) sobre um determinado suporte. Escrita é uma definição ampla, que engloba, por exemplo, o conceito de caligrafia.

Espécimen

Mostruário tipográfico.

Família tipográfica

É o conjunto de todas as formas dos caracteres em suas distintas variantes, desenhadas segundo um critério formal comum e agrupadas sob uma denominação genérica. Em linhas gerais, uma família tipográfica é um conjunto de fontes.

Fonte

(*font*) Expressão que designa, em tipografia digital, um arquivo de computador que contém os dados relativos a uma única variante de uma família tipográfica, sob uma codificação determinada: OpenType, PostScript, TrueType, etc.

Fundição tipográfica

(*foundry, type foundry*) Empresa que distribui e/ou produz tipos digitais. Esta denominação remete às antigas fábricas que fundiam tipos móveis de metal.

Glifo

(*glyph*) A aparência gráfica de um caractere, enquanto este último é uma unidade conceitual. Por exemplo, o glifo que representa o caractere A maiúsculo do alfabeto latino também pode representar o caractere Alfa maiúsculo do alfabeto grego.

Grade

(*grelha, malha, retícula, grid*) Estrutura construtiva que permite organizar os elementos compositivos seguindo uma pauta em um espaço determinado, comum a todos os elementos de um mesmo sistema, como os caracteres de um alfabeto ou as páginas de um livro. Não costuma ser visível no resultado final.

Haste

(*stem*) Traços estruturais que configuram um caractere. Geralmente, as retas verticais.

Hinting

Instruções que se aplicam aos caracteres para melhorar a sua visualização em tela ou

em baixa resolução, e que são determinadas pela posição dos pontos. N.T. Não existe tradução precisa deste termo para o castelhano, nem para o português.

Humanista

Modelo tipográfico que faz referência aos tipos do renascimento italiano, procedentes da escrita carolíngia. Também denominado veneziano ou garalde na classificação Vox-ATypI.

Ink traps

Literalmente, “armadilhas de tinta”. Essa forma de compensação ótica consiste no aumento dos espaços em branco, especialmente nos pontos em que dois traços se conectam em um ângulo muito agudo. Ao imprimir a letra, estes espaços compensam o excesso de tinta e mantêm a conexão limpa e nítida.

Itálica

(cursiva) Variante de uma família tipográfica, com forma e estrutura diferente da romana, com certa inclinação natural, que faz referência à escrita cursiva ou contínua.

Kerning

(crenagem) Compensação de espaçamento entre pares de caracteres específicos.

Legibilidade

(*legibility*) Grau de reconhecimento dos caracteres de uma determinada fonte. Embora muitos estudos e experimentos científicos tenham sido realizados a esse respeito, nenhum é conclusivo, e por isso continua a ser um parâmetro subjetivo.

Leiturabilidade

(*readability*) Grau de conforto experimentado pelo leitor durante o processo de leitura.

Lettering

(letreiramento) Técnica de produção de letras com ênfase no desenho de contornos, em oposição à escrita, onde cada traço é produzido com um único gesto manual. Faz referência àquelas letras que são desenhadas, e não escritas ou caligrafadas.

Ligatura

Glifo resultante da união de, geralmente, dois caracteres. Algumas ligaturas como o fi e ff são padrão na maioria das fontes (*Standard Ligatures*). Algumas fontes possuem formas especiais, denominadas ligaturas discricionárias (*Discretionary Ligatures*), como Th, ct, ch, ck e st.

Linha de base

Linha sobre a qual se apoiam os caracteres do alfabeto, correspondente ao limite inferior da altura-x de uma fonte.

Mapa de caracteres

Conjunto de caracteres na codificação de uma fonte.

Modulação

Variação da espessura e da simetria dos traços que definem um traçado tipográfico. O grau de modulação abrange uma gama muito ampla, desde a ausência de modulação dos tipos monolineares, passando pela variação sutil que ocorre no eixo de um tipo sem serifa ou linear, até os traços altamente modulados dos tipos góticos. A modulação também está relacionada com a propriedade cursiva ou a condição da letra cursiva.

Monolinear

Modelo tipográfico caracterizado por possuir traços sem modulação ou contraste aparentes e ausência de terminais, parecendo ter sido desenhado por uma única linha uniforme. Na classificação Vox-ATypI é denominado simplesmente linear.

Monoespaçados

Modelo tipográfico no qual todos os caracteres ocupam o mesmo espaço horizontal, com espaçamento idêntico entre eles.

Multiple Master

Esta tecnologia, criada pela Adobe, parte do princípio de que as fontes digitais – e os contornos glifos propriamente ditos – são números e equações matemáticas. Portanto, se o desenho de dois pesos extremos é executado, é possível obter estilos ou pesos intermediários por um processo chamado interpolação.

Negrito

(preto, *bold*) Variante de uma família de tipos com mais peso do que a regular.

Numerais alinhados

(numerais de caixa alta, numerais de título, lining figures) Aqueles que têm todos a mesma altura, como as letras de caixa alta.

Numerais desalinhados

(em estilo antigo, de texto, elzevirianos, de caixa baixa, *oldstyle figures*, *text figures*) Se distinguem por ter altura variável. Normalmente, o, 1 e 2 vão apenas da linha de base até a altura-x; 3, 4, 5, 7 e 9, avançam abaixo da linha de base; enquanto 6 e 8 chegam aproximadamente até a altura das letras de caixa alta.

Numerais proporcionais

(*proportional figures*) Numerais que se caracterizam por ter o espaçamento proporcional à largura de seu desenho. Por exemplo, o numeral 1 é naturalmente mais estreito que o 8; logo, a largura total desses caracteres (glifo + espaçamento) é diferente.

Numerais tabulares

(*tabular figures*) Numerais que se caracterizam por ocupar, todos, um espaço horizontal

idêntico, a fim de facilitar a composição de tabelas nas quais os algarismos devem ser alinhados uns abaixo dos outros.

Numerais em versalete

Numerais com o mesmo tamanho e alinhamento dos versaletes de uma fonte.

Olho

Parte impressa de um caractere. Faz referência à impressão com tipos de chumbo, na qual correspondia à porção saliente do tipo que recebia a tinta.

Ombro

Em anatomia tipográfica, traço curvo horizontal que se une a outro vertical, como no n, r e u.

Open Type

Formato vetorial de fontes tipográficas, essencialmente uma extensão do TrueType que permite maior flexibilidade e o armazenamento de curvas de Bézier cúbicas ou quadráticas. Pode ter tanto a extensão ttf quanto otf.

Overshoot

(compensação ótica) O grau em que traços curvos e diagonais se projetam além da linha de base ou da altura-x, a fim de compensar o efeito óptico de encolhimento que resultaria se elas fossem da mesma dimensão, já que, frente a uma altura igual, o círculo e o triângulo parecem menores do que o quadrado.

Peso

(*weight*) Expressão pela qual se denominam as variantes de espessura de uma família de fontes.

Ponto

Unidade básica de medida tipográfica em sistema duodecimal. Não existe um padrão universal, uma vez que existem dois sistemas diferentes com variação em suas medidas o anglo-americano, baseado na polegada, no qual 1 ponto corresponde a 0,3514729 mm; e o europeu ou Didot, no qual 1 ponto corresponde a 0,376065 mm. No primeiro, a medida de 12 pontos é chamada de paica, e, no sistema europeu, cícero. N.T. Nos sistemas de editoração eletrônica atuais, um ponto tipográfico equivale a 1/72 da polegada anglo-saxã estabelecida em 1959 como equivalente a 25,4mm, ou 0,3527 mm.

PostScript

É a linguagem de descrição de página (PDL) padrão dentro da indústria gráfica atual, desenvolvida na década de 1980 pela Adobe Systems. Com ela, apareceram também os formatos de fontes Type1 ou fontes PostScript.

Proteção lateral

(*sidebearing*) Espaço em branco que se encontra, no sentido horizontal, antes e depois do glifo. Consequentemente, a soma da proteção lateral direita do primeiro caractere e a proteção lateral esquerda do segundo formarão o espaço não impresso entre essas duas letras.

Swashes

(caudais, caudata) Traços prolongados ou adicionais de alguns caracteres alternativos, de caráter ornamental.

Romana

(normal, redonda, regular) Variante de uso corrente de uma família tipográfica, sem inclinação, com espessura proporcionada. É a mais comum e mais adequada para a composição de textos corridos. O termo não costuma ser adicionado ao nome da fonte por ser o estilo principal ou básico da família tipográfica.

Romana inclinada

(oblíqua, *oblique*, *slanted roman*) É uma letra com a mesma estrutura da romana, mas com inclinação.

Script¹

(caligráfica, escritural) Modelo tipográfico que se assemelha à escrita manual com pena.

Script²

Em linguagem computacional, arquivo de programação simples destinado a ações específicas, que complementa um programa de computador que o interpreta.

Sem serifa

(linear, sans serif) Modelo tipográfico cuja principal característica é a ausência de serifas.

Serifa

Elemento terminal, não construtivo, do traçado, em certas famílias tipográficas. Geralmente, está presente nos estilos que fazem referência à sua origem caligráfica ou lapidar (gravada em pedra), como reminiscência da ferramenta com a qual se escrevia, pintava, desenhava ou gravava letras.

Terminal

(remate) Acabamento na extremidade de um traço.

Tipo¹

(*type*, *sort*) Na impressão tipográfica com tipos móveis, cada uma das peças de metal utilizadas para compor textos.

Tipo²

(*typeface*) Coleção de caracteres tipográficos que compartilham as mesmas características de design, como peso, largura, inclinação, modulação de traço e espaçamento.

Tipografia

(*typography*) 1. O conjunto de práticas e processos envolvidos na utilização de símbolos visíveis relacionados aos caracteres ortográficos e para-ortográficos para fins de reprodução de textos. 2. Disciplina dedicada ao estudo, ao uso e à criação de letras, alfabetos e sistemas de escrita.

Tipógrafo

(*typographer*) Na impressão manual, o trabalhador que compõe com tipos de chumbo. Por extensão, qualquer pessoa que utilize tipos para compor um texto. É importante distinguir o designer de tipos (o profissional que desenha tipos) do tipógrafo ou designer tipográfico (o profissional que desenha com tipos; ou seja, cria e executa especificações de projetos gráficos).

Traço

Cada um dos elementos com os quais um caractere é construído. Faz referência à escrita manual.

True Type

Formato de fontes digitais desenvolvido pela Apple para competir com o PostScript. Ao contrário deste último, os glifos são desenhados com curvas quadráticas no lugar das curvas cúbicas e oferece maiores recursos de hinting.

Unicode

Versal

V. caixa alta

Versalete

(small caps) Letras que tem a forma das maiúsculas, mas o tamanho das minúsculas.

2. Documento enviado pela Professora Deborah Anderson

UNIVERSITY OF CALIFORNIA, BERKELEY

BERKELEY • DAVIS • IRVINE • LOS ANGELES • RIVERSIDE • SAN DIEGO • SAN FRANCISCO



SANTA BARBARA • SANTA CRUZ

DEPARTMENT OF LINGUISTICS

1203 DWINELLE HALL # 2650
BERKELEY, CA 94720-2650

6 July 2019

To Whom It May Concern:

I am writing to invite Rafael Dietzsch to be a Visiting Student Researcher in the Department of Linguistics at UC Berkeley as part of the Fulbright Doctoral Dissertation Research Award Program from August 2020 until May 2021. I would be very happy to act as his advisor. As a Visiting Student Researcher, he will have access to the UC Berkeley libraries, which have plentiful study spaces. There is a University Services Fee required for all Visiting Student Researchers of \$1000.

I have read the description of Rafael Dietzsch's proposed PhD project being undertaken at the Communication Postgrad Program at the University of Brasília, as well as his Master's thesis in Typeface Design at the University of Reading. The proposed work on the documentation of the typography for Brazilian indigenous languages is both timely and needed.

My project at UC Berkeley, the Script Encoding Initiative (SEI), focuses on helping to get modern minority languages and historic languages supported on computers and devices. The indigenous languages of Brazil are currently poorly represented on computers. As the UC Berkeley representative to the Unicode Consortium, I will introduce Mr. Dietzsch to those working on the Unicode Common Locale Data Project, an open-source project that collects data on the letters needed by languages of Brazil for software development, and to keyboard developers in Silicon Valley who work on creating keyboards for minority languages. It should also be noted that the UC Berkeley Department of Linguistics includes faculty who specialize in the languages of the Amazon, including documentation and revitalization of these languages, so the proposed work would be very fitting.

If you have any questions or concerns, please do not hesitate to contact me.

Sincerely,

A handwritten signature in cursive script, appearing to read "Deborah Anderson".

Deborah Anderson, Ph.D.

Researcher, Department of Linguistics, UC Berkeley

Project Head, Script Encoding Initiative, <http://linguistics.berkeley.edu/sei>

Email: dwanders@berkeley.net or dwanders@sonic.net