



Como assegurar a presença de uma língua no ciberespaço?

Marcel Diki-Kidiri



PIPT – Programa Informação para Todos
Sector da Comunicação e da Informação



Como assegurar a presença de uma língua no ciberespaço?

Marcel Diki-Kidiri

Investigador do CNRS,

Laboratório «Linguagem, Línguas e Culturas da África Negra»
(LLACAN)

PIPT – Programa Informação para Todos

Título original:

Comment assurer la présence d'une langue dans le cyberspace ?

de Marcel Diki-Kidiri

Publicado pela Organização das Nações Unidas para a Educação, Ciência e Cultura (UNESCO), Paris, França

© UNESCO 2007

© UNIÃO LATINA – DTIL 2007 para a tradução em língua portuguesa

A apresentação dos factos contidos nesta publicação, bem como as opiniões que aí se encontram expressas, não correspondem necessariamente às da UNESCO, e não comprometem esta mesma Organização. As denominações utilizadas nesta publicação e a apresentação dos dados que aí se encontram não implicam, por parte da UNESCO, nenhuma tomada de posição quanto ao estatuto jurídico dos países, territórios, cidades ou zonas mencionadas, nem quanto às suas autoridades ou traçado das suas fronteiras ou limites.

A tradução para a língua portuguesa foi realizada sob a responsabilidade da UNIÃO LATINA – DTIL.

Esta obra foi elaborada no âmbito do Estudo sobre o Multilinguismo no Ciberespaço, levado a cabo pela UNIÃO LATINA – DTIL, a pedido da UNESCO.

Tradutora: Vanda Dias

Revisoras: Rute Costa e Raquel Silva (Centro de Linguística da Universidade Nova de Lisboa)

Preâmbulo

As línguas são, antes de mais, instrumentos que permitem adquirir uma certa autonomia educativa e cultural. Para além de favorecerem a transmissão de conhecimentos entre gerações, são também meios de divulgação de culturas e tradições, entre e no seio de diferentes grupos étnicos, e em diferentes zonas geográficas. O fenómeno do desenvolvimento e o desaparecimento das línguas faz parte da evolução e do declínio das civilizações. O latim –língua morta– desempenha, ainda hoje, uma influência determinante nas línguas vivas, inclusive enquanto base inicial para o primeiro código padrão para computadores, o código ASCII. A língua materna é também o primeiro motor para a liberdade de expressão.

O desaparecimento de algumas línguas é um fenómeno presente ao longo da História. De facto, mesmo em países oficialmente monolíngues, as novas políticas vão no sentido de assegurar a expressão nas línguas endógenas como qualquer outro direito humano.

Segundo um estudo realizado pelo *Ethnologue*, o índice de diversidade linguística por região demonstra que a África é o continente que possui o indicador mais elevado no mundo. Alguns dados apontam para o facto de a diversidade linguística global estar, actualmente, em declínio. Outro factor preocupante: segundo algumas estimativas, metade das línguas desaparecerá por volta de 2050.

As tecnologias da informação e da comunicação desempenham um papel determinante nas transformações linguísticas ocorridas no mundo podendo, assim, vir a transformar-se num importante meio de comunicação entre as diferentes comunidades linguísticas. Paradoxalmente, elas poderão, também, ser um factor agravante da marginalização das línguas no ciberespaço. Existem cerca de 6 000 línguas no mundo, das quais 12 correspondem a

98 % das páginas web. O inglês, que detém 72 % dessas páginas é, indubitavelmente, o idioma dominante, segundo um estudo feito por O'Neill, Lavoie e Bennet, em 2003.

Assim, o desafio perante a comunidade internacional será o de ultrapassar os obstáculos, colossais, por forma a assegurar a criação de um ciberespaço multilíngue e culturalmente diversificado.

No intuito de alcançar este objectivo, a UNESCO – com a ajuda da União Latina e o contributo intelectual do perito Marcel Diki-Kidiri – publica o presente documento técnico, no âmbito da série de publicações do «Programa Informação para Todos».

De acordo com a «Recomendação sobre a promoção e uso do multilinguismo e o acesso universal ao ciberespaço», adoptada pela Conferência Geral da UNESCO aquando da 32ª Sessão, a presente publicação pretende contribuir para facilitar a tomada de decisões a favor da inclusão de novas línguas no espaço digital.

Abdul Waheed Khan
Vice-Director Geral
para a Comunicação e Informação

Índice

Introdução	7
1. Elaboração de recursos linguísticos	11
1.1. Da oral à escrita	11
1.2. Desenvolvimento terminológico	17
2. Elaboração dos recursos informáticos	19
2.1. Sistemas de escrita e ortografias	19
2.2. Jogos de caracteres e tipos de letra	21
2.3. Codificação dos caracteres	22
2.4. Primeiros recursos de base	26
2.5. Localização dos programas informáticos	32
3. Elaboração dos conteúdos culturais	35
3.1. Conteúdos culturais	35
3.2. Preparar as vias de acesso ao ciberespaço	37
4. Desenvolver a comunidade dos utilizadores	41
4.1. Criação de uma comunidade de utilizadores	41
4.2. Aspectos políticos e socioculturais	42
5. Conclusão	49
Referências	51
Anexo 1: Jogos de caracteres AFRFUL	53
Anexo 2: Jogos de caracteres AFRLIN	67



Introdução

O ciberespaço está aberto a todas as línguas do mundo, dado que a sua infra-estrutura não está submetida à autoridade de um poder central, que decidiria assim da sua utilização. Basta, em princípio, dispor de um computador ligado a um fornecedor de acesso à Internet para visualizar, na Rede, dados textuais, iconográficos ou sonoros na língua que se desejar. A concretização deste princípio, factor fundamental da democracia a nível mundial, necessita, no entanto, que estejam reunidas um certo número de condições técnicas e de recursos humanos e financeiros, que serão alvo de análise no presente estudo. Ao redigir este estudo, quisemos responder da forma mais simples possível à seguinte questão: como fazer com que uma língua menos dotada de recursos linguísticos e/ou informáticos e, até, de recursos humanos, consiga ocupar um lugar no ciberespaço e aí ser activa?

Por extensão, qualquer língua com mais ou menos recursos poderá encontrar algumas respostas neste estudo, se estiver pouco representada no ciberespaço. Descrevemos neste caso o percurso mais longo, o de uma língua que não tem sequer uma representação escrita, mas outras línguas mais favorecidas encontrarão aqui as etapas que lhes dizem respeito, no mesmo caminho.

Será necessário responder a uma questão preliminar. O que é uma língua menos favorecida? Trata-se de uma língua que não dispõe dos recursos essenciais, por serem insuficientes ou até por não existirem, de que são geralmente dotadas as grandes línguas do mundo, isto é: uma ortografia estável num determinado sistema de escrita, obras de referência (gramáticas, dicionários, obras literárias), obras de difusão massiva (imprensa escrita e audiovisual, filmes, canções e música), obras de cariz técnico e de aprendizagem (publicações técnicas e científicas, obras didácticas, manuais), diversos suportes de comunicação do quotidiano (cartazes, publicidades, correios, notícias, manuais de instruções, etc.), bem como um número significativo de aplicações informáticas nessa mesma língua. No que diz respeito aos recursos humanos, uma língua menos favorecida pode correr o

risco de desaparecer, se for usada apenas por um pequeno número de falantes. Para salvá-la, torna-se então necessário aumentar o número dos seus falantes, ensinando-a por todos os meios técnicos possíveis.

Felizmente, uma língua menos favorecida não acumula necessariamente todas estas lacunas em simultâneo. Ela pode ser maioritária, escrita, ensinada na escola e, no entanto, não possuir meios informáticos ou até recursos linguísticos em quantidade e qualidade suficientes. Será, então, mais justo dizer que as línguas menos favorecidas são-no de maneira muito diversa, indo desde as línguas que estão em perigo de extinção, até às línguas emergentes, as quais possuem já uma boa parte destes recursos, mas em número julgado insuficiente e incompleto. Em todos os continentes, há muitas línguas que correspondem à definição de línguas menos favorecidas. Na Europa podemos citar o bretão, o occitano ou ainda o basco, na América, a quase totalidade das línguas ameríndias, na Ásia, o myanmar¹, citando apenas uma entre centenas, na Oceânia, quase todas as línguas autóctones das Ilhas da Polinésia, da Micronésia e da Melanésia. Em África, onde se fala um terço das línguas do mundo, ou seja, cerca de 2000 línguas, as línguas mais favorecidas (afrikaans, kiswahili, hausa, etc.) contam-se pelos dedos das mãos e pertencem às línguas emergentes e, portanto, menos favorecidas.

Partiremos portanto da hipótese menos favorável, isto é, de uma língua que, outrora, foi portadora de uma cultura florescente, mas que actualmente só é falada por meia dúzia de pessoas idosas, numa pequena aldeia nos confins do mato africano, digamos na África Central, para estar o mais afastado possível da costa. Chamemos a esta língua o *ndeka*. Tal língua não tem, praticamente, nenhuma possibilidade de aceder um dia ao ciberespaço, dado que corre o risco de desaparecer para sempre com a morte do último falante, já que não são muito numerosos os que ainda a praticam. Um dia, eis que um jovem estudante oriundo daquela aldeia se depara com ferramentas como o SPIP, a Wikipédia e tantas outras, que poderiam constituir uma oportunidade excepcional para conservar a memória desta língua e,

1 Nova denominação da língua outrora chamada *língua birmana*.

por acréscimo, a da aldeia e da cultura dos seus antepassados. Seria até o meio mais seguro de ensinar esta língua aos jovens da sua geração e, assim, dar-lhe uma nova vida. Vem assim ter connosco para o guiarmos no seu projecto. Tendo esta hipótese em mente, o presente trabalho reveste-se de um cariz didáctico, no intuito de acompanhar, passo a passo, todos aqueles que se juntarão a nós em qualquer etapa do caminho para conduzir ao ciberespaço todas as línguas menos favorecidas, independentemente da sua riqueza.



1. Elaboração de recursos linguísticos

1.1. Da oral à escrita

O objectivo desta primeira etapa é dotar a língua de um mínimo de recursos linguísticos, ou seja, de uma ortografia recorrendo a um determinado sistema de escrita, de uma gramática escrita, de um dicionário e de uma recolha de textos, a maior possível. É, portanto, um trabalho de linguista que iremos aqui evocar nas grandes linhas, concentrando-nos sempre nos objectivos intermédios mais indispensáveis.

1.1.1. Recolha de textos

A primeira coisa a fazer é ir ao encontro dos falantes da língua e gravar o maior número de textos possível, relatos de vida, contos, provérbios, cantigas, poemas, lendas, conversas, narrações, etc. Para gravar estes dados, é aconselhável adquirir técnicas de investigação no terreno e utilizar sempre que possível aparelhos de qualidade profissional, a fim de obter uma maior qualidade sonora, dado que o som, assim gravado, será depois submetido a vários tratamentos. Seria uma pena que a qualidade do som se degradasse logo na segunda cópia! Em condições normais de trabalho, as gravações são logo traduzidas com a ajuda dos falantes da língua e transferidas para uma base de dados textual.

1.1.2. Transcrição fonética

Servindo-se do alfabeto fonético internacional (API), os linguistas podem transcrever, com a maior das fidelidades, qualquer som de qualquer língua do mundo. Os textos recolhidos podem, assim, ser transcritos mas é necessário ter em mente que a transcrição fonética reproduz fidedignamente cada som, tal como este é pronunciado num determinado momento. Ora, em francês, por exemplo, numa palavra como *papa* (papá), o primeiro *P* é ligeiramente mais explosivo do que o segundo, dado que está no início da palavra e o segundo no meio de duas vogais. Habitualmente, esta pequena diferença não

é tida em conta dado que não interfere no sentido da palavra, mas uma transcrição fonética minuciosa deverá ser assim: [p[<]apa]². A transcrição fonética é indispensável para permitir uma boa análise dos sons da língua, mas não será, com toda a certeza, a forma mais simples de a representar para um uso corrente.

1.1.3. *Análise e notação dos fonemas*

Se a transcrição fonética for correcta, isto é realizada com minúcia e fidelidade relativamente aos sons realmente pronunciados, será relativamente fácil proceder, posteriormente, a uma análise fonológica, cujo objectivo será identificar os sons úteis da língua, a que se dá o nome de *fonemas*. Trata-se de sons que produzem uma diferença de sentido, quando são trocados. Por exemplo, em francês, dir-se-á que /p/³ e /b/ são dois fonemas diferentes porque a sua diferença de pronúncia permite, por si só, distinguir as palavras «pain» (*pão*) e «bain» (*banho*).

Pelo contrário, certas pessoas pronunciam a palavra *roi* (*rei*) com um *R* enrolado e outras com um *R* gutural. Estes dois *R* são foneticamente diferentes e transcritos, respectivamente, [r] e [R]. No entanto, uma vez que esta diferença fonética não acarreta nenhuma diferença de sentido, estes dois sons representam apenas um único fonema que convencionalmente deverá ter a notação /r/. Um outro exemplo, o conjunto de sons [gz] que se encontra na palavra «exact» (*exacto*), pronunciada [egza] e o conjunto de sons [ks] da palavra «extra» (*extra*) pronunciada [ekstRa], representam ambos o mesmo fonema, que, convencionalmente, deverá ter a notação /x/.

Assim, a análise fonológica de uma língua leva ao estabelecimento da lista completa dos fonemas dessa mesma língua e à forma correcta de fazer a sua notação. Podemos assim utilizar esta lista de fonemas para escrever um texto. Fala-se então de *notação fonológica* por oposição a *transcrição fonética*. Numa notação fonológica, são os fonemas e não os sons que são representados. A notação é, portanto, mais económica e fornece uma boa base para a elaboração de uma ortografia.

2 Utilizamos os parênteses rectos para indicar que se trata de uma transcrição fonética.

3 Utilizamos as barras oblíquas para indicar um fonema ou uma sequência de fonemas.

1.1.4. Análise e notação dos tons

Podemos pronunciar uma sílaba num tom de voz (ou *registo*) relativamente grave, médio ou agudo, tal como as notas musicais *dó ré mi*. A maioria das línguas africanas explora estas variações de tom de voz para exprimir as diferenças de sentido de uma mesma combinatória de fonemas. Assim, em sango (ou sangho)⁴, as palavras [ká] (além), [kā] (ferida), e [kà] (e) são três palavras muito diferentes, e não se trata, de forma alguma, de homónimos. Estas línguas são qualificadas de *línguas tonais*.

Uma língua tonal pode explorar até cinco registos diferentes, mas a maioria das línguas africanas só explora dois ou três. Com dois registos, podemos distinguir um tom alto e um tom baixo. Com três registos, distinguimos um tom baixo, um tom médio e um tom alto. Para além destes, distinguimos, ainda, um som supra-alto e/ou um tom infra-baixo. É na sequência de uma análise dos tons que se poderá determinar se se trata de uma língua tonal ou não e, em caso afirmativo, o número de registos utilizados, a frequência dos tons, a melhor maneira de fazer a sua notação, etc.

Numa língua tonal, uma palavra é, necessariamente composta por, pelo menos, uma vogal e um tom. Assim, uma forma como *ka*, acima referida, só ganha sentido quando se explicita o tom sobre o qual deverá ser pronunciada. Logicamente, será então necessário transcrever os tons no sistema ortográfico da língua. Contudo, nem todas as línguas tonais fazem o mesmo uso dos tons. Algumas servem-se dos tons não só para distinguir palavras, a nível do léxico, mas também formas verbais e modalidades pessoais e aspectuais, portanto valores gramaticais. Nessas línguas, os tons ocupam um lugar demasiado importante para serem negligenciados. Pelo contrário, nas línguas que recorrem aos tons unicamente ao nível do léxico, e de maneira relativamente limitada (por exemplo, apenas para distinguir palavras breves onde a ambiguidade pode ser maior), considerou-se que não era fundamental fazer a notação dos tons na ortografia corrente, sendo

4 O *sango* ou *sangho* é a língua nacional da República Centro-Africana, com a qual o francês partilha o estatuto de língua oficial.

assim necessário decidir caso a caso. No entanto, convém sublinhar que, para trabalhos de investigação linguística sobre uma língua tonal, a transcrição dos tons é indispensável dado que, neste tipo de trabalho, deve-se considerar todas as informações sobre a realidade da língua, e os tons constituem uma dimensão importante e definitiva de uma língua tonal.

1.1.5. Elaboração de uma ortografia

Os estudos fonológicos e tonológicos conduzem a um sistema de transcrição da língua que representa correctamente os sons úteis, e isto da maneira mais económica possível. Permite escrever a 90 % o que se pronuncia, sem refinamento fonético, o suficiente para ter a tentação de se privilegiar apenas a notação fonológica, passando-se assim por cima de estudos ortográficos. No entanto, estes últimos tornam-se verdadeiramente indispensáveis para dotar uma língua de um sistema convencional, capaz de ter em conta o conjunto das necessidades de expressão escrita dos falantes. Com efeito, contrariamente à notação fonológica, que reflecte apenas os sons úteis da língua, uma ortografia bem concebida deverá também incluir as notações das relações gramaticais e das ideias. Consideremos, para o francês, os seguintes exemplos:

- a) /ɛl ʃãt/
- b) elle chante (*ela canta*)
- c) elles chantent (*elas cantam*)

A frase a) apresenta a notação fonológica que corresponde tanto à frase b) como à frase c). Se o francês se escrevesse unicamente com uma notação fonológica, não haveria nenhuma maneira de distinguir aqui o singular do plural, tal como o permite a convenção ortográfica.

Consideremos outro exemplo em sango, uma língua que distingue um tom alto, um tom médio e um tom baixo. Numa notação fonológica, estes tons são respectivamente notados com um acento agudo [á], um mácron [ā] e um acento grave [à], mas estes acentos não se encontram nas máquinas correntes na África Central. Uma ortografia destina-se, no entanto, ao uso de um vasto público, e a língua deve poder ser escrita tanto à mão como à máquina, ou no computador. É

por esta razão que é necessário elaborar uma convenção ortográfica que comporte um conjunto de regras que permitam escrever a língua de maneira mais prática.

O quadro seguinte mostra as diferentes etapas necessárias à elaboração de uma ortografia optimizada e estável para uma língua, neste caso para o sango :

Transcrição fonética	Notação fonológica	Ortografia 1 (47 %)	Ortografia 2 (53 %)	Tradução
1. [sùkùlà]	/sùkùlà	Sukûla	sukûla	lavar
2. [sã̃ ¹ gɔ̃ ¹]	/sãngō	Sängö	sahngo	sango (língua)
3. [mbásá ^m bálá]	/mbásámbáláa	Mbâsâmbâlâ	mbâssambala	sete
4. [mã̃ ¹ gɔ̃ ¹]	/mãingō	Mäingö	mayhngo	evolução
5. [k ^w à] ~ [k ^ù à]	/kùà	kua	kua	trabalho
6. [k ^w ã] ~ [k ^ù ã]	/kûã	küã	kwa	pêlo
7. [k ^w á] ~ [k ^ù á]	/kúá	kûâ	kwâ	morte

É de notar que a transcrição fonética reflecte a pronúncia das palavras tal como elas são registadas no gravador. Ela é, portanto, própria à investigação linguística, mas demasiado minuciosa para servir de base a uma ortografia corrente, destinada ao grande público. A notação fonológica da segunda coluna só retém os sons úteis, os fonemas e, desta forma, torna-se mais propícia para servir de base a uma ortografia corrente. Mas, a marcação de todos os tons constitui um obstáculo para uma escrita prática e uma leitura rápida. Além disso, os símbolos diacríticos utilizados para diferenciar os tons (acento agudo, mácron e acento grave) não estão disponíveis em todas as máquinas e só podem ser correctamente escritos no computador, facto que não está ao alcance de todos os utilizadores de uma língua menos favorecida. A convenção ortográfica da terceira coluna permite economizar 47 % da frequência dos tons, e de os escrever com o trema e o acento circunflexo, dois símbolos que figuram na tecla livre na maioria dos teclados ocidentais das máquinas de escrever e dos computadores utilizados na África Central. Apesar disto, com esta convenção subsistem 53 % da frequência dos tons para transcrever com acentos. Para reduzir ainda mais a taxa de frequência, efectuámos uma

reforma ortográfica cujos resultados se encontram na quarta coluna. Esta nova convenção permite economizar 53 % da frequência dos tons num texto. Nesta fase, podemos dizer que o sango (ou sahngo⁵) está finalmente dotado de uma ortografia otimizada e estável.

1.1.6. Elaboração de obras de referência

A análise dos textos recolhidos permitirá elaborar uma gramática completa, que cobrirá a fonologia, a sintaxe, os tipos de enunciados e o discurso. Geralmente, uma gramática de referência é concebida para servir de base a várias obras mais didáticas, tais como gramáticas de aprendizagem da língua e, se for caso disso, manuais escolares.

O estudo do vocabulário extraído dos textos recolhidos e, eventualmente, completado por inquéritos específicos, permitirá redigir um dicionário de referência tão completo quanto possível, o qual servirá de base a léxicos ou pequenos dicionários de aprendizagem. Esses mesmos textos serão os conteúdos dos livros de leitura, novelas, no intuito de incentivar a criação de outras obras literárias, se não for a partir da tradição oral, pelo menos a partir da observação da vida quotidiana na comunidade de falantes.

A realização de obras de referência e de obras práticas na língua não é uma etapa marginal que se possa ignorar. As línguas menos favorecidas necessitam sempre de obras deste género, mesmo quando fazem parte das línguas emergentes e, sobretudo, quando estão em perigo de desaparecer. Por outras palavras, estas obras são fundamentais para reforçar, e/ou criar, uma consciência comunitária nos falantes, sobretudo quando estes são escassos. São não só indispensáveis para ensinar a língua aos jovens, aumentando assim a sua capacidade de aprendizagem, mas também para alfabetizar os adultos e, por este meio, abrir-lhes as portas a conhecimentos mais alargados.

5 Aqui, o H inserido antes do NG significa que o tom médio do A é válido também para o O. É a aplicação de uma regra mais geral, que permite não ter que recorrer aos diacríticos para informar sobre os tons da palavra em determinados contextos.

1.2. Desenvolvimento terminológico

Para que uma comunidade de falantes utilize a sua língua como meio de comunicação no ciberespaço, é indispensável que esta língua seja enriquecida com os termos técnicos necessários para exprimir as realidades deste novo espaço. Por exemplo, termos como *correio electrónico*, *endereço electrónico*, *cópia conforme*, *conectar-se*, *ligar* (os aparelhos), *descarregar*, *publicar*, *redes*, *sítio*, *página Web* (página da Rede), *navegar*, etc. revelar-se-ão indispensáveis. O vocabulário necessário para nomear o material informático que utilizamos, constitui só por si já um léxico bastante considerável, que será necessário criar inteiramente, dado que, na maioria dos casos, este material não faz parte dos objectos culturais tradicionais e não é nem conhecido, nem nomeado. Sendo assim, devem ser postas em prática metodologias para a criação terminológica no seio da comunidade, a fim de dotar a língua com neologismos culturalmente aceitáveis. O desenvolvimento terminológico é uma actividade contínua e permanente, que será cada vez mais gerada pela comunidade de falantes, ao mesmo tempo que, ela própria, desenvolverá uma cultura crescente do ciberespaço. Não constitui portanto uma «etapa» no caminho de acesso ao ciberespaço, mas sim uma actividade de consolidação que é necessário iniciar a um dado momento, e prosseguir indefinidamente.



2. Elaboração dos recursos informáticos

2.1. Sistemas de escrita e ortografias

Antes de abordar esta etapa, convém explicitar certos conceitos, muitas vezes confundidos ou incompreendidos, sobretudo pelos não-especialistas, relativamente ao que se entende por sistema de escrita, sistema ortográfico (também chamado convenção ortográfica ou simplesmente ortografia) caractere, jogo de caracteres e tipo de letra.

2.1.1. *Os sistemas de escrita*

Consideremos a escrita latina, a escrita árabe, a escrita hebraica, a escrita chinesa, a escrita etíope, a escrita hieroglífica, a escrita cuneiforme e a escrita maia. Cada uma delas utiliza um conjunto de símbolos específicos com regras de construção que lhes são próprias. Trata-se, portanto, de sistemas de escrita diferentes. Contrariamente ao continente americano e ao continente europeu, onde os sistemas de escrita realmente utilizados contam-se pelos dedos das mãos, África conta com uma boa dúzia de sistemas de escrita, cujos mais conhecidos são a escrita latina, a escrita árabe, a escrita etíope, a escrita n'ko e a escrita tfinagh. Desta forma, quando nos propomos escrever uma língua que nunca foi escrita, convém saber em que tipo de escrita fazê-lo, para que a comunidade dos falantes possa imediatamente servir-se dela. Em geral, considera-se o sistema de escrita mais corrente na região.

2.1.2. *Os sistemas ortográficos*

Um mesmo sistema de escrita permite escrever línguas muito diferentes. Assim, a escrita latina serve para escrever a maioria das línguas da Europa Ocidental, da América e de África. No entanto, cada língua explora de maneira diferente os recursos desse sistema de escrita, em função das suas próprias estruturas fonológicas, sintáticas, enunciativas e semióticas. As regras de escrita daí

resultantes constituem um sistema ortográfico, próprio a cada uma das línguas. A título de exemplo, podemos considerar a duplicação de consoante simples nas quatro línguas seguintes:

- Em italiano, a geminação é importante porque pode produzir uma diferença de sentido. Por conseguinte, a duplicação de consoantes serve para transcrever a geminação e obriga a uma pronúncia acentuada da consoante dupla: *tutto* (tudo), *oggi* (hoje).
- Em francês, a duplicação de uma consoante tem apenas um valor etimológico ou estético, e não implica uma pronúncia específica da consoante. Comparemos, por exemplo, as palavras *addition* (adição), *apprécier* (apreciar), *atteler* (atrelar) com as palavras *adorer* (adorar), *apercevoir* (avistar) e *atelier* (atelier). Quer escrevamos estas palavras com as consoantes duplas, quer não, a pronúncia e o sentido permanecem os mesmos.
- Em inglês, mesmo não se tratando de uma regra geral, a duplicação da consoante depois de uma vogal serve, muitas vezes, para variar a pronúncia da vogal em questão. Assim, o /i/ pronuncia-se [ai] em *write* (escrever) e *hide* (esconder), mas [i] em *written* (escrito) e *hidden* (escondido). A duplicação do /t/ e do /d/ não serve para assinalar uma pronúncia marcada destas consoantes, como no caso do italiano, mas simplesmente para variar a pronúncia da vogal /i/ que as precede.
- Em sango (ou sangho), uma das regras da ortografia revista utiliza a duplicação da consoante simples para marcar o início de uma sequência tonal idêntica. Assim, em vez de se escrever *tāsēmā* (fila), onde o tom médio é assinalado com um trema em cada uma das sílabas da palavra, escreve-se *tassema*. A duplicação do /s/ é suficiente para indicar que o tom médio da primeira sílaba se repete nas outras sílabas que se seguem, até ao fim da palavra. Do mesmo modo, *mbâsâmbâlâ* (sete), que comporta uma sequência de tons altos, reescreve-se *mbâssambala*, aplicando assim a mesma regra. A duplicação do /s/ é suficiente para indicar que o tom alto do primeiro /â/ permanece nas sílabas seguintes até ao fim da palavra.

Estes quatro exemplos ilustram perfeitamente a que ponto as regras ortográficas são próprias de cada língua, mesmo quando utilizam o mesmo sistema de escrita.

2.2. Jogos de caracteres e tipos de letra

É denominado aqui *caractere* todo o símbolo discreto e significativo pertencente a um sistema de escrita. Notemos que, no sistema de escrita latino, o espaço que separa duas palavras é considerado um caractere, ao mesmo título que o acento agudo, uma letra, uma vírgula ou um número.

2.2.1. Os jogos de caracteres

Estabelecida a ortografia da língua, poder-se-á fazer a lista exaustiva dos caracteres necessários para escrever nessa língua, seguindo as regras da respectiva ortografia. Esta lista constitui o que se chama um *jogo de caracteres*. Comparando-o com os caracteres padrão utilizados em informática, podemos verificar se todos os caracteres do jogo estão aí presentes ou não. O caso de mais fácil tratamento é aquele em que os caracteres do jogo estão presentes nos caracteres padrão porque, nestes casos, não há nada de especial para fazer. Infelizmente, somos obrigados a constatar que, muitas vezes, os caracteres de que precisamos não se encontram nos jogos de caracteres padrão dos computadores. Fala-se então de *caracteres especiais*. Um jogo de caracteres pode comportar, por exemplo, sete vogais e vinte e quatro consoantes, alguns diacríticos (circunflexo, trema, til), dez dígitos (de 0 a 9), operadores matemáticos (+, -, %, <, >, =), sinais de pontuação e, por fim, símbolos como o asterisco (*), o parágrafo (§) ou, ainda, os símbolos monetários (\$, £, €). Outro jogo de caracteres poderia conter cinco vogais em vez de sete, e trinta ou quarenta consoantes em vez de vinte e quatro, etc.

2.2.2. Os tipos de letra

Um tipo de letra é, antes de mais, um conjunto de definições de formas de caracteres esboçados num mesmo estilo artístico. Os tipos de letra são, hoje em dia, programas que incluem, é certo, um stock de jogos de caracteres, mas também regras de representação

destes caracteres no ecrã, assim como funções informáticas para os gerir de maneira «inteligente». A criação de um tipo de letra com qualidade profissional requer conhecimentos aprofundados e uma grande competência técnica. É por esta razão que os tipos de letra são produtos industriais, protegidos por marcas registadas, tais como *Times New Roman*, *Garamond*, *Arial*, etc. Os tipos de letra gratuitos raramente são satisfatórios, se bem que existam alguns excelentes no domínio dos programas informáticos livres. Caso seja necessário criar caracteres especiais para escrever uma língua, depois de os ter procurado em vão nos tipos de letra já disponíveis, é necessário dirigir-se a uma pessoa competente ou a uma empresa especializada na área, para mandar fabricar tipos de letra com o jogo de caracteres escolhidos. O utilizador terá, ainda, a garantia de que os tipos em questão respeitam as normas internacionais.

2.3. Codificação dos caracteres

É importante perceber como é que os caracteres são codificados em informática para, pelo menos, poder colocar as questões mais pertinentes aos especialistas, sobretudo quando desejamos mandar criar tipos de letra com caracteres especiais para a sua língua. A maioria das línguas africanas recentemente escritas utiliza o alfabeto latino, enriquecido com vários caracteres, provenientes do stock de caracteres definidos pela Associação Internacional de Fonética (API) e adaptados pelo Instituto Africano Internacional (IAI). A título de exemplo, limitar-nos-emos a este caso de figura, sem ter em consideração outros sistemas de escrita. Aliás, a questão dos caracteres especiais não se coloca nos outros sistemas de escrita tais como no n'ko, no etíope, no tfinagh, dado que todos os caracteres destes sistemas estão normalizados e codificados, e os que não estão são variantes de caracteres normalizados e não caracteres especiais, com valores distintos.

2.3.1. A codificação num octeto

Imagine um fio eléctrico, no qual a única forma de criar um evento é desligar a corrente eléctrica e voltar a ligá-la. Obteríamos assim um valor nulo (0) quando a corrente está desligada e um valor positivo (1) quando se volta a ligá-la. Num dado ponto A, no fio, apenas estes

dois valores são possíveis e permitem responder «sim ou não» à pergunta «há corrente eléctrica?». Podemos então ver o ponto A como uma unidade mínima de armazenagem de informações que só pode assumir dois valores: 0 ou 1. Representa-se esta unidade através de um dígito binário (*binary digit* em inglês) chamado bit⁶. Se, com um único bit podemos exprimir 2 valores (0, 1), com 2 bits podemos exprimir 4 valores (00, 01, 10, 11) e, com 3 bits, 8 valores (000, 001, 010, 011, 100, 101, 110, 111). A progressão é, portanto, exponencial. O quadro seguinte ilustra o número de valores que se podem exprimir mediante o número de bits utilizados:

Número de bits	1	2	3	4	5	6	7	8
Número de valores	2	4	8	16	32	64	128	256

Se, por convenção, decidíssemos atribuir uma letra diferente a cada um dos valores obtidos com 3 bits, teríamos por exemplo o seguinte:

000 = a	001 = b	010 = c	011 = d	100 = e	110 = f	111 = g
---------	---------	---------	---------	---------	---------	---------

Não seria possível codificar todas as letras do nosso alfabeto. Na realidade, nunca se codificaram caracteres em 3 bits, mas primeiro em 7 bits e depois em 8 bits. Os caracteres codificados em 7 bits estão enumerados de 0 à 127 e constituem o jogo de caracteres chamado ASCII. Este jogo contém as 26 letras do alfabeto latino de base e nenhuma letra acentuada. O acréscimo de um só bit permitiu definir 128 caracteres suplementares, de 129 a 255. Esta extensão permitiu integrar os caracteres especiais necessários, tanto para a tipografia e desenho de gráficos como para diferentes línguas que utilizam os alfabetos europeus. Os jogos de caracteres obtidos desta forma contêm, cada um deles, 256 caracteres codificados de 0 a 255, mas são todos parcialmente diferentes. Correspondem à norma ISO-8859-*n* (onde *n* indica o número de um jogo em particular). Estes códigos são sempre representados, na máquina, por uma sequência de bits. Um conjunto de 8 bits chama-se um *octeto*.

São os tipos de letra que estabelecem a ligação entre os códigos (sequência de 0 e de 1), representando os caracteres no computador e os desenhos correspondentes (chamados *glifos*) que aparecem no

6 Vocábulo-mala formado a partir do início de *binary* e o fim de *digit*: *bi-* + *-it* > *bit*.

ecrã. Assim, o tipo *Times New Roman* lerá o código 097, correspondente ao «a minúsculo latino» e aparecerá no ecrã um dos seguintes «a»: a, **a**, a, **a**, mediante a escolha do utilizador. Estes quatro «a» têm em comum o estilo inerente à família tipográfica *Times New Roman*, que alia o aspecto cheio e desligado à pequena serifa (chamada *serif* em inglês), na base da letra. O mesmo acontece com o tipo de letra *Arial*, mas com letras mais «alongadas», sem serifas.

2.3.2. Limites da codificação num só octeto

Com a aceleração simultânea da potência dos computadores, das trocas internacionais e das necessidades de edição, os limites da codificação num octeto foram rapidamente atingidos. Um dos maiores inconvenientes da utilização deste tipo de codificação reside na necessidade de utilizar um jogo de caracteres específico para determinada língua que necessite de caracteres especiais. Assim, o jogo ISO-8859-1 (também chamado Latim-1) associa ao código 198 a letra «*Letra maiúscula latina æ*», isto é, a ligação «Æ». Mas esta letra pode não ser utilizada por uma língua menos favorecida que teria, pelo contrário, necessidade de um caractere inexistente tal como o «e aberto», ou seja «ε». Utilizando os programas adequados, um utilizador lambda tem a possibilidade de redefinir um caractere codificado. Poderia então substituir «Æ» por «ε» num tipo de letra local. Por conseguinte, para ler um texto escrito nesta língua com este tipo de letra local, o dito tipo deverá ser instalado na máquina. No entanto, um tipo de letra bem feito é um produto industrial, muitas vezes comercializado e que não se pode distribuir gratuitamente. O utilizador ocasional também não pretende investir num tipo de letra novo sempre que muda de língua de consulta! Mesmo dispondo de tipos de letra gratuitos, qualquer envio de documento (a um colega, a um editor ou tipógrafo) deverá, obrigatoriamente, ser acompanhado por um tipo de letra local «descartável»! Estes constrangimentos representam apenas uma pequena parte dos inconvenientes dos tipos de letra locais que, por outro lado, são de grande auxílio para os utilizadores quando localmente partilhadas. De qualquer forma, com um máximo de 256 códigos possíveis, a codificação num só octeto é, obviamente, insuficiente para codificar os 2 000 caracteres de base da escrita chinesa, e isto também é válido para o japonês e o coreano.

ISO 8859-1 Latin-1

0	32		64	@	96	`	128	160		192	À	224	à
1	33	!	65	A	97	a	129	161	ı	193	Á	225	á
2	34	"	66	B	98	b	130	162	ç	194	Â	226	â
3	35	#	67	C	99	c	131	163	£	195	Ă	227	ă
4	36	\$	68	D	100	d	132	164	¤	196	Ä	228	ä
5	37	%	69	E	101	e	133	165	¥	197	Å	229	å
6	38	&	70	F	102	f	134	166	ı	198	Æ	230	æ
7	39	'	71	G	103	g	135	167	§	199	Ç	231	ç
8	40	(72	H	104	h	136	168	¨	200	È	232	è
9	41)	73	I	105	i	137	169	©	201	É	233	é
10	42	*	74	J	106	j	138	170	₺	202	Ê	234	ê
11	43	+	75	K	107	k	139	171	«	203	Ë	235	ë
12	44	,	76	L	108	l	140	172	¬	204	Ì	236	ì
13	45	-	77	M	109	m	141	173	-	205	Í	237	í
14	46	.	78	N	110	n	142	174	®	206	Î	238	î
15	47	/	79	O	111	o	143	175	-	207	Ï	239	ï
16	48	0	80	P	112	p	144	176	°	208	Ð	240	ð
17	49	1	81	Q	113	q	145	177	±	209	Ñ	241	ñ
18	50	2	82	R	114	r	146	178	²	210	Ò	242	ò
19	51	3	83	S	115	s	147	179	³	211	Ó	243	ó
20	52	4	84	T	116	t	148	180	´	212	Ô	244	ô
21	53	5	85	U	117	u	149	181	µ	213	Õ	245	õ
22	54	6	86	V	118	v	150	182	¶	214	Ö	246	ö
23	55	7	87	W	119	w	151	183	·	215	×	247	÷
24	56	8	88	X	120	x	152	184	,	216	Ø	248	ø
25	57	9	89	Y	121	y	153	185	¹	217	Ù	249	ù
26	58	:	90	Z	122	z	154	186	ş-	218	Ú	250	ú
27	59	;	91	[123	{	155	187	»	219	Û	251	û
28	60	<	92	\	124		156	188	¼	220	Ü	252	ü
29	61	=	93]	125	}	157	189	½	221	Ý	253	ý
30	62	>	94	^	126	~	158	190	¾	222	Ť	254	ť
31	63	?	95	_	127		159	191	¿	223	ß	255	ÿ

Fonte: Programas integrados BPI - <http://www.progiciels-bpi.ca/tcao/apercu.html>

2.3.3. A codificação em vários octetos

Para ultrapassar os limites da codificação num só octeto, o consórcio Unicode e a Organização Internacional de Normalização (ISO) elaboraram a norma ISO/IEC-10646 que codifica em dois octetos um jogo de 65 536 caracteres, denominado «JUC-Unicode – Jogo Universal de Caracteres⁷» (em inglês: *Universal Characters Set – UCS*), o que dá para alojar todos os caracteres de todos os sistemas de escrita do mundo! É perfeito para os ideogramas chineses mas para os caracteres latinos que, até aqui, se contentavam com a codificação em um octeto, a codificação em dois octetos é demasiado ampla. Assim, para assegurar a compatibilidade com os documentos existentes, vários formatos de codificação foram propostos para os caracteres codificados em dois octetos. O mais conhecido daqueles formatos é o formato UTF-8.

No entanto, raramente é necessário fazer a conversão pessoalmente, dado que os programas informáticos de aplicação encarregam-se desta etapa. O que é realmente necessário compreender através da norma Unicode, é que cada um dos caracteres só pode ser definido uma vez, o que significa que um mesmo código digital só pode reenviar para uma definição de caracteres e vice-versa.

2.4. Primeiros recursos de base

2.4.1. Jogo de caracteres e tipos de letra

O primeiro recurso informático a elaborar para uma língua que não tem nenhum, é a definição de um jogo de caracteres a partir da lista dos caracteres retidos para a ortografia dessa mesma língua. Se o jogo de caracteres de que precisamos não contém nenhum caractere especial, podemos utilizar desde logo os programas informáticos existentes. Se, pelo contrário, são necessários caracteres especiais, neste caso, antes de se lançar na criação de um tipo particular, é necessário verificar se estes não se encontram já nos tipos de letra

7 Ver a página do sítio do consórcio Unicode: < <http://www.unicode.org/fr/charts/>>.

disponíveis na Internet, tais como *Lucida Sans Unicode*, *Gentium*, *Doulos SIL*⁸ ou nos tipos de letra africanos dos Programas integrados BPI⁹. Não havendo, será então necessário definir o seu próprio jogo de caracteres. É aconselhável verificar se este novo jogo pode também servir outras línguas da região, a fim de melhor rentabilizar os esforços empregues e garantir a possibilidade de produzir, posteriormente, textos multilingues sem haver necessidade de modificar os tipos de letra de uma língua para outra. Dois bons exemplos de definição de jogos de caracteres 8 bits (mas compatíveis com o Unicode) são fornecidos pelos Programas integrados BPI. Cada um destes jogos (*afrful* e *afrlin*) cobre várias línguas¹⁰.

Definido o jogo de caracteres, é necessário criar o tipo de letra correspondente. Hoje em dia, os peritos desaconselham a criação de novos tipos de letra 8 bits, devido às limitações acima evocadas. Será melhor dirigir-se a um especialista para criar um tipo de letra com os caracteres desejados e que respeitem a norma Unicode. Este recurso é tanto mais indispensável quando utilizado com um sistema de escrita não latino e ainda não representado no Unicode! O sítio <<http://www.freelang.com>> propõe um serviço gratuito de criação de tipos de letra a pedido dos seus visitantes. O tipo criado é livre, gratuito e partilhado no sítio. Várias ferramentas estão disponíveis na Internet para criar tipos de letra locais, nomeadamente nos sítios seguintes:

- 1) <http://scripts.sil.org/TypeDesignResources> (tipos de letra prontos a utilizar e ferramentas para criar outros);
- 2) <http://scripts.sil.org/SILFontList> (tipos de letra prontos a utilizar para descarregar);
- 3) http://scripts.sil.org/SILEncore_Glyphs (para visualizar os glifos dos tipos de letra).

8 Ver o sítio da SIL: <<http://scripts.sil.org/SILFontList>>.

9 Ver o sítio dos BPI: <<http://www.progiciels-bpi.ca/tcao/apercu.html>>.

10 Ver também o sítio burkinabe: <<http://www.abcburkina.net/sedelan/index.htm>> para as línguas do Burkina Faso.

2.4.2. Teclados virtuais

Dispondo já de um tipo de letra com os caracteres correspondentes ao jogo de caracteres que escolhemos, podemos instalá-lo no computador, seguindo o procedimento de instalação dos tipos de letra relativo ao sistema operativo utilizado. Depois é fácil utilizar estes tipos de letra, dado que a maioria dos processadores de texto modernos possuem uma função que permite a inserção de caracteres especiais. Por exemplo, no processador de texto do Word, abre-se uma janela de glifos e é só seleccionar o caractere desejado para o inserir no texto. Infelizmente, esta maneira de proceder é fastidiosa e atrasa consideravelmente a escrita do texto numa língua menos favorecida. É então indispensável elaborar combinatórias de teclas mais práticas que permitam aceder rapidamente aos caracteres especiais de determinado tipo de letra.

É interessante reparar que as línguas dominantes dos países industrializados têm um teclado especialmente criado para cada uma delas, até para certas variedades regionais de uma mesma língua. Por exemplo, o teclado francês do Quebec não é mesmo do francês de França. Não só todos estes teclados se distinguem pela posição das letras A Z e Q W mas, ainda, apresentam uma grande liberdade na disposição dos caracteres de pontuação e, naturalmente, dos caracteres especiais (diacríticos, caracteres acentuados, símbolos monetários, etc.). Em África, nenhuma das grandes línguas africanas (*kiswahili*, *zulu*, *hausa*, *fulfulde*, etc.) possui o seu próprio teclado. Assim, no caso das pequenas línguas menos favorecidas, não se espera dispor nos próximos tempos de um teclado físico. A solução mais simples será, então, a de elaborar um teclado virtual.

Baseando-se no teclado físico utilizado na região onde se fala a língua menos favorecida em questão, começamos por criar um quadro de combinatórias de teclas fáceis de memorizar. Por exemplo, decidimos que é preciso escrever a sequência de teclas «<» + «O» (menos as aspas e o +) para obter um «o aberto minúsculo», ou seja, «◊» e as teclas «Caps Lock » + «>» + «O» para obter a maiúscula correspondente «◊». Consideremos uma sequência com as seguintes combinatórias:

«<»+«O» = «◊»

«Caps Lock» + «>» + «O» = «◊»

Estas combinatórias só serão ergonómicas se os símbolos «<» e «>» estiverem sobrepostos numa mesma tecla, com o segundo situado por cima do primeiro. Serão então utilizados como teclas mortas. Assim, para poder utilizá-las novamente com os símbolos anteriores, basta carregar na «barra de espaço» depois de as ter accionado. É agora necessário acrescentar as seguintes linhas:

«<» + «espaço» = «<»

«Caps Lock» + «>» + «espaço» = «>»

Os caracteres «<» e «>» podem então ser utilizados como modificadores gerais para obter outros caracteres especiais em minúsculas e maiúsculas. O caso complica-se quando devemos acrescentar um ou dois diacríticos num dos caracteres já alterados, por exemplo, um acento agudo para indicar um tom alto e um til para indicar uma nasalação. Será então necessário assegurar-se que o programa informático utilizado para interpretar a sequência suporte uma sequência de comandos deste tipo: «acento agudo» + «til» + «modificador» + «caractere de base» = «caractere modificado com acentuação nasal». Actualmente, um dos melhores programas informáticos para a criação de teclados virtuais é o Keyman ^(TM) da empresa Travulsoft. Vários teclados virtuais para as línguas africanas realizados com este programa estão disponíveis no seguinte endereço: <http://scripts.sil.org/SILKeyboards>.

Basta instalar o Keyman no seu computador para gerar os teclados e tipos de letra correspondentes. É possível instalar vários teclados e passar de um para o outro através de uma simples combinação de teclas tal como «Alt+K». É só preciso assegurar-se que a combinação de teclas escolhida não é utilizada para outras funções por um dos programas informáticos instalados no computador.

2.4.3. Programas informáticos de tratamento de corpora assistido por computador

A partir do momento em que se utilizam caracteres especiais, é necessário pensar nas ferramentas de base que permitirão efectuar triagens alfabéticas, conversões de códigos de caracteres, alinhamento de textos, etc. Estas ferramentas revelar-se-ão muito úteis para a manipulação e processamento de textos a produzir na língua menos favorecida, tendo em vista a sua visualização na Rede. As ferramentas mais eficazes em termos de tratamento das línguas são os Programas integrados BPI abaixo citados.

O quadro 1 apresenta sequencialmente os cinco programas integrados para tratamento de *corpora* assistido por computador: **Alibi**, **Concorde**, **Recode**, **Ventile**, **Vocable**¹¹.

<i>Programas integrados</i>	QUADRO 1 – Descrição dos programas integrados para tratamento de corpora assistido por computador
Recode	Recode é um programa integrado de conversão automática de jogos de caracteres. Permite tratar e recodificar cerca de 175 jogos de caracteres diferentes e uma dúzia de <i>superfícies</i> de ficheiros. Mediante os pares de jogos de caracteres indicados no comando de chamada (de um jogo de partida para um jogo de chegada), Recode efectua uma recodificação do ou dos ficheiros de entrada. Cada jogo de caracteres pode ser convertido para a maioria dos outros 174, possibilitando assim vários milhões de conversões diferentes..
Ventile	Ventile é um programa integrado de produção de estatísticas textuais. Em matéria de elementos textuais, Ventile permite contabilizar o número de parágrafos, frases, palavras e caracteres por ficheiro. Em termos de estatísticas textuais, Ventile produz as medidas de frequência absolutas, três medidas de tendência central (o modo, a mediana e a média aritmética) e cinco medidas de dispersão (o mínimo, o máximo,

11 Fonte: <http://www.progiciels-bpi.ca/tcao/apercu.html>.

a diferença quartil, a diferença média e a diferença tipo). Os resultados estatísticos são visualizados digitalmente sob a forma de um quadro e, graficamente, sob a forma de um histograma.

Vocable Vocable é um programa integrado de produção de listas de vocabulário. Esta ferramenta permite explorar textos e fabricar a lista dos vocábulos actualizados nos textos. Estas listas podem ser organizadas pela ordem alfabética habitual (da esquerda para a direita), por ordem alfabética invertida (da direita para a esquerda) ou por ordem de frequência decrescente. Cada um dos vocábulos pode ser acompanhado da frequência das suas ocorrências e da lista de todas as referências textuais para cada ocorrência.

Concorde Concorde é um programa integrado de produção de concordâncias de palavras em contexto. Esta ferramenta permite explorar ficheiros textuais e produzir uma lista alfabética de elementos-vedetas com o respectivo contexto imediato. O elemento vedeta de uma concordância poderá ser baseado nas palavras, nos seus respectivos contextos frásicos, ou nos caracteres nos seus respectivos contextos de palavras. As palavras-vedetas podem assim ser organizadas pela ordem alfabética habitual ou invertida.

Alibi Alibi é um programa integrado de alinhamento bi-textual (ou bilingue). Esta ferramenta permite explorar paralelamente dois ficheiros e alinhar automaticamente os componentes textuais provenientes desse par de ficheiros. Como componentes textuais, Alibi pode alinhar parágrafos, frases ou palavras desses dois textos.

A realização de tais programas informáticos não está ao alcance de qualquer um. Com efeito, revelam-se ser muito eficazes e têm a vantagem de funcionar bem com os sistemas de exploração mais conhecidos e utilizados (Windows e Linux). Associados aos jogos de caracteres, aos tipos de letra e aos teclados virtuais, os programas informáticos em questão constituem recursos de base para o tratamento de uma língua menos favorecida.

2.5. Localização dos programas informáticos

Entende-se por localização de programas informáticos a actividade que consiste em pegar num programa informático concebido em determinada língua e num país estrangeiro e a adaptá-lo à língua e cultura de um outro país. Qualquer programa informático é composto por duas partes: o código informático, e a interface homem-máquina. Regra geral, o código informático permanece inalterável, salvo em casos excepcionais, tal como a necessidade de modificar o sentido da escrita ou a ordenação alfabética, quando os programas ainda não eram concebidos para serem localizáveis. Hoje em dia, a maioria dos programas informáticos podem ser localizados e, por conseguinte, a interface homem-máquina é de fácil identificação e acesso, graças a um ficheiro previsto para o efeito, ou através de ferramentas apropriadas.

Dizem respeito à localização de programas informáticos os menus (de desenrolar, fixos, flutuantes ou contextuais), as caixas de diálogo, as notificações, as informações de ajuda em linha, os botões e comandos de navegação, os grafismos (logótipos, bandeiras, brasões, ilustrações, aparência do sítio, *ouskin* em inglês e design), as unidades de medida (distância, pés, volume, tempo, moeda, densidade, etc. a converter em função das normas regionais), os tutoriais e a documentação (manuais de instalação, de utilização e de referência). O objectivo da localização é, obviamente, permitir ao utilizador trabalhar num ambiente informático que lhe é linguística e culturalmente familiar e, como tal, fácil de manusear. Consequentemente, a sua língua torna-se uma língua de trabalho das tecnologias da informação e da comunicação. Assim, a localização de programas informáticos numa língua menos favorecida contribui para valorizar esta língua aos olhos dos utilizadores e, nomeadamente, dos seus próprios falantes.

Quando desejamos traduzir programas informáticos para uma língua menos favorecida, a não ser que sejamos proprietário, é aconselhável tentar localizar programas que não têm direitos de autor, como os que são distribuídos com a licença GPL (*General Public Licence*)¹². A distribuição GNU /Linux (Ubuntu Linux¹³, por exemplo, para os países

12 Para mais informações sobre as licenças dos programas informáticos livres, ver <http://www.gnu.org/licenses/license-list.fr.html>.

13 Ver <http://www.ubuntu.com>.

africanos) proporciona ao mundo inteiro uma plataforma colaborativa em linha chamada *Rosetta*¹⁴, que permite a qualquer pessoa escolher um programa informático livre e traduzi-lo para a sua língua. Um conjunto de ferramentas de gestão de projectos é fornecido sob o nome de *Launchpad*¹⁵ e constitui uma ajuda preciosa para o utilizador que deseje lançar-se na localização de um programa informático. Tal dispositivo, de acesso gratuito, disponível para todas as línguas do mundo é, verdadeiramente, uma iniciativa extremamente rara e generosa, um verdadeiro achado para as línguas menos favorecidas que podem tirar daí partido para aceder mais facilmente ao estatuto de língua de trabalho no ciberespaço. No entanto, a localização de programas informáticos não é uma etapa indispensável para aceder ao ciberespaço. Uma língua menos favorecida em recursos pode ter acesso aos mesmos sem, contudo, ter arquitectado um projecto de localização. Desde que a língua disponha dos primeiros recursos de base (cf. ponto 2.4.), poder-se-á começar a elaboração dos conteúdos culturais e linguísticos para o ciberespaço.

14 Ver <https://launchpad.net/rosetta>.

15 Ver <http://launchpad.net>.



3. Elaboração dos conteúdos culturais

3.1. Conteúdos culturais

Os conteúdos culturais são de natureza textual, sonora ou iconográfica. Todos são passíveis de serem digitalizados e transferidos para a Rede, ou divulgados no ciberespaço de forma a serem aí partilhados.

3.1.1. O texto

Para além do *corpus* utilizado para descrever a língua e criar as primeiras obras de referência, é conveniente recolher ou criar ainda mais textos, de todos os géneros, desde os pequenos anúncios aos relatos de epopeias lendárias, às traduções de obras literárias ou, ainda, aos textos de imprensa. Todos estes textos serão em seguida objecto de diferentes tratamentos informáticos, graças aos programas informáticos de tratamento de *corpora* assistido por computador. Seria interessante divulgar os resultados assim obtidos na Rede, dado que poderiam contribuir para melhorar o próprio estudo da língua.

3.1.2. O som

Os documentos sonoros podem ser constituídos por gravações de listas de palavras ou de frases “vocalizadas”, de músicas, recitais, narrações de contos e lendas, entrevistas, reportagens mas, também, por música instrumental, sons metálicos, sons da vida e da natureza que sejam testemunhas do ambiente cultural local. A digitalização de todos estes documentos requer um material de gravação de qualidade profissional, o que, aos olhos de quem é oriundo dos países em vias de desenvolvimento e fala uma língua menos favorecida, pode ser muito dispendioso. A criação musical, em particular, pode implicar

a aquisição de aparelhos mais sofisticados e caros. O mesmo acontece para as reportagens da natureza, que captam o canto dos pássaros e os gritos de insectos nocturnos, com a preocupação de estudar o ambiente local e a ecologia natural. Seria suficiente que estes documentos sobre o ambiente fossem acompanhados por comentários na língua local, para que esta fosse valorizada.

Todos os ficheiros áudio assim obtidos deverão ser filtrados para serem limpos dos sopros e ruídos parasitas. Impõe-se, de seguida, um longo trabalho em estúdio para recortar e seccionar cada parte da gravação sonora, a fim de constituir ficheiros mais leves, mais facilmente transportáveis. Actualmente, é possível divulgar documentos sonoros na Internet e no ciberespaço, como se faz para uma estação de rádio ou para os filmes. No entanto, sem imagens, estes documentos sonoros, por mais interessantes que sejam, assemelham-se à televisão a preto e branco na era da cor.

3.1.3. A imagem

Os documentos iconográficos reúnem os desenhos, os logótipos, as fotos e o vídeo. Se os programas informáticos de desenho são de uso corrente, os aparelhos fotográficos e as câmaras de qualidade são um tanto ou quanto mais caros. Todavia, eles são essenciais para filmar uma dança, uma cerimónia, uma cena da vida corrente, uma vista panorâmica, uma reportagem, etc. Todas estas imagens podem ser montadas e retocadas para melhorar a sua qualidade. Os ficheiros imagem são hoje digitalizados no momento da sua criação. É no entanto necessário assegurar-se da boa qualidade da imagem, dado que estes tratamentos requerem várias cópias e isto pode, eventualmente, influenciar a qualidade da imagem para difusão.

Sempre que possível, é preferível associar som, imagens e texto na produção de documentos a serem divulgados na Internet. Mas, se uma língua pouco favorecida não for ainda uma língua escrita, é sempre possível criar documentos sonoros e iconográficos, à semelhança de um filme e/ou fotos comentadas, e de as divulgar na Rede.

3.2. Preparar as vias de acesso ao ciberespaço

3.2.1. O sítio Internet

O sítio Internet é a melhor maneira de afirmar a presença de uma língua no ciberespaço, dado que o podemos declinar infinitamente, desde a página pessoal até ao portal multimédia que propõe os serviços mais variados e diferentes. Antes de se lançar na criação de um sítio, convém considerar as seguintes observações:

Um sítio escrito numa língua menos favorecida só pode ser consultado pelos falantes dessa mesma língua. Isso torna-se difícil se os falantes não souberem ler nem escrever na sua própria língua. Aqui, o sítio pode servir de incentivo para os encorajar. Quando se acaba de elaborar uma ortografia para uma língua menos favorecida, é preferível conceber a criação de um sítio bilingue, utilizando simultaneamente a língua menos favorecida em questão e uma língua de maior alcance falada na região¹⁶. Neste caso, é importante zelar pela presença da língua menos favorecida nos botões e comandos de navegação, bem como em todas as mensagens destinadas ao visitante do sítio, a fim de evitar que ela se torne um mero objecto de curiosidade e perca todo o seu estatuto de língua de trabalho na Rede.

Os botões de navegação, os menus de comandos, os títulos, os logótipos, os textos decorativos, as animações e os estilos de apresentação do sítio são elementos que devem ser preparados com antecedência, com ajuda de programas informáticos especializados como *Button Studio*TM. Os ficheiros de texto, destinados a serem visualizados na Rede devem ser escritos em linguagem HTML (*HyperText Markup Language*). É possível escrever uma página na Rede com um simples programa de tratamento de texto como o *Open Office* (gratuito e de livre difusão), capaz de gravar um ficheiro em formato HTML. Existem também várias ferramentas mais adaptadas à criação de sítios, das mais simples e, muitas vezes, gratuitas como *Netscape*TM *Composer*,

16 Ver, por exemplo, o sítio da Associação YSB SAHNGO: <<http://sango.free.fr>>.

às mais sofisticadas e pagas como a ferramenta *Dreamweaver*TM. Para começar, é possível informar-se e formar-se no seguinte endereço: <http://www.unice.fr/urfist/ResInternet.html>.

A composição da página do sítio é uma questão de gosto e de cultura.

3.2.2. O correio electrónico e as mensagens instantâneas

É a ferramenta de comunicação mais utilizada na Internet. Mas, por razões de segurança, é muito difícil medir a taxa de utilização de uma ou de outra língua nas trocas mundiais por correio electrónico. A utilização de uma língua menos favorecida neste serviço passará, pois, despercebida. Pelo contrário, a localização de, pelos menos, um programa de mensagens e de mensagens instantâneas constitui um verdadeiro contributo para o equipamento da língua menos favorecida.

3.2.3. A telefonia IP e a telefonia móvel

A telefonia IP (*Internet Protocol*) permite comunicar por voz à distância, em tempo real, via Internet. Esta tecnologia pode portanto ser livremente utilizada pelos falantes de uma língua menos favorecida. Basta descarregar um programa adequado como o *Skype*TM, o mais conhecido neste momento, ou ainda o Google Talk, Voipbuster, Messenger, etc. e instalá-lo no computador. Somente a localização da interface utilizador do programa poderá constituir um verdadeiro recurso para a língua menos favorecida. O mesmo acontece com a telefonia móvel, cuja única maneira de contribuir para o enriquecimento da língua menos favorecida seria localizar a interface utilizador dos telefones, como é o caso na África do Sul.

3.2.4. Os fóruns e as listas de difusão

A instalação de um fórum ou de uma lista de difusão num sítio Internet permite à comunidade de falantes de uma língua menos favorecida dialogar entre si, por escrito. Sob reserva do ambiente do programa informático suportar os caracteres especiais necessários,

esta tecnologia poder-se-á revelar um excelente meio para dar vida à língua menos favorecida no ciberespaço. Com efeito, sem uma comunidade que anime e active as trocas nessa língua, esta corre o risco de ser apenas um elemento decorativo e inerte no ciberespaço. Obviamente, não será este o objectivo de todos os esforços realizados para levar uma língua menos favorecida para o ciberespaço. Vários portais comerciais e não comerciais oferecem a possibilidade de criar gratuitamente fóruns comunitários. Seguem-se aqui alguns dos mais conhecidos:

— <http://www.yahogroups.com> ;

— <http://www.google.com> ;

— <http://www.ubuntu.com> ;

— <http://www.free.fr> ;

— <http://lists.kabissa.org/mailman/options/a12n-entraide/mdkidiri%40free.fr>.



4. Desenvolver a comunidade dos utilizadores

4.1. Criação de uma comunidade de utilizadores

A existência de uma comunidade de utilizadores, instruída na sua própria língua, é absolutamente necessária para a sobrevivência dessa língua quando ela aceder ao ciberespaço. Com efeito, trabalhar no sentido de se conduzir as línguas menos favorecidas a ocupar um lugar no ciberespaço, não é de todo transformá-las em peças de mobiliário ou em figurantes mas, muito pelo contrário, é fazer com que elas aí existam e sejam produtivas para os seus locutores, os quais devem ser os primeiros interessados no domínio e promoção das suas línguas no ciberespaço. Torna-se assim primordial ensinar aos falantes de uma língua menos favorecida, nomeadamente aos jovens, a ler e a escrever nessa língua. O incentivo à informática e aos programas de burótica e comunicação deveria ser acompanhado pelo ensino da língua menos favorecida, por exemplo, no âmbito de actividades associativas, se não for na escola. As associações locais estão geralmente mais aptas a organizar formações deste tipo do que propriamente as escolas, que dependem bastante do sistema educativo nacional. Infelizmente, na maioria das vezes, as associações locais dispõem de poucos meios e não beneficiam de outros apoios, sobretudo quando se trata de uma língua menos favorecida. É óbvio que qualquer apoio prestado a associações locais empenhadas no desenvolvimento das línguas relacionadas com as TIC, contribuirá para a promoção da diversidade linguística e cultural no ciberespaço.

4.1.1. *Regeneração de uma comunidade*

Quando uma língua menos favorecida conta com poucos falantes, é indispensável consolidar o número desses falantes, ensinando essa língua aos mais jovens através de todos os meios disponíveis: actividades culturais, centros de leitura e de cultura, actividades

associativas e, se possível, na escola. Trata-se efectivamente de recriar um ambiente vital para a língua. Quanto mais a língua for ensinada e valorizada, mais ela produzirá conteúdo cultural que se poderá divulgar num sítio Internet. Devemos no entanto estar cientes do trabalho enorme que tal objectivo representa, bem como do investimento considerável que requer, tanto a nível de recursos humanos como financeiros, para além de uma determinação sem falhas a longo prazo. As associações culturais podem ser bastante activas localmente, no intuito de reabilitar uma língua em perigo e manter o interesse crescente a seu favor, nomeadamente junto dos decisores a nível político e administrativo.

4.1.2. Consolidação de uma comunidade

Tomemos o exemplo de uma língua menos favorecida mas que dispõe de uma população de falantes considerável. A educação das pessoas na sua própria língua é uma iniciativa ambiciosa, que exige a mobilização de vários recursos humanos e financeiros. Neste caso, o acesso da língua ao ciberespaço é um dos meios para mobilizar e dinamizar a comunidade de utilizadores dessa língua. A consolidação desta comunidade não poderá ser uma actividade anterior ao acesso da língua ao ciberespaço, mas uma actividade permanente que opera no sentido de dar vida à língua nesse espaço. Aqui, mais do que no caso das línguas em vias de extinção, que não desfrutam de um número tão importante de falantes, será necessário implicar quem decide, política e administrativamente, nos processos de valorização da língua menos favorecida. Sem isto, os esforços das associações não serão suficientes para garantir o seu desenvolvimento, por falta de meios a longo prazo.

4.2. Aspectos políticos e socioculturais

4.2.1. Argumentos para convencer

Como foi acima explicado, a acção de uma associação para o desenvolvimento de uma comunidade de utilizadores de uma língua menos favorecida, tem os seus limites nos meios financeiros que poderá mobilizar. Será então essencial sensibilizar os responsáveis

políticos, a todos os níveis, para todas as vantagens de apoiar a instrumentalização das línguas menos favorecidas e o reforço das capacidades dos seus falantes. Com o apoio das administrações locais, é mais fácil conceber o ensino das línguas na escola, nomeadamente na escola primária, onde a língua local é a mais indicada para servir de língua nas primeiras aquisições do saber pela criança. Um ensino bilingue só poderá ser bem sucedido, mesmo a nível local, se estiver inscrito no âmbito jurídico da Educação Nacional. Várias experiências de ensino bilingue implicando as línguas africanas no Burkina Faso, no Senegal, no Mali, no Burundi e no Ruanda demonstraram que as crianças aprendem muito melhor nas suas línguas maternas e obtêm melhores resultados escolares. A longo prazo, tornam-se cidadãos mais instruídos. Ora, um homem instruído é capaz de ter uma visão global e uma compreensão aprofundada dos problemas complexos do mundo moderno, e de contribuir com soluções duráveis na sua esfera de vida e actividades. Será também uma pessoa capaz de orientar um projecto socioeconómico tanto para ele como para a sua família, aldeia e região. A educação multilingue que implica a língua materna é a chave do desenvolvimento económico.

Para além das vantagens educativas, uma política que favoreça a diversidade linguística poderá criar empregos nas áreas da tradução, redacção de manuais, edição ou ainda animação. A diversidade linguística é, portanto, uma aposta económica e o desenvolvimento das línguas menos favorecidas contribui para essa diversidade. Quando uma língua menos favorecida diz respeito a uma região inteira de um país, é toda a política linguística nacional que está implicada no seu desenvolvimento. Este deverá então inscrever-se num plano nacional de ordenamento político, pondo em prática uma política linguística nacional afirmada em favor da diversidade linguística. Além disso, tal política deverá respeitar os direitos linguísticos que são parte integrante dos direitos fundamentais do homem.

4.2.2. Instrumentos de referência

Para além das autoridades políticas nacionais, recomenda-se também o pedido de apoio junto de organizações internacionais, que agem para a promoção e valorização da diversidade linguística no mundo, o que implica, pelo menos, uma atitude protectora para com os

esforços das associações locais no desenvolvimento das línguas menos favorecidas. Devemos sublinhar que, nos últimos quarenta anos, as organizações internacionais organizaram vários encontros sobre as línguas das populações locais, muitas vezes qualificadas de «línguas nacionais» por oposição às línguas europeias herdadas da colonização, nomeadamente em África e no continente americano. Estes encontros permitiram a elaboração e adopção de vários textos, chamados « instrumentos internacionais », nos quais nos podemos apoiar para desenvolver acções concretas. Entre as manifestações organizadas pelos Estados e Organizações internacionais, Amadou Touré¹⁷ sublinha, entre outras, as seguintes para África:

1. Reunião da UNESCO sobre o uso das línguas vernáculas no ensino (Monografias sobre a educação de base) – Paris, 1953;
2. Conferência Regional da UNESCO sobre a planificação e a organização dos programas de alfabetização em África – Abidjan, Março de 1964;
3. Congresso da Sociedade Africana de Linguística (SLAO) – Accra, 1965;
4. Reunião de peritos organizada pela UNESCO para a unificação dos alfabetos das línguas nacionais: *fulfulde, hawsa, kanuri, mandingue, songhay-zarma, tamasheq* – Bamako, 28 de Fevereiro – 05 de Março de 1966 ;
5. Conferência intergovernamental sobre as políticas culturais em África (UNESCO- OUA) – Accra, 1975;
6. Reunião organizada pela Agência de Cooperação Cultural e Técnica, sobre a promoção das línguas nacionais – Yaounde, 1977;
7. Conferência internacional da ACCT sobre «a investigação linguística, o uso e o ensino das línguas em África: os meios para reforçar a cooperação entre os Estados» – Yaounde, 1983;
8. Reunião da UNESCO sobre as estratégias de promoção das línguas africanas – Conakry, 1984;
9. Reunião de peritos da UNESCO sobre a promoção das línguas africanas como instrumentos de cultura e educação permanente – Yaounde;

17 Documento de trabalho para os colóquios da Academia Africana das Línguas, 2006.

10. Reunião sobre o projecto de Carta de acção para a promoção e utilização das línguas africanas na educação – Accra, Agosto de 1996;
11. Conferência intergovernamental sobre as políticas linguísticas em África (UNESCO-OUA-ACCT) – Harare, 1997;
12. Consulta africana sobre a criação da Academia Africana das Línguas – Bamako, 25-27 de Maio de 2001;
13. Conferência sobre o multilinguismo para a diversidade cultural e participação de todos no ciberespaço – Bamako, Maio de 2005;
14. 33ª Sessão da Conferência Geral da UNESCO que adoptou o projecto de resolução iniciado pela ACALAN, apresentado pelo Benin e relativo à organização de cinco colóquios regionais sobre as línguas transfronteiriças e as línguas de menor difusão em África – Paris, Outubro de 2005.

Vários grandes encontros permitiram a elaboração de instrumentos normativos e de textos regulamentares, tais como:

1. Carta da OUA (1963);
2. Manifesto Cultural Pan-africano de Argel (1969);
3. Carta Cultural de África – OUA (1976, Port-Louis);
4. Resultados da 1ª Conferência dos Ministros Africanos da Cultura da OUA (Port-Louis, 1986);
5. Plano de acção de Lagos da OUA para o desenvolvimento económico da África (1980);
6. Plano de Acção Linguística para África (Adis-Abeba, 21-25 de Julho de 1986);
7. Plano decenal da OUA sobre as línguas e tradições orais (1987);
8. Tratado de Abuja criando a Comunidade Económica Africana (1991);
9. Plano regional de recolha das tradições orais na África Austral (Harare, 1993);

10. Plano de ordenamento linguístico do espaço francófono 1990-2000 (ACCT 1993);
11. Programa de acção do decénio da educação em África (Harare, 1999);
12. Tratado fundador da União Africana (Lomé, 2000);
13. Resolução da 31ª sessão da Conferência Geral da UNESCO sobre a Academia Africana das Línguas (2001);
14. Decisão CM/Dez.613 (LXXIV) da 37ª Cimeira dos Chefes de Estado e de Governo da OUA, fundadora da Academia Africana das Línguas (Julho de 2001);
15. Declaração Universal sobre a Diversidade Cultural, adoptada pela 31ª sessão da Conferência Geral da UNESCO (2001);
16. Recomendações sobre a promoção e uso do multilinguismo e o acesso universal ao ciberespaço, adoptadas pela 32ª sessão da Conferência Geral da UNESCO (Outubro de 2003);
17. Textos fundadores da Cimeira Mundial sobre a Sociedade da Informação (Genebra, 2003 e Tunes, 2005);
18. Decisão AU/Dez.92 (VI) da sexta sessão ordinária da Conferência da União Africana relativa a um segundo decénio da educação para a África (2006-2015) e ao quadro de acção do segundo decénio (Doc. EX.CL/224-VIII Rev.2) (Khartoum, 2006);
19. Decisão AU/Dez.94(VI) da sexta sessão ordinária da Conferência da União Africana sobre a Carta revista do renascimento cultural africano (Khartoum 2006);
20. Decisão AU/Dez.95(VI) da sexta sessão ordinária da Conferência da União Africana sobre os estatutos da Academia Africana das Línguas (Khartoum, 2006);
21. Decisão AU/Dez.96(VI) da sexta sessão ordinária da Conferência da União Africana sobre a ligação entre a cultura e a educação (Khartoum, 2006);
22. Decisão AU/Dez.98(VI) da sexta sessão ordinária da Conferência da União Africana relativa à proclamação de 2006 como Ano das Línguas Africanas (Khartoum 2006).

Várias instituições foram criadas, entre outras:

- O ICA (Instituto Cultural Africano) – Dakar, Senegal;
- O CELHTO (Centro de Estudos Linguísticos e Históricos de Tradição Oral) – Niamey, Nigéria;

- O CERDOTOLA (Centro Regional de Documentação sobre as Tradições Orais e as Línguas Africanas) – Yaounde, Camarões;
- O EACROTANAL (*East African Center of Research on Oral Traditional and National Languages*) – Zanzibar, Tanzânia;
- O CICIBA (Centro Internacional das Civilizações bantu) – Libreville, Gabão;
- O BASE (Repartição Africana das Ciências da Educação), hoje IPED (Instituto Pan-africano da Educação para o Desenvolvimento) – Kinshasa, República Democrática do Congo.

A maioria destas instituições, devemos reconhecê-lo, ficaram aquém das expectativas, por falta de meios financeiros para funcionar plenamente. Assim, não será fácil poder contar com o seu apoio para levar a cabo uma acção local de valorização de uma língua menos favorecida. No entanto, com um pouco de perseverança e muita boa vontade, poderemos conseguir fazer delas umas aliadas nesta luta, o que não é negligenciável! Com efeito, um inventário dos instrumentos normativos existentes e um exame dos relatórios dos diferentes encontros organizados durante este longo espaço de tempo, demonstram que todos os países de África tiveram, a um dado momento e a diferentes graus, iniciativas para a promoção das suas línguas nacionais. É por esta razão, que será conveniente provocar novos estímulos, no sentido de os levar a implicarem-se ainda mais em acções concretas a favor da diversidade linguística. Apenas uma vontade política forte e permanente levará ao desbloqueio dos fundos necessários para organizar a educação multilingue e o desenvolvimento e uso das línguas na vida da nação, para que sejam criados bens culturais, económicos e sociais que proporcionem o bem-estar de todos.

5. Conclusão

No intuito de promover e valorizar a diversidade linguística e cultural no ciberespaço, será conveniente ajudar as línguas menos favorecidas a aceder a este espaço. Se é possível fazê-lo com uma pequena língua oral não escrita e ameaçada de extinção, tanto mais fácil será com todas as línguas menos favorecidas em melhor situação.

A primeira etapa consiste em empreender os estudos necessários para a elaboração de recursos linguísticos indispensáveis, ou seja: uma lista dos fonemas, um alfabeto e uma ortografia, uma gramática, um dicionário e uma recolha de textos.

A segunda etapa deve ser dedicada ao trabalho de tratamento informático da língua, a fim de identificar ou elaborar os recursos informáticos compatíveis, isto é: um jogo de caracteres, pelos menos um tipo de letra, um teclado virtual, programas informáticos para tratamento de *corpora*, os quais poderão também ser utilizados para refinar a análise da língua e melhorar os recursos linguísticos.

A terceira etapa consiste em elaborar os recursos culturais e acondicioná-los para os tornar partilháveis no ciberespaço. Trata-se aqui de gravar e digitalizar o maior número possível de dados textuais, sonoros e iconográficos e de os preparar para serem visualizados num sítio Internet. Também será necessário criar os componentes de um sítio Internet tais como os menus, as barras de navegação, os títulos e outros textos de comunicação homem-máquina. Em certos casos, a localização de programas informáticos será necessária para contribuir para a valorização da língua enquanto língua de trabalho e para dotá-la de recursos informáticos complementares.

Finalmente, será conveniente aprender a criar um sítio Internet na língua menos favorecida, eventualmente em complementaridade com outra língua de maior expansão. Encontramos disponíveis na Internet todas as ferramentas necessárias para assegurar tal aprendizagem, bem como para criar um fórum e localizar programas informáticos

livres. Com um sítio, um fórum, uma lista de difusão, uma telefonia IP, música, imagens fixas e vídeos, a pequena língua está doravante instalada no ciberespaço. No entanto, para que ela aí sobreviva, será necessário criar uma comunidade de utilizadores capazes de se servirem dela intensivamente. Ajudar as associações locais a desenvolver tais comunidades, é contribuir para a promoção e valorização da diversidade das línguas e culturas no ciberespaço.

Referências

Consórcio Unicode: <http://www.unicode.org/fr/charts>

Criação de tipos de letra em linha: <http://www.freelang.com>

Plataforma de tradução e localização de programas informáticos livres: <https://launchpad.net/rosetta>

Lista de tipos de letras: <http://scripts.sil.org/SILFontList>

Programas integrados BPI:
<http://www.progiciels-bpi.ca/tcao/apercu.html>

SEDELAN (Serviço de Edição em Línguas Nacionais de Koudougou, Burkina Faso): <http://www.abcburkina.net/sedelan/index.htm>

Sítio Internet bilingue sangho-francês da Associação YSB SAHNGO para a promoção da língua sangho: <http://sango.free.fr>

Universidade de Nice – Recursos multimédia, formação para a criação de sítios Internet: <http://www.unice.fr/urfist/ResInternet.html>

Anexo 1: Jogos de caracteres AFRFUL¹⁸

Os jogos de caracteres africanos *afrful* permitem cobrir os alfabetos do bambara, do ewondo, do francês e do fulfulde. O jogo de caracteres *AFRFUL-102-BPI_OCIL* é um jogo de caracteres em versão codificada, utilizado para a visualização das seguintes línguas africanas no ecrã: o bambara, o ewondo, o fulfulde (ou peulh). Aliás, os nomes aceitáveis para designar este jogo de caracteres codificado são *afrful102bpiocil*, *bambara*, *ewondo*, *fulfulde* ou *bra*.

O jogo de caracteres *AFRFUL-103-BPI_OCIL* é um jogo de caracteres em versão transliterada, utilizado para codificar as seguintes línguas africanas no teclado: o bambara, o ewondo, o fulfulde (ou peulh). Aliás, os nomes aceitáveis para designar este jogo de caracteres transliterados são *afrful103bpiocil*, *tbambara*, *tewondo*, *tfulfulde* ou *tbra*. Para cada um destes 255 caracteres apresentados na lista abaixo indicada, segue-se o significado de cada coluna:

Dec	Indica o valor decimal do caractere
Oct	Indica o valor octal do caractere
Hex	Indica o valor hexadecimal do caractere
Mne	Fornece o código mnemotécnico de duas letras da norma <i>RFC 1345</i>
UCS2	Indica o valor UCS do caractere
Kbd	Fornece a convenção de codificação do teclado desse caractere
AFRFUL-102-BPI_OCIL	Dá o nome do caractere, se possível, em francês, conforme à norma <i>ISO 10646</i>

18 Fonte: Programas integrados BPI <<http://www.progiciels-bpi.ca/tcao/apercu.html>>.

Dec	Oct	Hex	Mne	UCS2	Kbd	AFRFUL-102-BPI_OCIL
0	000	00	NU	0000		nul nulo
1	001	01	SH	0001		soh início de cabeçalho
2	002	02	SX	0002		stx início de texto
3	003	03	EX	0003		etx fim de texto
4	004	04	ET	0004		eot fim de transmissão
5	005	05	EQ	0005		enq interrogação
6	006	06	AK	0006		ack aviso de recepção
7	007	07	BL	0007		bel campainha
8	010	08	BS	0008		bs espaço para trás
9	011	09	HT	0009		ht tabulador horizontal
10	012	0a	LF	000A		lf interlinha
11	013	0b	VT	000B		vt tabulador vertical
12	014	0c	FF	000C		ff nova página
13	015	0d	CR	000D		cr retorno do carreto
14	016	0e	SO	000E		so fim de modificador (shift out)
15	017	0f	SI	000F		si início de modificador (shift in)
16	020	10	DL	0010		dle escape de ligação de dados
17	021	11	D1	0011		dc1 controlo de dispositivo um
18	022	12	D2	0012		dc2 controlo de dispositivo dois
19	023	13	D3	0013		dc3 controlo de dispositivo três
20	024	14	D4	0014		dc4 controlo de dispositivo quatro
21	025	15	NK	0015		nak aviso de recepção negativo
22	026	16	SY	0016		syn inação síncrona
23	027	17	EB	0017		etb fim de transmissão de bloco
24	030	18	CN	0018		can cancelar
25	031	19	EM	0019		em fim de meio

Dec	Oct	Hex	Mne	UCS2	Kbd	AFRFUL-102-BPI_OCIL
26	032	1a	SB	001A		sub substituto
27	033	1b	EC	001B		esc sair
28	034	1c	FS	001C		is4 separador de ficheiro
29	035	1d	GS	001D		is3 separador de grupo
30	036	1e	RS	001E		is2 separador de registo
31	037	1f	US	001F		is1 separador de unidade
32	040	20	SP	0020	<>	espaço
33	041	21	!	0021	!	ponto de exclamação
34	042	22	«	0022	«	aspa
35	043	23	Nb	0023	#	cardinal
36	044	24	DO	0024	\$	dólar
37	045	25	%	0025	%	percentagem
38	046	26	&	0026	&	e comercial
39	047	27	‘	0027	‘	apóstrofe
40	050	28	(0028	(parêntese esquerdo
41	051	29)	0029)	parêntese direito
42	052	2a	*	002A	*	asterisco
43	053	2b	+	002B	+	sinal de adição
44	054	2c	,	002C	,	vírgula
45	055	2d	–	002D	–	hífen, sinal de subtracção
46	056	2e	.	002E	–	ponto
47	057	2f	/	002F	/	barra oblíqua
48	060	30	0	0030	0	dígito zero
49	061	31	1	0031	1	dígito um
50	062	32	2	0032	2	dígito dois
51	063	33	3	0033	3	dígito três
52	064	34	4	0034	4	dígito quatro
53	065	35	5	0035	5	dígito cinco
54	066	36	6	0036	6	dígito seis
55	067	37	7	0037	7	dígito sete
56	070	38	8	0038	8	dígito oito
57	071	39	9	0039	9	dígito nove

Dec	Oct	Hex	Mne	UCS2	Kbd	AFRFUL-102-BPI_OCIL
58	072	3a	:	003A	:	dois pontos
59	073	3b	;	003B	;	ponto e vírgula
60	074	3c	<	003C	<	menor
61	075	3d	=	003D	=	igual
62	076	3e	>	003E	>	maior
63	077	3f	?	003F	?	ponto de interrogação
64	100	40	At	0040	@	arroba
65	101	41	A	0041	A	letra maiúscula latina A
66	102	42	B	0042	B	letra maiúscula latina B
67	103	43	C	0043	C	letra maiúscula latina C
68	104	44	D	0044	D	letra maiúscula latina D
69	105	45	E	0045	E	letra maiúscula latina E
70	106	46	F	0046	F	letra maiúscula latina F
71	107	47	G	0047	G	letra maiúscula latina G
72	110	48	H	0048	H	letra maiúscula latina H
73	111	49	I	0049	I	letra maiúscula latina I
74	112	4a	J	004A	J	letra maiúscula latina J
75	113	4b	K	004B	K	letra maiúscula latina K
76	114	4c	L	004C	L	letra maiúscula latina L
77	115	4d	M	004D	M	letra maiúscula latina M
78	116	4e	N	004E	N	letra maiúscula latina N
79	117	4f	O	004F	O	letra maiúscula latina O
80	120	50	P	0050	P	letra maiúscula latina P
81	121	51	Q	0051	Q	letra maiúscula latina Q
82	122	52	R	0052	R	letra maiúscula latina R
83	123	53	S	0053	S	letra maiúscula latina S
84	124	54	T	0054	T	letra maiúscula latina T
85	125	55	U	0055	U	letra maiúscula latina U
86	126	56	V	0056	V	letra maiúscula latina V
87	127	57	W	0057	W	letra maiúscula latina W
88	130	58	X	0058	X	letra maiúscula latina X
89	131	59	Y	0059	Y	letra maiúscula latina Y
90	132	5a	Z	005A	Z	letra maiúscula latina Z

Dec	Oct	Hex	Mne	UCS2	Kbd	AFRFUL-102-BPI_OCIL
91	133	5b	<(005B	[parêntese recto esquerdo
92	134	5c	//	005C	\	barra reversa
93	135	5d)>	005D]	parêntese recto direito
94	136	5e	'>	005E	^	acento circunflexo
95	137	5f	_	005F	_	sublinhado
96	140	60	'!	0060	`	acento grave
97	141	61	a	0061	a	letra minúscula latina a
98	142	62	b	0062	b	letra minúscula latina b
99	143	63	c	0063	c	letra minúscula latina c
100	144	64	d	0064	d	letra minúscula latina d
101	145	65	e	0065	e	letra minúscula latina e
102	146	66	f	0066	f	letra minúscula latina f
103	147	67	g	0067	g	letra minúscula latina g
104	150	68	h	0068	h	letra minúscula latina h
105	151	69	i	0069	i	letra minúscula latina i
106	152	6a	j	006A	j	letra minúscula latina j
107	153	6b	k	006B	k	letra minúscula latina k
108	154	6c	l	006C	l	letra minúscula latina l
109	155	6d	m	006D	m	letra minúscula latina m
110	156	6e	n	006E	n	letra minúscula latina n
111	157	6f	o	006F	o	letra minúscula latina o
112	160	70	p	0070	p	letra minúscula latina p
113	161	71	q	0071	q	letra minúscula latina q
114	162	72	r	0072	r	letra minúscula latina r
115	163	73	s	0073	s	letra minúscula latina s
116	164	74	t	0074	t	letra minúscula latina t
117	165	75	u	0075	u	letra minúscula latina u
118	166	76	v	0076	v	letra minúscula latina v
119	167	77	w	0077	w	letra minúscula latina w
120	170	78	x	0078	x	letra minúscula latina x
121	171	79	y	0079	y	letra minúscula latina y
122	172	7a	z	007A	z	letra minúscula latina z
123	173	7b	(!	007B	{	chaveta esquerda

Dec	Oct	Hex	Mne	UCS2	Kbd	AFRFUL-102-BPI_OCIL
124	174	7c	!!	007C		barra vertical
125	175	7d	!)	007D	}	chaveta direita
126	176	7e	'?	007E	~	til
127	177	7f	DT	007F		del eliminar
128	200	80	PA	0080		pad chave aleatória
129	201	81	HO	0081		hop octeto superior predefinido
130	202	82	BH	0082		bph paragem autorizada aqui
131	203	83	NH	0083		nbh paragem proibida aqui
132	204	84	IN	0084		ind índice
133	205	85	NL	0085		nel parágrafo
134	206	86	SA	0086		ssa início de zona seleccionada
135	207	87	ES	0087		esa fim de zona seleccionada
136	210	88	HS	0088		hts fim de tabulador horizontal
137	211	89	HJ	0089		htj tabulador horizontal com justificação
138	212	8a	VS	008A		vts fim de tabulador vertical
139	213	8b	PD	008B		pld entrelinha parcial para baixo
140	214	8c	PU	008C		plu entrelinha parcial para cima
141	215	8d	RI	008D		ri índice invertido
142	216	8e	S2	008E		ss2 substituição única dois
143	217	8f	S3	008F		ss3 substituição única três
144	220	90	DC	0090		dcs sequência de comando de aparelho
145	221	91	P1	0091		pu1 uso privado um
146	222	92	P2	0092		pu2 uso privado dois

Dec	Oct	Hex	Mne	UCS2	Kbd	AFRFUL-102-BPI_OCIL
147	223	93	TS	0093		sts modo de transmissão activado
148	224	94	CC	0094		cch anulação do caractere precedente
149	225	95	MW	0095		mw mensagem a aguardar
150	226	96	SG	0096		sga início de zona protegida
151	227	97	EG	0097		ega fim de zona protegida
152	230	98	SS	0098		sos início de sequência
153	231	99	GC	0099		sgci introdutor de caractere gráfico único
154	232	9a	SC	009A		sci introdutor de caractere único
155	233	9b	CI	009B		csi introdutor de sequência de comando
156	234	9c	ST	009C		st fim de sequência
157	235	9d	OC	009D		osc comando de sistema de exploração
158	236	9e	PM	009E		pm mensagem privada
159	237	9f	AC	009F		apc comando de pacote de programas
160	240	a0	NS	00A0	/_	espaço inquebrável
161	241	a1		0181	B[letra maiús. latina B com parêntese recto esquerdo
162	242	a2		018A	D[letra maiús. latina D com parêntese recto esquerdo
163	243	a3		0194	G]	letra maiús. latina gama
164	244	a4		0198	K[letra maiús. latina K com parêntese recto esquerdo
165	245	a5		01A4	P[letra maiús. latina P com parêntese recto esquerdo

Dec	Oct	Hex	Mne	UCS2	Kbd	AFRFUL-102-BPI_OCIL
166	246	a6		01AC	T[letra maiús. latina T com parêntese recto esquerdo
167	247	a7		01A9	S]	letra maiús. latina ESH
168	250	a8		019D	N]	letra maiús. latina N anzol (enj)
169	251	a9		014A	N[letra maiús. latina ENG (Sami)
170	252	aa		0189	D]	letra maiús. latina D africano (edh)
171	253	ab	<<	00AB	<<	abertura de aspas angulares esquerda
172	254	ac		0191	F]	letra maiús. latina F anzol
173	255	ad	--	00AD	\-	traço condicional (hífen virtual)
174	256	ae		01B3	Y[letra maiús. latina Y com parêntese recto esquerdo
175	257	af		0193	G[letra maiús. latina G com parêntese recto esquerdo
176	260	b0	DG	00B0	_DG	símbolo graus
177	261	b1		0253	b[letra min. latina b com parêntese recto esquerdo
178	262	b2		0257	d[letra min. latina d com parêntese recto esquerdo
179	263	b3		0263	g]	letra min. latina gama
180	264	b4		0199	k[letra min. latina k com parêntese recto esquerdo
181	265	b5		01A5	p[letra min. latina p com parêntese recto esquerdo

Dec	Oct	Hex	Mne	UCS2	Kbd	AFRFUL-102-BPI_OCIL
182	266	b6		01AD	t[letra min. latina t com parêntese recto esquerdo
183	267	b7		0283	s]	letra min. latina esh
184	270	b8		0272	n]	letra min. latina n com anzol (enj)
185	271	b9		014B	n[letra min. latina eng (Sami)
186	272	ba		0256	d]	letra min. latina d com anzol retroflexo
187	273	bb	>>	00BB	>>	fecho de aspas angulares para a direita
188	274	bc		0192	f]	letra min. latina f anzol
189	275	bd		0294	?^	letra latina gutural
190	276	be		01B4	y[letra min. latina y parêntese recto esquerdo
191	277	bf		0260	g[letra min. latina g parêntese recto esquerdo
192	300	c0	A!	00C0	A`	letra maiús. latina A com acento grave
193	301	c1	A'	00C1	A\'	letra maiús. latina A com acento agudo
194	302	c2	A>	00C2	A^	letra maiús. latina A com acento circunflexo
195	303	c3	A?	00C3	A~	letra maiús. latina A com til
196	304	c4	A:	00C4	A»	letra maiús. latina A com trema
197	305	c5		018E	E<	letra maiús. latina E reflexo (schwa)
198	306	c6	AE	00C6	AE+	letra maiús. latina AE (ligação AE maiúscula)
199	307	c7	C,	00C7	C\,	letra maiús. latina C com cedilha

Dec	Oct	Hex	Mne	UCS2	Kbd	AFRFUL-102-BPI_OCIL
200	310	c8	E!	00C8	E`	letra maiús. latina E com acento grave
201	311	c9	E'	00C9	E\'	letra maiús. latina E com acento agudo
202	312	ca	E>	00CA	E^	letra maiús. latina E com acento circunflexo
203	313	cb	E:	00CB	E»	letra maiús. latina E com trema
204	314	cc	l!	00CC	l`	letra maiús. latina l com acento grave
205	315	cd	l'	00CD	l\'	letra maiús. latina l com acento agudo
206	316	ce	l>	00CE	l^	letra maiús. latina l com acento circunflexo
207	317	cf	l:	00CF	l»	letra maiús. latina l com trema
208	320	d0		0190	E[letra maiús. latina E aberto
209	321	d1	N?	00D1	N~	letra maiús. latina N com til
210	322	d2	O!	00D2	O`	letra maiús. latina O com acento grave
211	323	d3	O'	00D3	O\'	letra maiús. latina O com acento agudo
212	324	d4	O>	00D4	O^	letra maiús. latina O com acento circunflexo
213	325	d5	O?	00D5	O~	letra maiús. latina O com til
214	326	d6	O:	00D6	O»	letra maiús. latina O com trema
215	327	d7		0152	OE+	digrama ligado maiús. latim OE (ligação OE)
216	330	d8		0186	O[letra maiús. latina O aberto
217	331	d9	U!	00D9	U`	letra maiús. latina U com acento grave

Dec	Oct	Hex	Mne	UCS2	Kbd	AFRFUL-102-BPI_OCIL
218	332	da	U'	00DA	U\'	letra maiús. latina U com acento agudo
219	333	db	U>	00DB	U^	letra maiús. latina U com acento circunflexo
220	334	dc	U:	00DC	U»	letra maiús. latina U com trema
221	335	dd		018E +0301	E<'	letra maiús. latina E reflexo com acento agudo
222	336	de		0186 +0301	O['	letra maiús. latina O aberto com acento agudo
223	337	df		01B2	V[letra maiús. latina V com parêntese recto esquerdo
224	340	e0	a!	00E0	a`	letra maiús. latina a com acento grave
225	341	e1	a'	00E1	a\'	letra maiús. latina a com acento agudo
226	342	e2	a>	00E2	a^	letra maiús. latina a com acento circunflexo
227	343	e3	a?	00E3	a~	letra min. latina a com til
228	344	e4	a:	00E4	a”	letra min. latina a com trema
229	345	e5		0259	e<	letra min. latina schwa (e invertido)
230	346	e6	ae	00E6	ae+	letra min. latina ae
231	347	e7	c,	00E7	c\,	letra min. latina c com cedilha
232	350	e8	e!	00E8	e`	letra min. latina e com acento grave
233	351	e9	e'	00E9	e\'	letra min. latina e com acento agudo
234	352	ea	e>	00EA	e^	letra min. latina e com acento circunflexo

Dec	Oct	Hex	Mne	UCS2	Kbd	AFRFUL-102-BPI_OCIL
235	353	eb	e:	00EB	e»	letra min. latina e com trema
236	354	ec	i!	00EC	i`	letra min. latina i com acento grave
237	355	ed	i'	00ED	i´	letra min. latina i com acento agudo
238	356	ee	i>	00EE	i^	letra min. latina i com acento circunflexo
239	357	ef	i:	00EF	i“	letra min. latina i com trema
240	360	f0		025B	e[letra min. latina e aberto
241	361	f1	n?	00F1	n~	letra min. latina n com til
242	362	f2	o!	00F2	o`	letra min. latina o com acento grave
243	363	f3	o'	00F3	o´	letra min. latina o com acento agudo
244	364	f4	o>	00F4	o^	letra min. latina o com acento circunflexo
245	365	f5	o?	00F5	o~	letra min. latina o com til
246	366	f6	o:	00F6	o»	letra min. latina o com trema
247	367	f7		0153	oe+	digrama ligado min. latino oe (ligação œ)
248	370	f8		0254	o[letra min. latina o aberto
249	371	f9	u!	00F9	u`	letra min. latina u com acento grave
250	372	fa	u'	00FA	u´	letra min. latina u com acento agudo
251	373	fb	u>	00FB	u^	letra min. latina u com acento circunflexo
252	374	fc	u:	00FC	u»	letra min. latina u com trema

Dec	Oct	Hex	Mne	UCS2	Kbd	AFRFUL-102-BPI_OCIL
253	375	fd		0259 +0301	e<'	letra min. latina schwa com acento agudo
254	376	fe		0254 +0301	o['	letra min. latina o aberto com acento agudo
255	377	ff		028B	v[letra min. latina v circular

Anexo 2: Jogo de caracteres AFRLIN¹⁹

Os jogos de caracteres africanos **afrlin** permitem cobrir os alfabetos do francês, do lingala, do sangho e do wolof. O jogo de caracteres *AFRLIN-104-BPI_OCIL* é um jogo de caracteres em versão codificada, utilizado para a visualização das seguintes línguas africanas no ecrã: o lingala, o sahnho e o wolof. Aliás, os nomes aceitáveis para designar este jogo de caracteres codificados são *afrlin104bpiocil*, *lingala*, *sango*, *wolof* ou *lin*.

O jogo de caracteres *AFRLIN-105-BPI_OCIL* é um jogo de caracteres em versão transliterada, utilizado para codificar as seguintes línguas africanas no teclado: o lingala, o sahnho e o wolof. Aliás, os nomes aceitáveis para designar este jogo de caracteres transliterados são *afrlin105bpiocil*, *tlingala*, *tsango*, *twolof* ou *tlin*. Para cada um destes 255 caracteres apresentados na lista abaixo indicada, segue-se o significado de cada coluna:

Dec	Indica o valor decimal do caractere
Oct	Indica o valor octal do caractere
Hex	Indica o valor hexadecimal do caractere
Mne	Fornece o código mnemotécnico de duas letras da norma <i>RFC 1345</i>
UCS2	Indica o valor UCS do caractere
Kbd	Fornece a convenção de codificação do teclado desse caractere
AFRFUL-102-BPI_OCIL	Dá o nome do caractere, se possível, em francês, conforme à norma <i>ISO 10646</i>

19 Fonte: Programas integrados BPI <<http://www.progiciels-bpi.ca/tcao/apercu.html>>.

Dec	Oct	Hex	Mne	UCS2	Kbd	AFRFUL-102-BPI_OCIL
0	000	00	NU	0000		nul nulo
1	001	01	SH	0001		soh início de cabeçalho
2	002	02	SX	0002		stx início de texto
3	003	03	EX	0003		etx fim de texto
4	004	04	ET	0004		eot fim de transmissão
5	005	05	EQ	0005		enq interrogação
6	006	06	AK	0006		ack aviso de recepção
7	007	07	BL	0007		bel campainha
8	010	08	BS	0008		bs espaço para trás
9	011	09	HT	0009		ht tabulador horizontal
10	012	0a	LF	000A		lf nova linha
11	013	0b	VT	000B		vt tabulador vertical
12	014	0c	FF	000C		ff nova página
13	015	0d	CR	000D		cr enter
14	016	0e	SO	000E		so fim de modificador (shift out)
15	017	0f	SI	000F		si início de modificador (shift in)
16	020	10	DL	0010		dle escape de ligação de dados
17	021	11	D1	0011		dc1 controlo de dispositivo um
18	022	12	D2	0012		dc2 controlo de dispositivo dois
19	023	13	D3	0013		dc3 controlo de dispositivo três
20	024	14	D4	0014		dc4 controlo de dispositivo quatro
21	025	15	NK	0015		nak confirmação de recepção negativa
22	026	16	SY	0016		syn inacção síncrona

Dec	Oct	Hex	Mne	UCS2	Kbd	AFRFUL-102-BPI_OCIL
23	027	17	EB	0017		etb fim de transmissão de bloco
24	030	18	CN	0018		can cancelar
25	031	19	EM	0019		em fim de meio
26	032	1a	SB	001A		sub substituto
27	033	1b	EC	001B		esc sair
28	034	1c	FS	001C		is4 separador de ficheiro
29	035	1d	GS	001D		is3 separador de grupo
30	036	1e	RS	001E		is2 separador de artigo
31	037	1f	US	001F		is1 separador de unidade
32	040	20	SP	0020	<>	espaço
33	041	21	!	0021	!	ponto de exclamação
34	042	22	«	0022	«	aspa
35	043	23	Nb	0023	#	cardinal
36	044	24	DO	0024	\$	dólar
37	045	25	%	0025	%	percentagem
38	046	26	&	0026	&	e comercial
39	047	27	‘	0027	‘	apóstrofe
40	050	28	(0028	(parêntese esquerdo
41	051	29)	0029)	parêntese direito
42	052	2a	*	002A	*	asterisco
43	053	2b	+	002B	+	sinal de adição
44	054	2c	,	002C	,	vírgula
45	055	2d	-	002D	-	hífen, sinal de subtracção
46	056	2e	.	002E	.	ponto
47	057	2f	/	002F	/	barra oblíqua
48	060	30	0	0030	0	dígito zero
49	061	31	1	0031	1	dígito um

Dec	Oct	Hex	Mne	UCS2	Kbd	AFRFUL-102-BPI_OCIL
50	062	32	2	0032	2	dígito dois
51	063	33	3	0033	3	dígito três
52	064	34	4	0034	4	dígito quatro
53	065	35	5	0035	5	dígito cinco
54	066	36	6	0036	6	dígito seis
55	067	37	7	0037	7	dígito sete
56	070	38	8	0038	8	dígito oito
57	071	39	9	0039	9	dígito nove
58	072	3a	:	003A	:	dois pontos
59	073	3b	;	003B	;	ponto e vírgula
60	074	3c	<	003C	<	menor
61	075	3d	=	003D	=	igual
62	076	3e	>	003E	>	maior
63	077	3f	?	003F	?	ponto de interrogação
64	100	40	At	0040	@	arroba
65	101	41	A	0041	A	letra maiúscula latina A
66	102	42	B	0042	B	letra maiúscula latina B
67	103	43	C	0043	C	letra maiúscula latina C
68	104	44	D	0044	D	letra maiúscula latina D
69	105	45	E	0045	E	letra maiúscula latina E
70	106	46	F	0046	F	letra maiúscula latina F
71	107	47	G	0047	G	letra maiúscula latina G
72	110	48	H	0048	H	letra maiúscula latina H
73	111	49	I	0049	I	letra maiúscula latina I
74	112	4a	J	004A	J	letra maiúscula latina J
75	113	4b	K	004B	K	letra maiúscula latina K
76	114	4c	L	004C	L	letra maiúscula latina L
77	115	4d	M	004D	M	letra maiúscula latina M
78	116	4e	N	004E	N	letra maiúscula latina N
79	117	4f	O	004F	O	letra maiúscula latina O
80	120	50	P	0050	P	letra maiúscula latina P
81	121	51	Q	0051	Q	letra maiúscula latina Q
82	122	52	R	0052	R	letra maiúscula latina R

Dec	Oct	Hex	Mne	UCS2	Kbd	AFRFUL-102-BPI_OCIL
83	123	53	S	0053	S	letra maiúscula latina S
84	124	54	T	0054	T	letra maiúscula latina T
85	125	55	U	0055	U	letra maiúscula latina U
86	126	56	V	0056	V	letra maiúscula latina V
87	127	57	W	0057	W	letra maiúscula latina W
88	130	58	X	0058	X	letra maiúscula latina X
89	131	59	Y	0059	Y	letra maiúscula latina Y
90	132	5a	Z	005A	Z	letra maiúscula latina Z
91	133	5b	<(005B	[parêntese recto esquerdo
92	134	5c	//	005C	\	barra reversa
93	135	5d)>	005D]	parêntese recto direito
94	136	5e	'>	005E	^	acento circunflexo
95	137	5f	_	005F	_	sublinhado
96	140	60	'!	0060	`	acento grave
97	141	61	a	0061	a	letra minúscula latina a
98	142	62	b	0062	b	letra minúscula latina b
99	143	63	c	0063	c	letra minúscula latina c
100	144	64	d	0064	d	letra minúscula latina d
101	145	65	e	0065	e	letra minúscula latina e
102	146	66	f	0066	f	letra minúscula latina f
103	147	67	g	0067	g	letra minúscula latina g
104	150	68	h	0068	h	letra minúscula latina h
105	151	69	i	0069	i	letra minúscula latina i
106	152	6a	j	006A	j	letra minúscula latina j
107	153	6b	k	006B	k	letra minúscula latina k
108	154	6c	l	006C	l	letra minúscula latina l
109	155	6d	m	006D	m	letra minúscula latina m
110	156	6e	n	006E	n	letra minúscula latina n
111	157	6f	o	006F	o	letra minúscula latina o
112	160	70	p	0070	p	letra minúscula latina p
113	161	71	q	0071	q	letra minúscula latina q

Dec	Oct	Hex	Mne	UCS2	Kbd	AFRFUL-102-BPI_OCIL
114	162	72	r	0072	r	letra minúscula latina r
115	163	73	s	0073	s	letra minúscula latina s
116	164	74	t	0074	t	letra minúscula latina t
117	165	75	u	0075	u	letra minúscula latina u
118	166	76	v	0076	v	letra minúscula latina v
119	167	77	w	0077	w	letra minúscula latina w
120	170	78	x	0078	x	letra minúscula latina x
121	171	79	y	0079	y	letra minúscula latina y
122	172	7a	z	007A	z	letra minúscula latina z
123	173	7b	(!	007B	{	chaveta esquerda
124	174	7c	!!	007C		barra vertical
125	175	7d	!)	007D	}	chaveta direita
126	176	7e	“?	007E	~	til
127	177	7f	DT	007F		del eliminar
128	200	80	PA	0080		pad chave aleatória
129	201	81	HO	0081		hop octeto superior predefinido
130	202	82	BH	0082		bph paragem autorizada aqui
131	203	83	NH	0083		nbh paragem proibida aqui
132	204	84	IN	0084		ind índice
133	205	85	NL	0085		nel parágrafo
134	206	86	SA	0086		ssa início de zona seleccionada
135	207	87	ES	0087		esa fim de zona seleccionada
136	210	88	HS	0088		hts fim de tabulador horizontal
138	212	8a	VS	008A		vts fim de tabulador vertical
139	213	8b	PD	008B		pld entrelinha parcial para baixo

Dec	Oct	Hex	Mne	UCS2	Kbd	AFRFUL-102-BPI_OCIL
140	214	8c	PU	008C		plu entrelinha parcial para cima
141	215	8d	RI	008D		ri índice invertido
142	216	8e	S2	008E		ss2 substituição única dois
143	217	8f	S3	008F		ss3 substituição única três
144	220	90	DC	0090		dcs sequência de comando de aparelho
145	221	91	P1	0091		pu1 uso privado um
146	222	92	P2	0092		pu2 uso privado dois
147	223	93	TS	0093		sts modo de transmissão activado
148	224	94	CC	0094		cch anulação do caractere precedente
149	225	95	MW	0095		mw mensagem a aguardar
150	226	96	SG	0096		sga início de zona protegida
151	227	97	EG	0097		ega fim de zona protegida
152	230	98	SS	0098		sos início de sequência
153	231	99	GC	0099		sgci introdutor de caractere gráfico único
154	232	9a	SC	009A		sci introdutor de caractere único
155	233	9b	CI	009B		csi introdutor de sequência de comando
156	234	9c	ST	009C		st fim de sequência
157	235	9d	OC	009D		osc comando de sistema de exploração
158	236	9e	PM	009E		pm mensagem privada
159	237	9f	AC	009F		apc comando de pacote de programas

Dec	Oct	Hex	Mne	UCS2	Kbd	AFRFUL-102-BPI_OCIL
160	240	a0	NS	00A0	/_	espaço inquebrável
161	241	a1		0190 +0300	E[´	letra maiús.latina E aberto com acento grave
162	242	a2		0190 +0301	E[´	letra maiús.latina E aberto com acento agudo
163	243	a3		0190 +0302	E[^	letra maiús.latina E aberto com acento circunflexo
164	244	a4		0190 +030C	E[\v	letra maiús.latina E aberto caron
165	245	a5		0186 +0300	O[´	letra maiús.latina O aberto acento grave
166	246	a6		0186 +0301	O[´	letra maiús.latina O aberto acento agudo
167	247	a7		0186 +0302	O[^	letra maiús.latina O aberto acento circunflexo
168	250	a8		019D	N]	N com parêntese recto direito (ENJ)
169	251	a9		014A	N[ENG (Sami)
170	252	aa		004E +0302	N^	N com acento circunflexo
171	253	ab	<<	00AB	<<	aspas esquerdas
172	254	ac		004E +0308	N»	letra maiús. latina N com trema
173	255	ad	--	00AD	\-	hífen virtual (traço condicional)
174	256	ae		01B3	Y[letra maiús. latina Y com parêntese recto esquerdo
175	257	af		0186 +030C	O[\v	letra maiús. latina O aberto caron

Dec	Oct	Hex	Mne	UCS2	Kbd	AFRFUL-102-BPI_OCIL
176	260	b0		00B0	_DG	símbolo graus
177	261	b1		025B +0300	e[`	letra min. latina e aberto com acento grave
178	262	b2		025B +0301	e[´	letra min. latina e aberto com acento agudo
179	263	b3		025B +0302	e[^	letra min. latina e aberto com acento circunflexo
180	264	b4		025B +030C	e[\v	letra min. latina e aberto caron
181	265	b5		0254 +0300	o[`	letra min. latina o aberto com acento grave
182	266	b6		0254 +0301	o[´	letra min. latina o aberto com acento agudo
183	267	b7		0254 +0302	o[^	letra min. latina o aberto circunflexo
184	270	b8		0272	n]	letra min. latina n com parêntese recto direito (enj)
185	271	b9		014B	n[letra min. latina n anzol eng (Sami)
186	272	ba		006E +0302	n^	letra min. latina n circunflexo
187	273	bb	>>	00BB	>>	aspas direitas
188	274	bc		006E +0308	n»	letra min. latina n com trema
189	275	bd		0294	?^	letra latina oclusiva gutural

Dec	Oct	Hex	Mne	UCS2	Kbd	AFRFUL-102-BPI_OCIL
190	276	be		01B4	y[letra min. latina y com parêntese recto esquerdo
191	277	bf		0254 +030C	o[v	letra min. latina o aberto caron
192	300	c0	A!	00C0	A`	letra maiús. latina A com acento grave
193	301	c1	A'	00C1	A'	letra maiús. latina A com acento agudo
194	302	c2	A>	00C2	A^	letra maiús. latina A com acento circunflexo
195	303	c3		01CD	Av	letra maiús. latina A caron
196	304	c4	A:	00C4	A»	letra maiús. latina A com trema
197	305	c5		018E	E<	letra maiús. latina E reflexo
198	306	c6	AE	00C6	AE+	letra maiús. latina AE (ligação ae maiúscula)
199	307	c7	C,	00C7	C\,	letra maiús. latina C com cedilha
200	310	c8	E!	00C8	E`	letra maiús. latina E com acento grave
201	311	c9	E'	00C9	E'	letra maiús. latina E com acento agudo
202	312	ca	E>	00CA	E^	letra maiús. latina E com acento circunflexo
203	313	cb	E:	00CB	E»	letra maiús. latina E com trema
204	314	cc	!!	00CC	I`	letra maiús. latina I com acento grave
205	315	cd	I'	00CD	I'	letra maiús. latina I com acento agudo

Dec	Oct	Hex	Mne	UCS2	Kbd	AFRFUL-102-BPI_OCIL
206	316	ce	I>	00CE	I^	letra maiús. latina I com acento circunflexo
207	317	cf	I:	00CF	I»	letra maiús. latina I trema
208	320	d0		0190		letra maiús. latina E aberto
209	321	d1	N?	00D1	N~	letra maiús. latina N til
210	322	d2	O!	00D2	O`	letra maiús. latina O acento grave
211	323	d3	O'	00D3	O\`	letra maiús. latina O acento agudo
212	324	d4	O>	00D4	O^	letra maiús. latina O acento circunflexo
213	325	d5	O?	01D1	O\v	letra maiús. latina O caron
214	326	d6	O:	00D6	O”	letra maiús. latina O trema
215	327	d7		0152	OE+	digrama ligado maiús. latino OE (ligação OE)
216	330	d8		0186	O[letra maiús. latina O aberto
217	331	d9	U!	00D9	U`	letra maiús. latina U acento grave
218	332	da	U'	00DA	U\`	letra maiús. latina U acento agudo
219	333	db	U>	00DB	U^	letra maiús. latina U acento circunflexo
220	334	dc	U:	00DC	U»	letra maiús. latina U trema
221	335	dd		011A	E\v	letra maiús. latina E caron
222	336	de		01CF	I\v	letra maiús. latina I caron

Dec	Oct	Hex	Mne	UCS2	Kbd	AFRFUL-102-BPI_OCIL
223	337	df		01D3	Uv	letra maiús. latina U caron
224	340	e0	a!	00E0	a`	letra min. latina a acento grave
225	341	e1	a'	00E1	a'	letra min. latina a acento agudo
226	342	e2	a>	00E2	a^	letra min. latina a acento circunflexo
227	343	e3		01CE	a\	letra min. latina a caron
228	344	e4	a:	00E4	a»	letra min. latina a trema
229	345	e5		0259	e<	letra minúscula latina schwa (e invertido)
230	346	e6	ae	00E6	ae+	letra min. latina ae (ligação ae)
231	347	e7	c,	00E7	c\,	letra min. latina c cedilha
232	350	e8	e!	00E8	e`	letra min. latina e acento grave
233	351	e9	e'	00E9	e'	letra min. latina e acento agudo
234	352	ea	e>	00EA	e^	letra min. latina e acento circunflexo
235	353	eb	e:	00EB	e»	letra min. latina e trema
236	354	ec	i!	00EC	i`	letra min. latina i acento grave
237	355	ed	i'	00ED	i'	letra min. latina i acento agudo
238	356	ee	i>	00EE	i^	letra min. latina i acento circunflexo
239	357	ef	i:	00EF	i"	letra min. latina i trema
240	360	f0		025B	e[letra min. latina e aberto
241	361	f1	n?	00F1	n~	letra min. latina n til

Dec	Oct	Hex	Mne	UCS2	Kbd	AFRFUL-102-BPI_OCIL
242	362	f2	o!	00F2	o`	letra min. latina o acento grave
243	363	f3	o'	00F3	o\'	letra min. latina o acento agudo
244	364	f4	o>	00F4	o^	letra min. latina o acento circunflexo
245	365	f5		01D2	o\v	letra min. latina o caron
246	366	f6	o:	00F6	o»	letra min. latina o trema
247	367	f7		0153	oe+	digrama ligado min. latino oe (ligação œ)
248	370	f8		0254	o[letra min. latina o aberta
249	371	f9	u!	00F9	u`	letra min. latina u acento grave
250	372	fa	u'	00FA	u\'	letra min. latina u acento agudo
251	373	fb	u>	00FB	u^	letra min. latina u acento circunflexo
252	374	fc	u:	00FC	u»	letra min. latina u trema
253	375	fd		011B	e\v	letra min. latina e caron
254	376	fe		01D0	i\v	letra min. latina i caron
255	377	ff		01D4	u\v	letra min. latina u caron



UNESCO

Sector da Comunicação e da Informação
Divisão da Sociedade da Informação
1, rue Miollis
75732 Paris cedex 15
França

Tel. + 33.1.45.68.45.00

Fax + 33.1.45.68.55.83

<http://www.unesco.org/webworld>

UNIÃO LATINA

Direcção Terminologia
e Indústrias da Língua
131, rue du Bac
75007 Paris
França

Tel. + 33.1.45.49.60.62

Fax + 33.1.45.49.67.39

<http://dtil.unilat.org>