

Próximo:[Caracteres não imprimíveis](#), Anterior:[Sintaxe para Strings](#), Acima:[Tipo de string](#) [[Conteúdo](#)][[Índice](#)]

2.4.8.2 Caracteres Não ASCII em Strings

Existem duas representações de texto para caracteres não ASCII em strings do Emacs: multibyte e unibyte (consulte [Representações de texto](#)). Grosso modo, strings unibyte armazenam bytes brutos, enquanto strings multibyte armazenam texto legível por humanos. Cada caractere em uma string unibyte é um byte, ou seja, seu valor está entre 0 e 255. Por outro lado, cada caractere em uma string multibyte pode ter um valor entre 0 e 4194303 (consulte [Tipo de caractere](#)). Em ambos os casos, os caracteres acima de 127 não são ASCII .

Você pode incluir um caractere não ASCII em uma constante de string escrevendo-o literalmente. Se a constante de string for lida de uma fonte multibyte, como um buffer ou string multibyte, ou um arquivo que seria visitado como multibyte, o Emacs lê cada caractere não ASCII como um caractere multibyte e automaticamente torna a string uma string multibyte. Se a constante de string for lida de uma fonte unibyte, o Emacs lerá o caractere não ASCII como unibyte e tornará a string unibyte.

Em vez de escrever um caractere literalmente em uma string multibyte, você pode escrevê-lo como seu código de caractere usando uma sequência de escape. Consulte [Sintaxe de escape geral](#) , para obter detalhes sobre sequências de escape.

Se você usar qualquer sequência de escape no estilo Unicode '\uNNNN' ou '\U00NNNNNN' em uma constante de string (mesmo para um caractere ASCII), o Emacs assume automaticamente que é multibyte.

Você também pode usar sequências de escape hexadecimais ('\ xn') e sequências de escape octais ('\ n') em constantes de string. **Mas cuidado:** se uma constante de string contiver sequências de escape hexadecimais ou octais, e todas essas sequências de escape especificarem caracteres unibyte (ou seja, menos de 256), e não houver outros caracteres literais não ASCII ou sequências de escape no estilo Unicode na string, então o Emacs assume automaticamente que é uma string unibyte. Ou seja, ele assume que todos os caracteres não ASCII que ocorrem na string são bytes brutos de 8 bits.

Em sequências de escape hexadecimais e octais, o código de caractere de escape pode conter um número variável de dígitos, de modo que o primeiro caractere subsequente que não seja um dígito hexadecimal ou octal válido encerra a sequência de escape. Se o próximo caractere em uma string puder ser interpretado como um dígito hexadecimal ou octal, escreva '\ ' (barra invertida e espaço) para encerrar a sequência de escape. Por exemplo, '\xe0\ ' representa um caractere, 'uma' com sotaque grave. '\ ' em uma constante de string é como backslash-newline; ele não contribui com nenhum caractere para a string, mas encerra qualquer escape hexadecimal anterior.

Próximo:[Caracteres não imprimíveis](#), Anterior:[Sintaxe para Strings](#), Acima:[Tipo de string](#) [[Conteúdo](#)][[Índice](#)]