

UTF-8 SAMPLER 𐀀 𐀁 𐀂 𐀃 𐀄 𐀅 𐀆 𐀇 𐀈 𐀉 𐀊 𐀋 𐀌 𐀍 𐀎 𐀏 𐀐 𐀑 𐀒 𐀓 𐀔 𐀕 𐀖 𐀗 𐀘 𐀙 𐀚 𐀛 𐀜 𐀝 𐀞 𐀟 𐀠 𐀡 𐀢 𐀣 𐀤 𐀥 𐀦 𐀧 𐀨 𐀩 𐀪 𐀫 𐀬 𐀭 𐀮 𐀯 𐀰 𐀱 𐀲 𐀳 𐀴 𐀵 𐀶 𐀷 𐀸 𐀹 𐀺 𐀻 𐀼 𐀽 𐀾 𐀿 𐁀 𐁁 𐁂 𐁃 𐁄 𐁅 𐁆 𐁇 𐁈 𐁉 𐁊 𐁋 𐁌 𐁍 𐁎 𐁏 𐁐 𐁑 𐁒 𐁓 𐁔 𐁕 𐁖 𐁗 𐁘 𐁙 𐁚 𐁛 𐁜 𐁝 𐁞 𐁟 𐁠 𐁡 𐁢 𐁣 𐁤 𐁥 𐁦 𐁧 𐁨 𐁩 𐁪 𐁫 𐁬 𐁭 𐁮 𐁯 𐁰 𐁱 𐁲 𐁳 𐁴 𐁵 𐁶 𐁷 𐁸 𐁹 𐁺 𐁻 𐁼 𐁽 𐁾 𐁿 𐂀 𐂁 𐂂 𐂃 𐂄 𐂅 𐂆 𐂇 𐂈 𐂉 𐂊 𐂋 𐂌 𐂍 𐂎 𐂏 𐂐 𐂑 𐂒 𐂓 𐂔 𐂕 𐂖 𐂗 𐂘 𐂙 𐂚 𐂛 𐂜 𐂝 𐂞 𐂟 𐂠 𐂡 𐂢 𐂣 𐂤 𐂥 𐂦 𐂧 𐂨 𐂩 𐂪 𐂫 𐂬 𐂭 𐂮 𐂯 𐂰 𐂱 𐂲 𐂳 𐂴 𐂵 𐂶 𐂷 𐂸 𐂹 𐂺 𐂻 𐂼 𐂽 𐂾 𐂿 𐃀 𐃁 𐃂 𐃃 𐃄 𐃅 𐃆 𐃇 𐃈 𐃉 𐃊 𐃋 𐃌 𐃍 𐃎 𐃏 𐃐 𐃑 𐃒 𐃓 𐃔 𐃕 𐃖 𐃗 𐃘 𐃙 𐃚 𐃛 𐃜 𐃝 𐃞 𐃟 𐃠 𐃡 𐃢 𐃣 𐃤 𐃥 𐃦 𐃧 𐃨 𐃩 𐃪 𐃫 𐃬 𐃭 𐃮 𐃯 𐃰 𐃱 𐃲 𐃳 𐃴 𐃵 𐃶 𐃷 𐃸 𐃹 𐃺 𐃻 𐃼 𐃽 𐃾 𐃿 𐄀 𐄁 𐄂 𐄃 𐄄 𐄅 𐄆 𐄇 𐄈 𐄉 𐄊 𐄋 𐄌 𐄍 𐄎 𐄏 𐄐 𐄑 𐄒 𐄓 𐄔 𐄕 𐄖 𐄗 𐄘 𐄙 𐄚 𐄛 𐄜 𐄝 𐄞 𐄟 𐄠 𐄡 𐄢 𐄣 𐄤 𐄥 𐄦 𐄧 𐄨 𐄩 𐄪 𐄫 𐄬 𐄭 𐄮 𐄯 𐄰 𐄱 𐄲 𐄳 𐄴 𐄵 𐄶 𐄷 𐄸 𐄹 𐄺 𐄻 𐄼 𐄽 𐄾 𐄿 𐅀 𐅁 𐅂 𐅃 𐅄 𐅅 𐅆 𐅇 𐅈 𐅉 𐅊 𐅋 𐅌 𐅍 𐅎 𐅏 𐅐 𐅑 𐅒 𐅓 𐅔 𐅕 𐅖 𐅗 𐅘 𐅙 𐅚 𐅛 𐅜 𐅝 𐅞 𐅟 𐅠 𐅡 𐅢 𐅣 𐅤 𐅥 𐅦 𐅧 𐅨 𐅩 𐅪 𐅫 𐅬 𐅭 𐅮 𐅯 𐅰 𐅱 𐅲 𐅳 𐅴 𐅵 𐅶 𐅷 𐅸 𐅹 𐅺 𐅻 𐅼 𐅽 𐅾 𐅿 𐆀 𐆁 𐆂 𐆃 𐆄 𐆅 𐆆 𐆇 𐆈 𐆉 𐆊 𐆋 𐆌 𐆍 𐆎 𐆏 𐆐 𐆑 𐆒 𐆓 𐆔 𐆕 𐆖 𐆗 𐆘 𐆙 𐆚 𐆛 𐆜 𐆝 𐆞 𐆟 𐆠 𐆡 𐆢 𐆣 𐆤 𐆥 𐆦 𐆧 𐆨 𐆩 𐆪 𐆫 𐆬 𐆭 𐆮 𐆯 𐆰 𐆱 𐆲 𐆳 𐆴 𐆵 𐆶 𐆷 𐆸 𐆹 𐆺 𐆻 𐆼 𐆽 𐆾 𐆿 𐇀 𐇁 𐇂 𐇃 𐇄 𐇅 𐇆 𐇇 𐇈 𐇉 𐇊 𐇋 𐇌 𐇍 𐇎 𐇏 𐇐 𐇑 𐇒 𐇓 𐇔 𐇕 𐇖 𐇗 𐇘 𐇙 𐇚 𐇛 𐇜 𐇝 𐇞 𐇟 𐇠 𐇡 𐇢 𐇣 𐇤 𐇥 𐇦 𐇧 𐇨 𐇩 𐇪 𐇫 𐇬 𐇭 𐇮 𐇯 𐇰 𐇱 𐇲 𐇳 𐇴 𐇵 𐇶 𐇷 𐇸 𐇹 𐇺 𐇻 𐇼 𐇽 𐇾 𐇿 𐈀 𐈁 𐈂 𐈃 𐈄 𐈅 𐈆 𐈇 𐈈 𐈉 𐈊 𐈋 𐈌 𐈍 𐈎 𐈏 𐈐 𐈑 𐈒 𐈓 𐈔 𐈕 𐈖 𐈗 𐈘 𐈙 𐈚 𐈛 𐈜 𐈝 𐈞 𐈟 𐈠 𐈡 𐈢 𐈣 𐈤 𐈥 𐈦 𐈧 𐈨 𐈩 𐈪 𐈫 𐈬 𐈭 𐈮 𐈯 𐈰 𐈱 𐈲 𐈳 𐈴 𐈵 𐈶 𐈷 𐈸 𐈹 𐈺 𐈻 𐈼 𐈽 𐈾 𐈿 𐉀 𐉁 𐉂 𐉃 𐉄 𐉅 𐉆 𐉇 𐉈 𐉉 𐉊 𐉋 𐉌 𐉍 𐉎 𐉏 𐉐 𐉑 𐉒 𐉓 𐉔 𐉕 𐉖 𐉗 𐉘 𐉙 𐉚 𐉛 𐉜 𐉝 𐉞 𐉟 𐉠 𐉡 𐉢 𐉣 𐉤 𐉥 𐉦 𐉧 𐉨 𐉩 𐉪 𐉫 𐉬 𐉭 𐉮 𐉯 𐉰 𐉱 𐉲 𐉳 𐉴 𐉵 𐉶 𐉷 𐉸 𐉹 𐉺 𐉻 𐉼 𐉽 𐉾 𐉿 𐊀 𐊁 𐊂 𐊃 𐊄 𐊅 𐊆 𐊇 𐊈 𐊉 𐊊 𐊋 𐊌 𐊍 𐊎 𐊏 𐊐 𐊑 𐊒 𐊓 𐊔 𐊕 𐊖 𐊗 𐊘 𐊙 𐊚 𐊛 𐊜 𐊝 𐊞 𐊟 𐊠 𐊡 𐊢 𐊣 𐊤 𐊥 𐊦 𐊧 𐊨 𐊩 𐊪 𐊫 𐊬 𐊭 𐊮 𐊯 𐊰 𐊱 𐊲 𐊳 𐊴 𐊵 𐊶 𐊷 𐊸 𐊹 𐊺 𐊻 𐊼 𐊽 𐊾 𐊿 𐋀 𐋁 𐋂 𐋃 𐋄 𐋅 𐋆 𐋇 𐋈 𐋉 𐋊 𐋋 𐋌 𐋍 𐋎 𐋏 𐋐 𐋑 𐋒 𐋓 𐋔 𐋕 𐋖 𐋗 𐋘 𐋙 𐋚 𐋛 𐋜 𐋝 𐋞 𐋟 𐋠 𐋡 𐋢 𐋣 𐋤 𐋥 𐋦 𐋧 𐋨 𐋩 𐋪 𐋫 𐋬 𐋭 𐋮 𐋯 𐋰 𐋱 𐋲 𐋳 𐋴 𐋵 𐋶 𐋷 𐋸 𐋹 𐋺 𐋻 𐋼 𐋽 𐋾 𐋿 𐌀 𐌁 𐌂 𐌃 𐌄 𐌅 𐌆 𐌇 𐌈 𐌉 𐌊 𐌋 𐌌 𐌍 𐌎 𐌏 𐌐 𐌑 𐌒 𐌓 𐌔 𐌕 𐌖 𐌗 𐌘 𐌙 𐌚 𐌛 𐌜 𐌝 𐌞 𐌟 𐌠 𐌡 𐌢 𐌣 𐌤 𐌥 𐌦 𐌧 𐌨 𐌩 𐌪 𐌫 𐌬 𐌭 𐌮 𐌯 𐌰

[illegible]

characters, depending on one's personal preference or etymological theory. In this sentence, for example, "-tàng", "chiàh", "mā", and "bē" are problematic using Chinese characters. "Góá" (I/me) and "po-lē" (glass) are as written in other Sinitic languages (e.g. Mandarin, Hakka)." Wagner Amaral of Pinese & Amaral Associados notes that the Brazilian Portuguese sentence for "I can eat glass" should be identical to the Portuguese one, as the word "machuca" means "inflict pain", or rather "injuries". The words "faz mal" would more correctly translate as "cause harm".

Burmese: In English the first person pronoun "I" stands for both genders, male and female. In Burmese (except in the central part of Burma) kyundaw (ကွယ်တော်မူ) for male and kyanma (ကွယ်တော်မူ) for female. Using here a fully-compliant Unicode Burmese font -- sadly one and only Padauk Graphite font exists -- rendering using graphite engine. [CLICK HERE](#) to test Burmese characters. The Quick Brown Fox The "I can eat glass" sentences do not necessarily show off the orthography of each language to best advantage. In many alphabetic written languages it is possible to include all (or most) letters (or "special" characters) in a single (often nonsense) pangram. These were traditionally used in typewriter instruction: now they are useful for stress-testing computer fonts and keyboard input methods. Here are a few examples (SEND MORE):

English: The quick brown fox jumps over the lazy dog. Jamaican: Chruu, a kwik di kwik brong fox a jomp huova di liezi daag de, yu no siit? Irish: "An bfuil do croí ag bualad ó faitíos an grá a m'eall lena póg éada ó slí do leasa tú?" "D'fhuascail íosa Úrmaic na hÓige Beannaite pór Éava agus Ádaim." Dutch: Pa's wijze lynx bezag vroom het fikse aquaduct. German: Falsches Üben von Xylophonmusik quält jeden größeren Zwerg. (1) German: Im finfteren Jagdfehloß am offenen Felsquellwaffer patzte der affig-flatterhafte kauzig-höfliche Bäcker über feinem veriffften kniffiligen C-Xylophon. (2) Swedish: Flygande bäckasiner söka strax hwila på mjuka tuvor. Icelandic: Sævör grét áðan því úlpan var ónýt. Polish: Pchnąć w tę łódź jeża lub ośm skrzyń fig. Czech: Přiliš žluťoučký kůň úpěl ďábelské kódy. Slovak: Starý kôň na hŕbe kníh žuje tiško povädnuté ruže, na stípe sa datel' učí kvákať novú ôdu o živote. Greek (monotonic): ξεσκεπάζω την ψυχοφθόρα βεβλυγμία Greek (polytonic): ξεσκεπάζω τὴν ψυχοφθόρα βεβλυγμία Russian: Съешь же ещё этих мягких французских булок да выпей чаю. Russian: В чашах юга жил-был цитрус? Да, но фальшивый экземпляр! ё. Bulgarian: Жълтата дюля беше щастлива, че пухът, който цъфна, замръзна като гъон. Sami (Northern): Vuol Ruota geđggiid leat mángga luosa ja čuovvža. Hungarian: Árvíztűrő tükörfúrógép. Spanish: El pingüino Wenceslao hizo kilómetros bajo exhaustiva lluvia y frío, añoraba a su querido cachorro. Portuguese: O próximo vôo à noite sobre o Atlântico, põe frequentemente o único médico. (3) French: Les naïfs ægithales hâtifs pondant à Noël où il gèle sont sûrs d'être déçus et de voir leurs drôles d'œufs abimés. Esperanto: Eĥoŝanĝo ĉiujŝaŭde. Hebrew: הוּגְבֹוּט קוֹץ דְּפָרָק חֲצַנְתָּ דִּיאַ עוֹמָשׁל חֶסֶת דִּיכּ הוּזְ (4) Notes: Other phrases commonly used in Germany include: "Ein wackerer Bayer vertilgt ja bequem zwei Pfund Kalbshaxe" and, more recently, "Franz jagt im komplett verwahrlosten Taxi quer durch Bayern", but both lack umlauts and esszet. Previously, going for the shortest sentence that has all the umlauts and special characters, I had "Grüße aus Bärenhöfe (und Öechtringen)!" Acute accents are not used in native German words, so I was surprised to discover "Öechtringen" in the Deutsche Bundespost Postleitzahlenbuch: It's a small village in eastern Lower Saxony. The "oe" in this case turns out to be the Lower Saxon "lengthening e" (Dehnungs-e), which makes the previous vowel long (used in a number of Lower Saxon place names such as Soest and Itzehoe), not the "e" that indicates umlaut of the preceding vowel. Many thanks to the Öechtringen-Namenschreibungsuntersuchungskomitee (Alex Bochannek, Manfred Erren, Asmus Freytag, Christoph Päper, plus Werner Lemberg who serves as Öechtringen-Namenschreibungsuntersuchungskomiteerechtschreibungsprüfer) for their relentless pursuit of the facts in this case. Conclusion: the accent almost certainly does not belong on this (or any other native German) word, but neither can it be dismissed as dirt on the page. To add to the mystery, it has been reported that other copies of the same edition of the PLZB do not show the accent! UPDATE (March 2006): David Krings was intrigued enough by this report to contact the mayor of Ebstorf, of which Oechtringen is a borough, who responded: Sehr geehrter Mr. Krings, wenn Oechtringen irgendwo mit einem Akzent auf dem O geschrieben wurde, dann kann das nur ein Fehldruck sein. Die offizielle Schreibweise lautet jedenfalls „Oechtringen“. Mit freundlichen Grüßen Der Samtgemeindebürgermeister i.A. Lothar Jessel From Karl Pentzlin (Kochel am See, Bavaria, Germany): "This German phrase is suited for display by a Fraktur (broken letter) font. It contains: all common three-letter ligatures: ffi ffl fft and all two-letter ligatures required by the Duden for Fraktur typesetting: ch ck ff fi fl ft ll fch fi ff ft tz (all in a manner such they are not part of a three-letter ligature), one example of f-l where German typesetting rules prohibit ligating (marked by a ZWNJ), and all German letters a...z, ä,ö,ü,ß, f [long s] (all in a manner such that they are not part of a two-letter Fraktur ligature)." Otto Stolz notes that "Schloß" is now spelled 'Schloss', in contrast to 'größer' (example 4) which has kept its 'ß'. Fraktur has been banned from general use, in 1942, and long-s (ſ) has ceased to be used with Antiqua (Roman) even earlier (the latest Antiqua-ſ I have seen is from 1913, but then I am no expert, so there may well be a later instance." Later Otto confirms the latter theory, "Now I've run across a book "Deutsche Rechtschreibung" (edited by Lutz Mackensen) from 1954 (my reprint is from 1956) that has kept the Antiqua-ſ in its dictionary part (but neither in the preface nor in the appendix)." Diaeresis is not used in Iberian Portuguese. From Yurio Miyazawa: "This poetry contains all the sounds in the Japanese language and used to be the first thing for children to learn in their Japanese class. The Hiragana version is particularly neat because it covers every character in the phonetic Hiragana character set." Yurio also sent the Kanji version: 色は匂へど 散りぬるを 我が世誰ぞ 常ならむ 有爲の楽山 今日越えて 浅き夢見じ 酔ひもせず Accented Cyrillic: (This section contributed by Vladimir Marinov.) In Bulgarian it is desirable, customary, or in some cases required to write accents over vowels. Unfortunately, no computer character sets contain the full repertoire of accented Cyrillic letters. With Unicode, however, it is possible to combine any Cyrillic letter with any combining accent. The appearance of the result depends on the font and the rendering engine. Here are two examples. Той видя бялата коса по главата ѝ и коса на рамото ѝ, и рече да ѝ рече: "Парата по пари от парата, не ща пари!", но си помисли: "Хей, помисли си! А ѝ река, а е скочила в тази река, която щеше да течѐ, а не тѐче." По пътя пътуват кюрди и югославяни. HTML Features Here is the Russian alphabet (uppercase only) coded in three different ways, which should look identical: АБВГДЕЖЗИЙКЛМНОПРСТУФХЦЧШЩЪЫЬЭЮЯ (Literal UTF-8) АБВГДЕЖЗИЙКЛМНОПРСТУФХЦЧШЩЪЫЬЭЮЯ (Decimal numeric character reference) АБВГДЕЖЗИЙКЛМНОПРСТУФХЦЧШЩЪЫЬЭЮЯ (Hexadecimal numeric character reference) In another test, we use HTML language tags to distinguish Bulgarian, Russian, and Serbian, which have different italic forms for lowercase б, г, д, п, and/or т: Bulgarian: [бгдп] [бгдпт] Мога да ям стъкло и не ме боли. Russian: [бгдпт] [бгдпт] Я могу есть стекло, это мне не вредит. Serbian: [бгдпт] [бгдпт] Могу јести стакло а да ми не штоди. Credits, Tools, and Commentary Credits: The "I can eat glass" phrase and the initial collection of translations: Ethan Mollick. Transcription / conversion to UTF-8: Frank da Cruz. Albanian: Sindi Keesan. Afrikaans: Johan Fourie, Kevin Poalses. Anglo Saxon: Frank da Cruz. Arabic: Najib Tounsi. Armenian: Vage Kundakçi. Belarusian: Alexey Chernyak. Bengali: Somnath Purkayastha, Deepayan Sarkar. Bislama: Dan McGarry. Braille: Frank da Cruz. Bulgarian: Sindi Keesan, Guentcho Skordev, Vladimir Marinov. Burmese: "cetanapa". Cabo Verde Creole: Cláudio Alexandre Duarte. Catalán: Jordi Bancells. Chinese: Jack Soo, Wong Pui Lam. Chinook Jargon: David Robertson. Cornish: Chris Stephens. Croatian: Marjan Baće. Czech: Stanislav Pecha, Radovan Garabík. Dutch: Peter Gotink. Pim Blokland, Rob Daniel, Rob de Wit. Erzian: Jack Rueter. Esperanto: Franko Luin, Radovan Garabík. Estonian: Meelis Roos. Faroese: Jón Gaasedal. Farsi/Persian: Payam Elahi. Finnish: Sampsa Toivanen. French: Luc Carissimo, Anne Colin du Terrail, Sean M. Burke. Galician: Laura Probaos. Georgian: Giorgi Lebanidze. German: Christoph Päper, Otto Stolz, Karl Pentzlin, David Krings, Frank da Cruz. Gothic: Aurélien Coudurier. Greek: Ariel Glenn, Constantine Stathopoulos, Siva Nataraja, Christos Georgiou. Hebrew: Jonathan Rosenne, Tal Barnea. Hausa: Malami Buba, Tom Gewecke. Hawaiian: na Hau'oli Motta, Anela de Rego, Kaliko Trapp. Hindi: Shirish Kalele, Nitin Dahra. Hungarian: András Rácz, Mark Holczhammer. Icelandic: Andrés Magnússon, Sveinn Baldursson. International Phonetic Alphabet (IPA): Siva

Nataraja / Vincent Ramos. Irish: Michael Everson, Marion Gunn, James Kass, Curtis Clark. Italian: Thomas De Bellis. Jamaican: Stephen J. Cherin. Japanese: Makoto Takahashi, Yurio Miyazawa. Karelian: Aleksandr Semakov. Khmer: Tola Sann. Kirchröadsj: Roger Stoffers. Kreyòl: Sean M. Burke. Korean: Jungshik Shin. Langenfelder Platt: David Krings. Lao: Tola Sann. Lëtzebuergesch: Stefaan Eeckels. Lingala: Denis Moyogo Jacquerey (Nkóta ya Kóngó mibalé). (Nkóta ya Kóngó mibal Lithuanian: Gediminas Grigas. Lojban: Edward Cherlin. Lusatian: Ronald Schaffhirt. Macedonian: Sindi Keesan. Malay: Zarina Mustapha. Maltese: Kenneth Joseph Vella. Manx: Éanna Ó Brádaigh. Marathi: Shirish Kalele. Marquesan: Kaliko Trapp. Middle English: Frank da Cruz. Milanese: Marco Cimarosti. Mongolian: Tom Gewecke. Napoletano: Diego Quintano. Navajo: Tom Gewecke. Nórdicg: Ywlyan Rott. Norwegian: Herman Ranes. Odenwälderisch: Alexander Heß. Old Irish: Michael Everson. Old Norse: Andrés Magnússon. Papiamentu: Bianca and Denise Zanardi. Pashto: N.R. Liwal. Pfälzisch: Dr. Johannes Sander. Picard: Philippe Mennecier. Polish: Juliusz Chroboczek, Paweł Przeradowski. Portuguese: "Cláudio" Alexandre Duarte, Bianca and Denise Zanardi, Pedro Palhoto Matos, Wagner Amaral. Québécois: Laurent Detillieux. Roman: Pierpaolo Bernardi. Romanian: Juliusz Chroboczek, Ionel Mugurel. Romansch: Alexandre Suter. Ruhrdeutsch: "Timwi". Russian: Alexey Chernyak, Serge Nesterovitch. Sami: Anne Colin du Terrail, Luc Carissimo. Sanskrit: Siva Nataraja / Vincent Ramos. Sächsisch: André Müller. Schwäbisch: Otto Stolz. Scots: Jonathan Riddell. Serbian: Sindi Keesan, Ranko Narancic, Boris Daljevic, Szilvia Csorba, O. Dag. Slovak: G. Adam Stanislav, Radovan Garabík. Slovenian: Albert Kolar. Spanish: Aleida Muñoz, Laura Probaos. Swahili: Ronald Schaffhirt. Swedish: Christian Rose, Bengt Larsson. Taiwanese: Henry H. Tan-Tenn. Tagalog: Jim Soliven. Tamil: Vasee Vaseeharan. Tibetan: D. Germano, Tom Gewecke. Thai: Alan Wood's wife. Turkish: Vage Kundakçı, Tom Gewecke, Merlign Olnon. Ukrainian: Michael Zajac. Urdu: Mustafa Ali. Vietnamese: Dixon Au, [James] Đỗ Bá Phước 杜伯福. Walloon: Pablo Saratxaga. Welsh: Geiriadur Prifysgol Cymru (Andrew). Yiddish: Mark David. Zeneise: Angelo Pavese. Tools Used to Create This Web Page: The UTF8-aware Kermit 95 terminal emulator on Windows, to a Unix host with the EMACS text editor. Kermit 95 displays UTF-8 and also allows keyboard entry of arbitrary Unicode BMP characters as 4 hex digits, as shown HERE. Hex codes for Unicode values can be found in The Unicode Standard (recommended) and the online code charts. When submissions arrive by email encoded in some other character set (Latin-1, Latin-2, KOI, various PC code pages, JEUC, etc), I use the TRANSLATE command of C-Kermit on the Unix host (where I read my mail) to convert the character set to UTF-8 (I could also use Kermit 95 for this; it has the same TRANSLATE command). That's it -- no "Web authoring" tools, no locales, no "smart" anything. It's just plain text, nothing more. By the way, there's nothing special about EMACS -- any text editor will do, providing it allows entry of arbitrary 8-bit bytes as text, including the 0x80-0x9F "C1" range. EMACS 21.1 actually supports UTF-8; earlier versions don't know about it and display the octal codes; either way is OK for this purpose. Commentary: Date: Wed, 27 Feb 2002 13:21:59 +0100 From: "Bruno DEDOMINICIS" <b.dedominicis@cite-sciences.fr> Subject: Je peux manger du verre, cela ne me fait pas mal. I just found out your website and it makes me feel like proposing an interpretation of the choice of this peculiar phrase. Glass is transparent and can hurt as everyone knows. The relation between people and civilisations is sometimes effusional and more often rude. The concept of breaking frontiers through globalization, in a way, is also an attempt to deny any difference. Isn't "transparency" the flag of modernity? Nothing should be hidden any more, authority is obsolete, and the new powers are supposed to reign through loving and smiling and no more through coercion... Eating glass without pain sounds like a very nice metaphor of this attempt. That is, frontiers should become glass transparent first, and be denied by incorporating them. On the reverse, it shows that through globalization, frontiers undergo a process of displacement, that is, when they are not any more speakable, they become repressed from the speech and are therefore incorporated and might become painful symptoms, as for example what happens when one tries to eat glass. The frontiers that used to separate bodies one from another tend to divide bodies from within and make them suffer.... The chosen phrase then appears as a denial of the symptom that might result from the destitution of traditional frontiers. Best, Bruno De Dominicis, Paris, France Other Unicode pages onsite: Peace in All Languages Frank's Compulsive Guide to Postal Addresses (especially the Index) Representing Middle English on the Web with UTF-8 The Kermit Bibliography (in UTF-8) Interchange of Non-English Computer Text (UTF-8 math and box-drawing) Unicode Table (in UTF-8) Unicode samplers and resources offsite: Unicode Code ConverterUnicode Code Conversion (converts among different Unicode encoding forms and notations). Michael Everson's Bibliography of Typography and Scripts Does your browser support Unicode English? (James Kass) I don't know, I only work here Anyone can be provincial! Transcriptions of "Unicode" Example Unicode Usage for Business Applications UTF-8 and Unicode FAQ for Unix/Linux Unicode fonts: Code 2000 (James Kass) Unicode Fonts for Windows Computers (Alan Wood) Unicode Fonts and Tools for X11 (Markus Kuhn) Everson Mono (Michael Everson) Agfa Monotype [Kermit 95] [K95 Screen Shots] [C-Kermit] [Kermit Home] [Display Problems?] [The Unicode Consortium] UTF-8 Sampler / The Kermit Project / Columbia University / kermit@columbia.edu