

Characters, Fonts, and Languages in MathBook XML

1 Introduction

This is a sample MathBook XML document to demonstrate and test different fonts used for different languages. For PDF output, and best results, process with `xelatex` (or `lualatex`, unsupported). If processed with `pdflatex`, be prepared for errors, and maybe try running with the command-line option `-interaction batchmode` or `-interaction nonstopmode`. Commentary here, and in the original source file, should help you track down what might be missing in your system if the `xelatex` results are not satisfactory. You can also compare PDF output with the HTML output, which should work well for many different characters tested here.

1. Your source XML file should always have

```
<?xml version="1.0" encoding="UTF-8" ?>
```

as the first line of the file to specify that your file contains Unicode characters.

2. You will need to determine how to make your operating system, keyboard and editor cooperate to create the Unicode characters you need. The [Full Unicode Input utility](#) can be useful for generating short snippets or you can copy/paste from the source of this document.
3. Generating variants of HTML for web browsers should work smoothly. You can test your system and browser for font support by viewing samples at Alan Wood's [Unicode Resources](#).

Not finding what you are looking for? Send us a request. Even better, send us a snippet of text that I can use as a sample. If you know of any subtleties in the use of \LaTeX for your desired font, please include those hints.

2 Language Examples

This next sections contain examples from various languages. The characters should render in HTML, but without any special support for subtle variations of how a language is used (e.g. date formats). For \LaTeX output, the `polyglossia` package may handle details like hyphenation. Here we mostly test characters and fonts. Improvements from those knowledgeable about the use of particular languages within \LaTeX are especially welcome, as are requests for particular languages.

1. Greek (modern), 2016-07-25
2. Korean, 2016-12-08
3. Hungarian (Magyar), 2016-07-25
4. Spanish, 2016-07-25
5. Vietnamese, 2016-07-25

List 2.1: Supported Languages

3 Greek (Modern)

- Language code: el
- Font: CMU Serif (Computer Modern Unicode)
- Source: NA

ἄνδρα μοι ἔννεπε, μοῦσα, πολύτροπον, ὃς μάλα πολλὰ πλάγχθη, ἐπεὶ Τροίης ἱερὸν πτολίεθρον ἔπερσεν: πολλῶν δ' ἀνθρώπων ἴδεν ἄστεα καὶ νόον ἔγνω, πολλὰ δ' ὃ γ' ἐν πόντῳ πάθεν ἄλγεα ὃν κατὰ θυμόν, ἀρνύμενος ἦν τε ψυχὴν καὶ νόστον ἐταίρων. ἀλλ' οὐδ' ὧς ἐτάρους ἐρρύσατο, ἰέμενός περ: αὐτῶν γὰρ σφετέρῃσιν ἀτασθαλίῃσιν ὄλοντο, νήπιοι, οἳ κατὰ βοῦς Ὑπερίονος Ἥελιοιο ἦσθιον: αὐτὰρ ὁ τοῖσιν ἀφείλετο νόστιμον ἦμαρ. τῶν ἀμόθεν γε, θεά, θύγατερ Διός, εἰπέ καὶ ἡμῖν. ἐνθ' ἄλλοι μὲν πάντες, ὅσοι φύγον αἰπὺν ὄλεθρον, οἴκοι ἔσαν, πόλεμόν τε πεφρευγότες ἠδὲ θάλασσαν: τὸν δ' οἶον νόστου κεχρημένον ἠδὲ γυναικὸς νύμφῃ πτόνι ἔρυκε Καλυψὼ δῖα θεάων ἐν σπέσσι γλαφυροῖσι, λιχαιομένη πόσιν εἶναι. ἀλλ' ὅτε δὴ ἔτος ἦλθε περιπλομένων ἐνιαυτῶν, τῷ οἳ ἐπεκλώσαντο θεοὶ οἰκόνδε νέεσθαι εἰς Ἰθάκην, οὐδ' ἐνθα πεφυγμένος ἦεν ἀέθλων καὶ μετὰ οἷσι φίλοισι. θεοὶ δ' ἐλέαιρον ἅπαντες νόσφι Ποσειδάωνος: ὃ δ' ἀσπερχές μενέαινε ἀντιθέῳ Ὀδυσῆϊ πάρος ἦν γαῖαν ἰκέσθαι. ἀλλ' ὃ μὲν Αἰθίοπας μετεκίαθε τηλόθ' ἐόντας, Αἰθίοπας τοὶ διχθὰ δεδαίεσθαι, ἔσχατοι ἀνδρῶν, οἳ μὲν δυσσομένου Ὑπερίονος οἳ δ' ἀνιόντος, ἀντιῶν ταύρων τε καὶ ἀρνεῶν ἐκατόμβης. ἐνθ' ὃ γ' ἐτέρπετο δαιτὶ παρῆμενος: οἳ δὲ δὴ ἄλλοι Ζητὸς ἐνὶ μεγάροισιν Ὀλυμπίου ἀθροοὶ ἦσαν. τοῖσι δὲ μύθων ἦρχε πατήρ ἀνδρῶν τε θεῶν τε: μνήσατο γὰρ κατὰ θυμόν ἀμύμονος Αἰγίσθιοιο, τὸν ῥ' Ἀγαμέμνονιδης τηλεκλυτὸς ἔκταν Ὀρέστης: τοῦ δ' ὃ γ' ἐπιμνησθεὶς ἔπε' ἀθανάτοισι μετηΰδα: ὦ πόποι, οἶον δὴ νῦν θεοὺς βροτοὶ αἰτιώωνται: ἐξ ἡμέων γὰρ φασὶ κάκ' ἔμμεναι, οἳ δὲ καὶ αὐτοὶ σφῆσιν ἀτασθαλίῃσιν ὑπὲρ μόρον ἄλγε' ἔχουσιν, ὧς καὶ νῦν Αἰγίσθος ὑπὲρ μόρον Ἀτρεΐδαο γῆμ' ἄλοχον μνηστῆν, τὸν δ' ἔκτανε νοστήσαντα, εἰδὼς αἰπὺν ὄλεθρον, ἐπεὶ πρό οἳ εἶπομεν ἡμεῖς, Ἑρμείαν πέμψαντες, εὐσχοπὸν ἀργεῖφρόντην, μήτ' αὐτὸν κτείνειν μήτε μνάσθαι ἄκοιτιν: ἐκ γὰρ Ὀρέσταο τίσις ἔσσεται Ἀτρεΐδαο, ὅππότε ἂν ἠβήσῃ τε καὶ ἦς ἱμεῖρεται αἴης. ὧς ἔφαθ' Ἑρμείας, ἀλλ' οὐ φρένας Αἰγίσθιοιο πεῖθ' ἀγαθὰ φρονέων: νῦν δ' ἀθρόα πάντ' ἀπέτισεν.

τὸν δ' ἡμείβετ' ἔπειτα θεά, γλαυκῶπις Ἀθήνη: ὦ πάτερ ἡμέτερε Κρονίδη, ὕπατε κρειόντων, καὶ λίην κείνός γε εὐοκίτι κείται ὀλέθρῳ: ὧς ἀπόλοιτο καὶ ἄλλος, ὅτις τοιαῦτά γε ῥέζοι: ἀλλὰ μοι ἄμφ' Ὀδυσῆϊ δαίφροني δαίεται ἦτορ, δυσμόρῳ, ὃς δὴ δηθὰ φίλων ἄπο πῆματα πάσχει νήσῳ ἐν ἀμφιρῦτῃ, ὅθι τ' ὀμφαλὸς ἐστὶ θαλάσσης. νῆσος δεινδρήεσσα, θεὰ δ' ἐν δώματι ναίει, Ἄτλαντος θυγάτηρ ὀλοόφρονος, ὃς

τε θαλάσσης πάσης βένθεα οἶδεν, ἔχει δέ τε κίονας αὐτὸς μακράς, αἱ γαῖάν τε καὶ οὐρανὸν ἀμφὶς ἔχουσιν. τοῦ θυγάτηρ δῦστηνον ὄδυρόμενον κατερύκει, αἶει δὲ μαλακοῖσι καὶ αἰμυλίοισι λόγοισιν θέλγει, ὅπως Ἴθάκης ἐπιλήσεται: αὐτὰρ Ὀδυσσεύς, ἰέμενος καὶ καπνὸν ἀποθρῶσκοντα νοῆσαι ἦς γαίης, θανέειν ἰμείρεται. οὐδέ νυ σοὶ περ ἐντρέπεται φίλον ἦτορ, Ὀλύμπιε. οὐ νύ τ' Ὀδυσσεύς Ἀργείων παρὰ νηυσὶ χαρίζετο ἱερὰ ῥέζων Τροίῃ ἐν εὐρείῃ; τί νύ οἱ τόσον ὠδύσσαο, Ζεῦ; τὴν δ' ἀπαμειβόμενος προσέφη νεφεληγερέτα Ζεὺς: τέκνον ἐμόν, ποῖόν σε ἔπος φύγεν ἕρκος ὀδόντων. πῶς ἂν ἔπειτ' Ὀδυσῆος ἐγὼ θεῖοιο λαθοίμην, ὃς περὶ μὲν νόον ἐστὶ βροτῶν, περὶ δ' ἱρὰ θεοῖσιν ἀθανάτοισιν ἔδωκε, τοὶ οὐρανὸν εὐρὺν ἔχουσιν; ἀλλὰ Ποσειδάων γαίηχος ἀσκελὲς αἶει Κύκλωπος κεχόλωται, ὃν ὀφθαλμοῦ ἀλάωσεν, ἀντίθειον Πολύφημον, ὅου κράτος ἐστὶ μέγιστον πᾶσιν Κυκλώπεσσι: Θόωσα δὲ μιν τέκε νύμφη, Φόρκυνος θυγάτηρ ἄλῶς ἀτρυγέτοιο μέδοντος, ἐν σπέσσι γλαφυροῖσι Ποσειδάωνι μιγεῖσα. ἐκ τοῦ δὴ Ὀδυσῆα Ποσειδάων ἐνοσίχθων οὐ τι κατακτείνει, πλάζει δ' ἀπὸ πατρίδος αἴης. ἀλλ' ἄγεθ', ἡμεῖς οἶδε περιφραζόμεθα πάντες νόστον, ὅπως ἔλθῃσι: Ποσειδάων δὲ μεθήσει ὃν χόλον: οὐ μὲν γὰρ τι δυνήσεται ἀντία πάντων ἀθανάτων ἀέκητι θεῶν ἐριδαινέμεν οἶος.

4 Korean

- Language code: ko-KR
- Font: NanumMyeongjo
- Source: [Wikipedia Seoul entry](#), accessed 2016-12-08.

서울특별시(서울□□□)는 대한민국의 수도이자 최대 도시이다. 기원전 18년부터 475년까지 백제의 첫 수도인 위례성(□□□)이었으며, 조선 건국 2년 후인 1394년 조선의 수도가 된 이후로 대한민국의 정부중앙청사, 청와대의 소재지가 되기까지 1,000년 이상 백제, 조선, 대한제국, 일제 강점기, 미 군정기, 대한민국의 정치·경제·사회·문화의 중심지 역할을 해왔다. 도시의 중앙으로 한강이 흐르고, 북한산, 관악산, 도봉산, 불암산, 인능산, 인왕산, 청계산 등 산으로 둘러싸인 분지 지형의 도시이다. 동서 간의 거리는 36.78km, 남북 간의 거리는 30.3km이며, 넓이는 605.25km²이다. 면적은 대한민국 국토의 0.6%에 불과하지만, 대한민국 인구의 1/5 정도 되는 약 990만여 명이 살고 있어서 인구밀도가 매우 높다. 1986년 아시안 게임과 1988년 하계 올림픽, 2002년 FIFA 월드컵, 2010년 G20 정상회의, 2012년 핵안보정상회의가 개최된 국제적인 도시이다. 시청과 정부서울청사가 소재하는 중구와 종로구의 세종대로를 중심으로 도심이 형성하며, 행정구역은 25개 구로 이루어진다. 2013년 예산은 20조 6287억 원으로 대한민국 지방자치단체 중 예산 규모가 가장 크다.

5 Hungarian (Magyar)

- Language code: hu-HU
- Font: Latin Modern

- Source: [Wikipedia Budapest entry](#), accessed 2016-07-25.

Budapest Magyarország fővárosa, egyben legnagyobb és legnépesebb városa, az Európai Unió kilencedik legnépesebb városa. Budapest az ország politikai, kulturális, kereskedelmi, ipari és közlekedési központja. Emellett Pest megye székhelye is, bár közigazgatásilag nem része annak. 2011-ben regisztrált lakónépessége meghaladta az 1,7 (elővárosokkal együtt pedig a 2,5) millió főt. A város lakossága 1989-ben volt a legnagyobb, az akkori lélekszám 2,1 millió fő volt, és az ezt követő szuburbanizáció után is megmaradt az ország legnépesebb városának. Területe mintegy 525 négyzetkilométer, ebben a tekintetben is első az ország települései között.

Budapest története a keltákig nyúlik vissza, mivel a város eredetileg kelta település volt. A rómaiak által létrehozott, a Danubius folyó jobb partján fekvő Aquincum és a bal parti Contra-Aquincum volt Pannonia provincia székhelye. A magyarok a 9. században érkeztek erre a területre. Az első általuk létrehozott települést a mongolok 1241 és 1242 között, a tatárjárás során teljesen kifosztották és elpusztították. A 15. században a helyreállított város lett a magyarországi reneszánsz humanizmus központja. Ezt követően százötven évnyi török hódoltság telepedett a városra és csak a 18-19. században tudott újra fejlődni, ezt fokozta, hogy a kiegyezést követően Bécs mellett az Osztrák–Magyar Monarchia fővárosa volt. 1873-ban egyesítették Budát, Pestet és Óbudát, ekkor jött létre Budapest. Ebben az időben épültek a város leghíresebb épületei, és ekkor nőtte ki magát világvárossá.

Budapest kedvelt idegenforgalmi célpont is, 2012-ben 4 821 600 turista kereste fel, ezzel az Euromonitor International felmérése szerint a világ 27. és Európa 7. leglátogatottabb városa. Budapesten több világörökségi helyszín is található, többek között a Duna-part látképe, a Budai Várnegyed, az Andrásy út, a Hősök tere és a Millenniumi Földalatti Vasút, amely a második legrégebbi a világon a londoni után. Turisztikailag jelentősek még a város Duna-hídjai, valamint gyógyfürdői, mivel Budapest a világ legtöbb gyógyfürdőjével rendelkező fővárosa. Itt található még a világ legnagyobb termásvizes barlangrendszere, Európa legnagyobb zsinagógája (a Dohány utcai), valamint a magyar Országház is, amely a világ harmadik legnagyobb parlamenti épülete, és a Szent Korona jelenlegi otthona.

Budapest Közép-Európa egyik pénzügyi központja és a világ 100. legnagyobb GDP-jét termelő városa. Az amerikai Forbes magazin szerint Budapest a 7. „legidillibb európai város”. A UCityGuides a világ 9. legszebb városának sorolja. A befolyásos amerikai Condé Nast Traveler utazási magazin által megkérdezett 1,3 millió személy véleménye alapján „Budapest a világ második legjobb városa”. A magyar főváros továbbá egyike Európa 10 legkedveltebb bevásárlóvárosának az Economist magazin kutatásai szerint, ugyanis itt található Közép- és Kelet-Európa legtöbb bevásárlóközpontja.

6. Russian

- Language code: ru-RU
- Font: CMU Serif (Computer Modern Unicode)
- Source: [Wikipedia Moscow entry](#), accessed 2016-12-09.

Москва — столица Российской Федерации, город федерального значения, административный центр Центрального федерального округа и центр Московской области, в состав которой не входит. Крупнейший по численности населения

город России и её субъект — 12 330 126 чел. (2016), самый населённый из городов, полностью расположенных в Европе, входит в первую десятку городов мира по численности населения. Центр Московской городской агломерации.

Историческая столица Великого княжества Московского, Русского царства, Российской империи (в 1728—1730 годах), Советской России и СССР. Город-герой. В Москве находятся федеральные органы государственной власти Российской Федерации (за исключением Конституционного суда), посольства иностранных государств, штаб-квартиры большинства крупнейших российских коммерческих организаций и общественных объединений.

Расположена на реке Москве в центре Восточно-Европейской равнины, в междуречье Оки и Волги. Как субъект федерации Москва граничит с Московской и Калужской областями.

Москва — важный туристический центр России. Московский Кремль, Красная площадь, Новодевичий монастырь и Церковь Вознесения в Коломенском входят в список Всемирного наследия ЮНЕСКО. Она является важнейшим транспортным узлом. Город обслуживают 5 аэропортов, 9 железнодорожных вокзалов, 3 речных порта (имеется речное сообщение с морями бассейнов Атлантического и Северного Ледовитого океанов). С 1935 года в Москве работает метрополитен.

7 Spanish

- Language code: es-ES
- Font: Latin Modern
- Source: [Wikipedia Madrid entry](#), accessed 2016-07-25.

Madrid es un municipio y ciudad de España. La localidad, con categoría histórica de villa, es la capital del Estado y de la Comunidad de Madrid. También conocida como la Villa y Corte, es la ciudad más poblada del país, con 3 165 235 habitantes empadronados según datos del INE de 2014 mientras que, con la inclusión de su área metropolitana la cifra de población asciende a 6 543 031 habitantes, siendo por ello la tercera o cuarta área metropolitana de la Unión Europea, según la fuente, por detrás de las de París y Londres, y en algunas fuentes detrás también de la Región del Ruhr, así como la tercera ciudad más poblada de la Unión Europea, por detrás de Berlín y Londres. Madrid ocupa el puesto nº 38 en la lista Economist Intelligence Unit de ciudades con mejor calidad de vida del mundo.

Como capital del Estado, Madrid alberga las sedes del Gobierno, las Cortes Generales, ministerios, instituciones y organismos asociados, así como la residencia oficial de los reyes de España¹² y del presidente del Gobierno. En el plano económico, es la cuarta ciudad más rica de Europa, tras Londres, París y Moscú. Para el 2009, el 50,1 % de los ingresos de las 5000 principales empresas españolas son generados por sociedades con sede social en Madrid, las cuales representan el 31,8 % de ellas. Es sede del 3.er mayor mercado de valores de Europa, y 2^a en el ámbito latinoamericano (Latibex) y de varias de las más grandes corporaciones del mundo. Es la 8.^a ciudad del mundo con mayor presencia de multinacionales, tras Pekín y por delante de Dubái, París y Nueva York.

En el plano internacional acoge la sede central de la Organización Mundial del Turismo (OMT), perteneciente a la ONU, la sede de la Organización Internacional de Comisiones de Valores (OICV), la sede de la Secretaría General

Iberoamericana (SEGIB), la sede de la Organización de Estados Iberoamericanos para la Educación, la Ciencia y la Cultura (OEI), la Organización Iberoamericana de Juventud (OIJ), y la sede de Public Interest Oversight Board (PIOB). También alberga las principales instituciones internacionales reguladoras y difusoras del idioma español: la Comisión Permanente de la Asociación de Academias de la Lengua Española, y sedes centrales de la Real Academia Española (RAE), del Instituto Cervantes y de la Fundación del Español Urgente (Fundéu). Madrid organiza ferias como FITUR, Madrid Fusión, ARCO, SIMO TCI, el Salón del Automóvil y la Cibeles Madrid Fashion Week.

Es un influyente centro cultural y cuenta con museos de referencia internacional, entre los que destacan el Museo del Prado, el Museo Nacional Centro de Arte Reina Sofía, el Thyssen-Bornemisza y CaixaForum Madrid, que ocupan, respectivamente, el 12º, 18º, 67º y 80º puesto entre los museos más visitados del mundo.

Los orígenes de la ciudad son objeto de revisión tras recientes hallazgos de enterramientos visigodos así como de restos que se remontan a los carpetanos o periodo prerromano. Las excavaciones arqueológicas también arrojan restos que se atribuyen al Madrid romano. Estos hallazgos de época visigoda han venido a confirmar que el posterior asentamiento fortificado musulmán de Mağriṭ (del siglo IX) se había asentado sobre un vicus visigodo del siglo VII llamado Matrice o matriz, arroyo. (AFI [ma ri t]),

No sería hasta el siglo XI cuando Madrid fue incorporada a la Corona de Castilla, tras su conquista por Alfonso VI de León y Castilla en 1083. Fue designada como sede de la Corte por el rey Felipe II en 1561, convirtiéndose en la primera capital permanente de la monarquía española. Desde el Renacimiento hasta la actualidad ha sido capital de España y sede del Gobierno y la administración del Estado salvo breves intervalos de tiempo: entre los años de 1601 y 1606 la capitalidad pasó a Valladolid; durante la Guerra de la Independencia Española, en que la capital se trasladó a Sevilla; en 1808 y en 1810 a Cádiz; durante la Guerra Civil, cuando el Gobierno republicano se trasladó primero a Valencia y después a Barcelona, durante la Guerra Civil el bando nacional se estableció en Burgos, y tras finalizar esta se estableció allí la capital hasta el 18 de octubre de 1939 que se volvió a trasladar a Madrid.

8 Vietnamese

- Language code: vi-VN
- Font: Latin Modern
- Source: [Wikipedia Hà Nội entry](#), accessed 2016-07-25.

Hà Nội là thủ đô của nước Cộng hoà Xã hội chủ nghĩa Việt Nam và cũng là kinh đô của rất nhiều vương triều Việt cổ. Do đó, lịch sử Hà Nội gắn liền với sự thăng trầm của lịch sử Việt Nam qua các thời kỳ. Hà Nội là thành phố lớn nhất Việt Nam về diện tích với 3328,9 km² sau đợt mở rộng hành chính năm 2008, đồng thời cũng là địa phương đứng thứ nhì về dân số với 7.500.000 người (năm 2015). Hiện nay, thủ đô Hà Nội và thành phố Hồ Chí Minh là đô thị loại đặc biệt của Việt Nam.

Hà Nội nằm giữa đồng bằng sông Hồng trù phú, nơi đây đã sớm trở thành 1 trung tâm chính trị, kinh tế và văn hóa ngay từ những buổi đầu của lịch sử Việt Nam. Năm 1010, Lý Công Uẩn, vị vua đầu tiên của nhà Lý, quyết định xây dựng kinh đô mới ở vùng đất này với cái tên Thăng Long. Trong suốt thời kỳ của các triều đại Lý, Trần, Lê, Mạc, kinh thành Thăng Long là trung tâm

văn hóa, giáo dục và buôn bán của cả nước. Khi Tây Sơn rời nhà Nguyễn lên nắm quyền trị vì, kinh đô được chuyển về Huế và Thăng Long bắt đầu mang tên Hà Nội từ năm 1831, dưới thời vua Minh Mạng. Năm 1902, Hà Nội trở thành thủ đô của Liên bang Đông Dương và được người Pháp xây dựng, quy hoạch lại, được coi là tiểu Paris của Phương Đông thời bấy giờ. Trải qua hai cuộc chiến tranh, Hà Nội là thủ đô của nước CHXHCN Việt Nam thống nhất và giữ vai trò này cho tới ngày nay.

Sau đợt mở rộng địa giới hành chính vào tháng 8 năm 2008, Hà Nội hiện nay gồm 12 quận, 1 thị xã và 17 huyện ngoại thành. Hiện nay, Hà Nội và Thành phố Hồ Chí Minh là hai trung tâm kinh tế - xã hội đặc biệt quan trọng của Việt Nam.

9 General Language and Script Support

2016-12-08: this section is preliminary.

HTML output works well if your source includes Unicode characters. Hyphenation is a non-issue, since text reflows and there is no flush-right support.

It is a different story for L^AT_EX, PDF, and print. So that is the focus in this sample document. You may not recognize the next subsection, but an explanation follows.

9.1

ὁ δὴ ἦτοι πρῶτος ἐστὶν ἡ οὓ.

9.2 Languages, Scripts, Subdivisions

The previous subsection has an `xml:lang` attribute on the `subsection` element, whose value is `el` indicating the Modern Greek language. So the entire subsection is being treated as characters from that language. There is a title, but it is not being affected now, and so is empty.

The behavior should be entirely similar for `part`, `chapter`, `section`, and `subsubsection`.

9.3 Language and Script Examples

Other elements have language support. We are starting with `foreign`, which also italicizes the content.

From the [Kermit Project](#) a [startling phrase](#) in many languages.

- Modern Greek (monotonic), Italic: *Μπορώ να φάω σπασμένα γυαλιά χωρίς να πάθω τίποτα.*
- Modern Greek (polytonic), Italic: *Μπορῶ νά φάω σπασμένα γυαλιά χωρίς νά πάθω τίποτα.*
- Korean, Italic: *나는 유리를 먹을 수 있어요. 그래도 아프지 않아요*
- Hungarian, Italic: *Meg tudom enni az üveget, nem lesz tőle bajom.*
- Russian, Italic: *Я могу есть стекло, оно мне не вредит.*

- Spanish, Italic: *Puedo comer vidrio, no me hace daño.*
- Vietnamese, Italic: *Tôi có thể ăn thủy tinh mà không hại gì.*

List 9.1: I Can Eat Glass

10 Font Tests

We place various blocks of Unicode characters here to determine the minimum configuration necessary to make them render. Alan Wood’s [Unicode Resources](#) site has been helpful in formulating these tests.

Basic Latin, U+0000–U+007F These 95 characters are the most basic, and should all render using `xelatex` with no special setup. U+0000 to U+001F are control codes and not used here. U+007F is also a control code and so is excluded. We have also replaced reserved \LaTeX characters by their MathBook XML equivalent empty elements

	0	1	2	3	4	5	6	7	8	9	A	B	C	D	E	F
002_	!	"	#	\$	%	&	'	()	*	+	,	-	.	/	
003_	0	1	2	3	4	5	6	7	8	9	:	;	<	=	>	?
004_	@	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
005_	P	Q	R	S	T	U	V	W	X	Y	Z	[\]	^	_
006_	`	a	b	c	d	e	f	g	h	i	j	k	l	m	n	o
007_	p	q	r	s	t	u	v	w	x	y	z	{		}	~	

Table 10.1: Basic Latin, Regular

Monospace, Basic Latin, U+0000–U+007F These are exactly the same characters as above, but now we wrap them in the `<c>` element intended for inline use. This does not test all verbatim situations but is a good simple first test.

	0	1	2	3	4	5	6	7	8	9	A	B	C	D	E	F
002_	!	"	#	\$	%	&	'	()	*	+	,	-	.	/	
003_	0	1	2	3	4	5	6	7	8	9	:	;	<	=	>	?
004_	@	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
005_	P	Q	R	S	T	U	V	W	X	Y	Z	[\]	^	_
006_	`	a	b	c	d	e	f	g	h	i	j	k	l	m	n	o
007_	p	q	r	s	t	u	v	w	x	y	z	{		}	~	

Table 10.2: Basic Latin, Monospace

Note that the single and double quotes are upright and dumb, not curly and smart: " ' " ' " '. The zero is distinguished from the capital “oh”: 0 0 0 0. And the numeral one is slightly different from the lower-case “ell”: 1 l 1 l 1 l. The hyphen should be short and not expanded into some other kind of dash: - - -. These characters should all cut/paste out of a PDF into a text editor with no conversion to other characters.

Note also that we have entered *all* these characters into the source with the `�NN`; XML notation, but needed to “protect” the question mark, since we use it in the \LaTeX verbatim routine to mark the beginning and end.

Latin-1 Supplement, U+0080–U+00FF These 94 characters should all render using either `pdflatex` or `xelatex` with no special setup. U+0080 to U+009F are control codes and not used here. U+00A0 (non-breaking space) and U+00AD (soft hyphen) are also excluded.

	0	1	2	3	4	5	6	7	8	9	A	B	C	D	E	F
00A_		ı	ċ	£	¤	¥	¦	§	¨	©	ª	«	¬		®	¯
00B_	°	±	²	³	´	µ	¶	·	¸	¹	º	»	¼	½	¾	¿
00C_	À	Á	Â	Ã	Ä	Å	Æ	Ç	È	É	Ê	Ë	Ì	Í	Î	Ï
00D_	Ð	Ñ	Ò	Ó	Ô	Õ	Ö	×	Ø	Ù	Ú	Û	Ü	Ý	Þ	ß
00E_	à	á	â	ã	ä	å	æ	ç	è	é	ê	ë	ì	í	î	ï
00F_	ð	ñ	ò	ó	ô	õ	ö	÷	ø	ù	ú	û	ü	ý	þ	ÿ

Table 10.3: Latin-1 Supplement, Regular

Monospace, Latin-1 Supplement, U+0080–U+00FF The same 94 characters as above, wrapped in a `<c>` element as if being used inside a sentence. These will all render with `xelatex` and none will render with `pdflatex` (so there is just blank space below). If we improve the latter, then these will get duplicated into the sample article.

	0	1	2	3	4	5	6	7	8	9	A	B	C	D	E	F
00A_		ı	ċ	£	¤	¥	¦	§	¨	©	ª	«	¬		®	¯
00B_	°	±	²	³	´	µ	¶	·	¸	¹	º	»	¼	½	¾	¿
00C_	À	Á	Â	Ã	Ä	Å	Æ	Ç	È	É	Ê	Ë	Ì	Í	Î	Ï
00D_	Ð	Ñ	Ò	Ó	Ô	Õ	Ö	×	Ø	Ù	Ú	Û	Ü	Ý	Þ	ß
00E_	à	á	â	ã	ä	å	æ	ç	è	é	ê	ë	ì	í	î	ï
00F_	ð	ñ	ò	ó	ô	õ	ö	÷	ø	ù	ú	û	ü	ý	þ	ÿ

Table 10.4: Latin-1 Supplement, Monospace

Latin Extended-A, U+0100–U+017F Good success rendering with `xelatex` and no extra setup. About 25% of these are missing when rendered with `pdflatex`.

	0	1	2	3	4	5	6	7	8	9	A	B	C	D	E	F
010_	Ā	ā	Ă	ă	Ą	ą	Ć	ć	Ĉ	ĉ	Č	č	Ď	ď		
011_	Đ	đ	Ē	ē	Ĕ	ĕ	Ė	ė	Ę	ę	Ě	ě	Ĝ	ĝ	Ğ	ğ
012_	Ġ	ġ	Ģ	ģ	Ĥ	ĥ	Ħ	ħ	Ĩ	ĩ	Ī	ī	Ĵ	ĵ	Ĳ	ĳ
013_	Ĭ	ĭ	Ĵ	ĵ	Ķ	ķ			Ĺ	ĺ	Ł	ł	Ł	ł	Ł	ł
014_	Ŧ	ŧ	Ń	ń	Ņ	ņ	Ň	ň			Đ	đ	Ō	ō	Ŏ	ŏ
015_	Ŕ	ŕ	Œ	œ	Ř	ř	Ŗ	ŗ	Ř	ř	Ś	ś	Ŝ	ŝ	Ş	ş
016_	Š	š	Ţ	ţ	Ť	ť			Ū	ū	Ū	ū	Ŭ	ŭ	Ů	ů
017_	Ű	ű	Ū	ū	Ŵ	ŵ	Ŷ	ŷ	Ÿ	Ž	ž	Ž	ž	Ž	ž	ƒ

Table 10.5: Latin Extended-A

Rendered with `xelatex` and no special setup, I seem to be missing only four characters:

- U+0138 (LATIN SMALL LETTER KRA, Greenlandic, removed 1973)

- U+0149 (LATIN SMALL LETTER N PRECEDED BY APOSTROPHE, Afrikaans, deprecated as of Unicode version 5.2.0)
- U+0166 (LATIN CAPITAL LETTER T WITH STROKE, Northern Sámi alphabet, used in northern parts of Norway, Sweden and Finland)
- U+0167 (LATIN SMALL LETTER T WITH STROKE, Northern Sámi alphabet, used in northern parts of Norway, Sweden and Finland)

Latin Extended-B, U+0180–U+024F Rendering with `xelatex` and no extra setup, maybe 50% missing, and some constructions of accents are clearly wrong. Almost none of these appear when rendered with `pdfLatex`. (When processed with `luaLatex` the incorrectly accented characters are not even visible.)

	0	1	2	3	4	5	6	7	8	9	A	B	C	D	E	F
018_																Ǝ
019_			f													
01A_	Ó	ó														Ů
01B_	ư															
01C_														Ǻ	ǻ	Ǫ
01D_	ı	Ŏ	ö	Ů	ů	Ů	ū	Ú	ú	Ů	ű	Ù	ù	ə	Ǻ	ǻ
01E_	À	ā	Æ	æ			Ĝ	ğ	K	k	Q	q	Q	q̄		
01F_	Ĵ				Ĝ	g			N	n	Á	á	É	é	Ó	ó
020_	Ǻ	ǻ	À	á	È	è	E	ê	Ì	ì	I	î	Ö	ö	O	ô
021_	Ř	ř	R	ŕ	Ů	ů	U	û	Ş	ş	Ṭ	ṭ			H	ħ
022_							A	á			Ö	ö	Ő	ő	O	ó
023_	O	ō	Y	ȳ				J								

Table 10.6: Latin Extended-B

This table is left-over and will be redone, perhaps.

Name	Range	Samples	LaTeX
Latin Extended-A	U+0100 – U+017F	Ą ą Ĳ ĳ	TeXLive
Latin Extended Additional	U+1E00 – U+1EFF	Ẁ Ẃ Ẅ	≈TeXLive
Cyrillic	U+0400 – U+04FF		polyglossia
Arabic	U+0600 – U+06FF		polyglossia
General Punctuation	U+2000 – U+206F	— “ ‰ ※	TeXLive
Letterlike Symbols	U+2100 – U+214F	°C № ™	≈TeXLive

Table 10.7: Sample Unicode Characters