

The World's Writing Systems: A XeLaTeX document showing typesetting in a variety of writing systems

Jake Warde

August 12, 2013

Sections 1-5 show examples of in different writing systems. The sections are organized according to recognized categories of writing systems. The XeTeX Companion (Section 3.1.3) contains an overview of types of writing systems.¹ Another recommended resource on writing systems is *Omniglot: The online encyclopedia of writing systems and languages*² Sections 6-7 are a summary of the technologies used to typeset text for all the world's languages. The document may also be a resource to someone learning to use XeLaTeX. XeLaTeX source is here: myprojectnotebook.com. No claims to being a XeLaTeX expert are made. In fact the document was created while learning to use XeLaTeX.

1 Abjads-Consonant Alphabets

Abjads, or consonant alphabets, represent consonants only, or consonants plus some vowels. Full vowel indication (vocalisation) can be added, usually by means of diacritics, but this is not usually done. Most of abjads, with the exception of Ugaritic, are written from right to left. There are three abjads currently in use: Arabic, Hebrew and Syriac. Abjads, or consonant alphabets, represent consonants only, or consonants plus some vowels. For more about abjads visit: <http://www.omniglot.com/writing/alphabetic.htm>

1.1 Arabic

عَلَى يَعْطَلُ وَأَنْ الْوَطَنَ شَانَ إِعْلَاءِ فِيهِ جُضِدٌ كُلُّ يَبْدُلُ وَأَنْ وَجِيرَانِهِ أَهْلِهِ وَمَعَ نَفْسِهِ مَعَ وَصَادِقًا أَمِينًا يَكُونُ أَنْ الْإِنْسَانَ عَلَى يَجِبُ
لِلتَّصْفِيَةِ مَثَالٌ وَهَذَا الْخَاصَّةِ الْمُنَوَّعَةِ عَلَى الْعَامَّةِ الْمُنَوَّعَةِ يَفْعَلُ وَأَنْ إِلَّا ذَلِكَ لَهُ يَتَمَّ وَلَنْ . لِلنَّاسِ السَّعَادَةَ يَجْلِبُ مَا

¹<http://xml.web.cern.ch/XML/lgc2/xetexmain.pdf>

²<http://www.omniglot.com/index.htm>

Note: The sample text in Arabic is from the following website offering Modern Standard Classical Arabic courses.<http://www.dalilusa.com>

1.2 Hebrew

רב־לשוני. ובמהשוב טקסט בפריסת בגופנים, וביישומים, הפעלה

Note: The web page from which the above text was copied provides information about the characteristics of the script used to write Hebrew. <http://rishida.net/scripts/samples/hebrew>

1.3 Syriac

ܫܘܪܝܝܬܐ ܕܥܘܢܝܢܐ ܕܩܘܪܕܐ ܕܩܘܪܕܐ ܕܩܘܪܕܐ ܕܩܘܪܕܐ ܕܩܘܪܕܐ

1.4 Phoenician

The Phoenician alphabet, called by convention the Proto-Canaanite alphabet for inscriptions older than around 1200 BCE, is also an abjad. It was used for the writing of Phoenician, a Northern Semitic language, used by the civilization of Phoenicia.

Phoenician became one of the most widely used writing systems, spread by Phoenician merchants across the Mediterranean world, where it was assimilated by many other cultures and evolved. The Aramaic alphabet, a modified form of Phoenician, was the ancestor of modern Arabic script, while Hebrew script is a stylistic variant of the Aramaic script. The Greek alphabet (and by extension its descendants such as the Latin, the Cyrillic and the Coptic), was a direct successor of Phoenician, though certain letter values were changed to represent vowels.³

𐤀𐤁𐤂𐤃𐤄𐤅𐤆𐤇𐤈𐤉𐤊𐤋𐤌𐤍𐤎𐤏𐤐𐤑𐤒𐤓𐤔𐤕𐤖𐤗𐤘𐤙𐤚𐤛𐤜𐤝𐤞𐤟𐤠𐤡𐤢𐤣𐤤𐤥𐤦𐤧𐤨𐤩𐤪𐤫𐤬𐤭𐤮𐤯𐤰𐤱𐤲𐤳𐤴𐤵𐤶𐤷𐤸𐤹𐤺𐤻𐤼𐤽𐤾𐤿𐥀𐥁𐥂𐥃𐥄𐥅𐥆𐥇𐥈𐥉𐥊𐥋𐥌𐥍𐥎𐥏𐥐𐥑𐥒𐥓𐥔𐥕𐥖𐥗𐥘𐥙𐥚𐥛𐥜𐥝𐥞𐥟𐥠𐥡𐥢𐥣𐥤𐥥𐥦𐥧𐥨𐥩𐥪𐥫𐥬𐥭𐥮𐥯𐥰𐥱𐥲𐥳𐥴𐥵𐥶𐥷𐥸𐥹𐥺𐥻𐥼𐥽𐥾𐥿𐧀𐧁𐧂𐧃𐧄𐧅𐧆𐧇𐧈𐧉𐧊𐧋𐧌𐧍𐧎𐧏𐧐𐧑𐧒𐧓𐧔𐧕𐧖𐧗𐧘𐧙𐧚𐧛𐧜𐧝𐧞𐧟𐧠𐧡𐧢𐧣𐧤𐧥𐧦𐧧𐧨𐧩𐧪𐧫𐧬𐧭𐧮𐧯𐧰𐧱𐧲𐧳𐧴𐧵𐧶𐧷𐧸𐧹𐧺𐧻𐧼𐧽𐧾𐧿𐨀𐨁𐨂𐨃𐨄𐨅𐨆𐨇𐨈𐨉𐨊𐨋𐨌𐨍𐨎𐨏𐨐𐨑𐨒𐨓𐨔𐨕𐨖𐨗𐨘𐨙𐨚𐨛𐨜𐨝𐨞𐨟𐨠𐨡𐨢𐨣𐨤𐨥𐨦𐨧𐨨𐨩𐨪𐨫𐨬𐨭𐨮𐨯𐨰𐨱𐨲𐨳𐨴𐨵𐨶𐨷𐨹𐨺𐨸𐨻𐨼𐨽𐨾𐨿𐩀𐩁𐩂𐩃𐩄𐩅𐩆𐩇𐩈𐩉𐩊𐩋𐩌𐩍𐩎𐩏𐩐𐩑𐩒𐩓𐩔𐩕𐩖𐩗𐩘𐩙𐩚𐩛𐩜𐩝𐩞𐩟𐩠𐩡𐩢𐩣𐩤𐩥𐩦𐩧𐩨𐩩𐩪𐩫𐩬𐩭𐩮𐩯𐩰𐩱𐩲𐩳𐩴𐩵𐩶𐩷𐩸𐩹𐩺𐩻𐩼𐩽𐩾𐩿𐪀𐪁𐪂𐪃𐪄𐪅𐪆𐪇𐪈𐪉𐪊𐪋𐪌𐪍𐪎𐪏𐪐𐪑𐪒𐪓𐪔𐪕𐪖𐪗𐪘𐪙𐪚𐪛𐪜𐪝𐪞𐪟𐪠𐪡𐪢𐪣𐪤𐪥𐪦𐪧𐪨𐪩𐪪𐪫𐪬𐪭𐪮𐪯𐪰𐪱𐪲𐪳𐪴𐪵𐪶𐪷𐪸𐪹𐪺𐪻𐪼𐪽𐪾𐪿𐫀𐫁𐫂𐫃𐫄𐫅𐫆𐫇𐫈𐫉𐫊𐫋𐫌𐫍𐫎𐫏𐫐𐫑𐫒𐫓𐫔𐫕𐫖𐫗𐫘𐫙𐫚𐫛𐫜𐫝𐫞𐫟𐫠𐫡𐫢𐫣𐫤𐫦𐫥𐫧𐫨𐫩𐫪𐫫𐫬𐫭𐫮𐫯𐫰𐫱𐫲𐫳𐫴𐫵𐫶𐫷𐫸𐫹𐫺𐫻𐫼𐫽𐫾𐫿𐬀𐬁𐬂𐬃𐬄𐬅𐬆𐬇𐬈𐬉𐬊𐬋𐬌𐬍𐬎𐬏𐬐𐬑𐬒𐬓𐬔𐬕𐬖𐬗𐬘𐬙𐬚𐬛𐬜𐬝𐬞𐬟𐬠𐬡𐬢𐬣𐬤𐬥𐬦𐬧𐬨𐬩𐬪𐬫𐬬𐬭𐬮𐬯𐬰𐬱𐬲𐬳𐬴𐬵𐬶𐬷𐬸𐬹𐬺𐬻𐬼𐬽𐬾𐬿𐭀𐭁𐭂𐭃𐭄𐭅𐭆𐭇𐭈𐭉𐭊𐭋𐭌𐭍𐭎𐭏𐭐𐭑𐭒𐭓𐭔𐭕𐭖𐭗𐭘𐭙𐭚𐭛𐭜𐭝𐭞𐭟𐭠𐭡𐭢𐭣𐭤𐭥𐭦𐭧𐭨𐭩𐭪𐭫𐭬𐭭𐭮𐭯𐭰𐭱𐭲𐭳𐭴𐭵𐭶𐭷𐭸𐭹𐭺𐭻𐭼𐭽𐭾𐭿𐮀𐮁𐮂𐮃𐮄𐮅𐮆𐮇𐮈𐮉𐮊𐮋𐮌𐮍𐮎𐮏𐮐𐮑𐮒𐮓𐮔𐮕𐮖𐮗𐮘𐮙𐮚𐮛𐮜𐮝𐮞𐮟𐮠𐮡𐮢𐮣𐮤𐮥𐮦𐮧𐮨𐮩𐮪𐮫𐮬𐮭𐮮𐮯𐮰𐮱𐮲𐮳𐮴𐮵𐮶𐮷𐮸𐮹𐮺𐮻𐮼𐮽𐮾𐮿𐯀𐯁𐯂𐯃𐯄𐯅𐯆𐯇𐯈𐯉𐯊𐯋𐯌𐯍𐯎𐯏𐯐𐯑𐯒𐯓𐯔𐯕𐯖𐯗𐯘𐯙𐯚𐯛𐯜𐯝𐯞𐯟𐯠𐯡𐯢𐯣𐯤𐯥𐯦𐯧𐯨𐯩𐯪𐯫𐯬𐯭𐯮𐯯𐯰𐯱𐯲𐯳𐯴𐯵𐯶𐯷𐯸𐯹𐯺𐯻𐯼𐯽𐯾𐯿𐰀𐰁𐰂𐰃𐰄𐰅𐰆𐰇𐰈𐰉𐰊𐰋𐰌𐰍𐰎𐰏𐰐𐰑𐰒𐰓𐰔𐰕𐰖𐰗𐰘𐰙𐰚𐰛𐰜𐰝𐰞𐰟𐰠𐰡𐰢𐰣𐰤𐰥𐰦𐰧𐰨𐰩𐰪𐰫𐰬𐰭𐰮𐰯𐰰𐰱𐰲𐰳𐰴𐰵𐰶𐰷𐰸𐰹𐰺𐰻𐰼𐰽𐰾𐰿𐱀𐱁𐱂𐱃𐱄𐱅𐱆𐱇𐱈𐱉𐱊𐱋𐱌𐱍𐱎𐱏𐱐𐱑𐱒𐱓𐱔𐱕𐱖𐱗𐱘𐱙𐱚𐱛𐱜𐱝𐱞𐱟𐱠𐱡𐱢𐱣𐱤𐱥𐱦𐱧𐱨𐱩𐱪𐱫𐱬𐱭𐱮𐱯𐱰𐱱𐱲𐱳𐱴𐱵𐱶𐱷𐱸𐱹𐱺𐱻𐱼𐱽𐱾𐱿𐲀𐲁𐲂𐲃𐲄𐲅𐲆𐲇𐲈𐲉𐲊𐲋𐲌𐲍𐲎𐲏𐲐𐲑𐲒𐲓𐲔𐲕𐲖𐲗𐲘𐲙𐲚𐲛𐲜𐲝𐲞𐲟𐲠𐲡𐲢𐲣𐲤𐲥𐲦𐲧𐲨𐲩𐲪𐲫𐲬𐲭𐲮𐲯𐲰𐲱𐲲𐲳𐲴𐲵𐲶𐲷𐲸𐲹𐲺𐲻𐲼𐲽𐲾𐲿𐳀𐳁𐳂𐳃𐳄𐳅𐳆𐳇𐳈𐳉𐳊𐳋𐳌𐳍𐳎𐳏𐳐𐳑𐳒𐳓𐳔𐳕𐳖𐳗𐳘𐳙𐳚𐳛𐳜𐳝𐳞𐳟𐳠𐳡𐳢𐳣𐳤𐳥𐳦𐳧𐳨𐳩𐳪𐳫𐳬𐳭𐳮𐳯𐳰𐳱𐳲𐳳𐳴𐳵𐳶𐳷𐳸𐳹𐳺𐳻𐳼𐳽𐳾𐳿𐴀𐴁𐴂𐴃𐴄𐴅𐴆𐴇𐴈𐴉𐴊𐴋𐴌𐴍𐴎𐴏𐴐𐴑𐴒𐴓𐴔𐴕𐴖𐴗𐴘𐴙𐴚𐴛𐴜𐴝𐴞𐴟𐴠𐴡𐴢𐴣𐴤𐴥𐴦𐴧𐴨𐴩𐴪𐴫𐴬𐴭𐴮𐴯𐴰𐴱𐴲𐴳𐴴𐴵𐴶𐴷𐴸𐴹𐴺𐴻𐴼𐴽𐴾𐴿𐵀𐵁𐵂𐵃𐵄𐵅𐵆𐵇𐵈𐵉𐵊𐵋𐵌𐵍𐵎𐵏𐵐𐵑𐵒𐵓𐵔𐵕𐵖𐵗𐵘𐵙𐵚𐵛𐵜𐵝𐵞𐵟𐵠𐵡𐵢𐵣𐵤𐵥𐵦𐵧𐵨𐵩𐵪𐵫𐵬𐵭𐵮𐵯𐵰𐵱𐵲𐵳𐵴𐵵𐵶𐵷𐵸𐵹𐵺𐵻𐵼𐵽𐵾𐵿𐶀𐶁𐶂𐶃𐶄𐶅𐶆𐶇𐶈𐶉𐶊𐶋𐶌𐶍𐶎𐶏𐶐𐶑𐶒𐶓𐶔𐶕𐶖𐶗𐶘𐶙𐶚𐶛𐶜𐶝𐶞𐶟𐶠𐶡𐶢𐶣𐶤𐶥𐶦𐶧𐶨𐶩𐶪𐶫𐶬𐶭𐶮𐶯𐶰𐶱𐶲𐶳𐶴𐶵𐶶𐶷𐶸𐶹𐶺𐶻𐶼𐶽𐶾𐶿𐷀𐷁𐷂𐷃𐷄𐷅𐷆𐷇𐷈𐷉𐷊𐷋𐷌𐷍𐷎𐷏𐷐𐷑𐷒𐷓𐷔𐷕𐷖𐷗𐷘𐷙𐷚𐷛𐷜𐷝𐷞𐷟𐷠𐷡𐷢𐷣𐷤𐷥𐷦𐷧𐷨𐷩𐷪𐷫𐷬𐷭𐷮𐷯𐷰𐷱𐷲𐷳𐷴𐷵𐷶𐷷𐷸𐷹𐷺𐷻𐷼𐷽𐷾𐷿𐸀𐸁𐸂𐸃𐸄𐸅𐸆𐸇𐸈𐸉𐸊𐸋𐸌𐸍𐸎𐸏𐸐𐸑𐸒𐸓𐸔𐸕𐸖𐸗𐸘𐸙𐸚𐸛𐸜𐸝𐸞𐸟𐸠𐸡𐸢𐸣𐸤𐸥𐸦𐸧𐸨𐸩𐸪𐸫𐸬𐸭𐸮𐸯𐸰𐸱𐸲𐸳𐸴𐸵𐸶𐸷𐸸𐸹𐸺𐸻𐸼𐸽𐸾𐸿𐹀𐹁𐹂𐹃𐹄𐹅𐹆𐹇𐹈𐹉𐹊𐹋𐹌𐹍𐹎𐹏𐹐𐹑𐹒𐹓𐹔𐹕𐹖𐹗𐹘𐹙𐹚𐹛𐹜𐹝𐹞𐹟𐹠𐹡𐹢𐹣𐹤𐹥𐹦𐹧𐹨𐹩𐹪𐹫𐹬𐹭𐹮𐹯𐹰𐹱𐹲𐹳𐹴𐹵𐹶𐹷𐹸𐹹𐹺𐹻𐹼𐹽𐹾𐹿𐺀𐺁𐺂𐺃𐺄𐺅𐺆𐺇𐺈𐺉𐺊𐺋𐺌𐺍𐺎𐺏𐺐𐺑𐺒𐺓𐺔𐺕𐺖𐺗𐺘𐺙𐺚𐺛𐺜𐺝𐺞𐺟𐺠𐺡𐺢𐺣𐺤𐺥𐺦𐺧𐺨𐺩𐺪𐺫𐺬𐺭𐺮𐺯𐺰𐺱𐺲𐺳𐺴𐺵𐺶𐺷𐺸𐺹𐺺𐺻𐺼𐺽𐺾𐺿𐻀𐻁𐻂𐻃𐻄𐻅𐻆𐻇𐻈𐻉𐻊𐻋𐻌𐻍𐻎𐻏𐻐𐻑𐻒𐻓𐻔𐻕𐻖𐻗𐻘𐻙𐻚𐻛𐻜𐻝𐻞𐻟𐻠𐻡𐻢𐻣𐻤𐻥𐻦𐻧𐻨𐻩𐻪𐻫𐻬𐻭𐻮𐻯𐻰𐻱𐻲𐻳𐻴𐻵𐻶𐻷𐻸𐻹𐻺𐻻𐻼𐻽𐻾𐻿𐼀𐼁𐼂𐼃𐼄𐼅𐼆𐼇𐼈𐼉𐼊𐼋𐼌𐼍𐼎𐼏𐼐𐼑𐼒𐼓𐼔𐼕𐼖𐼗𐼘𐼙𐼚𐼛𐼜𐼝𐼞𐼟𐼠𐼡𐼢𐼣𐼤𐼥𐼦𐼧𐼨𐼩𐼪𐼫𐼬𐼭𐼮𐼯𐼰𐼱𐼲𐼳𐼴𐼵𐼶𐼷𐼸𐼹𐼺𐼻𐼼𐼽𐼾𐼿𐽀𐽁𐽂𐽃𐽄𐽅𐽆𐽇𐽋𐽍𐽎𐽏𐽐𐽈𐽉𐽊𐽌𐽑𐽒𐽓𐽔𐽕𐽖𐽗𐽘𐽙𐽚𐽛𐽜𐽝𐽞𐽟𐽠𐽡𐽢𐽣𐽤𐽥𐽦𐽧𐽨𐽩𐽪𐽫𐽬𐽭𐽮𐽯𐽰𐽱𐽲𐽳𐽴𐽵𐽶𐽷𐽸𐽹𐽺𐽻𐽼𐽽𐽾𐽿𐾀𐾁𐾃𐾅𐾂𐾄𐾆𐾇𐾈𐾉𐾊𐾋𐾌𐾍𐾎𐾏𐾐𐾑𐾒𐾓𐾔𐾕𐾖𐾗𐾘𐾙𐾚𐾛𐾜𐾝𐾞𐾟𐾠𐾡𐾢𐾣𐾤𐾥𐾦𐾧𐾨𐾩𐾪𐾫𐾬𐾭𐾮𐾯𐾰𐾱𐾲𐾳𐾴𐾵𐾶𐾷𐾸𐾹𐾺𐾻𐾼𐾽𐾾𐾿𐿀𐿁𐿂𐿃𐿄𐿅𐿆𐿇𐿈𐿉𐿊𐿋𐿌𐿍𐿎𐿏𐿐𐿑𐿒𐿓𐿔𐿕𐿖𐿗𐿘𐿙𐿚𐿛𐿜𐿝𐿞𐿟𐿠𐿡𐿢𐿣𐿤𐿥𐿦𐿧𐿨𐿩𐿪𐿫𐿬𐿭𐿮𐿯𐿰𐿱𐿲𐿳𐿴𐿵𐿶𐿷𐿸𐿹𐿺𐿻𐿼𐿽𐿾𐿿𐀀𐀁𐀂𐀃𐀄𐀅𐀆𐀇𐀈𐀉𐀊𐀋𐀌𐀍𐀎𐀏𐀐𐀑𐀒𐀓𐀔𐀕𐀖𐀗𐀘𐀙𐀚𐀛𐀜𐀝𐀞𐀟𐀠𐀡𐀢𐀣𐀤𐀥𐀦𐀧𐀨𐀩𐀪𐀫𐀬𐀭𐀮𐀯𐀰𐀱𐀲𐀳𐀴𐀵𐀶𐀷𐀸𐀹𐀺𐀻𐀼𐀽𐀾𐀿𐁀𐁁𐁂𐁃𐁄𐁅𐁆𐁇𐁈𐁉𐁊𐁋𐁌𐁍𐁎𐁏𐁐𐁑𐁒𐁓𐁔𐁕𐁖𐁗𐁘𐁙𐁚𐁛𐁜𐁝𐁞𐁟𐁠𐁡𐁢𐁣𐁤𐁥𐁦𐁧𐁨𐁩𐁪𐁫𐁬𐁭𐁮𐁯𐁰𐁱𐁲𐁳𐁴𐁵𐁶𐁷𐁸𐁹𐁺𐁻𐁼𐁽𐁾𐁿𐂀𐂁𐂂𐂃𐂄𐂅𐂆𐂇𐂈𐂉𐂊𐂋𐂌𐂍𐂎𐂏𐂐𐂑𐂒𐂓𐂔𐂕𐂖𐂗𐂘𐂙𐂚𐂛𐂜𐂝𐂞𐂟𐂠𐂡𐂢𐂣𐂤𐂥𐂦𐂧𐂨𐂩𐂪𐂫𐂬𐂭𐂮𐂯𐂰𐂱𐂲𐂳𐂴𐂵𐂶𐂷𐂸𐂹𐂺𐂻𐂼𐂽𐂾𐂿𐃀𐃁𐃂𐃃𐃄𐃅𐃆𐃇𐃈𐃉𐃊𐃋𐃌𐃍𐃎𐃏𐃐𐃑𐃒𐃓𐃔𐃕𐃖𐃗𐃘𐃙𐃚𐃛𐃜𐃝𐃞𐃟𐃠𐃡𐃢𐃣𐃤𐃥𐃦𐃧𐃨𐃩𐃪𐃫𐃬𐃭𐃮𐃯𐃰𐃱𐃲𐃳𐃴𐃵𐃶𐃷𐃸𐃹𐃺𐃻𐃼𐃽𐃾𐃿𐄀𐄁𐄂𐄃𐄄𐄅𐄆𐄇𐄈𐄉𐄊𐄋𐄌𐄍𐄎𐄏𐄐𐄑𐄒𐄓𐄔𐄕𐄖𐄗𐄘𐄙𐄚𐄛𐄜𐄝𐄞𐄟𐄠𐄡𐄢𐄣𐄤𐄥𐄦𐄧𐄨𐄩𐄪𐄫𐄬𐄭𐄮𐄯𐄰𐄱𐄲𐄳𐄴𐄵𐄶𐄷𐄸𐄹𐄺𐄻𐄼𐄽𐄾𐄿𐅀𐅁𐅂𐅃𐅄𐅅𐅆𐅇𐅈𐅉𐅊𐅋𐅌𐅍𐅎𐅏𐅐𐅑𐅒𐅓𐅔𐅕𐅖𐅗𐅘𐅙𐅚𐅛𐅜𐅝𐅞𐅟𐅠𐅡𐅢𐅣𐅤𐅥𐅦𐅧𐅨𐅩𐅪𐅫𐅬𐅭𐅮𐅯𐅰𐅱𐅲𐅳𐅴𐅵𐅶𐅷𐅸𐅹𐅺𐅻𐅼𐅽𐅾𐅿𐆀𐆁𐆂𐆃𐆄𐆅𐆆𐆇𐆈𐆉𐆊𐆋𐆌𐆍𐆎𐆏𐆐𐆑𐆒𐆓𐆔𐆕𐆖𐆗𐆘𐆙𐆚𐆛𐆜𐆝𐆞𐆟𐆠𐆡𐆢𐆣𐆤𐆥𐆦𐆧𐆨𐆩𐆪𐆫𐆬𐆭𐆮𐆯𐆰𐆱𐆲𐆳𐆴𐆵𐆶𐆷𐆸𐆹𐆺𐆻𐆼𐆽𐆾𐆿𐇀𐇁𐇂𐇃𐇄𐇅𐇆𐇇𐇈𐇉𐇊𐇋𐇌𐇍𐇎𐇏𐇐𐇑𐇒𐇓𐇔𐇕𐇖𐇗𐇘𐇙𐇚𐇛𐇜𐇝𐇞𐇟𐇠𐇡𐇢𐇣𐇤𐇥𐇦𐇧𐇨𐇩𐇪𐇫𐇬𐇭𐇮𐇯𐇰𐇱𐇲𐇳𐇴𐇵𐇶𐇷𐇸𐇹𐇺𐇻𐇼𐇽𐇾𐇿𐈀𐈁𐈂𐈃𐈄𐈅𐈆𐈇𐈈𐈉𐈊𐈋𐈌𐈍𐈎𐈏𐈐𐈑𐈒𐈓𐈔𐈕𐈖𐈗𐈘𐈙𐈚𐈛𐈜𐈝𐈞𐈟𐈠𐈡𐈢𐈣𐈤𐈥𐈦𐈧𐈨𐈩𐈪𐈫𐈬𐈭𐈮𐈯𐈰𐈱𐈲𐈳𐈴𐈵𐈶𐈷𐈸𐈹𐈺𐈻𐈼𐈽𐈾𐈿𐉀𐉁𐉂𐉃𐉄𐉅𐉆𐉇𐉈𐉉𐉊𐉋𐉌𐉍𐉎𐉏𐉐𐉑𐉒𐉓𐉔𐉕𐉖𐉗𐉘𐉙𐉚𐉛𐉜𐉝𐉞𐉟𐉠𐉡𐉢𐉣𐉤𐉥𐉦𐉧𐉨𐉩𐉪𐉫𐉬𐉭𐉮𐉯𐉰𐉱𐉲𐉳𐉴𐉵𐉶𐉷𐉸𐉹𐉺𐉻𐉼𐉽𐉾𐉿𐊀𐊁𐊂𐊃𐊄𐊅𐊆𐊇𐊈𐊉𐊊𐊋𐊌𐊍𐊎𐊏𐊐𐊑𐊒𐊓𐊔𐊕𐊖𐊗𐊘𐊙𐊚𐊛𐊜𐊝𐊞𐊟𐊠𐊡𐊢𐊣𐊤𐊥𐊦𐊧𐊨𐊩𐊪𐊫𐊬𐊭𐊮𐊯𐊰𐊱𐊲𐊳𐊴𐊵𐊶𐊷𐊸𐊹𐊺𐊻𐊼𐊽𐊾𐊿𐋀𐋁𐋂𐋃𐋄𐋅𐋆𐋇𐋈𐋉𐋊𐋋𐋌𐋍𐋎𐋏𐋐𐋑𐋒𐋓𐋔𐋕𐋖𐋗𐋘𐋙𐋚𐋛𐋜𐋝𐋞𐋟𐋠𐋡𐋢𐋣𐋤𐋥𐋦𐋧𐋨𐋩𐋪𐋫𐋬𐋭𐋮𐋯𐋰𐋱𐋲𐋳𐋴𐋵𐋶𐋷𐋸𐋹𐋺𐋻𐋼𐋽𐋾𐋿𐌀𐌁𐌂𐌃𐌄𐌅𐌆𐌇𐌈𐌉𐌊𐌋𐌌𐌍𐌎𐌏𐌐𐌑𐌒𐌓𐌔𐌕𐌖𐌗𐌘𐌙𐌚𐌛𐌜𐌝𐌞𐌟𐌠𐌡𐌢𐌣𐌤𐌥𐌦𐌧𐌨𐌩𐌪𐌫𐌬𐌭𐌮𐌯𐌰𐌱𐌲𐌳𐌴𐌵𐌶𐌷𐌸𐌹𐌺𐌻𐌼𐌽𐌾𐌿𐍀𐍁𐍂𐍃𐍄𐍅𐍆𐍇𐍈𐍉𐍊𐍋𐍌𐍍𐍎𐍏𐍐𐍑𐍒𐍓𐍔𐍕𐍖𐍗𐍘𐍙𐍚𐍛𐍜𐍝𐍞𐍟𐍠𐍡𐍢𐍣𐍤𐍥𐍦𐍧𐍨𐍩𐍪𐍫𐍬𐍭𐍮𐍯𐍰𐍱𐍲𐍳𐍴𐍵𐍶𐍷𐍸𐍹𐍺𐍻𐍼𐍽𐍾𐍿𐎀𐎁𐎂𐎃𐎄𐎅𐎆𐎇𐎈𐎉𐎊𐎋𐎌𐎍𐎎𐎏𐎐𐎑𐎒𐎓𐎔𐎕𐎖𐎗𐎘𐎙𐎚𐎛𐎜𐎝𐎞𐎟𐎠𐎡𐎢𐎣𐎤𐎥𐎦𐎧𐎨𐎩𐎪𐎫𐎬𐎭𐎮𐎯𐎰𐎱𐎲𐎳𐎴𐎵𐎶𐎷𐎸𐎹𐎺𐎻𐎼𐎽𐎾𐎿𐏀𐏁𐏂𐏃𐏄𐏅𐏆𐏇𐏈𐏉𐏊𐏋𐏌𐏍𐏎𐏏𐏐𐏑𐏒𐏓𐏔𐏕𐏖𐏗𐏘𐏙𐏚𐏛𐏜𐏝𐏞𐏟𐏠𐏡𐏢𐏣𐏤𐏥𐏦𐏧𐏨𐏩𐏪𐏫𐏬𐏭𐏮𐏯𐏰𐏱𐏲𐏳𐏴𐏵𐏶𐏷𐏸𐏹𐏺𐏻𐏼𐏽𐏾𐏿𐐀𐐁𐐂𐐃𐐄𐐅𐐆𐐇𐐈𐐉𐐊𐐋𐐌𐐍𐐎𐐏𐐐𐐑𐐒𐐓𐐔𐐕𐐖𐐗𐐘𐐙𐐚𐐛𐐜𐐝𐐞𐐟𐐠𐐡𐐢𐐣𐐤𐐥𐐦𐐧𐐨𐐩𐐪𐐫𐐬𐐭𐐮𐐯𐐰𐐱𐐲𐐳𐐴𐐵𐐶𐐷𐐸𐐹𐐺𐐻𐐼𐐽𐐾𐐿𐑀𐑁𐑂𐑃𐑄𐑅𐑆𐑇𐑈𐑉𐑊𐑋𐑌𐑍𐑎𐑏𐑐𐑑𐑒𐑓𐑔𐑕𐑖𐑗𐑘𐑙𐑚𐑛𐑜𐑝𐑞𐑟𐑠𐑡𐑢𐑣𐑤𐑥𐑦𐑧𐑨𐑩𐑪𐑫𐑬𐑭𐑮𐑯𐑰𐑱𐑲𐑳𐑴𐑵𐑶𐑷𐑸𐑹𐑺𐑻𐑼𐑽𐑾𐑿𐒀𐒁𐒂𐒃𐒄𐒅𐒆𐒇𐒈𐒉𐒊𐒋𐒌𐒍𐒎𐒏𐒐𐒑𐒒𐒓𐒔𐒕𐒖𐒗𐒘𐒙𐒚

2 Alphabets

Alphabets, or phonemic alphabets, are sets of letters, usually arranged in a fixed order, each of which represents one or more phonemes, both consonants and vowels, in the language they are used to write. In some cases combinations of letters are used to represent single phonemes, as in the English sh, ch and th. For a history of alphabets, lists of alphabets currently in use, alphabets used to a limited extent, and alphabets that are no longer used visit: <http://www.omniglot.com/writing/alphabets.htm>.

Most specialists believe that the Phoenician alphabet, an abjad, was adopted for Greek during the early 8th century BC, perhaps in Euboea. The earliest known fragmentary Greek inscriptions date from this time, 770–750 BC⁴, and they match Phoenician letter forms of c. 800–750 BC.⁵ The innovation of the Greek writing system was that symbols were used to represent vowels. The introduction of vowels into the script (vocalization) ”and the abstraction of consonants as a process in speech – meant that the number of signs (or characters) actually used to represent spoken language could be reduced to a relative handful (around 30) because even in complex languages, such as some Inuit dialects, the number of actual consonantal and vocal sounds is finite. <http://people.ucalgary.ca/~mwolfe/Syllabary,alphabet.html>

2.1 Greek

μῆνιν ἄειδε θεὰ Πηληϊάδεω Ἀχιλῆος
οὐλομένην, ἣ μυρὶ Ἄχαιοῖς ἄλγε’ ἔθηκε,
πολλὰς δ’ ἰφθίμους ψυχὰς Ἄϊδι προΐαψεν
ἠρώων, αὐτοὺς δὲ ἐλώρια τεῦχε κύνεσσιν
οἰωνοῖσί τε πᾶσι, Διὸς δ’ ἐτελείετο βουλή,
ἐξ οὗ δὴ τὰ πρῶτα διαστήτην ἐρίσαντε
Ἄτρεΐδης τε ἄναξ ἀνδρῶν καὶ δῖος Ἀχιλλεύς.
τίς τ’ ἄρ σφωε θεῶν ἔριδι ζυνέηκε μάχεσθαι;

Note: The above text is the first 8 lines of the Illiad by Homer. See: <http://www.sacred-texts.com/cla/homer/greek/ili01.htm>

⁴The date of the earliest inscribed objects; A.W. Johnston, ”The alphabet”, in N. Stampolidis and V. Karageorghis, eds, Sea Routes from Sidon to Huelva: Interconnections in the Mediterranean 2003:263-76, summarizes the present scholarship on the dating

⁵Pierre Swiggers, Transmission of the Phoenician Script to the West, in Daniels and Bright, The World’s Writing Systems, 1996

2.2 Cyrillic Script Languages

В этой статье рассматриваются две конструкции современного русского языка: конструкция идентификации с существительным в именительном падеже (NomC) и сравнительная конструкция с существительным в творительном падеже (InsC). Эти две конструкции рассматриваются в условиях, когда они максимально сближаются: в контексте связки быть. Материал Национального корпуса русского языка показывает, что, вопреки встречающимся в литературе утверждениям, NomC все еще частотна в современном русском языке. Хотя процент InsC и возрастает, тем не менее распределение NomC и InsC до и после 1950 г. изменилось не очень сильно. В работе показывается, что распределение двух конструкций зависит от грамматической формы глагола. При употреблении связки настоящего времени используется исключительно NomC, в то время как при употреблении связки прошедшего и будущего времени превалирует конструкция InsC. Для разных именных групп распределение употреблений NomC и InsC различается. Процент NomC больше для тех именных групп, которые по своей семантике удовлетворяют требованиям NomC, а процент InsC больше у тех именных групп, которые по своей семантике удовлетворяют требованиям InsC.⁶

3 Syllabic Alphabets

Syllabic alphabets, alphasyllabaries or abugidas consist of symbols for consonants and vowels. The consonants each have an inherent vowel which can be changed to another vowel or muted by means of diacritics. Vowels can also be written with separate letters when they occur at the beginning of a word or on their own.

When two or more consonants occur together, special conjunct symbols are often used which add the essential parts of first letter or letters in the sequence to the final letter. Sample syllabic alphabets are shown at: <http://www.omniglot.com/writing/syllabic.htm>. Two syllabic alphabets are shown.

3.1 Devanāgarī

Devanagari, also called Nagari (Nāgarī, the name of its parent writing system, is an abugida alphabet of India and Nepal. Because of its use to write the Hindi language, devanagari is one of the most used and adopted writing systems in the world. The use of the name devanāgarī is relatively recent, and the older term nāgarī is still common. The rapid spread of the term devanāgarī may be related to the almost exclusive use of this script to publish sacred Sanskrit texts. This has led to such a close connection between

⁶Sample text from: Kuznetzova,J;Diachronic distribution of predicate nouns in Russian: Russian Linguistics 2013 <http://link.springer.com/article/10.1007/s11185-012-9107-1>

devanagari and Sanskrit that devanagari is now widely thought to be the Sanskrit script; however, before the colonial period there was no standard script for Sanskrit, which was written in whatever script was familiar to the local populace.⁷

संस्कृतम् जालपत्रे सुशवागतम्

संस्कृतभाषा भारतीय परम्परा तथा वृधेः आधारः । अस्माकं वृधेः, परम्परा, जीवनमुल्यवोध नमितिं प्रतदिनि जीवने संस्कृतभाषायाः आवश्यकम् अस्ति । वेदः, रामायणं, महाभारतं, श्रीमद्भगवद्गीता आदि धर्मग्रन्थाः संस्कृतभाषायां लिखितम् । इमे ग्रन्थाः नसिन्देह अन्यग्रन्थाणाम् अपेक्षा महतरं भवन्ति ।

संस्कृतभाषा दशसहस्र वर्षादपि पूरातनम् । सहस्रवर्षव्यापि इयं भाषा अस्माकं कथितभाषा आसीत् । इयं संस्कृतभाषा सर्वभाषाणां जननी अपि भवति । कथं इयं मुख्यभाषा तथा महत्वपूर्ण भवति सा स्वयम् कथयति । यदा अस्माकं भाषाज्ञानम् भवति तदा वयम् अस्माकं भाषायाः परम्परायाः ग्रन्थाणां च प्रशंसा तथा ज्ञान्तुं समर्थाः भविष्यामः ।

अतः आगच्छन्तु प्रतदिनि-जीवने वयं शुद्धं संस्कृतं कथयिष्यामः ।

Note: The web page from which the above text was copied supports the use of Sanskrit in daily life in the cultural life of India. The text is a welcome to Sanscrutam.com <http://www.sanscrutam.com>

3.2 Bengali

The Bengali alphabet is the writing system for the Bengali language. The script with variations is shared by Assamese and is basis for Meitei, Bishnupriya Manipuri, Kokborok, Garo and Mundari alphabets. All these languages are spoken in the eastern region of South Asia. Historically, the script has also been used to write the Sanskrit language in the same region. It is written from left to right and lacks distinct letter cases. It is recognizable by a distinctive horizontal line running along the tops of the letters that links them together, a property it shares with two other popular Indian scripts: Devanagari (used for Hindi, Marathi and Nepali) and Gurmukhi (used for Punjabi). The Bengali script is, however, less blocky and presents a more sinuous shape. Because of the large population of literate Bengali speakers, Bengali script is also one of the more widely used writing systems in the world.

ধারা ১: সমস্ত মানুষ স্বাধীনভাবে সমান মর্যাদা এবং অধিকার নিয়ে জন্মগ্রহণ করে। তাঁদের ববিকে এবং বুদ্ধি আছে; সুতরাং সকলেরই একে অপরের প্রতি ভ্রাতৃত্বসুলভ মন ভাব নিয়ে আচরণ করা উচিত।

Note: The sample text in Bengali is from *WikiTravel: The Free Travel Guide* http://wikitravel.org/en/Bengali_prasebook. The text is Article 1 of the Universal Declaration of Human Rights (by the United Nations): Article 1: All human beings

⁷<http://en.wikipedia.org/wiki/Devanagari>

symbols in some scripts, such as Chinese. These scripts could also be called logophonic, morphophonemic, logographic or logosyllabic. Two semanto-phonetic scripts are currently in use, namely Chinese and Japanese (Nihongo). For more on semanto-phonetic writing systems visit: <http://www.omniglot.com/writing/semanto-ponetic.php>

5.1 Chinese

审讯揭秘伊以欧洲秘战

以向欧盟施压定性真主党恐怖组织

NICHOLAS KULISH 报道

本周在塞浦路斯一家小法庭里举行的一场少人问津的审讯中，一名24岁的真主党地下成员罕见地向人们揭示了以色列与伊朗在全球范围内进行的一场秘密战争真主党成员涉嫌炸以色列客车

Note: The text excerpt is from the New York Times Chinese Edition, February 23, 2013. Nicholas Kulish covers the upcoming elections in Cyprus.⁹

5.2 Japanese (Nihongo)

6 Technologies used to set the type in the worlds writing systems

6.1 Overview of Technical Issues in Supporting Scripts

This section is based on a discussion of technical challenges in typesetting the worlds scripts presented by SIL International <http://scripts.sil.org>

”There are a number of challenges in the typesetting of non-Roman scripts. These include problems of interaction between the font and typesetting system, problems of very large character sets, and considerations of typesetting the particular script. Happily, the advent of new computer technology has considerably lessened some of these problems.”

⁹<http://cn.nytimes.com/>

6.2 Font Basics

6.3 Overview of Unicode Character Set

00	01	02	03	04	05	06	07	08	09	0A	0B	0C	0D	0E	0F
10	11	12	13	14	15	16	17	18	19	1A	1B	1C	1D	1E	1F
20	21	22	23	24	25	26	27	28	29	2A	2B	2C	2D	2E	2F
30	31	32	33	34	35	36	37	38	39	3A	3B	3C	3D	3E	3F
40	41	42	43	44	45	46	47	48	49	4A	4B	4C	4D	4E	4F
50	51	52	53	54	55	56	57	58	59	5A	5B	5C	5D	5E	5F
60	61	62	63	64	65	66	67	68	69	6A	6B	6C	6D	6E	6F
70	71	72	73	74	75	76	77	78	79	7A	7B	7C	7D	7E	7F
80	81	82	83	84	85	86	87	88	89	8A	8B	8C	8D	8E	8F
90	91	92	93	94	95	96	97	98	99	9A	9B	9C	9D	9E	9F
A0	A1	A2	A3	A4	A5	A6	A7	A8	A9	AA	AB	AC	AD	AE	AF
B0	B1	B2	B3	B4	B5	B6	B7	B8	B9	BA	BB	BC	BD	BE	BF
C0	C1	C2	C3	C4	C5	C6	C7	C8	C9	CA	CB	CC	CD	CE	CF
D0	D1	D2	D3	D4	D5	D6	D7	D8	D9	DA	DB	DC	DD	DE	DF
E0	E1	E2	E3	E4	E5	E6	E7	E8	E9	EA	EB	EC	ED	EE	EF
F0	F1	F2	F3	F4	F5	F6	F7	F8	F9	FA	FB	FC	FD	FE	FF

- Latin scripts and symbols
- Linguistic scripts
- Other European scripts
- African scripts
- Middle Eastern and Southwest Asian scripts
- Central Asian scripts
- South Asian scripts
- Southeast Asian scripts
- East Asian scripts
- Unified CJK Han
- American scripts
- Symbols
- Diacritics
- UTF-16 surrogates and private use
- Miscellaneous characters
- Unallocated code points

6.4 Codepages

More to be developed.