IMPE LaTeX System

Canonical Multilingual and Feature Showcase

Kelvin Yang

v0.1.3

Abstract

This document is the canonical visual specimen for the current IMPE LaTeX System public interface. It combines a local-family style audit, a deliberately small Unicode-range routing regression, dense multilingual paragraph composition with explicit local overrides, right-to-left text, vertical writing, generic complex shaping, CJK and Han family selection, and the principal document features provided by IMPE. It is intended both as a public showcase and as a release-time visual regression specimen.[1]

Contents

1 Local Font Families and Styles **3**

2 Multilingual Routing and Mixed-Script Composition **3**

2.1 Mixed-Script Paragraph 3

2.2 Minimal Automatic Unicode-Range Regression 7

3 Right-to-Left Scripts **7**

3.1 Arabic 7

3.2 Urdu 8

3.3 Hebrew 8

3.4 Syriac 8

[1] The showcase is intentionally broader than the small regression examples under `examples/`; those examples remain useful for isolating individual failures.

4 Vertical Writing **9**

4.1 Mongolian 9

4.2 Manchu 10

4.3 Old Uyghur 10

5 Complex Shaping and Stacked Sequences **10**

5.1 Devanagari 10

5.2 Tibetan 11

5.3 Brahmi, Kharoshthi, and Thai 12

5.4 Khitan Small Script: current linear fallback 12

6 East Asian Writing and CJK Family Selection **12**

6.1 Chinese 13

6.2 Korean and Historical Hangul 14

6.3 Japanese and Ryukyuan 16

6.4 Vietnamese 17

6.5 Tangut 18

6.6 WenJin Multi-Plane Coverage 19

6.7 Regional Forms of Shared Han Characters 19

7 Document Features **19**

7.1 Mathematics 21

7.2 Tables 21

7.3 Lists and Example Blocks 21

7.4 Images 22

7.5 Citations and Bibliography 22

7.6 Index Terms 23

7.7 Hyperlinks, Footnotes, and Running Headers 23

References **23**

Index **24**

1 Local Font Families and Styles

This section presents IMPE's explicit local font commands and compares their available serif, bold, italic, and sans-serif styles. The complete font-family matrix appears on the following landscape page. Where a requested face is unavailable, IMPE falls back according to the family configuration rather than requiring special-case markup.

2 Multilingual Routing and Mixed-Script Composition

Multilingual composition (combined routing and local overrides) is tested here in two deliberately different ways. The main specimen is a dense paragraph in which explicit local families, right-to-left runs, complex shaping, shared Han choices, and inline vertical writing are allowed to interact. A later compact regression enables automatic Unicode-range ownership for only a small set of ordinary scripts.

2.1 Mixed-Script Paragraph

The following specimen is intentionally one TEX paragraph. It contains no English transition sentences: every script that appears contributes a complete sentence or clause. Armenian, Georgian, Coptic, Glagolitic, and Tamil are left as raw Unicode to exercise the small automatic-routing set; all other script families in the paragraph use explicit local overrides.

Հայերեն այս ամբողջական նախադասությունը գրված է նույն բազմալեզու պարբերության մեջ: ეს სრული ქართული წინადადება იმავე მრავალენოვან აბზაცში სხვა დამწერლობებთან ერთად გვხვდება. Ⲟⲩⲟϩ ⲡⲓⲗⲟⲅⲟⲥ ⲁϥⲱⲱⲡⲓ Ⲛ̀ⲟⲩⲥⲁⲣⲝ ⲟⲩⲟϩ ⲁϥⲱⲱⲡⲓ Ⲛ̀ϧⲏⲧⲉⲛ. 𐌼𐌰 𐍈𐌰𐍅𐌰𐌽𐌴 𐌷𐌰𐍅𐌰𐍃 𐌲𐌰𐌼𐌰𐍂𐍃𐌰, 𐍅𐌴 𐌲𐌰𐌼𐌰𐍂𐍃 𐌷𐌰𐍅𐌰𐍃 𐍆𐌰𐍄𐌴𐍃. सर्वे भवन्तु सुखिनः, सर्वे सन्तु निरामयाः। தமிழில் எழுதப்பட்ட இந்த முழு வாக்கியம் அதே பலமொழிப் பத்திக்குள் தொடர்ந்து வருகிறது. ཤེས་རབ་དང་སྙིང་རྗེ་མེད་ཚོས་ཀྱི་རྒྱ་བ་ཡིན། هذه جملة عربية كاملة تظهر في وسط الفقرة نفسها وتختبر الانتقال إلى الكتابة من اليمين إلى اليسار. یہ ایک مکمل اردو جملہ ہے جو اسی کثیر لسانی پیراگراف کے اندر نستعلیق رسم الخط میں لکھی گئی ہے۔ זהו משפט עברי שלם שממשיך באותה פסקה לאחר הטקסט הערבי והאורדו. 𑖨𑖰𑖬𑖰𑖟 𑖨𑖰𑖬𑖰𑖟 𑖨𑖰𑖬𑖰𑖟

Family	Command	Regular	Bold	Italic	Bold italic	Sans	Sans bold	Sans italic	Sans bold italic
Latin and general-purpose families									
Noto	\NOT	Abg fi fl	**Abg fi fl**	*Abg fi fl*	***Abg fi fl***	Abg fi fl	**Abg fi fl**	*Abg fi fl*	***Abg fi fl***
Times	\TIM	Abg fi fl	**Abg fi fl**	*Abg fi fl*	***Abg fi fl***	Abg fi fl	**Abg fi fl**	*Abg fi fl*	***Abg fi fl***
Gentium Plus	\GEN	Ārya ṛ ṃ ḥ	**Ārya ṛ ṃ ḥ**	*Ārya ṛ ṃ ḥ*	***Ārya ṛ ṃ ḥ***	Ārya ṛ ṃ ḥ	Ārya ṛ ṃ ḥ	*Ārya ṛ ṃ ḥ*	*Ārya ṛ ṃ ḥ*
Charis SIL	\CHA	Ārya ṛ ṃ ḥ	**Ārya ṛ ṃ ḥ**	*Ārya ṛ ṃ ḥ*	***Ārya ṛ ṃ ḥ***	Ārya ṛ ṃ ḥ	Ārya ṛ ṃ ḥ	*Ārya ṛ ṃ ḥ*	*Ārya ṛ ṃ ḥ*
Libertinus	\LIB	Abg fi fl	**Abg fi fl**	*Abg fi fl*	***Abg fi fl***	Abg fi fl	**Abg fi fl**	*Abg fi fl*	**Abg fi fl**
European and ancient alphabetic scripts									
Carian	\CA	AΩN	AΩN	AΩN	AΩN	AΩN	AΩN	AΩN	AΩN
Coptic	\CO	ⲛⲟⲩⲧⲉ	ⲛⲟⲩⲧⲉ	ⲛⲟⲩⲧⲉ	ⲛⲟⲩⲧⲉ	ⲛⲟⲩⲧⲉ	ⲛⲟⲩⲧⲉ	ⲛⲟⲩⲧⲉ	ⲛⲟⲩⲧⲉ
Bopomofo	ㄓㄨˋㄧㄣ	ㄓㄨˋㄧㄣ	ㄓㄨˋㄧㄣ	ㄓㄨˋㄧㄣ	ㄓㄨˋㄧㄣ	ㄓㄨˋㄧㄣ	ㄓㄨˋㄧㄣ	ㄓㄨˋㄧㄣ	ㄓㄨˋㄧㄣ
Cuneiform	𒀭𒌋𒆠	𒀭𒌋𒆠	𒀭𒌋𒆠	𒀭𒌋𒆠	𒀭𒌋𒆠	𒀭𒌋𒆠	𒀭𒌋𒆠	𒀭𒌋𒆠	𒀭𒌋𒆠
Glagolitic	ⰽⰻⰴⰰ	ⰽⰻⰴⰰ	ⰽⰻⰴⰰ	ⰽⰻⰴⰰ	ⰽⰻⰴⰰ	ⰽⰻⰴⰰ	ⰽⰻⰴⰰ	ⰽⰻⰴⰰ	ⰽⰻⰴⰰ
Old Italic	𐌊𐌁𐌂	𐌊𐌁𐌂	𐌊𐌁𐌂	𐌊𐌁𐌂	𐌊𐌁𐌂	𐌊𐌁𐌂	𐌊𐌁𐌂	𐌊𐌁𐌂	𐌊𐌁𐌂
Old Hungarian	𐲀𐲁𐲁	𐲀𐲁𐲁	𐲀𐲁𐲁	𐲀𐲁𐲁	𐲀𐲁𐲁	𐲀𐲁𐲁	𐲀𐲁𐲁	𐲀𐲁𐲁	𐲀𐲁𐲁
Runic	ᚠᚢᚦ	ᚠᚢᚦ	ᚠᚢᚦ	ᚠᚢᚦ	ᚠᚢᚦ	ᚠᚢᚦ	ᚠᚢᚦ	ᚠᚢᚦ	ᚠᚢᚦ
South and Central Asian scripts									
Armenian	\HY	Հայերեն	**Հայերեն**	Հայերեն	***Հայերեն***	Հայերեն	**Հայերեն**	Հայերեն	***Հայերեն***
Hindi	\HI	हिन्दी	**हिन्दी**	हिन्दी	***हिन्दी***	हिन्दी	**हिन्दी**	हिन्दी	***हिन्दी***
Sanskrit	\SA	संस्कृतम्	संस्कृतम्	संस्कृतम्	संस्कृतम्	संस्कृतम्	संस्कृतम्	संस्कृतम्	संस्कृतम्
Devanagari	\DEV	देवनागरी	**देवनागरी**	देवनागरी	***देवनागरी***	देवनागरी	**देवनागरी**	देवनागरी	***देवनागरी***
Tamil	\TA	தமிழ்	**தமிழ்**	*தமிழ்*	***தமிழ்***	தமிழ்	**தமிழ்**	*தமிழ்*	***தமிழ்***
Brahmi	\BR	𑀢𑀫𑀺𑀵	𑀢𑀫𑀺𑀵	𑀢𑀫𑀺𑀵	𑀢𑀫𑀺𑀵	𑀢𑀫𑀺𑀵	𑀢𑀫𑀺𑀵	𑀢𑀫𑀺𑀵	𑀢𑀫𑀺𑀵
Georgian	\KA	ქართული	**ქართული**	ქართული	***ქართული***	ქართული	**ქართული**	ქართული	***ქართული***

Family	Command	Regular	Bold	Italic	Bold italic	Sans	Sans bold	Sans italic	Sans bold italic
Tibetan	`\TI`	བཤད་བདས་	**བཤད་བདས་**	ཤད་ཐལ།	**ཤད་ཐལ།**	བཤད་བདས་	**བཤད་བདས་**	ཤད་ཐལ།	**ཤད་ཐལ།**
RTL and West Asian scripts									
Arabic	`\AR`	العربية	**العربية**	*العربية*	***العربية***	العربية	**العربية**	العربية	**العربية**
Urdu	`\UR`	اُردُو	**اُردُو**	*اُردُو*	***اُردُو***	اُردُو	**اُردُو**	*اُردُو*	***اُردُو***
Imperial Aramaic	`\IA`	𐡀𐡓𐡌	**𐡀𐡓𐡌**	*𐡀𐡓𐡌*	***𐡀𐡓𐡌***	𐡀𐡓𐡌	**𐡀𐡓𐡌**	*𐡀𐡓𐡌*	***𐡀𐡓𐡌***
Nabataean	`\NB`	𐢘𐢚𐢡	**𐢘𐢚𐢡**	*𐢘𐢚𐢡*	***𐢘𐢚𐢡***	𐢘𐢚𐢡	**𐢘𐢚𐢡**	*𐢘𐢚𐢡*	***𐢘𐢚𐢡***
Hebrew	`\HE`	שלום	**שלום**	*שלום*	***שלום***	שלום	**שלום**	*שלום*	***שלום***
Syriac	`\SY`	ܫܠܡܐ	**ܫܠܡܐ**	*ܫܠܡܐ*	***ܫܠܡܐ***	ܫܠܡܐ	**ܫܠܡܐ**	*ܫܠܡܐ*	***ܫܠܡܐ***
Eastern Syriac	`\SYE`	ܫܠܡܐ	**ܫܠܡܐ**	*ܫܠܡܐ*	***ܫܠܡܐ***	ܫܠܡܐ	**ܫܠܡܐ**	*ܫܠܡܐ*	***ܫܠܡܐ***
Kharoshthi	`\KH`	𐨲	𐨲	𐨲	𐨲	𐨲	𐨲	𐨲	𐨲
Iranian and other historical scripts									
Inscriptional Parthian	`\PAR`	𐭠𐭫𐭥	𐭠𐭫𐭥	𐭠𐭫𐭥	𐭠𐭫𐭥	𐭠𐭫𐭥	𐭠𐭫𐭥	𐭠𐭫𐭥	𐭠𐭫𐭥
Inscriptional Pahlavi	`\PAH`	𐮀𐮅𐮇	𐮀𐮅𐮇	𐮀𐮅𐮇	𐮀𐮅𐮇	𐮀𐮅𐮇	𐮀𐮅𐮇	𐮀𐮅𐮇	𐮀𐮅𐮇
Psalter Pahlavi	`\PSP`	𐹩𐹬𐹥	𐹩𐹬𐹥	𐹩𐹬𐹥	𐹩𐹬𐹥	𐹩𐹬𐹥	𐹩𐹬𐹥	𐹩𐹬𐹥	𐹩𐹬𐹥
Avestan	`\AV`	𐬨𐬀𐬱𐬌	𐬨𐬀𐬱𐬌	𐬨𐬀𐬱𐬌	𐬨𐬀𐬱𐬌	𐬨𐬀𐬱𐬌	𐬨𐬀𐬱𐬌	𐬨𐬀𐬱𐬌	𐬨𐬀𐬱𐬌
Manichaean	`\MA`	𐫀𐫁𐫂𐫃	𐫀𐫁𐫂𐫃	𐫀𐫁𐫂𐫃	𐫀𐫁𐫂𐫃	𐫀𐫁𐫂𐫃	𐫀𐫁𐫂𐫃	𐫀𐫁𐫂𐫃	𐫀𐫁𐫂𐫃
Phoenician	`\PH`	𐤉𐤋𐤌	𐤉𐤋𐤌	𐤉𐤋𐤌	𐤉𐤋𐤌	𐤉𐤋𐤌	𐤉𐤋𐤌	𐤉𐤋𐤌	𐤉𐤋𐤌
Samaritan	`\SM`	ש׳ל׳ם	ש׳ל׳ם	ש׳ל׳ם	ש׳ל׳ם	ש׳ל׳ם	ש׳ל׳ם	ש׳ל׳ם	ש׳ל׳ם
Sogdian	`\SG`	𐼲𐼰𐼥	𐼲𐼰𐼥	𐼲𐼰𐼥	𐼲𐼰𐼥	𐼲𐼰𐼥	𐼲𐼰𐼥	𐼲𐼰𐼥	𐼲𐼰𐼥
Old Sogdian	`\SGO`	𐼓𐼰𐼥	𐼓𐼰𐼥	𐼓𐼰𐼥	𐼓𐼰𐼥	𐼓𐼰𐼥	𐼓𐼰𐼥	𐼓𐼰𐼥	𐼓𐼰𐼥
CJK and East Asian families									
Noto SC	`\SC`	汉字骨直	**汉字骨直**	汉字骨直	**汉字骨直**	汉字骨直	**汉字骨直**	汉字骨直	**汉字骨直**
Noto TC	`\TC`	漢字骨直	**漢字骨直**	漢字骨直	**漢字骨直**	漢字骨直	**漢字骨直**	漢字骨直	**漢字骨直**
Japanese	`\JP`	日本語漢字	**日本語漢字**	日本語漢字	**日本語漢字**	日本語漢字	**日本語漢字**	日本語漢字	**日本語漢字**

Family	Command	Regular	Bold	Italic	Bold italic	Sans	Sans bold	Sans italic	Sans bold italic
WenJin	\WJ	中國己丁	中國己丁	中國己丁	中國己丁	中國己丁	中國己丁	中國己丁	中國己丁
Korean	\KR	한국어漢字	**한국어漢字**	한국어漢字	**한국어漢字**	한국어漢字	**한국어漢字**	한국어漢字	**한국어漢字**
Tangut	\TG	𘗚𘗛𘗜	𘗚𘗛𘗜	𘗚𘗛𘗜	𘗚𘗛𘗜	𘗚𘗛𘗜	𘗚𘗛𘗜	𘗚𘗛𘗜	𘗚𘗛𘗜
Mongolian-family and other Asian scripts									
Khitan Small Script (linear)	\KHS	丐北妾	丐北妾	丐北妾	丐北妾	丐北妾	丐北妾	丐北妾	丐北妾
Mongolian	\MO	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
Mongolian Baiti	\MOb	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
Manchu	\MC	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
Segoe Historic	\SEG	𐤊𐤋𐤌	𐤊𐤋𐤌	𐤊𐤋𐤌	𐤊𐤋𐤌	𐤊𐤋𐤌	𐤊𐤋𐤌	𐤊𐤋𐤌	𐤊𐤋𐤌
Thai	\TH	ภาษาไทย	**ภาษาไทย**	ภาษาไทย	**ภาษาไทย**	ภาษาไทย	**ภาษาไทย**	ภาษาไทย	**ภาษาไทย**
Old Turkic	\OT	𐰋𐰇𐰼𐰴	𐰋𐰇𐰼𐰴	𐰋𐰇𐰼𐰴	𐰋𐰇𐰼𐰴	𐰋𐰇𐰼𐰴	𐰋𐰇𐰼𐰴	𐰋𐰇𐰼𐰴	𐰋𐰇𐰼𐰴
Old Uyghur	\UY	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
Vietnamese									
Quốc Ngữ	\VI	Tiếng Việt	**Tiếng Việt**	Tiếng Việt	**Tiếng Việt**	Tiếng Việt	Tiếng Việt	Tiếng Việt	Tiếng Việt
Hán-Nôm	\HN	𡨸喃馭歪	**𡨸喃馭歪**	𡨸喃馭歪	𡨸喃馭歪	𡨸喃馭歪	𡨸喃馭歪	𡨸喃馭歪	𡨸喃馭歪

前後不同文種保持在同一個段落中。 ᠮᠣᠩᠭᠣᠯ ᠪᠢᠴᠢᠭ᠄ ᠳᠤᠮᠳᠠ ᠳᠤᠷᠠᠳᠠᠨ ᠪᠣᠯᠤᠭᠰᠠᠨ ᠪᠣᠯᠪᠠᠴᠠᠭᠠᠨ 縦書きのモンゴル文字の直後にも日本語の完全な文章が続き、通常の横書き状態が回復したことを確認します。 ᠬᠣᠶᠠᠷ ᠳᠠᠮᠳᠠ ᠬᠣᠷᠣᠭᠳᠣᠯ᠂

أَنَّ الْفَقْرَةَ الْعَرَبِيَّةَ الْكَامِلَةَ تَبْقَى مُسْتَقِرَّةً دَاخِلَ وَثِيقَةٍ مُتَعَدِّدَةِ اللُّغَاتِ.

Inline Arabic inside an English sentence: العربية والكتابة في سطر واحد مع علامات الترقيم before returning to ordinary Latin text.

3.2 Urdu

اُردُو ایک ایسی زبان ہے جس کی تحریری روایت میں نستعلیق رسم الخط کو خاص اہمیت حاصل ہے۔ اس رسم الخط میں حروف صرف ایک دوسرے سے ملتے ہی نہیں بلکہ الفاظ کی مجموعی ساخت بھی سطر کی بصری صورت پر اثر ڈالتی ہے، اس لیے طویل عبارت مختصر نمونے سے کہیں زیادہ مفید آزمائش فراہم کرتی ہے۔ اس پیراگراف میں مختلف لمبائی کے الفاظ، مختلف رموزِ اوقاف، اور کئی مسلسل جملے شامل ہیں تاکہ دائیں سے بائیں بہاؤ، لفظی فاصلوں، اور سطر کے اختتام پر عبارت کی تقسیم کو واضح طور پر دیکھا جا سکے۔ جب عبارت ایک سطر میں پوری نہ آئے تو اگلی سطر میں منتقلی قدرتی ہونی چاہیے، الفاظ کی ترتیب برقرار رہنی چاہیے، اور نستعلیق کی تشکیل کسی غلط فونٹ یا غیر متعلقہ عربی طرز میں تبدیل نہیں ہونی چاہیے۔ یہ آخری جملہ خاص طور پر اس بات کی جانچ کرتا ہے کہ ایک مکمل اُردُو پیراگراف کئی سطروں میں تقسیم ہونے کے بعد بھی اپنی مقامی نستعلیق فیملی اور درست سمت برقرار رکھتا ہے۔

3.3 Hebrew

העברית נכתבת מימין לשמאל, ולכן פסקה ארוכה מאפשרת לבדוק לא רק את צורת האותיות אלא גם את התנהגות השורות, הרווחים וסימני הפיסוק. בטקסט הזה מופיעים משפטים באורכים שונים כדי לבדוק כיצד המערכת מתמודדת עם רצף עברי ארוך בתוך פסקה רגילה. כאשר השורה נעשית ארוכה מדי, המעבר לשורה הבאה צריך להתרחש במקום טבעי בלי לשנות את סדר המילים ובלי לפגוע במיקום הנקודות, הפסיקים והסוגריים. אפשר גם לכתוב מילים מנוקדות כמו שָׁלוֹם, עִבְרִית וְכְתִיבָה בתוך אותה פסקה, וכך לבדוק שהניקוד נשאר מחובר לאות הנכונה גם לאחר שבירת שורה. המשפט האחרון נועד לוודא שהפסקה כולה נשארת יציבה בתוך מסמך רב־לשוני ולאחריה חוזר הטקסט הסובב להתנהגות הרגילה.

3.4 Syriac

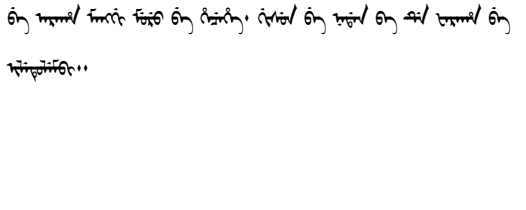
Western / general Syriac:

ܥܠܡܐ ܕܟܬܒܝܢ ܡܢ ܝܡܝܢܐ ܠܣܡܐܠܐ، ܘܠܗܢܐ ܦܣܩܐ ܕܐܪܘܟܐ ܡܫܟܚܐ ܠܡܒܕܩ ܠܐ ܪܩ ܐܬ ܐܘܬܝܘܬ ܗܘܪܝܬ ܐܠܐ ܐܦܠܘ ܐܬ ܗܢܓܢܘܬ ܗܪܘܚܝܢ، ܗܪܘܝܚܝܢ ܘܣܝܡܢܝ ܗܦܝܣܘܩ. ܒܛܩܣܛ ܗܙܐ ܡܘܦܝܥܝܢ ܡܫܦܛܝܢ ܒܐܘܪܟܝܢ ܫܘܢܝܢ ܟܕܝ ܠܒܕܩ ܟܝܨܕ ܗܡܥܪܟܬ ܡܬܡܘܕܕܬ ܥܡ ܪܨܦ ܥܒܪܝ ܐܪܘܟ ܒܬܘܟ ܦܣܩܐ ܪܓܝܠܐ. ܟܐܫܪ ܗܫܘܪܐ ܢܥܫܝܬ ܐܪܘܟܐ ܡܕܝ، ܗܡܥܒܪ ܠܫܘܪܐ ܗܒܐܐ ܦܪܝܟ ܠܗܬܪܚܫ ܒܡܩܘܡ ܛܒܥܝ ܒܠܝ ܠܫܢܘܬ ܐܬ ܣܕܪ ܗܡܝܠܝܡ ܘܒܠܝ ܠܦܓܘܥ ܒܡܝܩܘܡ ܗܢܩܘܕܘܬ، ܗܦܣܝܩܝܡ ܘܗܣܘܓܪܝܝܡ. ܐܦܫܪ ܓܡ ܠܟܬܘܒ ܡܝܠܝܡ ܡܢܘܩܕܘܬ ܟܡܘ ܫָׁלוֹם، ܥִׁבְרִׁית ܘְׁכְׁתִׁיבָׁה ܒܬܘܟ ܐܘܬܗ ܦܣܩܐ، ܘܟܟ ܠܒܕܩ ܫܗܢܝܩܘܕ ܢܫܐܪ ܡܚܘܒܪ ܠܐܘܬ ܗܢܟܘܢܐ ܓܡ ܠܐܚܪ ܫܒܝܪܬ ܫܘܪܐ. ܗܡܫܦ�מ ܗܐܚܪܘܢ ܢܘܥܕ ܠܘܘܕܐ ܫܗܦܣܩܐ ܟܘܠܐ ܢܫܐܪܬ ܝܨܝܒܐ ܒܬܘܟ ܡܣܡܟ ܪܒ־ܠܫܘܢܝ ܘܠܐܚܪܝܐ ܚܘܙܪ ܗܛܩܣܛ ܗܣܘܒܒ ܠܗܬܢܗܓܘܬ ܗܪܓܝܠܐ.

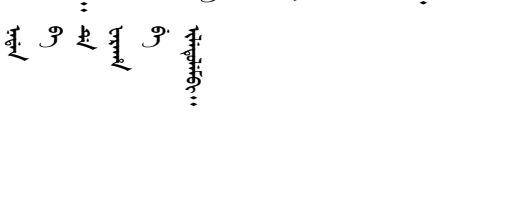
ܗܘܐ ܕܟܬܒܐ ܕܟܬܒܐ ܫܡܝܐ ܕܐܪܥܐ ܕܟܬܒܐ ܕܟܬܒܐ ܕܐܪܥܐ. ܡܢ ܗܘ ܕܟܬܒܐ ܕܟܬܒܐ ܕܟܬܒܐ

4.2 Manchu

Horizontal



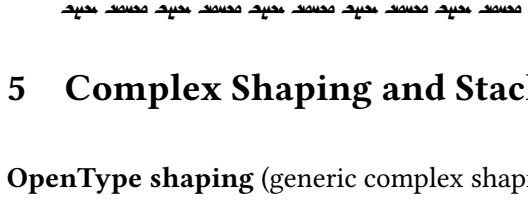
Vertical



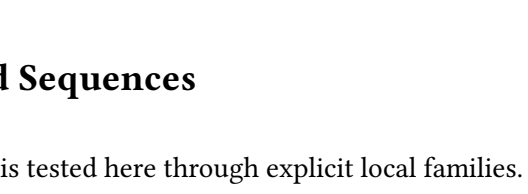
4.3 Old Uyghur

The Old Uyghur sample is deliberately extended so that several successive vertical columns must be produced. It is used as a layout stress string rather than as a continuous philological quotation.

Horizontal



Vertical



5 Complex Shaping and Stacked Sequences

OpenType shaping (generic complex shaping) is tested here through explicit local families. This deliberately separates shaping from Unicode-range ownership: if a local specimen is wrong, the failure belongs to the selected family or shaping path rather than to an automatic range transition. Egyptian Hieroglyphs remain outside this showcase, and Khitan Small Script is recorded only in its current linear fallback form.

5.1 Devanagari

The Sanskrit local family exercises conjuncts, virama behavior, pre-base and post-base vowel signs, and longer clusters without enabling a document-wide Devanagari owner.

A longer local specimen:

धर्मक्षेत्रे कुरुक्षेत्रे समवेता युयुत्सवः। संस्कृतवाक्यानि देवनागर्या लिखितानि सन्ति।

Table 2: Devanagari shaping specimens

Sequence	Specimen
kṣa	क्ष
jña	ज्ञ
śra	श्र
stra	स्त्र
ddhya	द्ध्य
dharma	धर्म
dharmakṣetre	धर्मक्षेत्रे
saṃskṛtam	संस्कृतम्
oṃ	ॐ

5.2 Tibetan

Tibetan provides a visible example of generic vertical stacking inside one local shaping run.

Table 3: Tibetan stack specimens

Type	Specimen
stack	བསྒྲུབས་
stack	བརྒྱུད་
stack	བསྟན་
stack	རྒྱལ་
stack	དབྱངས་

A longer local paragraph exercises Tibetan shaping and line breaking under ordinary document width:

བོད་ཡིག་ནི་བོད་ཀྱི་སྐད་ཡིག་འབྲི་བའི་ཡི་གེའི་མ་ལག་ཅིག་ཡིན་ཞིང་། ཚོག་དང་ཚོག་གི་བར་ལ་ཚེག་གིས་མཚམས་འབྱེད་པ་དང་། ཡི་གེ་མང་པོ་ཞིག་མགོ་ཅན་དང་འདོགས་ཅན་དུ་སྦྱར་ནས་ཚོག་བར་གཅིག་གི་ནང་དུ་སྦྱངས་ཏེ་འབྲི་དགོས།དཔེར་ན་བསྒྲུབས་པ་དང་། བརྒྱུད་པ། བསྟན་པ། རྒྱལ་ཁབ། དབྱངས་ཅན་སྡུ་བྱའི་ཚོག་ནང་དུ་ཡི་གེ་མང་པོ་སྦྱར་ཡོད་པས། ཡིག་གཟུགས་ཀྱི་སྦྱངས་སྟངས་དང་ཚེག་བར་གྱི་བར་ཐག་གཏིས་ཀ་ཡང་དག་པར་སྟོན་དགོས། ཚོག་ཕྲེང་རིང་པོ་ཞིག་ཏོག་ངོས་ཀྱི་ཞེང་ཚད་ལས་བརྒལ་བའི་སྐབས་སུ། ཚེག་གི་རྟེས་སུ་རང་བཞིན་གྱིས་ཕྲེང་བ་གསར་པར་འགྲོ་ཐུབ་པ་དང་། ལྟ་ཕྱིའི་ཡི་གེའི་སྦྱངས་སྟངས་དང་ཚོག་གི་གོ་རིམ་མི་འགྱུག་པ་གལ་ཆེ། དེའི་ཕྱིར་ཚོག་ཕྲེང་རིང་པོ་འདིས་བོད་ཡིག་གི་གཟུགས་སྒྲོར་དང་། ཚེག་བར། ཕྲེང་བར། ཤད་དང་ཉིས་ཤད་ཀྱི་གནས་སྟངས། དེ་བཞིན་དུ་ཕྲེང་བ་བརྟེ་བའི་སྐབས་ཀྱི་སྒྲིའི་སྒྲིག་སྟངས་བཅས་ལ་ཚོད་ལྟ་བྱེད་ཐུབ།།

5.3 Brahmi, Kharoshthi, and Thai

Brahmi is provided through two local families. The redistributable Noto family provides broad Brahmi coverage, but in the current IMPE/XeTEX setup it does not produce the expected stacked forms for all conjunct sequences. Segoe UI Historic is therefore exposed through `\SEG` as an alternative reference family and produces the intended stacking in these cases. Because Segoe UI Historic is a Microsoft Windows font, its font binary is not distributed with IMPE; the local command is available only when the user supplies or already has access to the font.

Table 4: Additional generic-shaping specimens

Script / family	Specimen
Brahmi / Noto	𑀲𑀻𑀘𑁆 𑀫𑀸, 𑀥𑀻𑀭𑁆 𑀤𑀺.
Brahmi / Segoe Historic	𑀲𑀻𑀘𑁆 𑀫𑀸, 𑀥𑀻𑀭𑁆𑀤𑀺.
Kharoshthi	𐨑𐨏𐨑𐨑𐨔𐨗𐨒𐨔𐨕𐨔
Thai	ภาษาไทย กำ กิ กี๊ เก้า

5.4 Khitan Small Script: current linear fallback

Khitan Small Script is already exposed through IMPE's public font interface, while its canonical stacked layout is not yet used in this showcase. The following specimen is taken from the Khitan Small Script epitaph of Xiao Hudujin and is deliberately long enough to exercise the present linear fallback across continuous text, cluster boundaries, spacing, and paragraph line breaking.

又 令兮分 矢关 几女村 刈公 力亚出爻村 令小岔当 乏又公 又雨 击几当 北小寸伏 本亦北 本丰 尖 吊为买 本几村 爻化 用卌 仝寻 几亦 业及困小爻 丙车爻火 壮 几寸 公爻 冈为丰 本化关 寸勿 丙气 仝关 本关 击亦本 关化 寸奈伏 本化当村 令小亦当 寸化 土小岔关 爻化卑 令交仝 小列 伏穴女 丞为市为 乏化劣 刹 又化关 几 几穴卡 列行仝 业丙小 化夯 业爻 本关 几 击夯 寸夸关 化 雨北

6 East Asian Writing and CJK Family Selection

CJK family selection (continuous text, shared Han ownership, and explicit regional overrides) distinguishes the document CJK family from language-specific local families. Unicode codepoints alone cannot reliably determine the intended regional form of shared Han characters, while several East Asian writing systems also require shaping behavior that extends beyond ordinary Han glyph selection.

The specimens in this section therefore use normal-width continuous text rather than isolated character strings. They exercise paragraph composition, line breaking, punctuation, regional glyph forms, Hangul composition, historical jamo, supplementary-plane ideographs, and mixed-script runs. A compact table at the end isolates shared Han codepoints for direct visual comparison.

6.1 Chinese

Mainland Simplified Chinese — Noto SC This paragraph uses the explicit Simplified Chinese family throughout.

现代中文排版并不只是把一个个汉字依次放进页面。一个真正的正文环境还需要同时处理标点、数字、书名号、引号、括号以及连续句子之间的换行关系，并且在较长的段落中保持稳定的字面和行距。对于共享的汉字码位，不同地区使用的字体还可能给出不同的字形，因此同一个“骨、直、令、具、角”在不同字体家族中会表现出细微而稳定的区域差异。这里使用一整段连续的简体中文来检查大陆简体字形、标点悬挂附近的布局、正常段落换行以及本地字体选择在长文本中是否能够始终保持有效，而不是只验证几个孤立字符能不能显示。

Mainland Traditional Chinese — Noto SC Traditional characters can also be deliberately typeset through the Mainland Chinese family. This separates character repertoire from regional glyph selection: traditional codepoints do not by themselves imply a Taiwan or Hong Kong font.

現代中文排版並不只是把一個個漢字依次放進頁面。一個真正的正文環境還需要同時處理標點、數字、書名號、引號、括號以及連續句子之間的換行關係，並且在較長的段落中保持穩定的字面和行距。對於共享的漢字碼位，不同地區使用的字體還可能給出不同的字形，因此同一個「骨、直、令、具、角」在不同字體家族中會表現出細微而穩定的區域差異。這裡特意使用傳統漢字而仍然選擇大陸字體家族，以檢查字符繁簡與地域字形選擇能否保持彼此獨立，並觀察長段落中共享漢字、繁體字和全角標點的連續排版。

Taiwan Traditional Chinese — Noto TC The Taiwan family is tested with continuous Taiwan Traditional Chinese rather than with a short regional-glyph sample.

正體中文的正文排版除了字符本身之外，也牽涉字型、標點、行距與段落節奏。當文件同時包含不同語言或不同地區的漢字時，程式不能僅依 Unicode 碼位猜測應採用哪一套地區字形，而應由明確的語言或字型選擇來決定。這個段落使用臺灣正體字型，

並刻意包含「骨、直、令、具、角、門、國、學、龍」等適合觀察地區字形差異的共享漢字，同時加入引號、頓號、逗號與句號，藉此確認較長的正體中文內容在正常欄寬中能自然換行，而且從段首到段末始終維持同一個本地 CJK family。

Document Traditional Chinese — Shanggu The following paragraph is deliberately left completely unwrapped. It therefore remains under the document CJK family, Shanggu, and provides the control against which the explicit Chinese, Japanese, and Korean families can be compared.

漢字進入現代數位排版之後，同一個編碼往往可以由許多不同的字體呈現，而字形的歷史層次、地域習慣和設計風格並不會因此消失。傳統中文尤其適合用來觀察這種情形：有些字在不同地區的正文中結構略異，有些字的筆形和部件位置不同，而另一些差別只有在一整段文字的灰度與節奏中才會顯現。這個段落不使用任何語言專用命令，使所有漢字直接落入文件的 Shanggu CJK family；因此它不只是檢查單字是否存在，也檢查文件級 CJK 選擇在較長的連續文本、標點和自動換行之間能否穩定維持。

6.2 Korean and Historical Hangul

Modern Korean — Hangul The first Korean specimen deliberately avoids Hanja. It tests ordinary modern Hangul composition, punctuation, word spacing, and paragraph wrapping.

한글로 쓰인 긴 문장은 여러 음절 블록이 끊임없이 이어지면서도 낱말 사이의 간격과 문장 부호의 위치가 자연스럽게 유지되어야 한다. 짧은 보기에서는 글꼴이 몇 글자를 제대로 보여 주는지만 확인할 수 있지만, 실제 문단에서는 줄 끝에서 낱말이 어떻게 배치되는지, 괄호와 따옴표가 앞뒤 글자와 어떻게 어울리는지, 여러 줄을 지나도 같은 글꼴 선택이 계속 유지되는지를 함께 살펴볼 수 있다. 이 문단은 완성형 한글이 이어지는 일반적인 한국어 본문을 충분히 길게 배치하여 문장 전체의 모양과 줄 간격을 확인하고, 뒤에 나오는 한자 혼용문과 옛한글 자료를 비교하기 위한 현대 한글 기준 표본으로 사용한다.

Modern Korean — mixed Hanja and Hangul The next paragraph keeps Hangul and Hanja inside a single explicit Korean family. Shared Han characters must therefore retain Korean regional forms without interrupting the surrounding Hangul text.

韓國語의 現代 文體는 오늘날 대체로 한글만으로 적지만, 歷史、言語、哲學、法律과 같은 분야에서는 漢字를 함께 적은 문헌을 어렵지 않게 만날 수 있다. 이런 漢字混用文을 조판할 때에는 한글 音節과 句讀點 사이에 들어가는 漢字가 갑자기 다른 地域 字形으로 바뀌지 않아야 하며, 韓國、朝鮮、漢字、文化、文字와 같은 어휘도 앞뒤의 한글 문맥과 한 family 안에서 자연스럽게 이어져야 한다. 이 문단은 國語의 읽기 흐름을 유지하면서 여

러 漢字語와 고유어를 섞어, 실제 문헌에서 볼 수 있는 긴 혼용문이 줄바꿈을 거쳐도 안정적으로 조판되는지를 시험한다.

Jeju language — corpus specimen Jeju is useful here because practical orthographies make active use of the arae-a and decomposed Hangul sequences that are rare in ordinary modern Korean. The following is a composite regression specimen assembled from short attested Jeju-language passages rather than a newly invented "standard" Jeju paragraph.

제주말 (濟州말) 은 제주섬서 골아지는 말이다. 제주말모음은 시간 지나멍 여러 가지 변화가 이섯다. 제주섬은 한라산광 ᄌᆞᆺ듸 오름덜록 일뤄진 섬이라. 엿적 화산활동으록 굼부리가 꿰쉉 곤데곤데 ᄋᆞ라 오름이 하영 삿다. 낭 (木) 은 나무질로 뒌 줄기를 가지고 잇인 여러헤살이 식물이다. 제주말 위키벡과 또시는 제주어 위키백과는 제주말로 운영헴신 위키벡과 언어판이라.

Middle Korean and Old Hangul — historical corpus specimen The historical specimen deliberately retains decomposed jamo and tone marks. Rather than normalizing several sources into modern spelling, short genuine historical fragments are placed consecutively to form a paragraph-level shaping regression.

나·랏:말ᄊᆞ·미中듀ᄋᆞ國·귁·에 달·아문ᄍᆞᆼ·와로서르ᄉᆞᄆᆞᆺ·디아·니ᄒᆞᆯ·씨 ·아기 ᄇᆡ·여셔 과골·이 가·ᄉᆞᆷ 알·파주·글 ·ᄃᆞᆺ·ᄒᆞ거·든 두들겐 開闢ᄒᆞᆯ 젯 므리 疏通ᄒᆞ고 남ᄀᆞᆫ 古今엣 남기 섯것도다 머·리 알ᄑᆞ·고 ·고히 막·고 머·리와·눈·괘 환·티 아·니커·든 ·긔운·이 우·흐로 올·아 머·리 알·포미 ᄠᆞ리ᄂᆞᆫ ·ᄃᆞᆺ거·든 넙고 큰 天造 ᄉᆞ이예 다ᄉᆞᆯ며 어즈러우민ᄃᆞᆯ 엇디 덛덛ᄒᆞᆫ 數ㅣ리오

The corpus paragraph above tests realistic historical material. The following synthetic boundary sequence has a different purpose: it explicitly crosses Hangul Jamo Extended-A, ordinary Jamo, and Extended-B while remaining inside one shaping run.

·딸

Modern decomposed control:

각

6.3 Japanese and Ryukyuan

Modern Japanese Modern Japanese combines kana, regional Han forms, punctuation, and frequent script changes inside the same sentence. The entire paragraph remains under the explicit Japanese family.

現代日本語の文章では、漢字、ひらがな、カタカナが一つの段落の中で何度も切り替わり、句読点や括弧、かぎ括弧も連続して現れる。短い文字列だけではフォントが表示できることしか分からないが、長い本文にすると、漢字と仮名の大きさの釣り合い、約物の位置、行末での折り返し、そして次の行へ移った後の字形の連続性まで確認できる。この段落では「日本語」「漢字」「文字」「学校」「文化」「骨」「直」「令」「具」「角」などを通常の文脈に入れ、日本語用の地域字形を保ったまま複数行の文章が安定して組版されるかを試している。

Historical kana orthography The opening sentence below is historical text; the continuation is a showcase-written regression passage that deliberately retains historical kana spellings such as ゐる, やう, and 言ふ. This keeps the specimen long enough for paragraph testing without presenting newly written material as an archival transcription.

真夏の宿屋は空虚であつた。窓の外には白い雲がゆつくり流れ、庭の木立では蟬が絶えず鳴いてゐる。旅人は古い椅子に腰をおろし、まだ來ない馬車の音を待ちながら、時計の針ばかりを眺めてゐた。誰も何も言はないので、午後の光だけが廊下を少しづつ移つて行くやうに見えた。かうして長い時間が過ぎても、玄關の戸は一度も開かず、遠くの道から聞えて來る物音だけが、時々靜かな宿の空氣を動かしてゐるのであつた。

Japanese Kanbun The first sentence is taken from Rai San'yō's *Nihon Gaishi*; the remainder is a showcase-written Kanbun-style regression continuation. The purpose here is typographic: Japanese Han glyphs must remain selected even when no kana occur in the run.

外史氏曰、吾讀舊志、見鳥羽帝時數下制符。蓋日本文章之傳、漢籍之學與國史之記相因而久矣。昔之學者讀經史、作序記、論人物、述治亂、多用漢文而各成其體。今取其字列於一段、所以驗日本字形之連續、句讀之位置、長行之折返、以及共享漢字在無假名介入時亦不離其本地字體也。文字雖同出一碼、而骨直令具角門國學龍諸字之形、隨地域字體而微異，故尤宜以連續文章觀之。

Okinawan — corpus specimen Okinawan is represented by a composite specimen from the Wikimedia Okinawan test project. Its mixed use of kana and Han characters makes it a

useful additional Japanese-family regression, while the orthography should not be read as a claim that one universal Okinawan spelling standard exists.

沖縄口ぇー琉球諸語ぬてぃ壱ちやいびーん。うちなーぐちぬ首里那覇方言ぉー琉球国ぬじょーま言葉やびたん。琉球王國や、昔ジジ中心んかい存在さる王国やたん。王や首里天加那志とぅ言ん。首里城や、沖縄県ぬ城やん。琉球王朝ぬ王城っし、沖縄県ぬいっちん大さるぐしくやん。漢字ぇー、大昔唐ぬ国んかいそーじー現りーる文字ぬてぃーちやん。やまとぅ、高麗、るーちゅー、ベトナムぉー、大昔唐ぬ国から漢字伝わてぃちゃん。

Ainu Ainu written in the modern katakana orthography makes extensive use of Katakana Phonetic Extensions for syllable-final consonants. The following specimen therefore exercises ordinary katakana together with characters such as ㇰ, ㇱ, ㇺ, ㇻ, and the combining form ㇷ゚ inside continuous text.

イタㇰプリカ　ソモネコㇺカ　シサムモシㇼ　モシㇼソカワ　チヌㇺケニㇱパ　チヌㇺケ　カッケマッ　ウタペラリワ　オカウㇱケタ　クニネネワ　アイヌイタㇰアニ　クイタㇰルウェ　ネワネヤクン　ラモッシワノ　クヤイライケㇷ゚　ネルウェタパンナ。エエパキタ　カニアナㇰネ　アイヌモシㇼ　シシㇼムカ　ニㇷ゚タニコタン　コアパマカ　クネルウェネ　ウッチケクニㇷ゚　ラカサㇰペ。

6.4 Vietnamese

The same constitutional passage is used throughout this subsection so that the specimens differ primarily in writing system or font selection rather than in linguistic content. Quốc Ngữ is first typeset with the document Latin family, Libertinus, and then with the dedicated Vietnamese family. The corresponding Hán-Nôm text uses the VinaWiki repertoire.

Quốc Ngữ — document Latin family The first specimen is deliberately unwrapped and therefore remains in the document Latin family, Libertinus. It tests Vietnamese diacritics without a language-specific font override.

Trải qua mấy nghìn năm lịch sử, Nhân dân Việt Nam lao động cần cù, sáng tạo, đấu tranh anh dũng để dựng nước và giữ nước, đã hun đúc nên truyền thống yêu nước, đoàn kết, nhân nghĩa, kiên cường, bất khuất và xây dựng nên nền văn hiến Việt Nam. Từ năm 1930, dưới sự lãnh đạo của Đảng Cộng sản Việt Nam do Chủ tịch Hồ Chí Minh sáng lập và rèn luyện, Nhân dân ta tiến hành cuộc đấu tranh lâu dài, đầy gian khổ, hy sinh vì độc lập, tự do của dân tộc, vì hạnh phúc của Nhân dân. Cách mạng tháng Tám thành công, ngày 2 tháng 9 năm 1945, Chủ

tịch Hồ Chí Minh đọc Tuyên ngôn độc lập, khai sinh ra nước Việt Nam dân chủ cộng hòa, nay là Cộng hòa xã hội chủ nghĩa Việt Nam.

Quốc Ngữ — Vietnamese family The same text is repeated through the explicit Vietnamese local family, allowing the dedicated Vietnamese font to be compared directly with Libertinus under identical paragraph conditions.

Trải qua mấy nghìn năm lịch sử, Nhân dân Việt Nam lao động cần cù, sáng tạo, đấu tranh anh dũng để dựng nước và giữ nước, đã hun đúc nên truyền thống yêu nước, đoàn kết, nhân nghĩa, kiên cường, bất khuất và xây dựng nên nền văn hiến Việt Nam. Từ năm 1930, dưới sự lãnh đạo của Đảng Cộng sản Việt Nam do Chủ tịch Hồ Chí Minh sáng lập và rèn luyện, Nhân dân ta tiến hành cuộc đấu tranh lâu dài, đầy gian khổ, hy sinh vì độc lập, tự do của dân tộc, vì hạnh phúc của Nhân dân. Cách mạng tháng Tám thành công, ngày 2 tháng 9 năm 1945, Chủ tịch Hồ Chí Minh đọc Tuyên ngôn độc lập, khai sinh ra nước Việt Nam dân chủ cộng hòa, nay là Cộng hòa xã hội chủ nghĩa Việt Nam.

Hán-Nôm — VinaWiki repertoire The Hán-Nôm specimen corresponds to the Quốc Ngữ passage above rather than serving as an unrelated coverage string. Because the local family uses the VinaWiki font, the specimen is taken from the VinaWiki transcription itself and therefore tests the repertoire for which that font is intended.

疎過尒酐薢歷史、人民越南勞動勤劬、創造、鬥爭英勇抵竫渃吧符渃、缶熏燽皝傳統慀渃、團結、仁義、堅強、不屈吧墡竫皝坪文獻越南。自薢1930、迂事領導貼黨共產越南由主席胡志明創立吧煩鍊、人民些進行局鬥爭數戣、涪艱苦、犧牲爲獨立、自由貼民族、爲幸福貼人民。革命胋磣成功、暝2胋9薢1945、主席胡志明讀宣言獨立、開生罨渃越南民主共和、肣羅共和社會主義越南。

6.5 Tangut

Wikisource provides a sizeable Unicode Tangut corpus, including poetry, Buddhist literature, maxims, and the Tangut text of the Liángzhōu bilingual stele. The following paragraph is therefore a composite corpus specimen: several short attested sequences from independent Tangut texts are joined only for visual regression. It is not presented as one continuous historical work.

䋣莥敘繖絀疋，猱猱繖虈敆緻，驦纁穉羝紓概蔆，衱孤妥豹豹，蘇餓蠢鯈礙剱螽，縩礣蔉厰蠡甕巃，鱀颯鼀隴概菽赧毅逡，韌宛敘撐弩傯岕岕孅概陲，蔭霈虈粗概羝赧

𪜀𩁹，𦉢𪚥𪘂𪘉𩂣𩡧𩠐，𪘋𢆶𢓊𢦏𪘒𪜀𪘂，𢦏𢦏𢖍𪘒𪘁𪘂𪘀𪜀𪜀，𪘁𢖍𢍰𪘂𪗋𪘀𪘂𪜀𪘂𪘂。

6.6 WenJin Multi-Plane Coverage

WenJin is different from the language-specific families above. It is exposed as one public local family whose internal fallback configuration spans multiple CJK planes. The paragraph below is therefore intentionally a coverage stress test rather than a linguistic specimen: ordinary Han text is interleaved with supplementary-plane characters while the public command remains unchanged.

中國文字的編碼範圍遠遠超過基本多文種平面，因此只測試一二三天地玄黃這類常用字並不足以確認大型漢字字體的覆蓋是否正常。這一段先由普通漢字開始，隨後插入第二平面的𠀀、吉、𡴀、堅、𠀁，再回到典籍文字與古今學術，之後繼續進入第三平面的𠀂、丰、𠀃、𠀄、巨。當不同平面的字符在同一個長段落中交錯出現時，使用者仍然只看見一個 `\WJ` family；Plane 0、Plane 2 和 Plane 3 的具體 fallback 應由內部配置完成，而不應要求正文作者為不同 Unicode plane 手動切換命令。

The following table complements the paragraph test with deterministic block coverage samples. For each CJK Unified Ideographs block, the first twenty encoded ideographs are shown. This makes missing coverage or an incorrect Plane 0, Plane 2, or Plane 3 fallback immediately visible and keeps the regression sample stable across releases.

6.7 Regional Forms of Shared Han Characters

The preceding specimens test normal paragraphs. Only here are shared codepoints deliberately isolated. Every cell below contains the same Unicode character; the visible difference should therefore come from the selected regional family rather than from different encoded variants.

7 Document Features

feature system (composable document modules) is demonstrated throughout this PDF, but this section gives each major public feature an explicit specimen as well. The showcase therefore tests not only isolated modules but also whether they coexist in a realistic document.

Table 5: WenJin CJK block coverage

Block	Range	Twenty-character specimen
CJK Unified Ideographs	U+4E00–U+9FFF	一丁丐七丄丅丆万丈三上下丌不与丏丐丑丞专
Extension A	U+3400–U+4DBF	丷丮乿乍乍乂肙乞乞乭乖乖缶乭乖乖弩奚乭乭
Extension B	U+20000–U+2A6DF	亡亏乙且工歹丰且丘亡亦丈正世丰屯屯丙兮引
Extension C	U+2A700–U+2B73F	丗丒瓦軣奇奔飆瀲氏岼冫夷屑乖瓶猝巫杏硿垃
Extension D	U+2B740–U+2B81F	丸弗所爾久夾今侖侲倉傳僾奘岡凗滲前勢卑挔
Extension E	U+2B820–U+2CEAF	七不卡东夯匝所轩皀兹趴堯秊杯到耔俄昱畬陯
Extension F	U+2CEB0–U+2EBEF	寸才几万刁亏可歹屮彐夸豕池旡电車芷玉卌襾
Extension I	U+2EBF0–U+2EE5F	正丹乍窔仉仿伀仅伤佩伪侄岱俱傂俞傣傻拿龠
Extension G	U+30000–U+3134F	乚丰匡冂巨丸立歹专申兮巡吊砑垚垠冤財鬲甄
Extension H	U+31350–U+323AF	元吉耿尮珑厥邹浮劲齣豸呇谷郤滓裳猃皋礪硇篡
Extension J	U+323B0–U+3347F	五北丏彐罕葙餑銑媩墎禾烺鉀嘲轇灺乖卒祈烝

Table 6: Regional forms of shared Han codepoints

Character	Shanggu	Noto SC	Noto TC	Japanese	Korean
骨	骨	骨	骨	骨	骨
直	直	直	直	直	直
令	令	令	令	令	令
具	具	具	具	具	具
角	角	角	角	角	角
門	門	門	門	門	門
國	國	國	國	國	國
學	學	學	學	學	學
龍	龍	龍	龍	龍	龍

7.1 Mathematics

Mathematical font selection is deliberately left at IMPE's default setting. The following single display combines several common mathematical alphabets, operators, delimiters, accents, matrices, integrals, sums, limits, radicals, and nested scripts as a compact stress specimen.

$$
\begin{aligned}
\mathcal{Z}_{\beta,\mu} &\coloneqq \operatorname{Tr}_{\mathscr{H}} \exp\Big[-\beta\left(\widehat{\mathbf{H}} - \mu\widehat{\mathbf{N}}\right)\Big] \\
&= \sum_{n=0}^{\infty} \frac{1}{n!} \int_{\mathbb{R}^{3n}} \prod_{j=1}^{n} \frac{d^3\mathbf{p}_j}{(2\pi\hbar)^3}\, \exp\left[-\sum_{j=1}^{n} \sqrt{m_j^2 c^4 + c^2\|\mathbf{p}_j\|^2}\right] \\
&\quad \times \det\begin{pmatrix} \dfrac{\partial^2 \mathcal{L}}{\partial q^i \partial q^j} & \nabla_\mu F^{\mu\nu} & \left\langle \psi \middle| \widehat{\mathbf{H}} \middle| \phi \right\rangle \\ \displaystyle\sum_{k=1}^{\infty} \frac{(-1)^{k+1}}{k^2} & \sqrt[3]{1+\sqrt{1+\lambda^2}} & \dfrac{\mathbb{E}[X^4] - 3\mathbb{E}[X^2]^2}{\operatorname{Var}(X)^2} \\ \displaystyle\oint_{|z|=1} \frac{\log(1+z)}{z^{r+1}}\, dz & \displaystyle\prod_{p\in\mathbb{P}} \left(1 - p^{-s}\right)^{-1} & \displaystyle\lim_{\varepsilon\to 0^+} \frac{f(x+\varepsilon) - f(x-\varepsilon)}{2\varepsilon} \end{pmatrix} \\
&\quad + \underbrace{\widetilde{\mathbf{A}}^{\dagger}{}_{\mu}{}^{\nu}\, \overline{\mathbf{B}}_{\nu}{}^{\rho}\, \widehat{\mathbf{C}}_{\rho}{}^{\mu}}_{\text{accented tensor contraction}} + \frac{\Gamma\left(\frac{1}{2} + i\alpha\right)\zeta(2s)}{\sqrt{2\pi\sigma^2}} \\
&\quad + \begin{cases} e^{-1/x^2}\sin(\pi/x), & x \neq 0, \\ 0, & x = 0, \end{cases} \\
&\quad \xrightarrow[\alpha\to 0]{\beta\to\infty} \left(\bigoplus_{j=1}^{r} \mathbb{C}^{\,n_j}\right) \otimes \bigwedge^{k} V \\
&\quad \Longrightarrow \mathbb{P}\left(|X - \mathbb{E}[X]| \geq t\right) \leq \frac{\operatorname{Var}(X)}{t^2}, \qquad t > 0, \\
&\quad \alpha, \beta, \gamma, \delta, \varepsilon, \theta, \vartheta, \lambda, \mu, \nu, \xi, \pi, \varpi, \rho, \varrho, \sigma, \phi, \varphi, \chi, \psi, \omega \in \mathbb{R}.
\end{aligned}
$$

7.2 Tables

Table 7 uses the IMPE book-table environment rather than a raw `tabular` wrapper.

7.3 Lists and Example Blocks

A normal list remains ordinary LaTeX:

1. automatic font ownership;

Table 7: Feature integration summary

Feature	Public route	Exercised here
Mathematics	`math`	yes
Tables	`tables`	yes
Images	`image`	yes
Citations	`citations`	yes
Index	`index`	yes
Hyperlinks	`hyperlinks`	yes
Running headers	`headers`	yes
List environments	`lists_envs`	yes

2. script-specific behavior where required;
3. reusable document features;
4. stable public interfaces.

The IMPE `ExampleBlock` environment provides a distinct indented specimen:

> *A canonical showcase should be readable as a document, useful as a visual regression specimen, and representative of the public interface that new documents are expected to use.*

7.4 Images

The image feature is exercised with the standard TEX distribution's example image when it is available. This keeps the canonical source self-contained without committing an otherwise meaningless binary illustration.

Figure 1 also exercises cross-reference hyperlinks.

7.5 Citations and Bibliography

Citation support is loaded through the feature system and configured before feature activation. A normal textual citation can therefore be written as Knuth (1984), while a parenthetical citation appears as (Knuth, 1984).

The bibliography is printed near the end of the document.

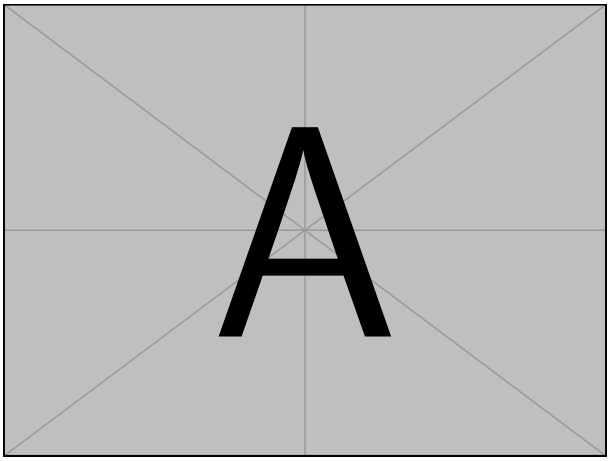


Figure 1: Image-feature specimen using the standard TEX example image.

7.6 Index Terms

Several concepts in earlier sections were inserted with the public `\Term` interface. For example, **canonical showcase** (release-time specimen) itself is now indexed. The generated index at the end of the PDF therefore also exercises indexed terms, optional descriptions, and hyperlinks back to their first occurrence.

7.7 Hyperlinks, Footnotes, and Running Headers

The table of contents is clickable, numbered headings carry stable targets, and section headings can link back toward their table-of-contents entries. The running header above this page is produced by the `headers` feature.

Footnote markers are also bidirectional.[2]

For an external-link specimen, the project repository can be represented with a normal hyperlink: IMPE LATEX System on GitHub.

References

Knuth, D. E. (1984). *The texbook*. Addison-Wesley.

[2]Clicking the footnote marker jumps to this note; the note marker can link back to the original position.

Index

canonical showcase (release-time specimen), 23
CJK family selection (continuous text, shared Han ownership, and explicit regional overrides), 12

feature system (composable document modules), 19

Multilingual composition (combined routing and local overrides), 3

OpenType shaping (generic complex shaping), 10

Right-to-left text (directional behavior), 7

Unicode-range routing (automatic font ownership), 7

Vertical writing (horizontal-layout-compatible vertical embedding), 9