

# Unicode と XSL による多言語組版

2003 年 12 月 29 日  
アンテナハウス株式会社

## コンピュータによる多言語組版の課題

最初に、コンピュータによる多言語組版に取り組むにあたって問題になると考えられる点を一覧してみる。これらはいずれも単独でも難しい課題だが、技術が日進月歩で進歩しているため、それを理解して使いこなすのはさらに難しくなる。

この資料は、コンピュータで多言語組版を行う方法について、最初に多言語組版の課題を整理する。次いで、Unicode、XML と XSL (Extensible Stylesheet Language) による多言語組版の現状を述べ、最後に実際の組版例を示す<sup>(1)</sup>。

## 組版対象のデータをどう作成するか？

コンピュータで文字情報を処理するには、まず、その文字情報が符号化されたデータとして作成されている必要がある。日本語のみを対象とするのと比べると、多言語のデータを作成するのは格段に難しくなる。

### 1. 文字符号化方式の選定

- 文字符号化方式 (いわゆる文字コード) は主として国単位で、地域ローカルな文字コード表として標準化されてきた。しかし、地域ローカルな文字コード表でデータを表す方法では、多言語混在の文書を簡単に扱うことはできない。多言語編集、特に多言語混在の文書を編集、組版しようとしたら Unicode が必須であろう。
- Unicode はどの言語まで使えるか？Unicode の最新の標準化状況とそれを実装した製品にどのようなものがあるか？特に Unicode は、か

なり早いスピードで進化しているので最新の情報を正しく把握する必要がある。

- Unicode にはどのような問題があるだろうか？
- 従来のシフト JIS や ASCII 符号化によるテキストファイルでは使えるコードの種類が限られていた。これに対して、Unicode のテキストでは、新しいコードがいろいろ定義されている。例えば、U+2000 から U+200F までの 16 文字に次のようなコードがある。こういうコードは組版上どういう意味をもつのか？どうやって使いこなすのか？

### U+2000 からの 16 文字

```
2000;N # EN QUAD
2001;N # EM QUAD
2002;N # EN SPACE
2003;N # EM SPACE
2004;N # THREE-PER-EM SPACE
2005;N # FOUR-PER-EM SPACE
2006;N # SIX-PER-EM SPACE
2007;N # FIGURE SPACE
2008;N # PUNCTUATION SPACE
2009;N # THIN SPACE
200A;N # HAIR SPACE
200B;N # ZERO WIDTH SPACE
200C;N # ZERO WIDTH NON-JOINER
200D;N # ZERO WIDTH JOINER
200E;N # LEFT-TO-RIGHT MARK
200F;N # RIGHT-TO-LEFT MARK
```

2. コンピュータの選定。ハードウェアと OS はどうやって選ぶ？

<sup>(1)</sup>この文書は、アンテナハウスの標準文書形式である SimpleDoc.dtd に準拠した XML で記述して、XSL Formatter V2.5 で組版して PDF 化したものである。この文書自体、Unicode、XML、XSL による多言語組版の実例でもある。

- Macintosh、Windows2000/XP、Unix（Solaris等）、Linux、JAVA、どのような環境を選ぶか？
- Windowsでは、Uniscribeという多言語処理層（ライブラリー）が整備されていて、これを使いこなすことでアジア圏の言語を含める多言語処理が可能になる。Internet ExplorerやMicrosoft WordはUniscribeを使って、多様なアジア圏の言語まで処理ができるようになっている。
- 多言語処理ではWindowsがもっとも進んでいるが、ではJAVAではアジア圏の言語をどこまで処理できるだろうか？また、LinuxやSolarisなどのUNIXでの多言語組版の現状は？

### 3. データをコンピュータにどうやって入力するか。

- データ入力用のソフトウェアにはどんなものがあるか？
- キーボードの選定、キーボードをどうやって用意するか？キーボードは各国で標準化されていて、各国で販売しているPCには、その国の方式のキーボードが付随している。特に日本で販売しているPCに付属のキーボードで他の国の言語を入力するのはどうしたら良いか？
- IMEが必要か？その選定方法は？周知のように日本語を入力するには、ローマ字入力、かな漢字変換を行う方式が主流である。しかし、漢字の日本語読みを知らない外国人が、漢字を入力するのにローマ字読みで入力するのは無理ではないかと思う。それと同じで、中国語の入力方式として中国人にはピンイン入力が自然かも知れないが、日本人がピンインで入力するのは難しいだろう。

### 4. データの表現法

- アプリケーション依存のバイナリかアプリケーション独立のXML形式にするか？
- 多言語処理を実現するにはXMLが一番良い。しかし、XMLを使いこなすのはハードルが高いのが事実である。XMLのタグはそれほど難しいものではないが、一般の人達はタグ

を異様に恐れる傾向がある。XMLへのハードルをどうやって下げたら良いか？

- XMLを使う場合、データ構造（スキーマ）を設計しなければならない。
- 新しくデータ構造を定義するのではなく、既存のDTD/スキーマを使えないか。例えば、DocBook.dtdは使えないのか？
- 新しい標準DTD/Schema定義の動向は？

### 5. 編集ソフトウェアの選定

- 使い慣れた編集ソフトを使えるかどうかは、文書作成の生産性に非常に大きな影響を与えるので、どのような編集ソフトを使うことができるかを選択することは非常に重要である。この観点からは、Microsoft Wordを使えれば非常に便利だ。Microsoft Wordは多言語編集ソフトとしてどこまで使えるか？
- 多言語対応を標榜している編集ソフトは多数ある。しかし、例えば、英語、西欧の言語、日本語、中国語、韓国語、アラビア語、ヘブライ語、タイ語をひとつのバージョンですべて編集できるソフトは数が少ない。もし、言語別に編集ソフトを切り替えなければならぬなら、多言語が混在する文書は作成できない。また、言語別に切り替えるとなると、操作を新しく覚えたり、データの互換性の問題も出るので、これは避けたい。しかし、どうしても言語別に編集ソフトの切り替えが必要になったらどうするか？
- 多言語をWYSIWYG、あるいはWYSIWYGに近い形で編集できるツールがあるか？あるとして、どのソフトを選択するのが最適か？
- XMLでデータを作成するには、スキーマに従ったデータの入力編集作業を支援するためのツールが必要である。そのようなソフトがあるか？
- 専門家が文書を作成する場合は、訓練、学習ができるので、タグを見せるタイプのXML編集ソフトでも使用できる。多言語でそのようなことのできるXML編集ツールがあるか？あるとして、どのソフトを選択するのが最適か？

## 組版の方法

1. レイアウトを頻繁に変えることができ、かつ、WYSIWYG で編集でき、レイアウト編集した結果を XML のソースデータにリアルタイムで反映できるような、本格的な多言語 XML 組版ソフトは存在しないのか？なぜ存在しないのか？
2. 文字を画面なり、紙、PDF なりに視覚化して表すには、フォントが必須である。では、Unicode 対応フォントにはどのようなものがあるか？
3. PDF を作成して配布したり、印刷しようとするとき、フォントのアウトラインの埋め込みが必須である。従って、多言語組版で使用するフォントは、アウトラインの埋め込みが許可されたフォントでないといけない。多言語の組版をしようとするとき、使えるフォントにはどんなフォントがあるか？
4. XSL-FO によるレイアウト指定方法
  - ・ XSL-FO はどこまで使えるか？どの程度まで、複雑なレイアウトが指定できるか？
  - ・ XSL Formatter は、XSL 仕様を満たす XML の多言語組版ソフトであるが、どんな特徴があるか？
  - ・ 組版規則が言語別に違うケースにも対処できているのか？
  - ・ 文章の中に異なる組版規則の言語が混在する時にも対処できるのか？

- ・ 右から左へ記述する言語と左から右へ記述する言語が混在しても対処できるのか？

## 印刷・PDF 作成方法

1. 多言語の印刷はどうやってやるか？
2. 印刷用 PDF と Web 用 PDF の違い、使い分け。

## その他

1. 目次の作成方法や索引の作成方法
2. 索引のソート順、言語別のソート規則、多言語混在時のソート規則

## 多言語組版の基礎知識

### 文字と言語

言語は文字によって記述される。言語をコンピュータで扱うためには、まずその前提として、言語を表記する文字を扱うことができなければならない。コンピュータで文字を扱う時は、普通は文字の種類を集合として規定し、番号付けした符号化文字集合を使う。従来は、各国・地域別に規定された符号化文字集合が使われてきた。次の表は主要な言語が主にどのような地域別の文字コードで表されるかを示す。

言語コード	言語名称	言語名称 (日)	文字の種類	地域別文字コード
ar	Arabic	アラビア語	Arabic	ASMO 449, Latin/Arabic Alphabet
bg	Bulgarian	ブルガリア語	Cyrillic	Latin/Cyrillic Alphabet
km	Cambodian	カンボジア語	Khmer	(Unicode V3.0 版で初登録)
zh-CN	Chinese (Simplified)	簡体字中国語	Simplified Chinese	GB2312, GBK, GB18030
zh-TW	Chinese (Traditional)	繁体字中国語	Traditinal Chinese	BIG5
hr	Croatian	クロアチア語	Latin	Latin Alphabet No.2,10
cs	Czech	チェコ語	Latin	Latin Alphabet No.2
da	Danish	デンマーク語	Latin	Latin Alphabet No.1,4,5,6,8,9
nl	Dutch	オランダ語	Latin	Latin Alphabet No.1,5,9
en	English	英語	Latin	Latin Alphabet No.1..10
et	Estonian	エストニア語	Latin	Latin Alphabet No.4,6,7,9
fi	Finnish	フィンランド語	Latin	Latin Alphabet No.4,6,7,9,10
fr	French	フランス語	Latin	Latin Alphabet No.9,10
de	German	ドイツ語	Latin	Latin Alphabet No.1..10 (7 除く)

言語コード	言語名称	言語名称 (日)	文字の種類	地域別文字コード
el	Greek	ギリシャ語	Greek	Latin/Greek Alphabet
he	Hebrew	ヘブライ語	Hebrew	Latin/Hebrew Alphabet
hi	Hindi	ヒンディ語	Devanagari	IS 13194 (ISCII) ほか
hu	Hungarian	ハンガリー語	Latin	Latin Alphabet No.2,10
is	Icelandic	アイスランド語	Latin	Latin Alphabet No.1,6,9
id	Indonesian	インドネシア語	Latin	Latin Characters
it	Italian	イタリア語	Latin	Latin Alphabet No.1,3,5,8,9,10
ja	Japanese	日本語	Latin、漢字、かな、カタカナ	JISX0201, JIS X0208, JIS X0212
kk	Kazakh	カザフ語	Cyrillic	Extended Latin/Cyrillic Alphabet (Cyrillic Asean)
ko	Korean	韓国語	ハングル、漢字	KS C5601, KS X1001, Johab
lv	Latvian	ラトビア語	Latin	Latin Alphabet No.4,7
ms	Malay	マレー語	Latin または Arabic	Latin Alphabet, Arabic Extended
	Lappish/Sami	ラップ語	Latin Extended	
lt	Lithuanian	リトアニア語	Latin	Latin Alphabet No.4,6,7
no	Norwegian	ノルウェー語	Latin	Latin Alphabet No.1,4..9
fa	Persian (Farsi)	ペルシャ語	Arabic	Extended Latin/Arabic Alphabet (Arabic Character 28+ Original 4 Characters)
pl	Polish	ポーランド語	Latin	Latin Alphabet No.2,7,10
pt	Portuguese	ポルトガル語	Latin	Latin Alphabet No.1,3,5,8,9
ro	Romanian	ルーマニア語	Latin	Latin Alphabet No.10
ru	Russian	ロシア語	Cyrillic	koi8-r, Latin/Cyrillic Alphabet 32 Chars (not compatible with Ukrainian)
sr	Serbian	セルビア語	Cyrillic	Latin/Cyrillic Alphabet (Serbian)
sk	Slovak	スロバキア語	Latin	Latin Alphabet No.2
sl	Slovenian	スロベニア語	Latin	Latin Alphabet No.2,4,6,10
es	Spanish	スペイン語	Latin	Latin Alphabet No.1,5,8,9
sv	Swedish	スウェーデン語	Latin	Latin Alphabet No.1,4,5,6,8,9
sw	Swahili	スワヒリ語	Latin	
tl	Tagalog/Takalog	タガログ語	Latin	
th	Thai	タイ語	Thai	TIS 620, Latin/Thai Alphabet
tr	Turkish	トルコ語	Latin	Latin Alphabet No.5
uk	Ukrainian	ウクライナ語	Cyrillic	koi8-u, Latin/Cyrillic Alphabet 33 Chars
ur	Urdu	ウルドゥー語	Arabic Extended	
vi	Vietnamese	ベトナム語	Latin	Extended Latin Characters
xh	Xhosa		Latin	
zu	Zulu	ズールー語	Latin	

## Unicode

現在、国別文字コードに代わり、Unicode が単一のアプリケーションで全世界のあらゆる言語の文字とテキストを取り扱う基盤を与えている。

## Unicode の歴史

1991 年 10 月 Unicode 1.0.0 版発行  
1996 年 7 月 Unicode 2.0.0 版発行  
1999 年 9 月 Unicode 3.0.0 版発行

2002年 3月 Unicode 3.2.0 版発行

2003年 4月 Unicode 4.0.0 版発行

Unicode は単に符号化文字集合を定義するだけではなく、テキスト処理のための各種の基準を定義している。①文字を書き進める方向を規定する Unicode 文字データベース、②禁則文字など文字の改行特性を規定する Unicode Line Breaking Properties、③双方向性処理を規定する UnicodeBIDI、などの技術レポートは Unicode の仕様の一部である。これらのレポートは、完全とは言えないこともあるが、アプリケーション・プログラムを作成する際に参照として貴重な情報源である。Unicode の仕様とこれらの技術レポートによって、多言語組版エンジンの開発が非常に容易になった。

### OS とアプリケーションの内部文字コード

1980年代から1990年代までのパーソナル・コンピュータの OS は各地域別の文字コード表に基づいていたため、OS の上で動作するアプリケーション・プログラムも地域別であった。例えば、日本語 Windows98/Me は、OS 内部ではシフト JIS (JISX0201 と JIS X0208) でテキストと文字処理を行っている。このため WindowsMe で動くアプリケーションでは、A with diacresis : Ä、O with diacresis : Ö、U with diacresis : Ü などのラテンの特殊文字を正しく処理するのが容易ではない。これらの文字は日本語の半角カタカナと文字コードがぶつかってしまうためである。

Microsoft の Windows2000/XP は、OS 内部の処理が Unicode が基本になっていて、多言語処理機能が飛躍的に強化されている。Microsoft の OS で多言語処理を行おうとしたら、Windows2000 または XP を選定する必要がある。

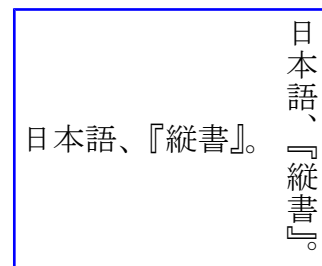
アプリケーション・ソフトウェアには、内部的に Unicode でデータを処理しているものと、ローカル文字コードで処理しているものがある。多言語の処理を行おうとしたら、内部的に Unicode 処理をしているアプリケーション・ソフトウェアを選択する必要がある。例えば、XSL Formatter や Microsoft Word2000/XP は、Unicode アプリケーションであるが、FrameMaker は、Unicode アプリケーションではない。

### アプリケーションの役割

アプリケーション・ソフトウェアが Unicode を扱えるだけでは多言語処理ができるとは言えない。Unicode と多言語処理の間には、超えなければならない問題が沢山ある。文字コードのレベルでいうと次のようなものである。

#### グリフ置換

日本語や中国語（繁体字）では同一文章を横書きと縦書きができる。横書きと縦書きでは句読点や括弧類は同じ文字コードを異なるグリフで表示・印刷する必要がある。組版エンジンは自動的にグリフの入れ替えを行わねばならない。



また、アラビア語では、同じ文字コードであっても単独の時、単語の開始位置に表れる時、中間、終了位置に表れる時でグリフが変化する。アラビア語を正しく扱うソフトウェアは文字の位置によるグリフの入れ替えを実現しなければならない。

#### 音節文字

タイ語、クメール語、ラオ語などの東南アジア大陸部の文字は子音字を中心に母音記号を上下、左右に配置する音節文字である。これらの言語の文字コードは、Unicode では子音字、母音記号、声調記号の記号レベルで規定されている。そこで、アプリケーションは複数の文字コード・ポイントのシーケンスをまとめて子音字の上下左右に配置して表示したり印刷する必要がある。

#### フォント

文字コードに次いで重要な基盤はフォント技術である。実際に使えるフォントがなければ、文字を画面に表示したり、印刷できない。次の表は Microsoft Windows2000/XP に標準で添付されているか、あるいは、インターネットで無償ダウンロードできる Windows2000/XP 用のフォントのリストである。

Arial Unicode MS は、Unicode の全域をカバーする唯一のフォントである。しかし、Unicode3.2 の全文字は揃っておらず、また、必ずしも品質は良くないという欠点がある。

言語別にみれば欧米の言語、スラブ系言語、日本語、中国語（繁体字、簡体字）、韓国語、アラビア語、ヘブライ語、タイ語等については TrueType または OpenType（TrueType フォーマット）のフォントを無償で揃えることができる。もちろん、デザインを重視した場合、このようなフォントのみでは不足であるが、技術説明書、製品のサービス説明書などの用途では十分実用的である。

Windows2000 を標準でインストールした時に、表で Windows2000 と示すすべてのフォントがインストールされる訳ではない。Angsana（タイ語用のフォント）や Mangal（ヒンディー語用のフォント）など Windows2000/XP を標準でインストールしたままではインストールされない。コントロールパネルの「地域のオプション」→「システムの言語設定」でタイ、インドなどを選択してシステムを再設定すると初めてインストールされる（次の図を参照）。

フォント・ファミリー	カバーする主要な文字	入手方法	分類
Arial Unicode MS	Unicode V2 全文字	Office2000/XP etc.	Sans-serif
Arial	Latin, Greek, Cyrillic, Arabic, Hebrew	2000/XP	Sans-serif
Courier New	Latin, Greek, Cyrillic, Arabic, Hebrew	2000/XP	Monospace
Lucida Console	Latin, Greek, Cyrillic	2000/XP	Monospace
Lucida Sans Unicode	Latin, Greek, Cyrillic, Hebrew, 記号類	2000/XP	Sans-serif
Microsoft Sans Serif	Latin, Greek, Cyrillic, Arabic, Hebrew, Thai	2000/XP	Sans-serif
Tahoma	Latin, Greek, Cyrillic, Arabic, Hebrew, Thai	2000/XP	Sans-serif
Times New Roman	Latin, Greek, Cyrillic	2000/XP	Serif
Vernada	Latin, Greek, Cyrillic	2000/XP	Sans-serif
Arabic Transparent	Arabic	2000/XP	Sans-serif (Latin), Cursive (Arabic)
Traditional Arabic	Arabic	2000/XP	Sans-serif (Latin), Cursive (Arabic)
Sylfaen	Latin, Greek, Cyrillic, Armenian, Georgian	XP	Serif
MS Hei	Simplified Chinese	IE5, Global IME5	Monospace (Latin), Sans-serif (Chinese)
MS Song	Simplified Chinese	IE5, Global IME5	Monospace (Latin), Serif (Chinese)
SimSun	Simplified Chinese	XP	Monospace (Latin), Serif (Chinese)
MingLiU	Traditional Chinese	2000/XP	Monospace (Latin), Serif (Chinese)
PMingLiU	Traditional Chinese	Office2000	Serif
Mangal	Devanagari	2000/XP	
Palatino Linotype	Greek Polytonic	2000/XP	Serif
Shruti	Gujarati	XP	
Raavi	Gurmukhi	XP	
David	Hebrew	2000/XP	Serif
David Transparent	Hebrew	2000/XP	Serif
Fixed Miriam Transparent	Hebrew	2000/XP	Monospace
Miriam	Hebrew	2000/XP	Sans-serif
Miriam Fixed	Hebrew	2000/XP	Monospace

フォント・ファミリー	カバーする主要な文字	入手方法	分類
Miriam Transparent	Hebrew	2000/XP	Sans-serif
Rod	Hebrew	2000/XP	Monospace
MS Gothic (MS ゴシック)	Japanese	2000/XP	Monospace (Latin), Sans-serif (Japanese)
MS Mincho (MS 明朝)	Japanese	2000/XP	Monospace (Latin), Serif (Japanese)
Tunga	Kannada	XP	
Batang	Korean	2000/XP	Serif
Gulim Che	Korean	IE5, Global IME5	Monospace (Latin), Sans-serif (Korean)
Estrangelo Edessa	Syriac	XP	
Latha	Tamil	2000/XP	
Gautami	Telugu	XP	
MV Boli	Thaana	XP	
Angsana New	Thai	2000/XP	Serif
Cordia New	Thai	2000/XP	Sans-serif
IrisUPC	Thai	2000/XP	Sans-serif



地域のオプションの設定

## PDF 技術

多言語組版のもう一つの推進要素は PDF 技術だろう。多言語の説明書を組版することができても、それを紙に印刷して持ち歩かなければならないということ

では、全地球をリアルタイムでカバーする電子化時代には間に合わない。

紙を電子化した媒体である PDF ができたことによって、多言語の組版結果をインターネット経由でどこにでも送信したり、CD-ROM などの電子媒体で流通させることができるようになった。

国境を越えて PDF を流通させたり、多言語組版をして、アラビア語やタイ語を含む文書の PDF を作成する場合は、アウトラインの埋め込みを許可されたフォントを指定し、PDF 生成時にフォントのアウトラインを埋め込む必要がある。

## XML

XML は、次のような点で多言語の文書のコンテンツを表現するのに最適である。

- XML は、UTF-8、UTF-16 という Unicode ベースの符号化方式を標準として採用している。従って文字コード変換をすることなく Unicode テキストを扱える。
- Microsoft Word のようなワープロを使えば、通常のビジネス文書レベルでの多言語文書作成と印刷、PDF 化は簡単にできる。しかし、膨大な量の製品マニュアルを作成したり、作成した文書をいろいろな形に加工したり、あるいはプロフェッショナルな品質をもつ印刷物を作成するためには、他のアプリケーションとデータを交換したり、他のア



アプリケーションと組み合わせて使う必要がある。このような時、情報を XML で表現しておくことで、アプリケーション間の連携が容易に実現できる。

- XML では文書ファイルを分割して作成したり、あるいは、画像ファイルは外部ファイルとして置くことができる。この仕組みを使って、文書作成時にいろいろな言語の文書で異なる部分は別ファイルとして作成、共通の部分は一つとし、画像はすべての言語で共通にするというような構成として、組版時にパーツ文書を組み立てて最終形にするような方法が取れる。

### 多言語 XML の編集

多言語の XML コンテンツを作成する方法は次の 3 通りがある。

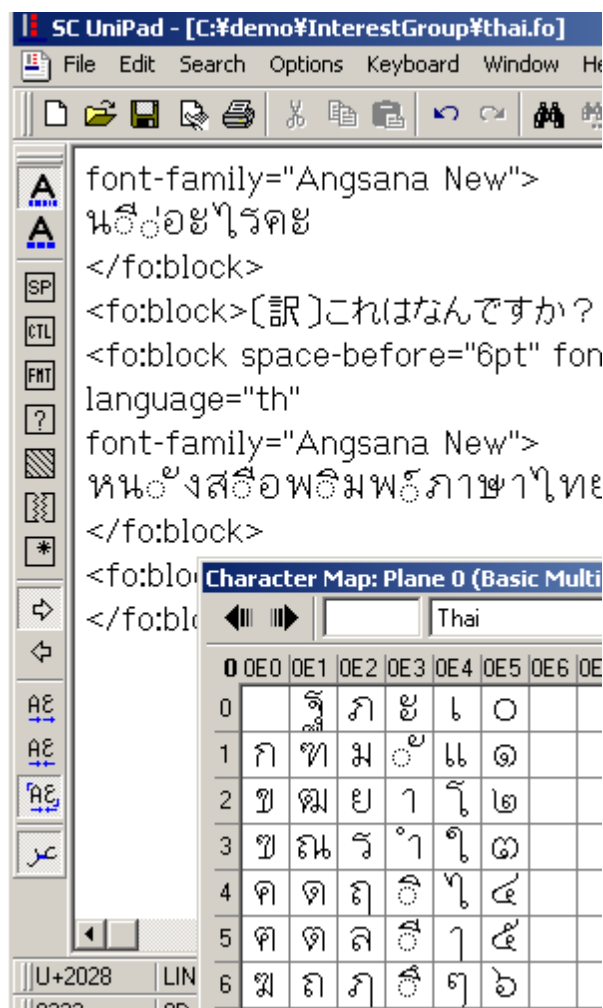
1. 多言語の扱えるテキスト・エディタを使う。
2. 多言語対応の XML エディタを使う。
3. 多言語を扱えるワープロ・ソフトを使う。

XML はテキストファイルなので、必要に応じてテキスト・エディタを使って記述できる。多言語を扱えるテキスト・エディタには、Windows の NotePad や UniPad などがある。特に、UniPad は Unicode4.0 の文字表を使ってタイ語などを Unicode のコードポイント単位で表示編集処理ができるので便利である（右の図を参照）。

XML エディタで多言語を本格的に扱えるものはまだないようだ。古くからある XML エディタの多くは、Unicode を扱えない。新しい XML エディタは Unicode を扱えるようになってきているが、既に述べたように単に Unicode を扱えるだけでは多言語対応の XML エディタということはいできない。

ひとつのバージョンで最も多くの言語を編集できるワープロソフトは Microsoft Word である。Microsoft Word2002 までの XML 機能は限定されたものであった。このため Word の文書を XML 化するには、RTF で保存し RTF から XML に変換するなどの方法がとられてきた。しかし、Microsoft Word2003 では XML Schema で定義したユーザスキーマによる XML 文書編集が可能になり、またユーザスキーマを使わない文書についても WordprocessorML という XML 形式で保存できるようになった。WordprocessorML 形式は、XML

形式なので任意の XML 形式に変換することが RTF よりも遥かに簡単である。アンテナハウスは、WordprocessorML を XSL-FO に変換するスタイルシートを世界で初めて商品化した。



UniPad の入力画面

また、2003 年の秋にリリースされた OpenOffice1.1/StarOffice7 でも多言語の編集機能が強化され、アラビア語やタイ語の編集ができるようになった。OpenOffice の保存形式は XML 形式なので多言語の XML コンテンツを作成する際の選択肢の一つと考えることができる。

### XSL による多言語組版

#### XSL

XSL は XML を紙のような用紙サイズ概念をもつメディアの上に組版するための仕様である。XSL の仕



様は、以下に述べるように国際化や多言語組版について配慮した設計になっている。

XSL では組版の対象となるオブジェクトの種類を定義している。オブジェクトの種類には、文字レベル、インライン・レベル、ブロック・レベル、リスト（箇条書き）、表、リンクなどのダイナミック・オブジェクト、ページの領域（本文、ヘッダ、フッタ、サイドバー、あるいは、脚注領域、前方フロート領域、サイド・フロート領域）などがある。

各オブジェクトに対してプロパティ（属性値）を指定することで、各オブジェクトのレイアウトや修飾の指定ができる。

XSL Formatter は、XSL の仕様に基づいてレイアウトを指定した XML ドキュメントを、指定した用紙の上に組版するための多言語組版エンジンである。多言語組版エンジンは、XSL の仕様では規定されていない機能についても、アンテナハウスの独自拡張などの方法で多言語機能を強化している。

## フォントの指定

段落やインライン・オブジェクトにフォント・ファミリーのプロパティを指定することで、組版フォントを設定する。データが正しい文字コードで作成されていても、指定を間違えると文字が表示されなかったり、場合によっては文字の字形が入れ替わってしまうことがあるため、多言語の組版にあたって、フォントファミリー名を適切に設定することは大変重要である。

フォントファミリー名は、Windows のメニューに表れる名前を指定する。例えば、FO に対して次のように指定する。

- font-family="MS 明朝"
- font-family="MS ゴシック"
- font-family="Arial"
- font-family="Times New Roman"

一般的なファミリー名（generic-font family）で指定することもできる。一般的なフォントファミリーには serif, sans-serif, cursive, fantasy, monospace の 5 種類がある。XSL Formatter は、フォントファミリー・プロパティの値を一般的なファミリー名で指定すると、組版を実行時に、稼動している Windows にインストールされて

いる実際のフォント名に対応させる。言語別の一般的なフォント・ファミリー→実際のフォント名の対応表を、XSL Formatter 「組版」メニューの「組版オプション」→「言語別設定・日本語処理」で設定できる。

フォントファミリー名を次のようにリストで指定すると、左から優先的に適用する。この機能を利用して、欧文と和文が混在する段落に対して、欧文フォントと和文フォントを一括指定することもできる。

### font-family の指定例

```
<fo:block
font-family="Arial, MS Gothic, sans-serif">
English is Arial. 日本語はゴシックになります。
</fo:block>
```

このように指定すると、XSL Formatter は Arial フォントにグリフがある文字は Arial を適用し、Arial フォントにグリフが無い時は MS ゴシックを適用する。組版結果は次のようになる。

English is Arial. 日本語はゴシックになります。

## 日本語と中国語の漢字混植

Unicode では、日本語、中国語（繁体字）、中国語（簡体字）、韓国語<sup>(2)</sup>について、Unification を適用し、字形の同じ文字に同一のコードを割り振られている。このため Unicode でこの 3 つの言語のグリフを使い分けするには、フォント・ファミリーの指定による方法しかない。しかし、Unification されている文字であっても、日本語、中国語では文字のグリフが異なることが良くある。また、フォントのデザインも違うため、日本語文書に中国語のフォント・ファミリーを指定すると読者には違和感がある。このため日本語、中国語（繁体字）、中国語（簡体字）を混在させるときは、特に注意して、フォント・ファミリーは generic-font family ではなく、各言語のフォント・ファミリー名を指定しなければならない。

### 日本語と中国語の font-family 指定例

```
<fo:block>
<fo:inline font-size="12pt" font-family="MS 明朝">
```

<sup>(2)</sup>現在、韓国語はハングル文字を使って表記することが多く、漢字の使用頻度は少ない。

```
日：浅与
</fo:inline>
、
<fo:inline font-size="12pt" font-
family="SimSun">
简体：浅与
</fo:inline>
、
<fo:inline font-size="12pt" font-
family="MingLiU">
繁体：浅与
</fo:inline>
</fo:block>
```

これは次のように組版される。

```
日：浅与、简体：浅与、繁体：浅与
```

### 多言語の混植

文章の段落に複数の言語が混在する、多言語混植の組版には、様々な難しい問題がある。

#### ベースライン調整

多言語を混植した場合、フォントのベースラインをどうやって揃えるかという点も大きな課題である。フォントには、ラテン文字のように文字の下方にベースラインがあるもの、インド系の文字のようにベースラインが上にある (hanging ベースライン) もの、漢字のように下辺がベースラインになるものがある。XSL では、それらの文字を横に配置したときベースラインを調整できる<sup>(3)</sup>。

#### 自動空白調整

日本語組版において日本語の中にラテン文字による欧文を混植する場合、和文と欧文の間には一定の空白を空けるのが一般的である。CSS3 にはこの動作が規定されているが XSL 仕様にはこの機能がない。アンテナハウスは XSL 拡張仕様で和欧文間の空白量を定義し、XSL Formatter は和欧文間の空白量の自動調整ができる。

#### 和欧文間の自動空白量

```
<fo:block font-size="12pt"
padding="4pt"
```

```
xmlns:fo="http://www.w3.org/1999/XSL/
Format"
xmlns:axf="http://www.antennahouse.com/
names/XSL/Extensions">
<fo:block axf:text-autospace="none">
漢字 English sentence かな 2004 二千四
</fo:block>
<fo:block axf:text-autospace="ideograph-
alpha">
漢字 English sentence かな 2004 二千四
</fo:block>
<fo:block axf:text-autospace="ideograph-
numeric, ideograph-alpha">
漢字 English sentence かな 2004 二千四
</fo:block>
<fo:block axf:text-autospace="ideograph-
numeric, ideograph-alpha" axf:text-
autospace-width="0.12em" >
漢字 English sentence かな 2004 二千四
</fo:block>
</fo:block>
```

これは次のように組版される。

```
漢字English sentenceかな2004二千四
漢字 English sentence かな2004二千四
漢字 English sentence かな 2004 二千四
漢字English sentenceかな2004二千四
```

### XSL の国際化機能

XSL では、行の進行方向、文字の進行方向等の既定値は欧文の横書きであるが、それ以外の進行方向も自在に指定できる。

#### Writing-mode

文書全体、あるいは特定の領域に writing-mode を指定することで文字を書き進める方向と行を積み上げる進行方向を指定できる。但し、writing-mode が指定できるのは、次の FO から生成されるエリアのみである。例えば、fo:block に対して writing-mode を指定して右から左に書き進めることはできないので、

<sup>(3)</sup>これについては、"Internationalized Text Formatting in CSS and XSL"(Steve Zilles)を参照。現時点で、XSL Formatter はベースライン調整機能の未実装である

その場合は、当該の fo:block を fo:block-container に入れる。

- fo:simple-page-master
- fo:region-body
- fo:region-before
- fo:region-after
- fo:region-start
- fo:region-end
- fo:table
- fo:block-container
- fo:inline-container

writing-mode="tb-rl"と指定すれば、日本語、中国語の縦書きを指定できる。また、writing-mode="rl-tb"を指定すればアラビア語やヘブライ語のように文字を右から左に書く言語の指定ができる。Writing-mode をページに指定すれば、例えば、多段組の段の進行方向を指定できるし、表に指定すれば表のセルの列を右から左に進むように指定できる。

#### UnicodeBIDI と fo:bidi-override

ラテン・アルファベットや日本語のような左から右へ記述する文字と、アラビア文字やヘブライ文字のように右から左へ記述する文字が混在すると、文字の進め方が曖昧になることがある。多言語混植の問題解決のために、Unicode では UnicodeBIDI という処理方法を定めている。XSL では fo:bidi-override を使って UnicodeBIDI を制御できる。UnicodeBIDI と fo:bidi-override については、後述の「アラビア語を使ってみる」の項で説明する。

#### 改行位置の決定

文字列を組版する上で最も重要な問題は改行位置の決定方法である。言語によって改行できる位置が異なる。大別すると単語で分かち書きをする言語と、単語で分かち書きをしない言語に分かれる。さらに単語で分かち書きをしない言語は、一般の文字の間のどこでも改行できる言語と、単語の区切りで改行する言語に分かれる。

##### 単語を分かち書きする言語

英語や西欧の言語、アラビア語、韓国語（ハングル）、現代インド文字

##### 単語を分かち書きしない言語

任意の音節文字間で改行できる言語

日本語、中国語（簡体字）、中国語（繁体字）

##### 単語の区切りで改行する言語

タイ語、クメール語、ラオ語

多言語組版エンジンは、上に述べた 3 種類の改行決定方法を実装し、言語によって使い分けることができなければならない。英語のような言語では、改行位置は原則として単語間の空白、または、センテンスの区切りの空白の位置となる。日本語や中国語の漢字やかなの文章は原則としてどの文字間でも改行できる。タイ語、クメール語やラオ語のような言語を組版する際、改行位置を決定するためにはある種の辞書を用意する必要がある。XSL Formatter は 3 通りの改行位置決定方法を実装している。タイ語についてのみ辞書を使って改行位置を決定できる。

なお、文章の中で改行位置を予め指定しておきたい時は、Unicode の U+200B（ゼロ幅スペース）を改行可能な位置に挿入しておけば、改行位置を決める際に、U+200B を改行候補位置として扱う。

#### ハイフネーション

単語を分かち書きする言語では改行位置が単語の区切りになるため、長い単語が行末にかかって次の行に送られると、その行は文字数が少なくなる。このようなことから単語の組み合わせによって行の長さが大幅に変動してしまう。そこで単語を行末で分割することで行の長さを平均化するハイフネーション機能が必要である。XSL ではハイフネーション機能の動作をオン/オフしたり、ハイフネーションの連続をコントロールするためのプロパティを定義しており、XSL Formatter ではこれらの機能をつかってハイフネーションの頻度を調整できる。

XSL Formatter は、TeX のハイフネーション・アルゴリズムを標準で実装しており、言語別の TeX のハイフネーション・パターン辞書（を XML 形式にしたもの）を使ってハイフネーションを行うことができる（ハイフネーションの設定例は、「主要言語別組版例」の英語の項を参照）。

さらに、Computer Hyphenation 社のハイフネーションプログラム "Hyphenologist" をオプションとして用意しており、Hyphenologist をインストールすると、40 カ国を超える言語のハイフネーション処理ができるようになる。

XSL の仕様では、fo:block、fo:character、fo:page-sequence に、country、language という属性を指定できる (xml:lang によって、country-language を一括指定も可) ので、ハイフネーション機能を、文書全体、ページ毎、段落単位で言語別に使い分けもできる。

### ジャスティフィケーション

XSL の仕様では text-align 属性でジャスティフィケーションを指定できる。ジャスティフィケーションを実現するには言語別処理が必要である。例えば、ジャスティフィケーションの際、英語ではワード・スペーシングが多少変化しても良いが、ハイフネーションを使って単語間の空白量ができるだけ変動しないようにするのが良い。

アラビア語では、ワード・スペーシングが変化しないように組むのが普通である。このためアラビア語の両端揃えでは、文字間にカシダというグリフを挿入して単語の長さを伸ばすことで両端揃えを実現している。

日本語や中国語は文字と文字の間の空き量を調整してジャスティフィケーションを行う。しかし、行の中に欧文が混在したときは欧文の部分はラテン文字の組版ルールに従う。

タイ語は改行位置が単語の区切りになるため行の長さが、変動しやすい。しかし、サンスクリットに由来する単語を除いてハイフネーションはしないのが普通である。このため、タイ語にジャスティフィケーションを指定すると非常に見づらい組版結果になりがちである。

XSL ではジャスティフィケーションを指定できるが、その実際の処理は組版エンジンが行うものであり、組版エンジンの優劣が表れやすい部分であろう。

### 記号類と英数字間の改行

Unicode には Line Breaking Properties (UAX#14) という仕様があり、文字に対して改行特性を定義している。UAX#14 は、U+200B (ゼロ幅スペース) や U+2060 (Word Joiner) などの必須の改行特性が規定されている。XSL Formatter はこれらの必須改行特性については UAX#14 に準拠している。

しかし、その他の文字については、UAX#14 の改行特性の規定は非常にルーズである。そこで、記号類と英数字間の改行については UAX#14 をカスタマイズし

てより改行を起こりにくくしている。記号類と英数字間の改行の起こりやすさについては、アンテナハウスの XSL 拡張 axf:line-break-at-punctuation-in-word によって指定できる。

### 日本語組版処理

日本語の組版には禁則処理、約物の詰めなどの独自のルールが沢山あり、日本独自の組版規則を実装しなければ、日本語組版エンジンとしては実用にならない。

現在、これらは XSL の仕様では規定されていないが、CSS3 では検討されている。アンテナハウスでは CSS3 を参考にしながら、XSL の仕様を独自に拡張して、XSL Formatter に実装している。

### タイ語を使ってみる

Windows2000 の標準フォントの中で、Tahoma、Microsoft Sans Serif の 2 つはタイ文字の範囲をサポートしている。Windows2000/XP で地域のオプション「タイ」を追加すると、次のタイ語のフォントが追加インストールされる。

- Angsana New
- AngsanaUPC
- Browallia New
- BrowalliaUPC
- Cordia New
- CordiaUPC
- DilleniaUPC
- EucrosiaUPC
- FreesiaUPC
- IriUPC
- JasmineUPC
- KodchiangUPC
- LilyUPC

タイ語を入力して組版してみる。ここでは、SC Unipad という Unicode 用のテキスト編集ソフトを使う。UniPad では、Unicode のコード表を見ながらタイ語のコードを入力できる。

タイ語には、Angsana New フォントで 16 ポイントを指定。

นี่อะไรคะ

〔訳〕 これはなんですか？

หนังสือพิมพ์ภาษาไทยครับ

〔訳〕 タイ語の新聞です。

タイ語は単語間で分かち書きをしない。しかし、改行位置は単語の区切りである。このため、改行位置を決定するには、辞書を引いて単語の区切りを調べる。XSL Formatter V2.5 では、Windows の Uniscribe という機能を使って、自動的に単語の区切りで改行できる機能が追加される。次の例では、学校という単語の区切り位置で改行が行われていることを示す。

単語 [学校]

โรงเรียน

โรงเรียนโรงเรียน

โรงเรียนโรงเรียนโรงเรียน

โรงเรียนโรงเรียนโรงเรียนโรงเรียน

โรงเรียนโรงเรียนโรงเรียนโรงเรียน

โรงเรียน

学校という単語の綴りの母音をひとつ間違えると行の区切り位置が変わることを示す。

โรงเรียน

โรงเรียนโรงเรียน

โรงเรียนโรงเรียนโรงเรียน

โรงเรียนโรงเรียนโรงเรียนโรงเรียน

โรงเรียนโรงเรียนโรงเรียนโรง

เรียนโรงเรียน

次は、日本語とタイ語の混じった文章の例である。

ศ、สの後のรはしばしば発音されません。

動詞の前にการ kaan やຄວາມ khwaam を付けると、動詞が名詞化されます。

## アラビア語を使ってみる

Windows2000 の標準フォントの中でアラビア文字の範囲をサポートしているのは次の5種類である。

- Arial
- Courier New
- Tahoma
- Microsoft Sans Serif
- Times New Roman

なお、Windows2000 の地域のオプションで追加される Andalus、Arabic Transparent、Simplified Arabic、Simplifie Arabic Fixed、Traditional Arabic は、フォントの埋め込みが禁止になっているので使えない。

最初に、アラビア語のみ（国連世界人権宣言の冒頭）を含む文書の例をあげる。アラビア文字は、Unicode のデータベースで文字自体が右から左に進むという特性が定義されているのでアラビア語を書いただけで、アラビア語の部分は右から左に書き進められる。

### アラビア語の例

```
<fo:block
  font-family="Tahoma"
  language="ar">
アラビア語（略）
</fo:block>
```

これは次のように組版される。この段落を含む文章の進行方向は、左から右へ書き進める設定になっているのでアラビア語の行が左寄せになってしまう。また、ピリオドが右端に配置されてしまう。

## الإعلان العالمي لحقوق

### الإنسان

الديباجة

لَمَّا كَانَ الاعتراف بالكرامة المتأصلة في جميع أعضاء الأسرة البشرية وبحقوقهم المتساوية الثابتة هو أساس الحرية والعدل والسلام في العالم.

ولما كان تناسي حقوق الإنسان وازدراؤها قد أفضيا إلى أعمال همجية أذت الضمير الإنساني. وكان غاية ما يرنو إليه عامة البشر انبثاق عالم يتمتع فيه الفرد بحرية القول والعقيدة ويتحرر من الغزق والفاقة.

XSL-FO では `writing-mode` で、領域の中で書き進める方向を変更できる。`writing-mode` は、リファレンス・エリア（参照エリア）を生成する領域に対してのみ設定できるので、アラビア語の段落を `block-container` に入れる。`block-container` に対して `writing-mode="rl-tb"` を指定するとこの領域全体が右から左へ書き進める設定になるので、段落も右から開始するようになる。ピリオドも左端に配置される。

### アラビア語の右から左への記述の指定例

```
<fo:block-container
  writing-mode="rl-tb"
  font-family="Tahoma"
  language="ar">
<fo:block>
  アラビア語（略）
</fo:block>
</fo:block-container>
```

これは次のように組版される。

## الإعلان العالمي لحقوق الإنسان

الديباجة  
لما كان الاعتراف بالكرامة المتأصلة في جميع أعضاء الأسرة البشرية وبحقوقهم المتساوية الثابتة هو أساس الحرية والعدل والسلام في العالم.

ولما كان تناسي حقوق الإنسان وازدراؤها قد أفضى إلى أعمال همجية أدت الضمير الإنساني. وكان غاية ما يرنو إليه عامة البشر انبثاق عالم يتمتع فيه الفرد بحرية القول والعقيدة ويتحرر من الفزع والفاقة.

次はアラビア語混じりの英語の例を紹介する。

باب **ab** means either *father* or *a father*, and باب **bab** either *door* or *a door*.

### 進行方向が曖昧になる時の指定方法

日本語や英語のような左から右に記述する文章の中にアラビア語やヘブライ語のような右から左に記述する文字が混在することを `BIDI`（双方向性）という。

進行方向の異なる文字が入れ子になっていると、文字の進行方向に関して曖昧さが生まれることがある。これを処理するため、Unicode は `BIDI` アルゴリズムを定義している。Unicode の `BIDI` は、文字の特性に基づく暗黙の部分と、埋め込み、オーバーライドを明示的に制御する部分がある。

XSL では、Unicode の方向整形コードまたは、`fo:bidirectional-override` を使って処理を指定する。XSL Formatter はこれを正しく処理できる。次に簡単な例を上げる。

FO の中でアラビア語を括弧で括った文字列の場合  
<fo:block>ضمض (ضمض) ENGLISH</fo:block>

括弧は、文字の方向性を持たない中立(ニュートラル)の文字である。一般的に、中立の文字は周辺の文字の方向性に影響を受け、`Left-to-Right` と `Left-to-Right` には含まれた文字は `Left-to-Right` になり、`Right-to-Left` と `Right-to-Left` には含まれた文字は `Right-to-Left` になる。この周辺の方向性が対立した場合は、上位の方向性にしたがうので、`fo:block` の `writing-mode` と同じになる。

したがって、上の `fo:block` は次のように表示される。

شخص (شخص) ENGLISH

これを回避する方法の 1 つは、Unicode の方向整形コード (`RLM`、`RLE`) を使うことである<sup>(4)</sup>。

#### RLM を使った例

```
<fo:block>ضمض (ضمض) &#x200F;ENGLISH</fo:block>
```

#### RLE を使った例

```
<fo:block>&#x202B;ضمض (ضمض) &#x202C;ENGLISH</fo:block>
```

この二つはいずれも次のように表示される。

شخص (شخص) ENGLISH

`fo:bidirectional-override` を使っても同じことを指定できる。

### おわりに

Unicode と XML、XSL は多言語組版の世界を大きく変えている。XSL Formatter は、世界のあらゆる言語を高品位に組版できるという目標に向けて改良を続けていきたいと考えている。専門家の皆様の教を請いたい。

<sup>(4)</sup>FO のデータの例では、Unicode の `LRO` (`U+202D`: 続く文字列は左から右方向として扱う。) を使ってアラビア語の文字を強制的に（入力順に）左から右に表示されるようにしている。組版後の表示順とは逆になっている。

## 主要言語別組版例

### 日本語

#### 海に沈む島

##### ツバルは今

今、南太平洋に浮かぶ小さな島ツバルが、危機にさらされている。地球の温暖化で、最初に海に沈む島と想像されている。1997年京都で環境に関する会議が開かれ2008年から2012年の間に先進国全体の温室効果ガスの排気量を、1990年の排気量と比較して5%以上減らすことを義務つけた。

##### 温暖化防止対策

チェック	事項	チェック	事項
	エアコンの使用を減らす		ごみを減らす
	テレビを付けっぱなしにしない		水を出しっぱなしにしない
	できるだけ車を使わず歩く		紙を再利用する

たと以比年をガ体間二〇議境七いとに暖るらがさ洋  
 を上較の、スのに〇〇がに年る想海化。さ、なに今  
 義減し排一の温先一人開関京。像にで地れ危島浮、  
 務らて気九排室進二年かす都一さ沈、球れ機ツか南  
 つす五量九気効国年かれるで九れむ最のてにバぶ太  
 けこ%と〇量果全のら二会環九て島初温いさル小平

### ヘブライ語

#### האי הטובע בים

##### מה קורה ב"טובל"

בימים אלה, האי הקטן "טובל" אשר בדרום הפסיפיק, עומד בפני סכנה. בעקבות התחממות כדור הארץ, נראה שטובל הוא האי הקרוב ביותר לטבוע בים. בשנת 1997 נערכה בקיוטו ועידה שעסקה בנושאים הקשורים באיכות הסביבה, ובה נקבע כי בין השנים: 2008-2012 יש להוריד את שיעור פליטת הפחמן הדו- חמצני במדינות המתקדמות בלפחות חמישה אחוזים (בהשוואה לשיעור פליטת הפחמן הדו- חמצני בשנת 1990).

כדי למנוע את התחממות כדור הארץ

פריט	בדיקה	פריט	בדיקה
לייצר פחות אשפה		להפחית את השימוש במזגנים	
לחסוך במים		לא להשאיר את הטלוויזיה דולקת כל הזמן	
למחזר נייר		להשתדל ללכת יותר, ופחות להשתמש במכונית	



## アラビア語

- アラビア語は右から左に書かれます。文字は単語中の、先頭、内部、終端で形が変わります。

### الغوص في البحر

ماذا يحصل في توفاليو الان؟

الان، تعتبر توفاليو من الجزر الصغيرة التي تتجه نحوها الانظار العالمية. من المعتقد بان توفاليو سوف تصبح البلد الاول الذي يغوص في البحر. في عام 1997 تم عقد مؤتمر في مدينة كيبوتو حول مشاكل البيئة. وفي هذا المؤتمر تم اقرار تقليل كمية ثاني اوكسيد الكربون (CO2) في الجو بنسبة اكثر من 5% خلال الفترة من عام 2008 الى 2012، مقارنة بعام 1990.

لمنع ارتفاع حرارة العالم

الفحص	الفقرة	الفحص	الفقرة
	التقليل من استخدام مكيف الهواء.		التقليل من القمامة.
	عدم ترك التلفزيون مفتوح.		الاقتصاد بالماء
	الاعتماد على السير بدلا من السيارة بقدر الامكان.		اعادة استخدام الورق

## タイ語

- タイ語は表音文字で、42の子音と32の母音と声調符号で表されます。

เกาะที่กำลังจะจม

เกาะตูลู...

เกาะเล็กๆที่อยู่ทางใต้ของทะเลแปซิฟิกกำลังอยู่ในภาวะอันตรายตามการคาดคะเนแล้ว เกาะตูลูจะเป็นประเทศแรกที่จมนหายไปใต้น้ำทะเลจากสภาวะโลกร้อน(Global Warming)จากการประชุมระดับโลกในด้านปัญหาสิ่งแวดล้อมที่เกี่ยวข้องเมื่อปีค.ศ.1997 ที่ประชุมได้มีมติให้ประเทศพัฒนาแล้วทั้งหมดลดปริมาณการระบายสารคาบอนไดออกไซด์ออกสู่บรรยากาศให้ได้มากกว่า 5% ในระหว่างปีค.ศ.2008 ถึง ค.ศ.2012 เมื่อเทียบกับปริมาณของสารดังกล่าวที่ระบายออกในปีค.ศ.1990

การหลีกเลี่ยงสภาวะโลกร้อน (Global Warming)

เครื่องหมาย	รายการ	เครื่องหมาย	รายการ
	ลดการใช้เครื่องปรับอากาศ		ลดปริมาณขยะ
	ไม่เปิดโทรทัศน์ทิ้งไว้โดยไม่จำเป็น		ไม่เปิดน้ำทิ้งไว้
	พยายามเดินแทนการใช้รถยนต์		นำกระดาษมารีไซเคิลใช้ใหม่

## 繁体中国語

### 沈下大海的島嶼

#### 現在的圖華路(Tuvalu)島

現在、浮在南太平洋上的小島圖華路濱臨于極大的危機。由于地球溫暖化的影響、它可能會成爲第一個沈下大海的島嶼。1997 年在日本京都召開的有關環境的會議上、就自 2008 年至 2012 年之間所有先進國家的溫室效應氣體的排氣量、做出了履行與 1990 年排氣量相比至少減少 5%義務的規定。

#### 溫暖化防止措施

檢查	事項	檢查	事項
	少用空調		減少垃圾
	不要將電視機開 不管		不要發生長流水現象
	儘量步行不用汽車		紙張再利用

定 5 比 0 履氣效進年年就境召年嶼下成響球危濱的在  
。%至年行量應國之至自的開在。大爲、溫機臨小南  
義務少排與、氣家間 2 2 會 的 日 1 海 一 可 化 由 極 圖 平 在  
的減量 9 出 的 溫 有 1 0 上 關 京 9 的 個 能 的 于 大 華 洋、  
規少相 9 了 排 室 先 2 8、環 都 7 島 沈 會 影 地 的 路 上 浮

## 簡体中国語

### 沉下大海的岛屿

#### 現在的圖華路(Tuvalu)島

現在、浮在南太平洋上的小島圖華路濱臨于極大的危機。由于地球溫暖化的影響、它可能會成爲第一個沉下大海的島嶼。1997 年在日本京都召開的有關環境的會議上、就自 2008 年至 2012 年之間所有先進國家的溫室效應氣體的排氣量、做出了履行與 1990 年排氣量相比至少減少 5%義務的規定。

#### 溫暖化防止措施

检查	事项	检查	事项
	少用空调		减少垃圾
	不要将电视机开着不管		不要发生长流水现象
	尽量步行不用汽车		纸张再利用

## 韓国語

### 바다 속으로 가라앉는 섬

#### 투발루는 지금

남태평양의 조그만 섬나라인 투발루는 지금 바다에 잠길 위기에 처해 있다. 지구 온난 현상으로 인해 최초로 바다 속으로 사라질 것으로 보인다. 1997년 교토에서 환경에 관한 회의가 열렸고, 이 회의에서 2008년에서 2012년 사이에 선진국 전체의 온실 효과를 일으키는 가스의 배기량을 1990년의 배기량에 비해 5% 이상 감소 시키는 것을 의무화 하였다.

#### 온난 현상 방지 대책

체크	사항	체크	사항
	에어콘 사용을 줄인다		쓰레기를 줄인다
	텔레비를 오래 켜두지 않는다		물을 절약한다
	가능한 한 자동차를 이용하지 않고 걷는다		종이를 재활용한다

## 英語 (The Chicago Manual of Style より抜粋)

### 13.2

<fo:block  
hyphenate="true"  
language="en">  
によりハイフネー  
ションを指定。

This chapter will describe some of the common problems that arise in setting technical material and will suggest ways in which these problems can be solved or circumvented. It is intended for authors unfamiliar with techniques of typesetting and for copyeditors not blessed with a mathematical background. For more on typesetting and printing in general see chapter 19.

### 13.3

The advent of sophisticated phototypesetting systems, including both photomechanical and CRT systems, has revolutionized the setting of mathematical copy in recent years. Many expressions and arrangements of expressions that formerly were impossible or very difficult to set are now relatively easy to achieve. Not every manuscript involving mathematical expressions is composed by such an advanced system, however, and authors and editors should have some idea what to expect of the particular typesetting system employed for the manuscript in hand.

### 13.4

Typesetting systems can be thought of as existing on four levels of sophistication in mathematical capabilities.

## 資料集

### **Extensible Stylesheet Language (XSL) Version 1.0 W3C Recommendation 15 October 2001**

<http://www.w3.org/TR/2001/REC-xsl-20011015/>

### **CSS3 Text Module W3C Candidate Recommendation 14 May 2003**

<http://www.w3.org/TR/2003/CR-css3-text-20030514/>

### **Antenna House による XSL 拡張仕様**

<http://www.antenna.co.jp/XML/axf-extension/axf-extension.htm>

### **Unicode**

<http://www.unicode.org/>

### **Internationalized Text Formatting in CSS and XSL**

<http://homepage.mac.com/thgewecke/.Public/SZillesPaper.pdf>

### **UniPad**

<http://www.unipad.org>

### **UnicodeFonts**

<http://www.alanwood.net/unicode/fonts.html>

### **Office 2003 XML Reference Schemas**

<http://www.microsoft.com/office/xml/default.msp>

### **FOP**

<http://xml.apache.org/fop/index.html>

### **TeX のハイフネーション辞書の入手先**

<http://www.ctan.org/tex-archive/language/hyphenation/?action=/tex-archive/language/>

### **World Script**

<http://www.omniglot.com>