

# **JATS XML for Japanese language**

**The 3<sup>rd</sup> Asian Science Editors'  
Conference and Workshop 2016**

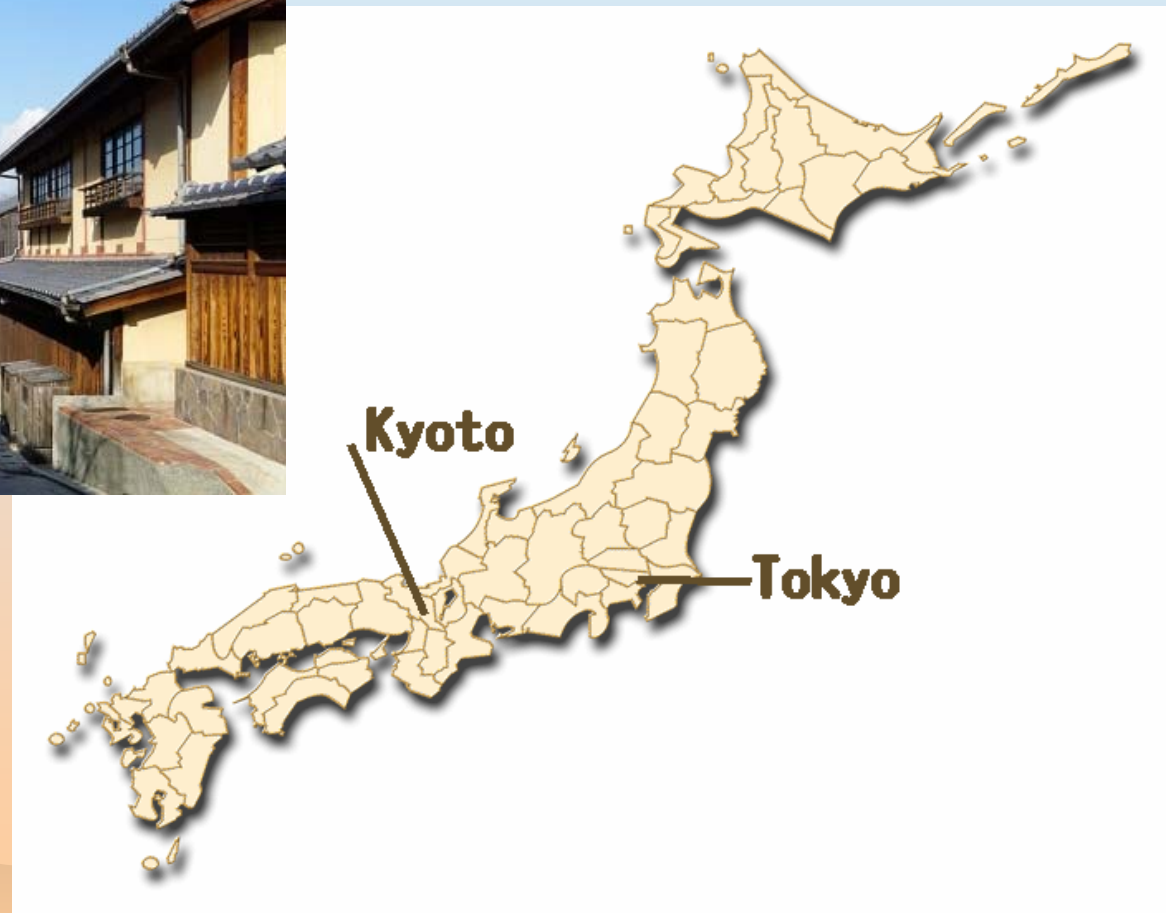
Hidehiko Nakanishi Ph.D.  
Nakanishi Printing Co., Ltd. Kyoto Japan



# I am from Kyoto, Japan



# Kyoto was a former capital



Where my company, Nakanishi Printing, is located.

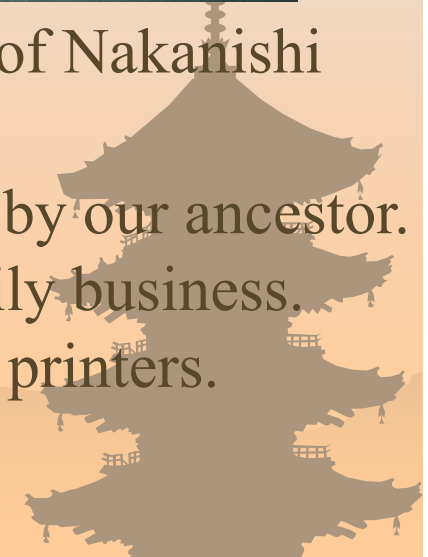
# Our Tradition

Former building of Nakanishi Printing in Taisho era (1912-1926)



Current building of Nakanishi printing

Founded in 1865 by our ancestor.  
150 year old family business.  
One of the oldest printers.



# Our history



A brazier made by Woodcut print plate in 19c



Type picker 1960's



Today



# Contents

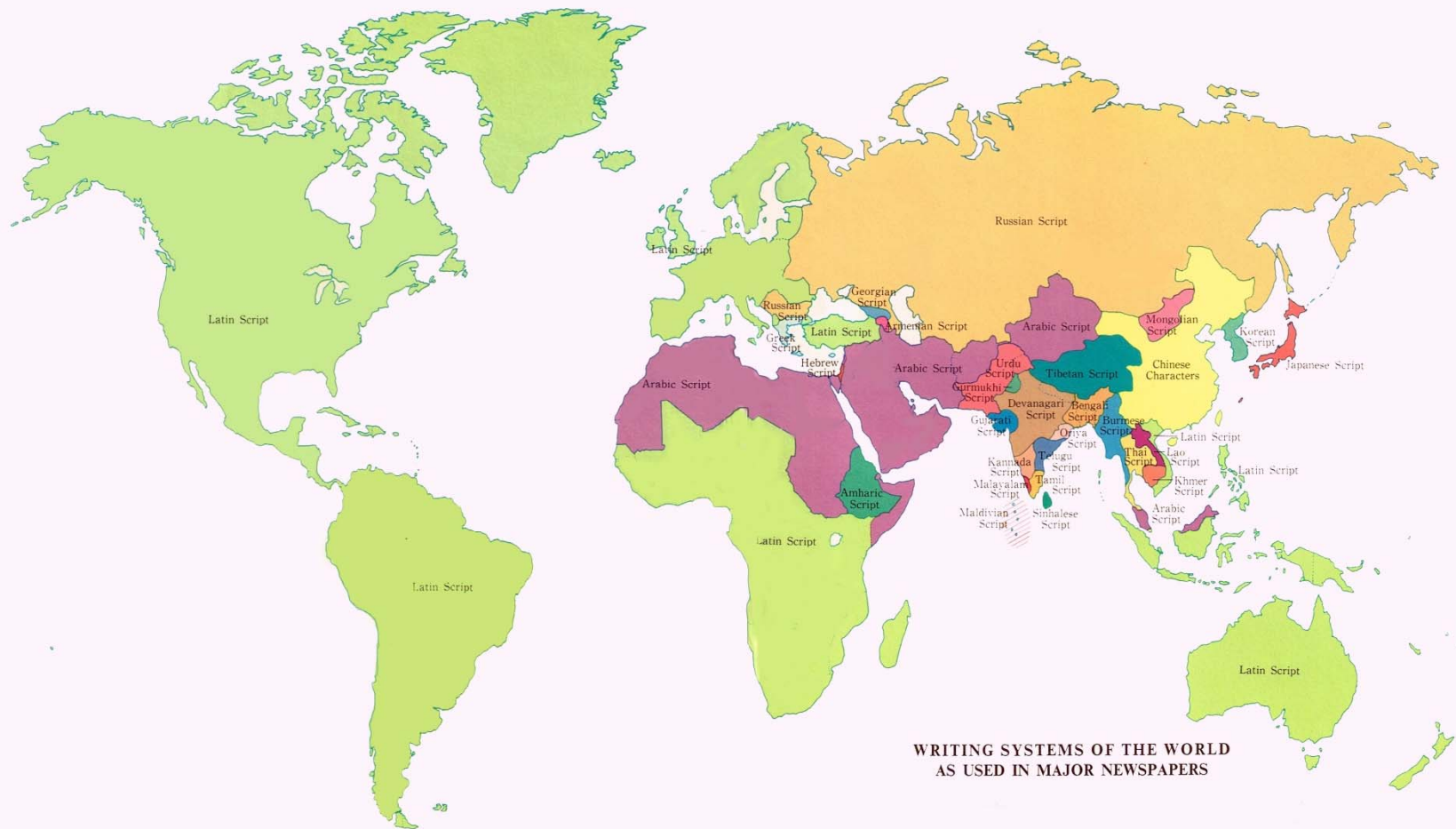
- 1. Introduction**
- 2. JATS adoption in Japanese XML expression**
- 3. Creating Japanese XML articles in JATS**
- 4. Creating PDF using AH Formatter**
- 5. Challenges of Applying JATS to Japanese language texts**
- 6. Future**
- 7. Conclusion**



# Introduction



# Many countries use Non-Latin script





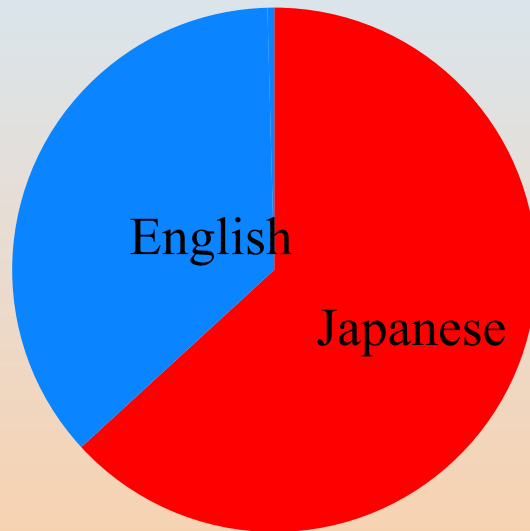
# Not all research articles are written in English.

- ❁ Many articles don't even use the Latin alphabets



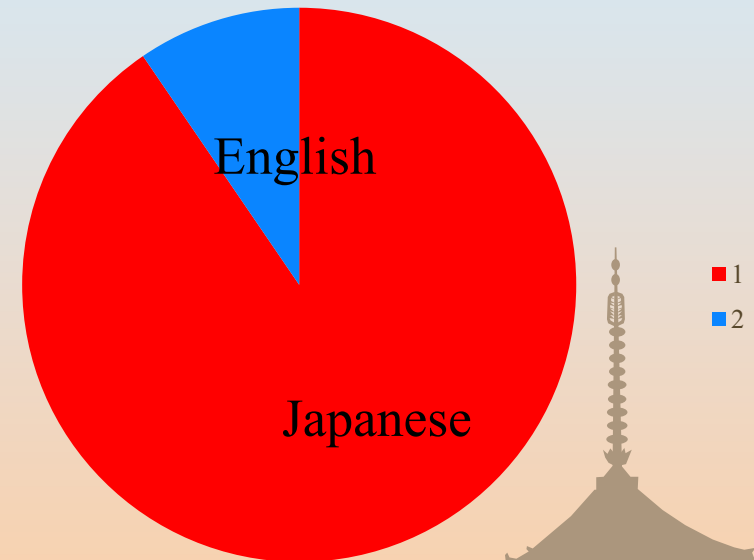
# What languages are used in articles written in Japan?

## STM



Articles published in J-Stage, E-journal platform operated by the Japan Science and Technology Agency (JST).

## All areas



University journal articles indexed in NDL-OPAC,



# The importance of articles in native tongue

- ❁ English is the world's common language.
- ❁ But there are many merits of describing and reading in one's native tongue.



# The change of world common language

## ❁ English

- As natural language, common language especially in STM domain.
- The evolution of translation software will change this position.

## ❁ XML

- Becoming the world common language.
- Prescribing common structure → Common understanding

# The importance of the XML in the native tongue

- ❁ Common structured articles in the native language → XML in the native language.
- ❁ So far English was thought to be essential for XML description, now it is not!
- ❁ XML articles in Japanese is ideal.
- ❁ We need suitable schema.



# **JATS adoption in Japanese XML expression**



# We wanted schema applicable to Japanese

- ❁ Even for Japanese-language articles, e-articles are essential.
- ❁ We were looking for schema for Japanese-language articles.
- ❁ Such schema had to accept English as well.



# JATS multi-language support

- ❁ In 2011 JATS 0.4 enabled to express Japanese-language articles in XML.
- ❁ J-STAGE supported JATS 0.4 immediately
- ❁ We started creating JATS XML for Japanese-language articles.





# This is a Japanese e-journal

**J-STAGE** | My J-STAGEとは? | ログイン | 新規登録 | ショッピングカート | ヘルプ | Japanese > English

**日本消化器外科学会雑誌**  
The Japanese Journal of Gastroenterological Surgery

ONLINE ISSN: 1348-9372 PRINT ISSN: 0386-9768 2014年12月30日現在 収録数: 9,537記事

記事 巻号頁 | DOI |

資料の中を検索します。 検索 詳細検索

閲覧する 論文投稿する このジャーナルについて 発行機関について

最新巻号 早期公開

J-STAGEトップ > 資料トップ > 全文HTML

日本消化器外科学会雑誌  
Vol. 45 (2012) No. 7 p. 697-707

DOI <http://dx.doi.org/10.5833/jjgs.45.697>

記事言語: Japanese

記事ツール

- お気に入り登録
- 被引用アラート
- 認証解除アラート
- 追加情報アラート
- URLコピー
- 著者にメール
- 書誌事項をダウンロード
- RIS
- BibTeX
- 問い合わせ

この記事を共有

- メールで知らせる

全文 図表(6) 引用文献(31) 被引用文献(1)

原著

**膵臓癌切除後腫瘍マーカー非正常化は重要な予後因子である**

元井 冬彦<sup>1,2</sup>, 島村 弘宗<sup>1,2</sup>, 石山 秀一<sup>1,4</sup>, 及川 昌也<sup>1,2</sup>, 桜井 直樹<sup>1,2</sup>, 阿南 陽二<sup>1,7</sup>, 中村 隆司<sup>1,2</sup>, 内山 哲之<sup>1,2</sup>, 片寄 友<sup>1,2,10</sup>, 海野 倫明<sup>1,2,10</sup>

1) 宮城肝胆膵癌化学療法研究会 2) 東北大学大学院消化器外科 3) 国立病院機構仙台医療センター外科 4) 仙台厚生病院消化器外科 5) 仙台オープン病院消化器外科 6) 山形県立中央病院外科 7) 坂総合病院外科 8) 東北厚生年金病院消化器外科 9) 石巻市立病院外科 10) 東北大学大学院統合がん治療外科

J-STAGE公開日 2012/07/18

本文PDF [1136K]

Index

- ▼ Abstract
- ▼ はじめに
- ▼ 対象と方法
- ▼ 結果
- ▼ 1. 腫瘍マーカー値測定

The Japanese Journal of Gastroenterological Surgery

# Same page expressed in English

ONLINE ISSN: 1348-9372 PRINT ISSN: 0386-9768 2014年12月30日現在 収録数: 9,537記事

閲覧する 論文投稿する このジャーナルについて 発行機関について

最新巻号 早期公開

J-STAGEトップ > 資料トップ > 書誌事項

The Japanese Journal of Gastroenterological Surgery  
Vol. 45 (2012) No. 7 p. 697-707

English 記事言語: English 次の記事

DOI http://dx.doi.org/10.5833/jjgs.45.697

ORIGINAL ARTICLE

ORIGINAL ARTICLE

**Sustained Elevation of Serum Tumor Markers after Resection is an Important Prognostic Factor for Pancreatic Cancer**

Fuyuhiko Motoi<sup>1,2)</sup>, Hiromune Shimamura<sup>1,2)</sup>, Shu-ichi Ishiyama<sup>1)4)</sup>, Masaya Oikawa<sup>1)4)</sup>, Naoki Sakurai<sup>1)4)</sup>, Youji Anami<sup>1)7)</sup>, Ryuji Nakamura<sup>1)8)</sup>, Tetsuyuki Uchiyama<sup>1)9)</sup>, Yu Katayose<sup>1)2)10)</sup>, Michiaki Unno<sup>1)2)10)</sup>

1) Miyagi-HBPCOG 2) Department of Surgery, Division of Gastroenterological Surgery, Tohoku University, Graduate School of Medicine 3) Department of Surgery, National Hospital Organization, Sendai Medical Center 4) Department of Gastroenterological Surgery, Sendai Kosei Hospital 5) Department of Gastroenterological Surgery, Sendai Open Hospital 6) Department of Surgery, Yamagata Prefectural Central Hospital 7) Department of Surgery, Saka General Hospital 8) Department of Gastroenterological Surgery, Tohoku Koseinenkin Hospital 9) Department of Surgery, Ishinomaki Municipal Hospital 10) Department of Surgery, Division of Surgery and Oncology, Tohoku University, Graduate School of Medicine

Released on J-STAGE 2012/07/18

キーワード: pancreatic cancer, surgery, tumor marker, postoperative normalization, prognostic factor

全文HTML 本文PDF [1136K]

抄録 引用文献(31) 被引用文献(1)

**Purpose:** Curative resection is the only chance for long-term survival for pancreatic carcinoma, however, the surgical results remain poor. Even after resection with a microscopically negative margin (R0), prognostic factors and predictors of recurrence are needed. Several reports revealed that sustained elevation of tumor markers (TM) after

記事ツール

- お気に入り登録
- 被引用アラート
- 認証解除アラート
- 追加情報アラート

URLコピー

著者にメール

書誌事項をダウンロード

RIS

BibTeX

[ヘルプ]

問い合わせ

この記事を共有

- メールで知らせる

### 3. 切除後腫瘍マーカーによる生存期間の比較

294例の生存曲線を、Fig. 1に示す。生存期間中央値は684日で、5年全生存率は24.4%であった。3群で生存曲線を比較すると、各群の1, 3, 5年全生存率は、非上昇群で84.8, 48.0, 43.6%, 正常化群で78.9, 42.0, 31.7%, 非正常化群で61.6, 18.4, 9.0%であり、生存期間中央値は、非上昇群で1,092日、正常化群で736日、非正常化群で505日であった。非正常化群で、有意に生存期間が短く、予後不良であった(Fig. 2,  $P < 0.0001$ , log-rank test)。

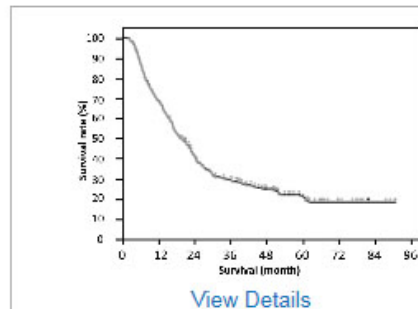


Fig. 1  
Kaplan-Meier survival curve of entire cohort.

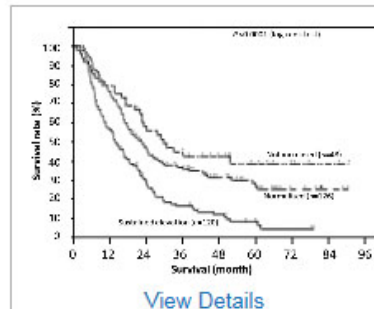


Fig. 2  
Kaplan-Meier survival curves by tumor marker status. There is significant difference between the survival curves by log-rank test ( $P < 0.0001$ ).

### 4. 多変量解析による予後因子の抽出

年齢・性・術前腫瘍マーカー・術式・進行度・組織型・癌遺残度・術後腫瘍マーカー・術後補助療法の各因子で、Cox比例ハザードモデルを用い解析した。単変量解析では、術前腫瘍マーカー(CA19-9上昇>非上昇例)、術式(臍全摘>臍頭切除)、進行度(stage 4>stage 3かつstage 1-2>stage 3)、癌遺残(R1-2>R0)、術後腫瘍マーカー(非正常化>非上昇例)が有意な予後因子として抽出された。それら単変量解析で有意であった因子を用いて多変量解析を行うと、進行度、術後腫瘍マーカーが有意な予後規定因子として選別された(Table 3)。



# Expressing Multiple Languages

- ❁ Alternate expressions for a single object are necessary
- ❁ Simple repetition of a tag can be confusing
  - Two name expressions of the same person?
  - Or two different persons?
- ❁ JATS introduced “alternatives” tags for such cases



# “Alternatives” Tags

- Two name expressions of a single person

<name-alternatives>

```
<name name-style="eastern" xml:lang="ja-Jpan">
```

```
<surname>中西</surname>
```

```
<given-name>秀彦</given-name>
```

```
</name>
```

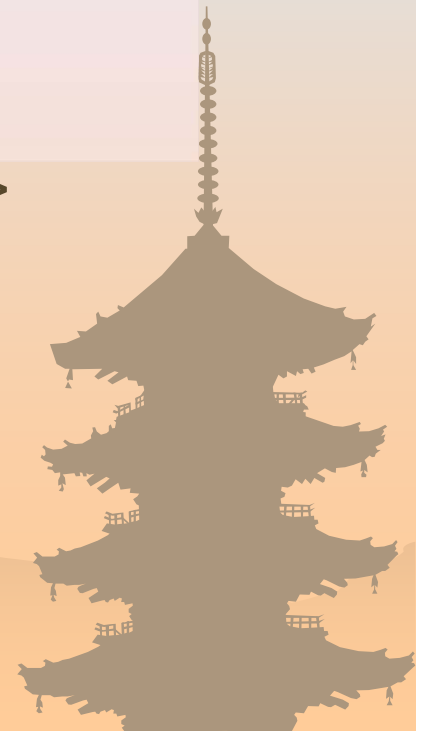
```
<name name-style="western" xml:lang="en">
```

```
<surname>Nakanishi</surname>
```

```
<given-name>Hidehiko</given-name>
```

```
</name>
```

```
</name-alternatives>
```



# “Alternatives” tags

```
- <contrib-group>
- <contrib contrib-type="author" xmlns:xlink="http://www.w3.org/1999/xlink" xlink:type="simple">
- <name-alternatives>
- <name name-style="eastern" xml:lang="ja-Jpan">
  <surname>元井</surname>
  <given-names>冬彦</given-names>
</name>
- <name name-style="western" xml:lang="en">
  <surname>Motoi</surname>
  <given-names>Fuyuhiko</given-names>
</name>
</name-alternatives>
<xref ref-type="aff" rid="aff1">1</xref>
<xref ref-type="aff" rid="aff2">2</xref>
</contrib>
- <contrib contrib-type="author" xmlns:xlink="http://www.w3.org/1999/xlink" xlink:type="simple">
- <name-alternatives>
- <name name-style="eastern" xml:lang="ja-Jpan">
  <surname>島村</surname>
  <given-names>弘宗</given-names>
</name>
- <name name-style="western" xml:lang="en">
  <surname>Shimamura</surname>
  <given-names>Hiromune</given-names>
</name>
</name-alternatives>
<xref ref-type="aff" rid="aff1">1</xref>
<xref ref-type="aff" rid="aff3">3</xref>
</contrib>
- <contrib contrib-type="author" xmlns:xlink="http://www.w3.org/1999/xlink" xlink:type="simple">
- <name-alternatives>
- <name name-style="eastern" xml:lang="ja-Jpan">
  <surname>石山</surname>
  <given-names>秀一</given-names>
</name>
- <name name-style="western" xml:lang="en">
  <surname>Ishiyama</surname>
  <given-names>Shu-ichi</given-names>
</name>
</name-alternatives>
<xref ref-type="aff" rid="aff1">1</xref>
<xref ref-type="aff" rid="aff4">4</xref>
</contrib>
```



# How multiple language can be expressed in JATS

element name	multi-language tag	Note
article title	<trans-title>	
article subtitle	<trans-subtitle>	
names	<name-alternatives>	
affiliations	<aff-alternatives>	
collaborators	<collab-alternatives>	
abstract	<abstract>	<abstract> is repeatable with different "xml:lang". <trans-abstract> is for articles later translated.
keyword group	<kwd-group>	<kwd-group> is repeatable with different "xml:lang".
generic	<alternatives>	any component which need multi-language data

# The proposal from Japan

- ❁ It is not enough to describe Japanese only by multiple language expansion of JATS
- ❁ XSPA propose some new elements for JATS standing committee. Some of them were adopted in JATS1.1
  - ❁ era elements
  - ❁ ruby elements

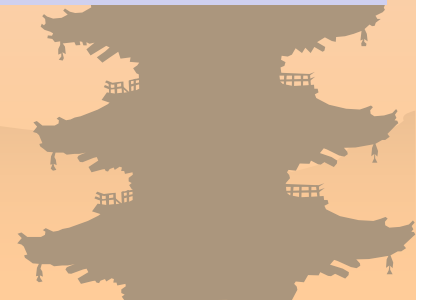




# era elements

## Example

```
...  
<date date-type="received" calendar="Japanese"  
iso-8601-date="1947-07-01">  
<day>1</day>  
<month>7</month>  
<year>22</year>  
<era>昭和</era>  
</date>  
...
```



# ruby element

## Example 1

畿内や西日本一帯では<sup>むぎ</sup>麦を<sup>うらさく</sup>裏作とする二毛作が普及して  
いった。

```
...  
<p>... 畿内や西日本一帯では<ruby><rb>麦</rb>  
<rt>むぎ</rt></ruby>を<ruby><rb>裏</rb><rt>うら</rt></ruby>  
<ruby><rb>作</rb><rt>さく</rt></ruby>とする二毛作が普及していった。  
...</p>  
...
```

# Creating Japanese XML articles in JATS



# Creating XML articles in JATS

- ❁ We don't have tools readily available for creating Japanese XML files.
  
- ❁ Our method
  1. Convert Microsoft Word to Microsoft Office Open XML
  2. Convert Microsoft Office Open XML to JATS XML
  3. Validate XML



# (1) Converting Microsoft Word to Microsoft Office Open XML

```
1 <?xml version="1.0" encoding="UTF-8" standalone="yes"?>
2 <?mso-application progid="Word.Document"?>
3 <w:wordDocument xmlns:w="http://schemas.microsoft.com/office/word/2003/wordml" xmlns:v="urn:schemas-microsoft-com:vml"
4 xmlns:w10="urn:schemas-microsoft-com:office:word" xmlns:s1="http://schemas.microsoft.com/schemaLibrary/2003/core"
5 xmlns:aml="http://schemas.microsoft.com/aml/2001/core" xmlns:wx="http://schemas.microsoft.com/office/word/2003/auxHint"
6 xmlns:o="urn:schemas-microsoft-com:office:office" xmlns:dt="uid:C2F41010-65B3-11d1-A29F-00AA00C14882"
7 xmlns:wsp="http://schemas.microsoft.com/office/word/2003/wordml/sp2" xmlns:st1="urn:schemas-microsoft-com:office:smrttags" w:macrosPresent="no"
8 w:embeddedObjPresent="no" w:ocxPresent="no" xml:space="preserve"><w:ignoreElements
9 w:val="http://schemas.microsoft.com/office/word/2003/wordml/sp2"/><o:SmartTagType o:namespaceuri="urn:schemas-microsoft-com:office:smrttags"
10 o:name="State"/><o:SmartTagType o:namespaceuri="urn:schemas-microsoft-com:office:smrttags" o:name="City"/><o:SmartTagType
11 o:namespaceuri="urn:schemas-microsoft-com:office:smrttags" o:name="place"/><o:DocumentProperties><o>Title>症例報
12 告</o>Title><o:Author>Norihiro</o:Author><o>LastAuthor>kamei</o>LastAuthor><o:Revision>2</o:Revision><o:TotalTime>1</o:TotalTime><o:Created>2012-06
13 -01T05:03:00Z</o:Created><o>LastSaved>2012-06-01T05:03:00Z</o>LastSaved><o:Pages>1</o:Pages><o:Words>2497</o:Words><o:Characters>14236</o:Character
14 s><o:Company>Toshiba</o:Company><o:Lines>118</o:Lines><o:Paragraphs>33</o:Paragraphs><o:CharactersWithSpaces>16700</o:CharactersWithSpaces><o:Versi
15 on>11.0000</o:Version></o:DocumentProperties><w:fonts><w:defaultFonts w:ascii="Century" w:fareast="MS 明朝" w:h-ansi="Century" w:cs="Times New
16 Roman"/><w:font w:name="Wingdings"><w:panose-1 w:val="050000000000000000"/><w:charset w:val="02"/><w:family w:val="Auto"/><w:pitch
17 w:val="variable"/><w:sig w:usb-0="00000000" w:usb-1="10000000" w:usb-2="00000000" w:usb-3="00000000" w:csb-0="80000000"
18 w:csb-1="00000000"/></w:font><w:font w:name="MS 明朝"><w:altName w:val="MS Mincho"/><w:panose-1 w:val="02020609040205080304"/><w:charset
19 w:val="80"/><w:family w:val="Roman"/><w:pitch w:val="fixed"/><w:sig w:usb-0="A00002BF" w:usb-1="68C7FCFB" w:usb-2="00000010" w:usb-3="00000000"
20 w:csb-0="0002009F" w:csb-1="00000000"/></w:font><w:font w:name="SimSun"><w:altName w:val="宋体"/><w:panose-1
21 w:val="02010600030101010101"/><w:charset w:val="86"/><w:family w:val="Auto"/><w:pitch w:val="variable"/><w:sig w:usb-0="00000003"
22 w:usb-1="080E0000" w:usb-2="00000010" w:usb-3="00000000" w:csb-0="00040001" w:csb-1="00000000"/></w:font><w:font w:name="Century"><w:panose-1
23 w:val="02040604050505020304"/><w:charset w:val="00"/><w:family w:val="Roman"/><w:pitch w:val="variable"/><w:sig w:usb-0="00000287"
24 w:usb-1="00000000" w:usb-2="00000000" w:usb-3="00000000" w:csb-0="0000009F" w:csb-1="00000000"/></w:font><w:font w:name="@MS 明朝"><w:panose-1
25 w:val="02020609040205080304"/><w:charset w:val="80"/><w:family w:val="Roman"/><w:pitch w:val="fixed"/><w:sig w:usb-0="A00002BF" w:usb-1="68C7FCFB"
26 w:usb-2="00000010" w:usb-3="00000000" w:csb-0="0002009F" w:csb-1="00000000"/></w:font><w:font w:name="@SimSun"><w:panose-1
```

MS Open XML tags

## (2) Converting Microsoft Office Open XML to JATS XML

- ❁ Through XSLT, removing unnecessary tags.
- ❁ Perl program processing.
- ❁ We faced the difficulty of Asian languages
  - A word connects to the next word without space.
  - Computer cannot distinguish word separation.
  - Even in given name and surname separation.

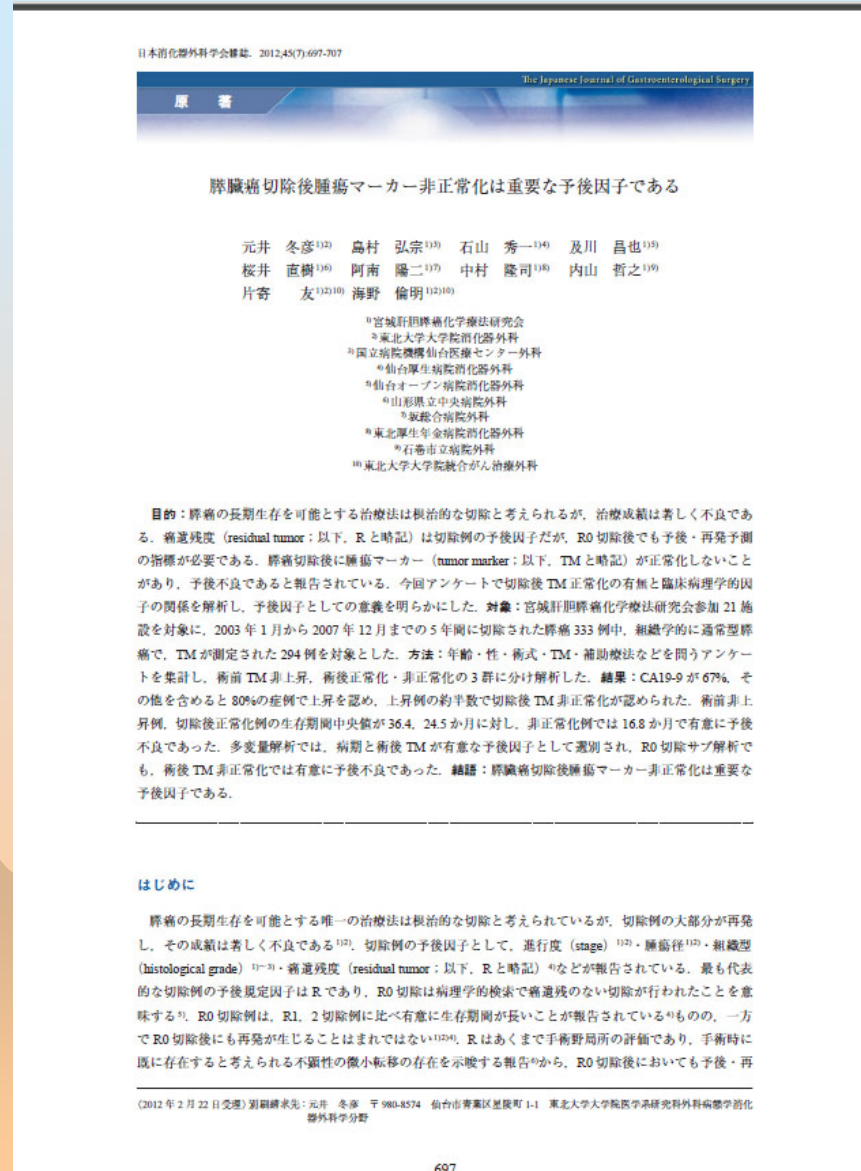


## (3) Validating XML

- ❁ Use the Oxygen XML editor
- ❁ Final JATS XML is obtained to be uploaded to J-STAGE



# PDF is still necessary



- ❁ For paper publishing.
- ❁ For readability.





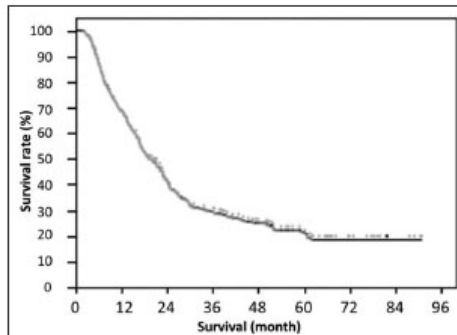


Fig. 1 Kaplan-Meier survival curve of entire cohort.

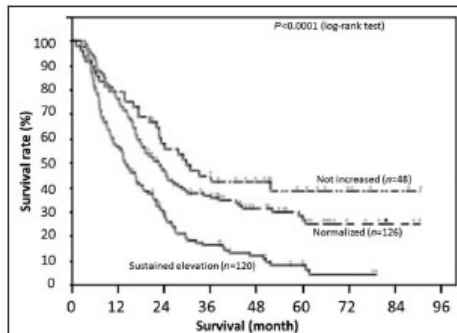


Fig. 2 Kaplan-Meier survival curves by tumor marker status. There is significant difference between the survival curves by log-rank test ( $P<0.0001$ ).

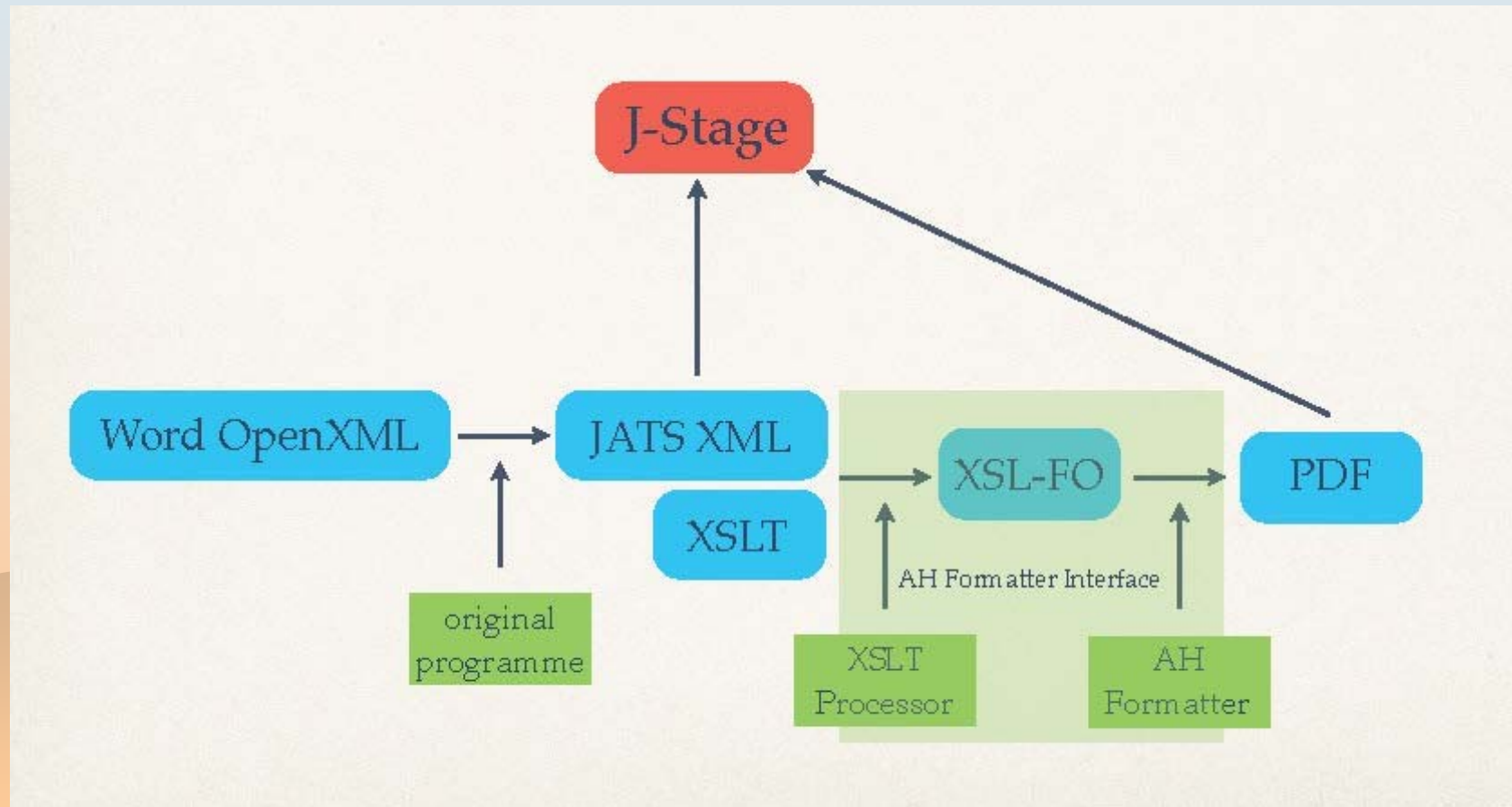
42.0, 31.7%, 非正常化群で 61.6, 18.4, 9.0%であり、生存期間中央値は、非上昇群で 1,092 日、正常化群で 736 日、非正常化群で 505 日であった。非正常化群で、有意に生存期間が短く、予後不良であった (Fig. 2,  $P < 0.0001$ , log-rank test)。

#### 4. 多変量解析による予後因子の抽出

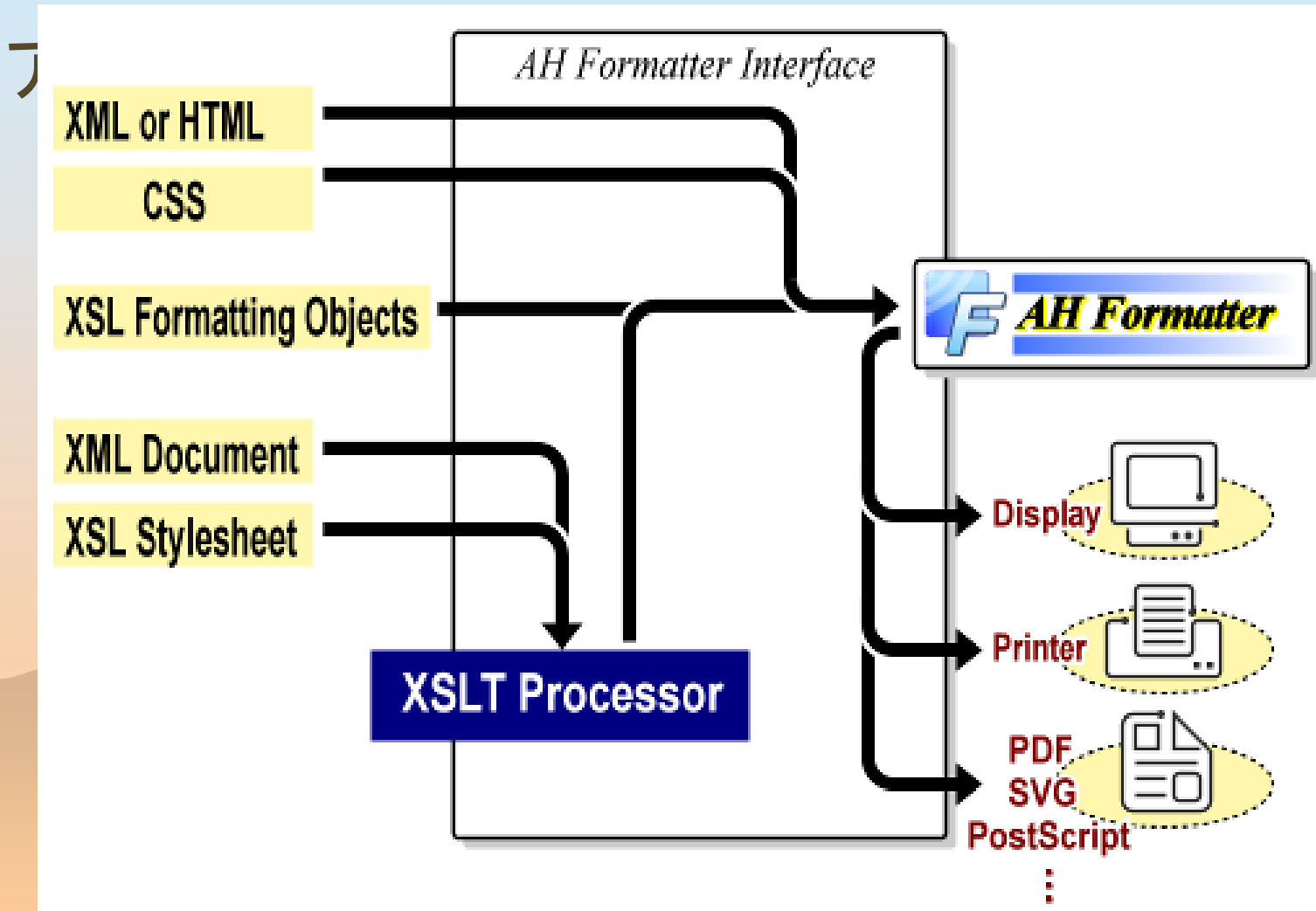
年齢・性・術前腫瘍マーカー・術式・進行度・組織型・癌遺残度・術後腫瘍マーカー・術後補助療法の各因子で、Cox 比例ハザードモデルを用い解析した。単変量解析では、術前腫瘍マーカー (CA19-9 上昇 > 非上昇例)、術式 (胃全摘 > 胃頭切除)、進行度 (stage 4 > stage 3 かつ stage 1-2 > stage 3)、癌遺残 (R1-2 > R0)、術後腫瘍マーカー (非正常化 > 非上昇例) が有意な予後因子として抽出された。それら単変量解析で有意であった因子を用いて多変量解析を行うと、進行度、術後腫瘍マーカーが有意な予後規定因子として選別された (Table 3)。



# Creating PDF using AH Formatter

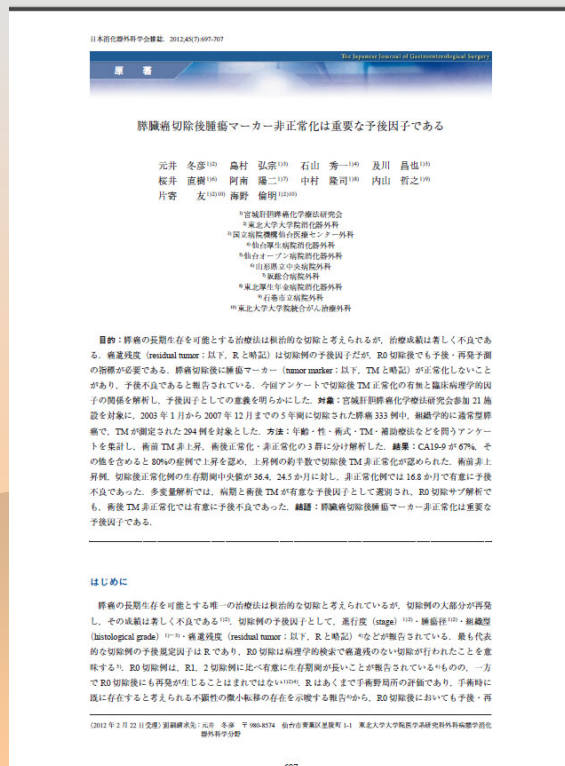


# Antenna House AH Formatter



# XSLT

❁ The XSLT converts a JATS file into XSL-FO which expresses page model format for PDF.

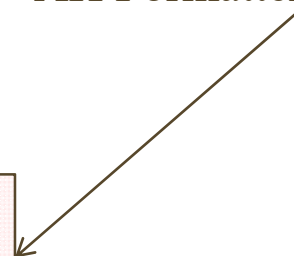


```

<?xml version="1.0" encoding="UTF-8"?>
<!-- edited with XMLSpy v2012 sp1 (http://www.altova.com) by Kenichi Kusunoki (Nakanishi Printing Co., Ltd.) -->
<xsl:stylesheet version="2.0" xmlns:xsl="http://www.w3.org/1999/XSL/Transform" xmlns:fo="http://www.w3.org/1999/XSL/Format" xmlns:xlink="http://www.w3.org/1999/xlink"
xmlns:mml="http://www.w3.org/1998/Math/MathML" xmlns:axf="http://www.antennahouse.com/names/XSL/Extensions" exclude-result-prefixes="xlink mml"
xmlns:kk="http://www.nacos.com/2012/kusunoki">
  <xsl:import href="kk_function.xsl"/>
  <!-- 『日本消化器外科学会雑誌』用スタイルシート -->
  <!-- by Antenna House, Inc., 2012-03-30 -->
  <!-- J-Stage3用スタイルシートをインポート -->
  <xsl:import href="jstage3-xslfo_new.xsl"/>
  <!-- この雑誌のためだけの設定をここに。 -->
  <!-- 本文の行送り [jpub3-xslfo] -->
  <xsl:variable name="text-leading">16.25pt</xsl:variable>
  <!-- ページの大きさと余白 -->
  <xsl:variable name="page-width">210mm</xsl:variable>
  <xsl:variable name="page-height">297mm</xsl:variable>
  <xsl:variable name="page-margin-inside">30.387mm</xsl:variable>
  <xsl:variable name="page-margin-outside">30.387mm</xsl:variable>
  <xsl:variable name="page-margin-top">9mm</xsl:variable>
  <xsl:variable name="page-margin-bottom">15mm</xsl:variable>
  <xsl:variable name="page-body-margin-top">17mm</xsl:variable>
  <xsl:variable name="page-body-margin-bottom">12mm</xsl:variable>
  <xsl:variable name="page-column-count">1</xsl:variable>
  <xsl:variable name="page-column-gap">0mm</xsl:variable>
  <xsl:variable name="first-page-header-extent">23.573mm</xsl:variable>
  <xsl:variable name="trans-abstract-header-extent">16mm</xsl:variable>
  <xsl:variable name="page-header-extent">10mm</xsl:variable>
  <xsl:variable name="page-footer-extent">4mm</xsl:variable>
  <!-- ページヘッダー [jpub3-xslfo] -->
  <xsl:attribute-set name="page-header">
    <xsl:attribute name="font-family"><xsl:value-of select="$textfont"/></xsl:attribute>
    <xsl:attribute name="font-size">7pt</xsl:attribute>
    <xsl:attribute name="line-height">7pt</xsl:attribute>
  </xsl:attribute-set>
  <!-- ノンブル領域 [jpub3-xslfo] -->
  <xsl:attribute-set name="page-header-vol-title-cell">
    <xsl:attribute name="width">auto</xsl:attribute>
    <xsl:attribute name="font-family">serif</xsl:attribute>
    <xsl:attribute name="font-weight">normal</xsl:attribute>
    <xsl:attribute name="font-size">7pt</xsl:attribute>
    <xsl:attribute name="start-indent">0pt</xsl:attribute>
    <xsl:attribute name="end-indent">0pt</xsl:attribute>
  </xsl:attribute-set>

```

For Japanese rendering  
AH Formatter extension



```

<!-- ページヘッダー [jpub3-xslfo] -->
<xsl:attribute-set name="page-header">
  <xsl:attribute name="font-family"><xsl:value-of select="$textfont"/></xsl:attribute>
  <xsl:attribute name="font-size">7pt</xsl:attribute>
  <xsl:attribute name="line-height">7pt</xsl:attribute>
</xsl:attribute-set>

```

# Challenges of Applying JATS to Japanese language texts

- ❁ But in Japan, exquisite type settings are requested.
- ❁ Automatic type setting by AH formatter may not be sufficient.



# Avoiding Word Breakup

- ❁ Some words, such as personal names shall not be broken-up between lines

(unacceptable)

この論文の主執筆者は中  
西秀彦でした。

(acceptable)

この論文の主執筆者は中西  
秀彦でした。

- ❁ We use “Zero Width Joiner” code (&#x200D;)

e.g. 中&#x200D;西

- ❁ 「中」 and 「西」 are torn up by “&#x200D”

→「中西」cannot be searched

# Avoiding Line-Top Punctuations

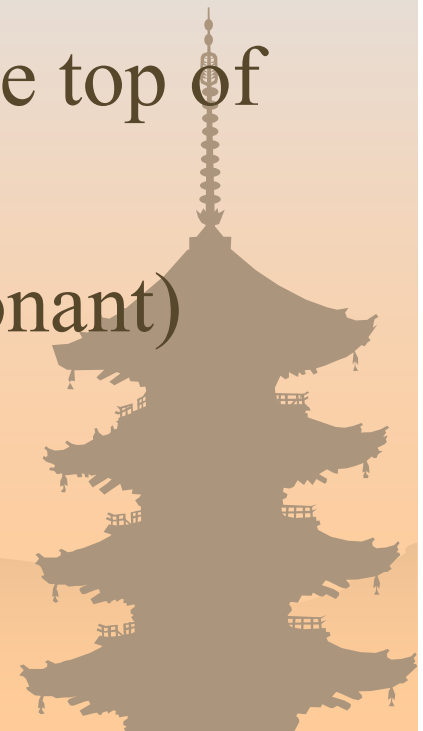
(unacceptable)

行頭・行末にあると体裁が悪い文字を調整することを「禁則処理」といいます。

(acceptable)

行頭・行末にあると体裁が悪い文字を調整することを「禁則処理」といいます。

- ❁ Punctuation marks shall not come at the top of a line ⇒ Also in English
- ❁ 「っ」or「ツ」 (to mark a geminate consonant) does not come in a head of a line ⇒ Japanese rule
- ❁ AH Formatter can handle these rules





# Positioning Figures/Tables

- ❁ Figures and tables should be positioned in the SAME page that the corresponding texts appear.
- ❁ This requires customized XSLT, sometimes for each figures and tables.
- ❁ This increases costs.



# Positioning Figures/Tables

```
<xsl:attribute-set name="float">
  <xsl:attribute name="axf:float">
    <xsl:choose>
      <xsl:when test="@id='F3'">page top</xsl:when>
      <xsl:when test="@id='T1'">page bottom</xsl:when>
      <xsl:otherwise>column top</xsl:otherwise>
    </xsl:choose>
  </xsl:attribute>
</xsl:attribute-set>

<xsl:attribute-set name="td">
  <xsl:attribute name="background-color">
    <xsl:choose>
      <xsl:when test="matches(.,'^¥+$')">#b3b3b3</xsl:when>
      <xsl:otherwise/>
    </xsl:choose>
  </xsl:attribute>
</xsl:attribute-set>
```

Every articles need these XSLTs



# Using Formatter for STM articles

- ❁ There are no major problems
- ❁ The basic style of writing STM papers do not differ greatly between western countries and Japan.
- ❁ Word separators should be inserted in XML in advance
- ❁ Possible if disregarding profitability



# Current Japanese XML journal



# List of Japanese XML Journals

1	日本コンピュータ化学会	Journal of Computer Chemistry, Japan	混
2	日本育種学会	育種学研究	和
3	日本熱帯医学会	Tropical Medicine and Health	混
4	社団法人 日本産業衛生学会	産業衛生学雑誌	和
5	特定非営利活動法人 日本歯周病学会	日本歯周病学会会誌	和
6	日本緩和医療学会	Palliative Care Research	混
7	日本応用糖質科学会	Journal of Applied Glycoscience	混
8	一般社団法人 日本鉄鋼協会	鉄と鋼	和
9	独立行政法人 海洋研究開発機構	JAMSTEC Report of Research and Development (JAMSTEC-R)	混
10	日本官能評価学会	日本官能評価学会誌	混
11	一般社団法人 日本消化器外科学会	日本消化器外科学会雑誌	混
12	公益社団法人 日本看護科学学会	日本看護科学会誌	和
13	日本行動医学会	行動医学研究	混
14	日本社会心理学会	社会心理学研究	混
15	日本公衆衛生看護学会	日本公衆衛生看護学会誌	和
16	科学技術振興機構	情報管理	和

混:English Japanese mixed up

和:Pure Japanese

April 2016 JST provided

# Few Japanese XML Journals

- ❁ The total number of pure Japanese journals in J-STAGE is 568
- ❁ The number of pure Japanese XML Journal is only 7
  - After 4 years have passed



# The difficulties of creating Japanese XML

- ❁ Labor and time are necessary
- ❁ Which require additional costs
- ❁ It is hard for publishers to find more merits than PDF journal



# The most difficult problem

## No separating spaces

- ❁ Typical in East Asian languages
- ❁ No separating spaces between words





# Inserting word separators.

- ❁ we insert separators manually.
  - surname, “中”, given name, “吉昭”, are attached as “中吉昭” in an article
  - It is separated as “中@吉昭”
- ❁ Possible alternatives are “中@吉昭”, and “中吉@昭”, but only human can eliminate them
- ❁ There is no algorithm to determine it correctly.



# Tools necessary for us

- ❁ XML tool of the East Asia cultural sphere
- ❁ The inputting tool is necessary at first
- ❁ Not MS WORD
- ❁ We need input tools for Japanese XML
- ❁ A common tools for the East Asia cultural sphere Japan, China and Korea is needed!



# Future

- ❁ What is to be done next
  - Vertical writing
  - Emphasis or “Kenten”
  - Warichu

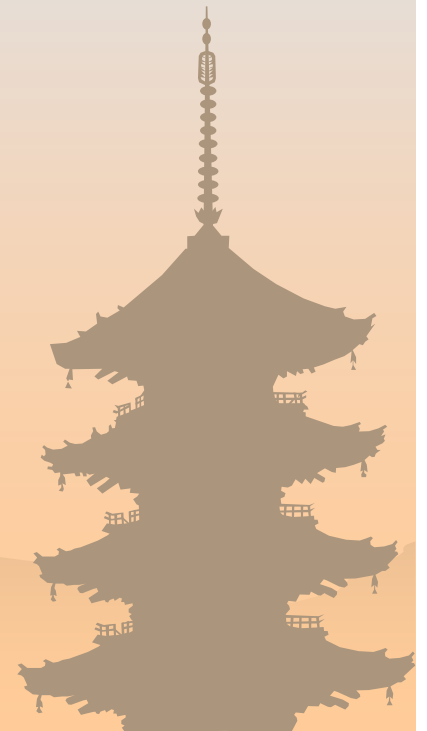


# Vertical writing

(あいうえお)

(かきくけこ)

- ❁ Traditionally, Japanese (and Chinese and Korean) writes from top to bottom



# Vertical Writing

1 2 3 4 5

1  
2  
3  
4  
5

1  
2  
3  
4  
5

一  
二  
三  
四  
五

- ❁ Vertical Writing causes some interesting problems, orientation of Arabic numerals and Latin alphabets

New element for direction is necessary.  
as `<writing-direction="vertical">`



# Emphasis

あ い う え お

か  
き  
く  
け  
こ

- ❁ Emphasis or “Kenten”
- ❁ It is like bold faces and italics in English
- ❁ We use <styled-content> for these expressions
- ❁ We need a generic tag, like as <emphasis>



# Warichu

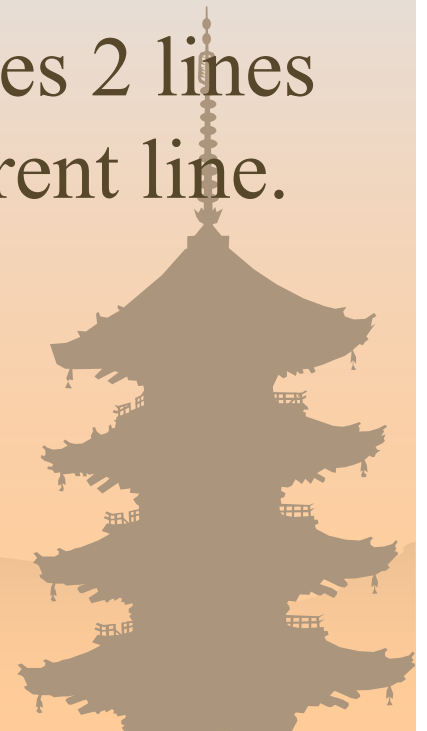
かきくけこ (これは見)  
本です。)

か  
き  
く  
け  
こ

(こ  
れ  
は  
見  
本  
で  
す。)

- ❁ Vertical writing texts sometimes contain notes called “Warichu”.

- ❁ Warichu uses 2 lines within a parent line.



# Warichu



Historical document example

在地立券一通

天養元年大和國御檢注時、依被免除寺領坪と檢注狀、在領主・作人連判、

負名作人注文一通

同年寺領坪と目錄并作人連判、在領主・作人連判、

負名と寄一通

同年本免之外、加納・余田・家地・作畠才可并所當之由、在領主・作人連判、

國郡立券狀一通

平治元年國使・寺使相共立券并檢注狀

坪と作人注文一通

同年負名作人檢注之狀、在作人連判



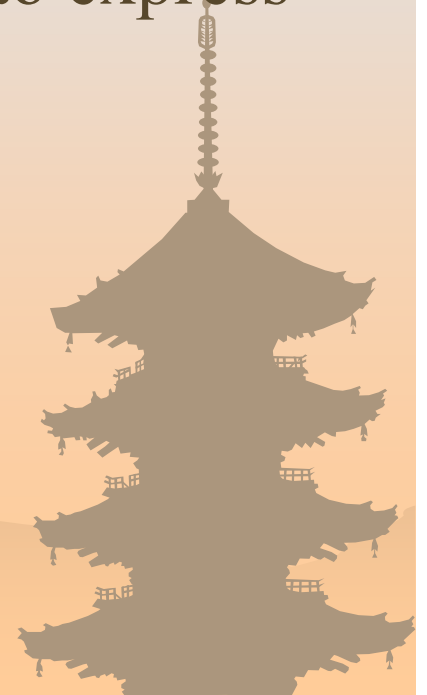
# Suggestion

- ❁ Additional tags for
  - Vertical writing
  - Emphasis or “Kenten”
  - Warichu



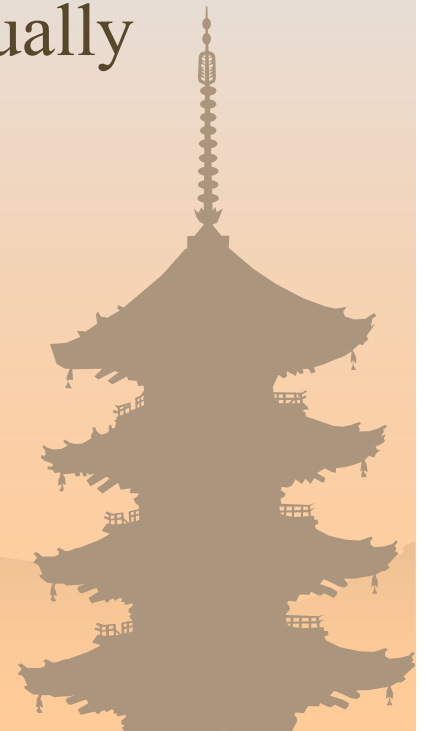
# Conclusion

- ❁ JATS opened a new horizon in processing Japanese-language articles
  - No major difficulties
  - UTF-8, encoding for XML, also enables to express most Japanese characters correctly



# Conclusion

- ❁ Still there are remaining issues in processing non-Latin languages such as Japanese.
- ❁ Challenges
  - Word separators have to be inserted manually
  - Line break issues
  - Positioning figures and tables correctly
- ❁ Some requests to JATS
  - for applying Asian languages





Only English is not globalization

이것을 영어로 말하는 것을 유감스럽게 생각합니다

Thank you