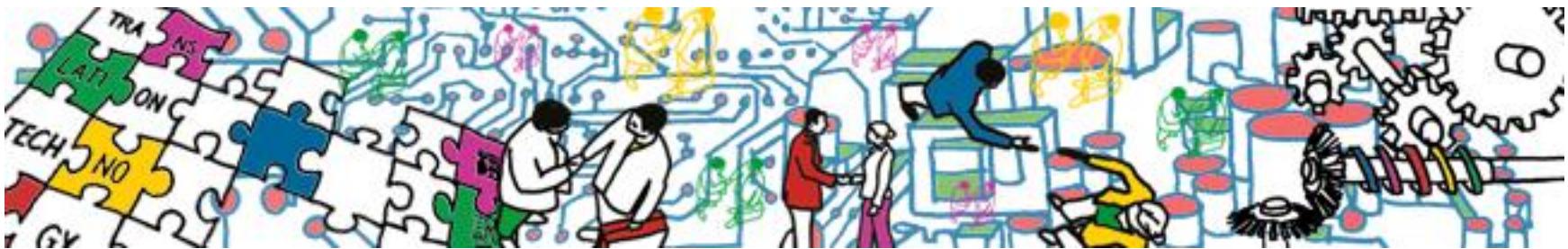

Bringing Procedural Knowledge to XLIFF

Prof. Dr. Klemens Waldhör
TAUS Labs & FOM University of Applied
Science

FEISGILTT
16 October 2012
Seattle, USA



Presentation Overview



- ✓ Why?
- ✓ How?
- ✓ Examples
- ✓ Summary and advantages



Beyond XLIFF 2.0

- ✓ Syntax
- ✓ Semantics
- ✓ Pragmatics



XLIFF Limitations

- XLIFF represents and communicates document related information
 - Text
 - Attributes, properties
- Semantic aspects are only partially supported
 - Constraining values with XSD
 - XLIFF 1.x implementer specific interpretation
 - XLIFF 2.0 with modules beyond core
- Complex tasks hard or impossible to realize
 - Work flow aspects
 - Limport
 - Requires representing procedural knowledge



The traditional approach

Segmentliste und Editor

Ansichtsmodus Bearbeitungsmodus Tags verbergen Tag-Kurzansicht Tag-Vollansicht Sortierung / Filterung zurücksetzen

Ausgangstext Zieltext Zieltext überarbeitet Status QM Matchrate Autostatus letzter Bearbeiter

Einstellungen

Segment-Metadaten

QM

Freigabe: Leichte Mängel: Muss überarbeitet werden:

Status

Status 1: Status 2: Status 3:

Terminologie

In Quelle: empfohlene Einstellung nahegelegte Einstellung Einstellung

recommended setup

prefork

Abbrechen Speichern

Dateien

Beispiel-Unterord... install-unix.ap... install-unix.apache...

For more information type `<1>/configure --help</1>` in the corresponding source tree.

service httpd restart Wenn Sie den obigen Anweisungen folgen, werden Sie einen laufenden Apache2 mit Unterstützung für PHP als SAPI-Modul erhalten.

Apache may be built multithreaded by selecting the `<1>worker</1>` MPM, rather than the standard `<2>prefork</2>` MPM, when Apache is built.

Natürlich existieren für Apache und PHP viele weitere Konfigurationseinstellungen. Verwenden Sie `./configure --help` im jeweiligen Quellcodeverzeichnis, um weitere Informationen zu erhalten.

This is done by adding the following option to the argument passed to `./configure`, in step 3 above:

`--with-mpm=worker`

This should not be undertaken without being aware of the consequences of this decision, and having at least a fair understanding of the implications.

The Apache documentation regarding `<1><2> MPM-Modules</2></1>` discusses MPMS in a great deal more detail.

`<1>Note</1>:` `<1>Hinweis</1>:`

The `<1>Apache MultiViews FAQ</1>` discusses using multiviews with PHP.

`<1>Note</1>:` `<1>Hinweis</1>:`

To build a multithreaded version of Apache, the target system must support threads.

In this case, PHP should also be built with experimental Zend Thread Safety (ZTS).

Under this configuration, not all extensions will be available.

The recommended setup is to build Apache with the default `<1>prefork</1>` MPM-Module.

service httpd restart Wenn Sie den obigen Anweisungen folgen, werden Sie einen laufenden Apache2 mit Unterstützung für PHP als SAPI-Modul erhalten.

Natürlich existieren für Apache und PHP viele weitere Konfigurationseinstellungen. Verwenden Sie `./configure --help` im jeweiligen Quellcodeverzeichnis, um weitere Informationen zu erhalten.

Falls Sie eine Multithreaded-Version von Apache2 bauen wollen, müssen Sie das Standard MPM-Modul `prefork` entweder durch `worker` oder `perchild` ersetzen.

Falls Sie eine Multithreaded-Version von Apache2 bauen wollen, müssen Sie das Standard MPM-Modul `prefork` entweder durch `worker` oder `perchild` ersetzen.

99

0

Marc Mittag

beoLogisch Editor 0.4.0 (Thu Oct 11 14:56:09 CEST 2012)

File Edit Go to Options Tasks Synchronize Data Source Functions Help

C:\Program Files\OpenTMS\test\tekom2009\sample.odt.xls

```

<sn 0>Montage-, Einbau- und Betriebsanleitung
<tn 0>
<sn 1>Radblocksystem DRS 250, 315, 8.800
<tn 1>
<sn 2>1830244.EPS
<tn 2>
<sn 3>Vorwort
<tn 4>
<sn 4>Sie haben ein Produkt der Mannesmann Dematic AG erworben, das nach den neusten Erkenntnissen der Technik hergestellt worden ist.
<tn 4>
<sn 5>
<sn 5>Mit dieser Betriebsanleitung möchten wir dem Benutzer zweckdienliche Anweisungen zum sicheren und sachgerechten Arbeiten geben sowie eine leichte Instandhaltung ermöglichen.
<tn 5>
<sn 6>Jede Person, die mit Transport, Aufstellung, Inbetriebnahme, Bedienung, Wartung und Reparatur unseres Radblocksystems DRS und dessen Zusatzeinrichtungen beauftragt ist, muß gelesen und verstanden haben:
<tn 6>
<sn 7>die Betriebsanleitung
<tn 7>
<sn 8>die Sicherheitsvorschriften
<tn 8>
<sn 9>die Sicherheitshinweise der einzelnen Kapitel und Abschnitte
<tn 9>
<sn 10>Um Bedienungsfehler zu vermeiden und einen störungsfreien Betrieb unserer Produkte zu gewährleisten, muß die Betriebsanleitung dem <ph id="0">&lt;text>&lt;span></span></ph><ph id="1">&lt;text>&lt;span>&lt;alphabetical-index-mark text="string-value='Bedienungspersonal'&/></span></ph><ph id="2">&lt;text>&lt;span>Bedienungspersonal stets zugänglich sein.</span></ph></ph>
<tn 10>
<sn 11>Weitere Unterlagen
<tn 11>

```

TM Data Sources MicrosoftTranslate

FUZZY 70 Auto Search

Data Source: GoogleTranslate

90: Vorwort Vorwort

Preface Preface

Imagine...

- ▶ Customer wants to enhance existing XLIFF Editor or application with his own small “XLIFF apps”
 - ▶ Cost computations
 - ▶ Own similarity measures
 - ▶ ...
- ▶ Translation service provider wants to add some additional operations through “XLIFF apps” as part of the XLIFF file
 - ▶ Constraint checks
 - ▶ Synchronizing TMs through web
 - ▶ Add a new MT system
- ▶ ...

Current Limitations...

- ▶ XLIFF does not support this
- ▶ Translation GUIs and application using XLIFF supply a basic functionality
 - ▶ Not extendible
 - ▶ Or proprietary approach for extension
- ▶ No common agreed way how running “apps” in translation applications



Extending XLIFF

- Use HTML approach
 - From passive presentation to active content
 - Proven approach
 - JavaScript – ECMA Script - standard
- Add Script support to XLIFF
 - <script...> tag represents semantic knowledge
 - Invoked on element basis
- Cross application functionality
 - Vendor independent
 - Interoperability supported



Advantages

- XLIFFs cross platform goals supported
 - Mappers between vendor specific XLIFF implementations
- User is not limited to the functionality supplied by the application vendor
 - Can extend functionality of editors
 - E.g. specialized word counting
 - Matching quality computation
 - Access external data sources like MT, term databases etc. using Web Service
 - May add his own work flow features



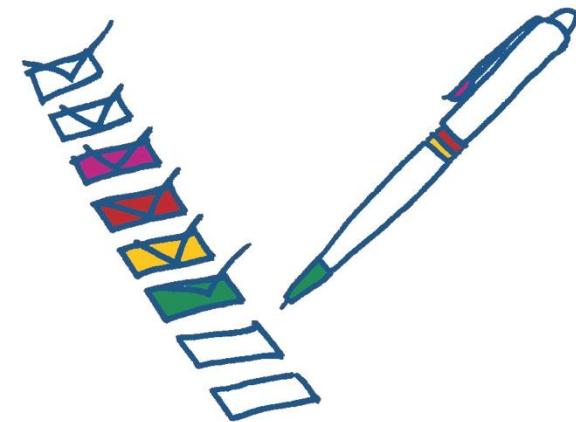
Components - Questions

- ▶ **Functions and Procedures**
 - ▶ How and where to define functions and procedures in XLIFF
- ▶ **Parameter Passing**
 - ▶ How to handle parameter passing?
- ▶ **Return Value Passing**
 - ▶ How to return values back to the application?
- ▶ **GUI Elements**
 - ▶ Which GUI elements should be supported?
- ▶ **Security**
 - ▶ How to avoid security leaks?



Functions, Procedures

- Following a programming language approach
 - Like HTML
 - Include scripting code in XLIFF elements
 - Script tag
 - Properties
- Scripting language
 - ECMA Script (“Java Script”)
 - Wide spread
 - Easy to implement
 - Security concept
 - Other languages like PHP, Perl?
 - Access available in languages like Java etc.



Types, Parameters, Return Values

- **Types**
 - JSON
- **Parameters**
 - Automatic passing of parameters
 - JSON based representation of element as a parameter for script
 - Additional parameters for script
 - Languages, file name, other elements...
- **Return Values**
 - JSON supports complex objects
 - Most modern programming languages support JSON



Script Integration

➤ Using the HTML based approach

➤ Script tag

```
<script type="text/javascript"> ... some JavaScript code </script>
<script src="javascript.js" type="text/javascript"></script>
```

➤ XLIFF

➤ Add to XLIFF Header

➤ Zero or more scripts allowed in header

```
<?xml version="1.0" encoding="UTF-8"?>
<xliff version="1.0">
<file datatype="XML" original="..." source-language="de" xml:space="default" target-
language="en">
<header>
  <script type="text/javascript">
    ... some JavaScript code
  </script>
  <script src="javascript.js" type="text/javascript"></script>
...
</header>
...
```



Script Invocation

- Differences to HTML
 - Script tag only in header, not elsewhere in the document
 - All functions etc. declared in header
- Script invocation in XLIFF as some kind of trigger
 - Attribute in element
 - on-enter
 - Run script when the element is entered by the application
 - on-exit
 - Run script when the element is left by the application
 - Additional triggers possible
 - E.g. some kind of trigger which reacts on changes to the target element
- Implementation of trigger invocation up to application
 - Editor, command line tools, ...



Informal Script Invocation Logics

- Application enters an element
 - Check for on-enter attribute
 - if exists,
 - execute script
 - replace trans-unit with return JSON string converted to xml
- Application leaves a element
 - Check for on-exit attribute
 - if exists
 - execute script
 - replace trans-unit with return JSON string converted to xml
 - on-exit is applied to modified trans-unit of on-enter



Default handling in Applications

➤ Problem

- XLIFF files will normally not contain any scripts
- Run scripts when application is used automatically

➤ Solution

- Support default handling for scripts
- Application tries to load default.js
 - Contains user written scripts
 - Read from directory of XLIFF file
- For each XLIFF element apply default functions
 - “on-enter-<element-name>” and “on-exit-<element-name>”
 - Example: on-enter-trans-unit
 - Function naming: Xpath?



Script Example

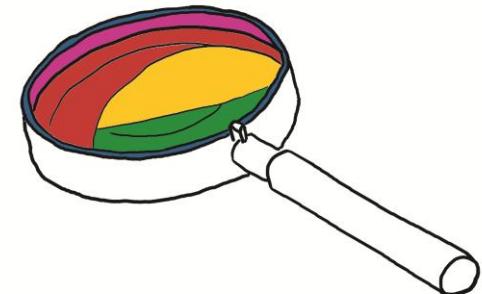
➤ Main usage in trans-unit element

```
<trans-unit  
    approved="no" id="3"  
    translate="yes"  
    on-enter="estimateTranslationTime"  
    on-exit="validateTranslationTime">  
    <source xml:lang="de"><g />haus</source>  
    <target xml:lang="en"><g>house</g></target>  
</trans-unit>
```

Call a JavaScript function **estimateTranslationTime** with trans-unit as JSON object when trans-unit entered

Call a JavaScript function **validate** with trans-unit as JSON object when trans-unit element is left

```
{  
  "trans-unit": {  
    "-approved": "no",  
    "-id": "3",  
    "-translate": "yes",  
    "-on-enter": "estimateTranslationTime",  
    "-on-exit": " validateTranslationTime ",  
    "source": {  
      "-xml:lang": "de",  
      "#text": "haus"  
    },  
    "target": {  
      "-xml:lang": "en",  
      "g": "house"  
    }  
  }  
}
```



Script Example Return Value

```
<trans-unit
    approved="no" id="3"
    translate="yes"
    on-enter="estimateTranslationTime"
    on-exit="validateTranslationTime" >
    <source xml:lang="de"><g />haus</source>
    <target xml:lang="en"><g>house</g></target>
</trans-unit>
```

Application modifies trans-unit (not required)

```
<trans-unit
    approved="no" id="3"
    translate="yes"
    on-enter="estimateTranslationTime"
    on-exit="validateTranslationTime" >
    <source xml:lang="de"><g />haus</source>
    <target xml:lang="en"><g>house</g></target>
    <prop-group>
        <prop prop-type="estimatedTime">00:00:10</prop>
        <prop prop-type="neededTime">00:00:12</prop>
    </prop-group>
</trans-unit>
```

JSON Return Value for validate

```
{
    "trans-unit": {
        "-approved": "no",
        "-id": "3",
        "-translate": "yes",
        "-on-enter": "estimateTranslationTime",
        "-on-exit": "validateTranslationTime",
        "source": {
            "-xml:lang": "de",
            "#text": "haus"
        },
        "target": {
            "-xml:lang": "en",
            "g": "house"
        },
        "prop-group": {
            "prop": [
                {
                    "-prop-type": "estimatedTime",
                    "#text": "00:00:10"
                },
                {
                    "-prop-type": "neededTime",
                    "#text": "00:00:12"
                }
            ]
        }
    }
}
```



GUI Elements, Forms

- Script might require user interaction
 - Extension of base application
 - Display of properties etc.
 - Input some data
- Problem
 - No GUI elements foreseen in scripting languages
- Possible solution (Java approach)
 - Support import of classes and access to classes / instances
 - Definition of some basic GUI element
 - Yes / No / Cancel / ok
 - Simple input dialogue
 - ...



Example application

- Translation Change measure for MT matches
- Prototype
 - MT Translation for a segment
 - Copied to target
 - Similarity based on source segment comparison
 - Modification done by translator
 - In order to compute a fair price for the target changes compare translators changes with MT translation based on Levenshtein

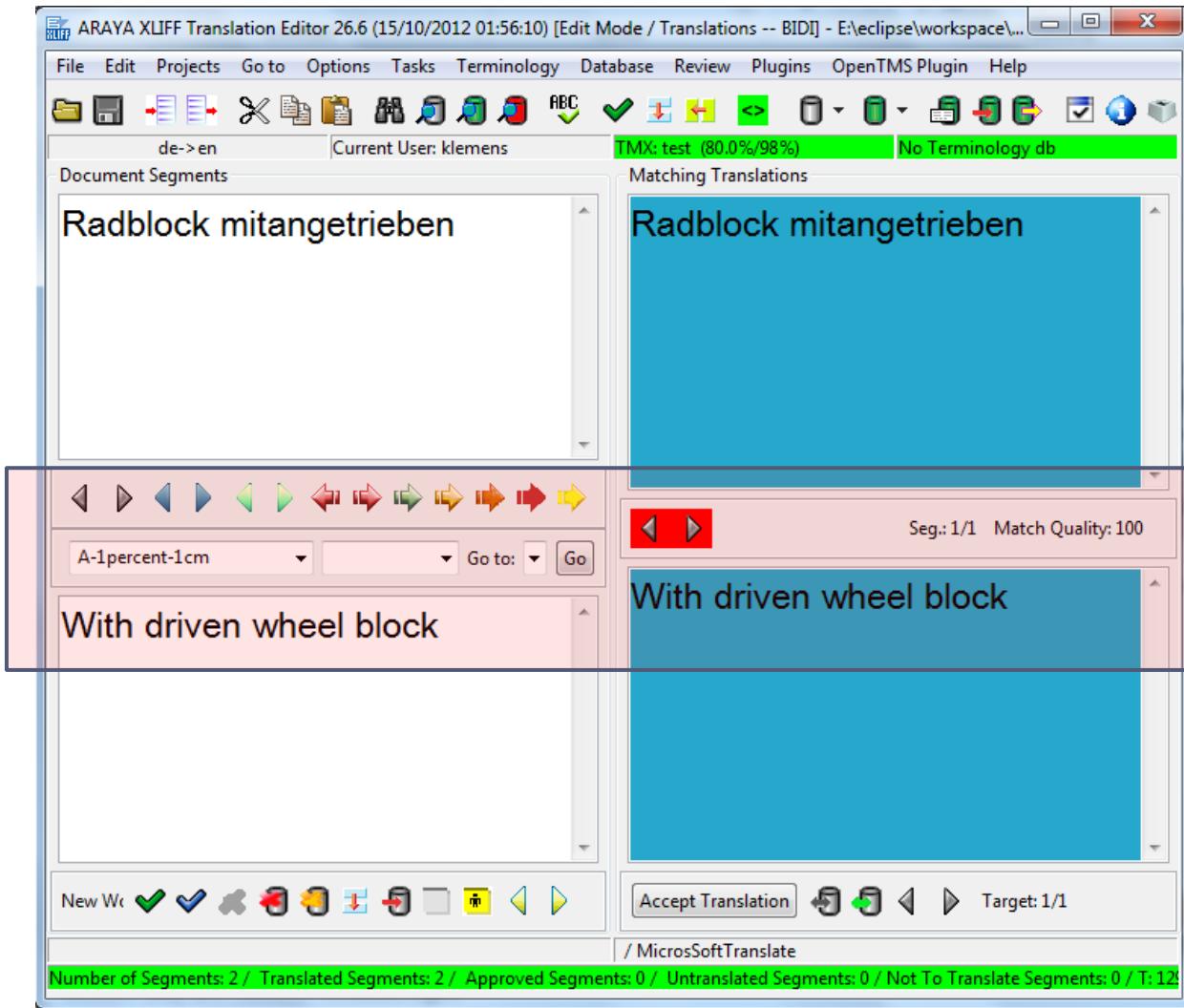


Example XLIFF File

```
<?xml version="1.0" encoding="UTF-8"?>
<xliff version="1.0">
  <file datatype="XML" original="C:\Program Files\OpenTMS\test\tekom2009\sample.odt.content.xml" source-language="de"
  xml:space="default" target-language="en">
    <header>
      <script src="file:///E:/eclipse/workspace/openTMS/test/ScriptTest/xml_for_script-
3.1/jsXMLLoader/xmlDom.js" />
      <script src="file:///E:/eclipse/workspace/openTMS/test/ScriptTest/testscript.js" />
    </header>
    <body>
      ...
      <trans-unit approved="no" help-id="0" id="0" reformat="yes" translate="yes"
      xml:space="preserve" on-enter="computeLevenOfTransunit(transunit);">
        <source xml:lang="de">Radblock mitangetrieben</source>
        <target xml:lang="en">With driven wheel block</target>
        <alt-trans match-quality="MT" id="054bdc3a-b361-4afc-b296-86ad2e819716"
        xml:space="preserve" origin="MicrosSoftTranslate">
          <source xml:lang="de">Radblock mitangetrieben</source>
          <target xml:lang="en">With driven wheel block</target></alt-trans>
      </trans-unit>
    </body>
  </file>
</xliff>
```



Example before changes by translator



Example after changes by translator

The screenshot shows the ARAYA XLIFF Translation Editor interface. The main window displays a document segment "Radblock mitangetrieben" in the source language (German) and its translation "With driven wheel block" in the target language (English). A status bar at the bottom indicates "Number of Segments: 2 / Translated Segments: 2 / Approved Segments: 0 / Untranslated Segments: 0 / Not To Translate Segments: 0 / T: 12". A blue callout box highlights the English translation "With driven wheel block" and contains the text: "MT translation changed by translator Recomputed similarity based on comparing the MT alt-trans target with the new target".



Summary and Chances

- Easy integration into existing applications
- Freedom for the user
 - Make XLIFF based applications more versatile and powerful
- XliffApp store
 - Market for application/vendor independent modules
 - Additional revenue for vendor independent developers and translators
- Next step
 - Building a demonstration prototype
 - e.g. for openTMS



For more information, please contact:

klemens@translationautomation.com

