Using TBX-Min with Spreadsheet Glossaries

Guide to trying TBX-Min

James Schofield Hayes

2014-08-19

This is intended to be a simple tutorial providing instructions on converting existing glossary in Spreadsheet format into TBX-Min. For more information on TBX and TBX-Min see http://tbxconvert.gevterm.net or http://www.tbxinfo.net/

Table of Contents

Using TBX-Min with Spreadsheet Glossaries

Converting a spreadsheet to TBX-Min:

Step 1: Acceptable Spreadsheet Formats

First, save the glossary you would like to convert to TBX-Min as one of the following:

- **.xlsx -** (Microsoft Excel 2007, 2010, or above) *recommended
- **.xls** - (Microsoft Excel 97-2007)
- **.csv** - Be sure it is encoded as UTF-8 if you use this option! .csv may sometimes be inaccurate
- **.txt** - like .csv, make sure this is UTF-8 (tab delimited is most tested) .txt may sometimes be inaccurate

Step 2: Open Browser to Conversion Webpage

Now open a web browser to http://tbxconvert.gevterm.net/tbx-min. At the bottom of this page is the "**Spreadsheet Glossary Converter**" section.

Spreadsheet Glossary Converter

You can convert glossaries stored in Spreadsheet format. An instruction guide is available for download. You will be asked which columns of the Spreadsheet represent which TBX-Min data categories (such as a 'Source Language' column or a 'Comment' column). Please note that in order for the converter to work, column headers of some sort **must** exist. You will also be asked about some information about the document in general (which gets used in the Header of the TBX-Min file).

TBX-Min supports single language as well as two language glossaries, rather than being strictly bilingual. This converter only requires a source language. This feature can be used to pull out individual language glossaries one (or two if you want language pairs) at a time. If you want to do this, simply state the desired language to be the "source" here and then label the appropriate column header on the next page as 'Source Language'.

Only Excel Spreadsheets or CSV files (.xlsx, .xls, .csv, or .txt *[tab delimited format]*) are supported currently, however .xlsx or .xls are recommended.

Convert Spreadsheet to TBX-Min: [Choose File] No file chosen

Source Language*: [_____]
Target Language: [_____]
License: [_____]
Creator: [_____]
Description: [_____]
Directionality*: [bidirectional ▼]
Subject: [_____]
Dictionary ID*: [_____]
*mandatory
[Convert]

Step 3: Upload Spreadsheet and Give Relevant Information

Now click on "Choose File" and select the file from your computer.   Then, fill out the header information (Source Language, Target Language, etc.). *only Source Language, Directionality, and Dictionary ID are required* When desired header boxes have been filled out, click "Convert".

Step 4: Select Header Columns

Now you should see a summary of your header information with a selection box and a text area at the bottom:

Given Header Information:

Source Language: en

Target Language: es

License: CC By*

Creator: JSH
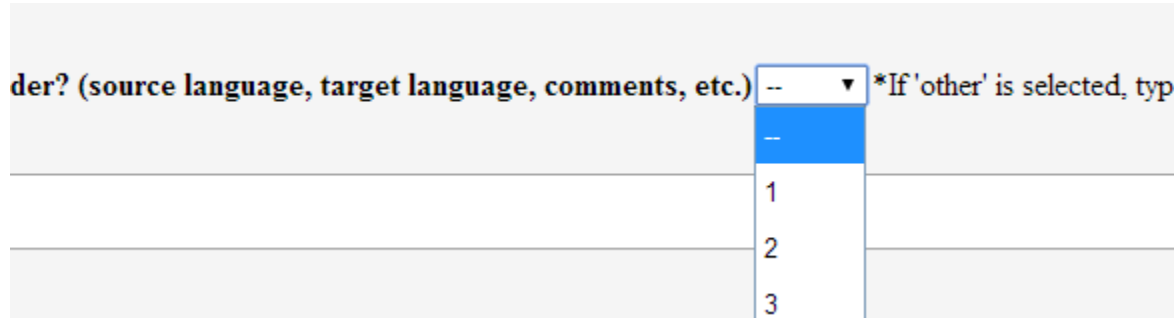
Description:

Directionality: bidirectional

Subject (leave blank if each entry is different):

Dictionary ID: D001

Which row contains the spreadsheet Header? (source language, target language, comments, etc.) [-- ▼] *If 'other' is selected, type custom row and press "Tab"
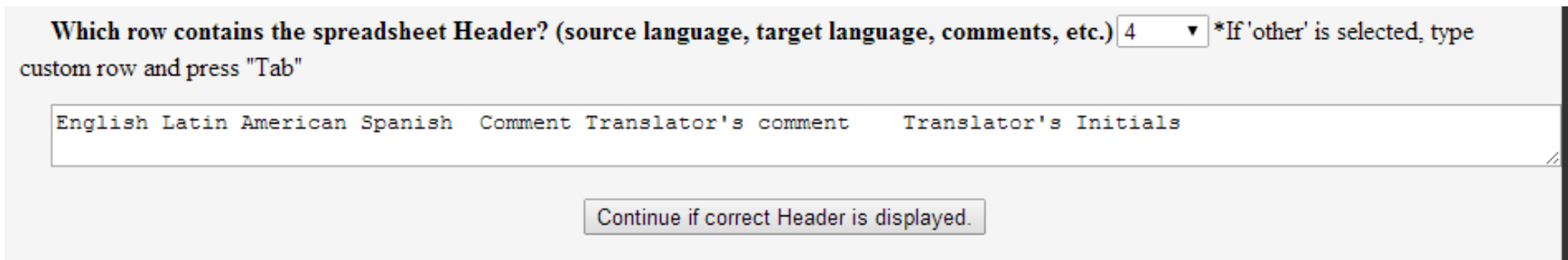
The "spreadsheet Header" is whatever you have at the top of your spreadsheet columns indicating what information is held in each column (i.e. "English Terms", "Spanish Terms", or "Source Terms", "Target Terms". Simply click on the drop down menu and select a number

(this number represents the number of the row of the header). If the row happens to be below row 10, select "Other" and type a custom number:

der? (source language, target language, comments, etc.) -- ▼ *If 'other' is selected, typ

--
1
2
3

You will see the header displayed in the text area when the row is correct (the example header was on row 4 of the spreadsheet):

Which row contains the spreadsheet Header? (source language, target language, comments, etc.) 4 ▼ *If 'other' is selected, type custom row and press "Tab"

English Latin American Spanish   Comment Translator's comment   Translator's Initials

Continue if correct Header is displayed.

If this is the correct header click "Continue if correct Header is displayed".

Step 5: Assign Header Columns and Convert

Now the header values must be assigned to their TBX-Min equivalents. The options are:

- Source Language
- Target Language

- Part of Speech: TBX-Min has specific values it allows here. Each term can only have one of **six** values (*you may need to change this in your spreadsheet*): *illegal values will be ignored and will not be converted
    - verb
    - adverb
    - noun
    - properNoun
    - other
- Term Status : TBX-Min has specific values it allows here. Each term can only have one of **four** values (*you may need to change this in your spreadsheet*): *illegal values will be ignored and will not be converted
    - preferred
    - admitted
    - notRecommended
    - deprecated
- Customer
- Note* (Note may be used more than once if your categories do not happen to match with what is available!)
- Subject (This is the subject, or domain of the entry, such as "Clothing Accessories")

- Concept ID - This will be generated if there are not already specific values in your glossary

TBX Convert: TBX-Min

Please asign each header to a corresponding TBX-Min element (leave blank if none fit):

*Please note that Source Language and Target Language are required.

English [Source Language ▼]

Latin American Spanish [Target Language ▼]

Comment [Note ▼]

Translator's comment [Note ▼]

Translator's Initials [Note ▼]

[Convert Spreadsheet]

Step 6: Complete and Download TBX-Min File

Once all the values (that you want to convert, others may just be left blank) are defined, click "Convert Spreadsheet". Your browser should now download a TBX-Min file converted from your spreadsheet based Glossary!

Validating the TBX-Min File

You will want to check to see if your file is valid TBX-Min. Assuming you successfully downloaded a TBX-Min file using the process described above, your file should be conformant! However, to check for sure, visit http://tbxconvert.gevterm.net/tbx-min/validate/validate.html:

Validate TBX-Min Files Against XSD/RNG Schemas

Validate your TBX-Min file against the current XSD/RNG schema. Schemas are available here.

XSD Validation:

Select TBX-Min file to validate: [Choose File] No file chosen   [Validate]

RNG Validation:

Select XML file to validate: [Choose File] No file chosen   [Validate]

Simply upload the TBX-Min file you just made and then click "Validate". If there are errors, it is best to contact an XML expert to inspect your TBX-Min file to see what is failing.

Viewing / Editing the new TBX-Min File

If you want to view the contents of your TBX-Min file in a user friendly way, and be able to edit the data, you may use the viewer found at http://viewer.tbxinfo.net/:

TBX Terminology Manager: Login

The online TBX terminology manager allows a user to view and edit TBX files dynamically. To get started, select one of the three options below.

Privacy disclaimer: This web page is experimental and all submitted data is collected for analysis. No personal identifiers are collected, but confidential data should not be submitted to this site. Users who do not wish to expose their data to the broader public should create a user account and log in, but staff may examine submitted data for diagnostic and developmental purposes. If you do not agree to these terms, please do not use this site.

Sign in:

Registered Users

User Name: [   ]

Password: [   ]

[Sign In]

Register for an account

First Name: [   ]

User Name: [   ]

Password: [   ]

Confirm Password: [   ]

[Create User]

Public access (does not require a user accont)

Note: All glossaries built while logged in as the public user are visible to the public.

Use public user: [Sign In]

Step 1: Sign In

You may register your own account or use the public Access account (Note that anything uploaded on this account is available to any user).

Step 2: Upload TBX-Min

Once you have signed in, you will have the option to upload your TBX-Min file(s). All previously uploaded files will be shown here as well.

Clicking "Show Samples" will give you access to some previously validated sample TBX-Min files, should you ever wish to look at those. Simply upload the TBX-Min file you want to view and click "Import TBX". Once the import is complete your file will be listed in the "**Saved Glossaries**" section.

You have several options here:

- "View and Edit" - This will take you to the manager, which allows you to view and edit TBX-Min
- "Rename" - renames the file
- "Delete" - delete the file
- "Clone" - copy the file with a new name of your choosing (**This is always recommended before editing TBX-Min files, to have a backup**)
- "Download" - download the TBX-Min file

Step 3: Create a Backup

If you plan to edit your TBX-Min file, you should clone it first for backup purposes. Click "Clone", then choose how many backups (clones) you want. We will just say "1". Now name your backup (the example was named to "SWLS - Backup"). Once this is done your management page should have your backup listed as well as the original:

- SWLS - Backup ➔   [View and Edit] [Rename] [Delete] [Clone] [Download]

- SWLS Legal - Technical Terminology Glossary for CEMEX-November 2011 ➔   [View and Edit] [Rename] [Delete] [Clone] [Download]

Step 4: View and Edit TBX-Min

We are now ready to safely look at our TBX-Min file. *__Note: only tested to work on Mozilla Firefox and Google Chrome, does not yet work with Internet Explorer__* Click "View and Edit" on the file you wish to look at. A new tab should come up as the viewer.

Step 5: Edit Header of TBX-Min

On the viewer screen you will see your Header Information (which you can edit by clicking on the existing values, or add values that don't exist by clicking "Add Header Info"

Upload: SWLS Legal - Technical Terminology Glossary for CEMEX-November 2011
id*: D001
creator*: SWLS
description*: Technical Terminology Glossar for CEMEX, Nov 2011 (This is a test file)
languages : source: 'en' target: 'es'
dateCreated*: 2014-08-11T23:55:03.9412
license*:
directionality*:

Deleting any value completely (making it completely blank) will delete it from the TBX-Min header. It may be added back at any time.

Step 6: Edit Term Information

Select a term from the lists under the "**Term Information**" section (either source or target). This will cause all of the related information to this term to display below:

Term Information:

Click on text to edit it (if the value is allowed to be edited).
Please select a term:

Source Terms:	Target Terms:
"fair and equitable treatment (""FET"")"	"auxiliar, suplementa
(tribunal) to challenge	"hormigón, concreto"
alkali sulfates	"mezcla, solución"
ancillary	"principio de ""reclam
Appraisal Committee	"proveedor, abastece
arbitral tribunal	"revisión, puesta a pu
attrition	"solidez, robustez"
authoritative official source	"trato justo y equitati
award	(sistema de) Reducc
ball mills	aceite combustible

Term*: **Appraisal Committee**   Entry ID: C011   [Delete Term]
Term Status*: ---
Part of Speech*: ---
Customer*: ---

Target term(s)*:

Term	Term Status
Comisión de Avalúos	---

Notes*: [Add Note]

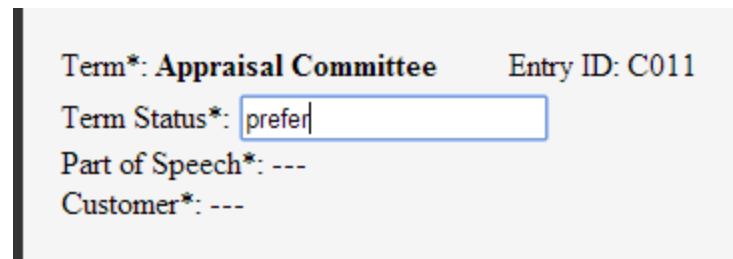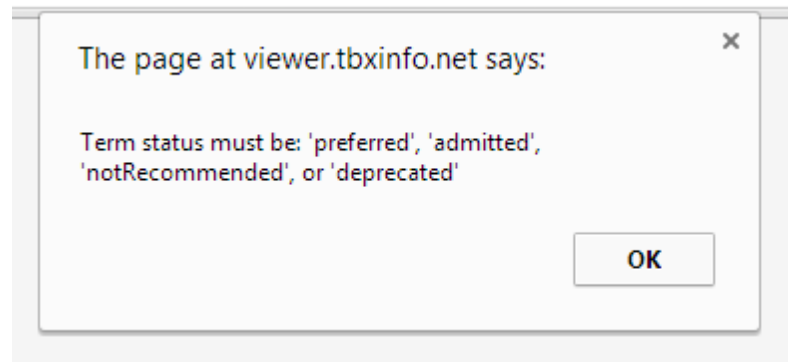| Note Type	Note
Comment	Cfr. Ley de Expropiaciones de Venezuela

Total number of entries: 167

Step 7: Notes on Term Status and Part of Speech

All values marked as "---" do not exist in your current TBX-Min file. These may be added or edited simply by clicking on them:

Again, "Term Status" and "Part of Speech" can only have specific values in TBX-Min. These are enforced if you try to edit these values and a list of valid options will pop up if you type it incorrectly.

Step 8: Adding and Deleting Information

Terms may be completely deleted via the "Delete" button. They may be edited by click on them and typing. New comments may be added by clicking "Add Note" and filling in the appropriate values. "**Note Type**" is set to "Comment" by default, but can be any value you wish "Comment Author" perhaps? Soon, the ability to add terms will also be available.

Step 9: Save and Download

When all changes look good you may either "**Save**", which saves your file and refreshes the page, or " **Download**", which will save any download the most recently saved version of the TBX-Min file to your computer.

Convert TBX-Min file to TBX-Basic

TBX-Min may be converted into its cousin format TBX-Basic using our converter at http://tbxconvert.gevterm.net/. This allows you to take advantage of any software that supports TBX-Basic. To do this, simply go to the mentioned website, and scroll to the "**TBX Basic Conversion**" section:

Next to "**Convert TBX-Min to TBX-Basic**", choose your TBX-Min file and upload it. Then click the "Convert" button to the right of your file.

(Very Near) Future of TBX-Min

One of the features which will soon be available is the ability to combine a TBX-Min file into an existing TBX-Basic master file.   The practical application of this is that a Project Manager could take out a single language set glossary (say "English" and "Spanish") from a TBX-Basic file (using the "**Convert TBX-Basic to TBX-Min**" option pictured above") in the form of a simple TBX-Min file. They could then give it to the translator assigned to this language pair. The translator could then use the English-Spanish TBX-Min glossary and edit it as necessary and return it to the Project Manager. The Project Manager would then simply combine the TBX-Min file back into the master TBX-Basic file.