LEXICALIZING THE FORUM

Spiros Doikas

(Paper submitted to the 10th Conference "Hellenic Language and Terminology",
Athens, Greece, 12–14 November 2015)

Abstract

Translators' websites with terminology modules, as well as translation fora, have been important tools of the trade for translators in providing assistance with terminology queries.

It is to be noted, that a terminology module is substantially different to a forum, and any forum script, if not heavily modified, is hardly the optimum medium for asking, answering and searching terminology.

To that end, a Greek forum based on SMF (an open-source forum script), was developed and modified over the years specifically to serve the above purpose. Various aspects of that customisation effort will be discussed here, such as:

Rules

Rules regarding not only user behaviour but the recommended way of posting and answering terminological queries.

Modifications of the SMF forum software

Php modifications—increase the number of characters for the subject field; multiple changes needed in various files; use URL trimming for improved display of long URLs; provide a Google BBC button to easily create search queries.
MySQL modifications—increase the number of characters for the subject field and optimise database.

Specific “mods” — “mods” are code modifications posted by users that provide new functionality or enhance an existing functionality and are available via the SMF site’s “Mods” section).

Search enhancement—there is Search enhancement mod, which has some bugs, and a custom mod which was developed to replicate part of that functionality.

Style guide

Rules for the submission of terminology questions and their editing by a moderator/user once the translation has been finalised.

Rules for delimiters/initialisms.

Maintenance of entries

Duplicates removal, adding translations to entries, fixing case (different rules for different languages, i.e. in French in names of organisations only first letter should be capital), initialisms, delimiters.

Mass export, import and update of terms—using php scripts and MySQL commands to export and then manipulate in Excel.

Λεξικοποιώντας το φόρουμ

Σπύρος Δόικας

Περίληψη

Οι ιστότοποι μεταφραστών με ορολογικές υπομονάδες, καθώς και τα μεταφραστικά φόρα, είναι σημαντικά εργαλεία για τους μεταφραστές, παρέχοντάς τους βοήθεια σε ερωτήματα ορολογίας.

Ωστόσο, πρέπει να σημειωθεί ότι μια υπομονάδα ορολογίας διαφέρει σημαντικά από ένα φόρουμ, και ένα φόρουμ, αν δεν υποστεί σημαντικές τροποποιήσεις, δεν είναι το ιδανικότερο μέσο για ερωτήσεις, απαντήσεις και αναζητήσεις όρων.

Για τον σκοπό αυτό, αναπτύχθηκε και τροποποιήθηκε ένα ελληνικό φόρουμ με βάση το SMF (ένα λογισμικό φόρουμ ανοιχτού κώδικα). Αναλύονται διάφορες πτυχές αυτής της προσπάθειας προσαρμογής, όπως:

Κανόνες

Κανόνες που δεν αφορούν μόνο τη συμπεριφορά των χρηστών αλλά και ο ενδεδειγμένος τρόπος δημοσίευσης και απάντησης ορολογικών ερωτημάτων.

Τροποποιήσεις του λογισμικού SMF

Τροποποιήσεις php – Αύξηση του αριθμού των χαρακτήρων για το πεδίο του θέματος (απαιτούνται πολλές αλλαγές σε διάφορα αρχεία)˙ λειτουργία περικοπής URL (μέσω αλλαγής κώδικα CSS) για βελτιωμένη εμφάνιση μεγάλων σε μήκος διευθύνσεων URL˙ κουμπί «Google» για εύκολη δημιουργία αναζητήσεων.

Τροποποιήσεις MySQL – Αύξηση του αριθμού των χαρακτήρων για το πεδίο του θέματος και τη βελτιστοποίηση της βάσης δεδομένων.

SMF Mods – Τροποποιήσεις πηγαίου κώδικα του προγράμματος που αναρτώντα σε μορφή «πακέτων» από τους χρήστες και παρέχουν νέα λειτουργικότητα ή τροποποιούν μια υφιστάμενη λειτουργικότητα του SMF και είναι διαθέσιμα μέσω του τμήματος «Mods» της ιστοσελίδας).

Βελτίωση αναζήτησης – Υπάρχει διαθέσιμο ένα «Mod» με τίτλο «Search Enhancement», το οποίο έχει κάποια προγραμματιστικά σφάλματα. Ως εκ τούτου, αναπτύχθηκε μια προσαρμοσμένη λειτουργία, η οποία αναπαράγει μέρος αυτής της λειτουργικότητας.

Χρήσιμες λειτουργίες του SMF

Υπάρχουν κάποιες λειτουργίες του SMF που βοηθούν ιδιαίτερα την επεξεργασία όρων. Ένα παράδειγμα είναι η δυνατότητα γρήγορης επεξεργασίας (που μπορεί να εφαρμοστεί τόσο σε προβολή θέματος όσο και σε προβολή πίνακα). Η γρήγορη επεξεργασία δίνει τη δυνατότητα επεξεργασίας χωρίς να χρειάζεται να γίνει φόρτωση νέας σελίδας. Σε προβολή πίνακα δίνει τη δυνατότητα γρήγορης διόρθωσης πολλών θεμάτων (αντίστοιχες διορθώσεις, π.χ. σε λογισμικό Mediawiki, το οποίο θεωρείται μια ενδεδειγμένη λύση ηλεκτρονικής-συνεργατικής λεξικογραφίας, θα απαιτούσαν πολύ περισσότερο χρόνο).

Οδηγός ύφους

Οδηγίες σχετικά με την υποβολή ερωτήσεων ορολογίας και την επεξεργασία τους από τον συντονιστή/ερωτώντα άπαξ και η μετάφραση έχει οριστικοποιηθεί.

Οδηγίες για οριοθέτες. Με το σκεπτικό ότι κάποια στιγμή θα γίνει εξαγωγή των όρων διαμορφώθηκαν κάποιες οδηγίες που θα διευκολύνουν την επεξεργασία τους. Για παράδειγμα ως οριοθέτης μεταξύ του όρου και των αποδόσεων έχουν οριστεί οι χαρακτήρες « -> ». Σε περίπτωση όμως που έστω και μία απόδοση εμπεριέχει το κόμμα ως μέρος της, τότε χρησιμοποιείται ως οριοθέτης το « –> », δηλαδή το en-dash (αντί του ενωτικού της προηγούμενης περίπτωσης) και το σύμβολο «μεγαλύτερο από». Ως οριοθέτης μεταξύ των αποδόσεων έχει οριστεί το κόμμα. Σε περίπτωση όμως που έστω και μία απόδοση εμπεριέχει το κόμμα ως μέρος της, τότε χρησιμοποιείται ως οριοθέτης το σύμβολο «|».

Οδηγίες για αρκτικόλεξα. Τα αρκτικόλεξα παρατίθενται μετά τον ανεπτυγμένο όρο σε παρένθεση (αν υπάρχει πάνω από ένα, διαχωρίζονται με κόμμα). Δεν χρησιμοποιούνται τελείες. Αποφεύγεται η χρήση αρκτικόλεξων που δεν ανήκουν στη γλώσσα του όρου καθώς και όσων αντικατοπτρίζουν μέρος του όρου.

Μαζική συντήρηση λημματολογίου

Κατάργηση διπλοτύπων, προσθήκη αποδόσεων, διόρθωση πεζοκεφαλαίων (υπάρχουν διαφορετικοί κανόνες για κάθε γλώσσα, π.χ. για τα ονόματα οργανισμών στα γαλλικά το πρώτο γράμμα είναι κεφαλαίο και τα υπόλοιπα πεζά), αρκτικόλεξα, οριοθέτες.

Μαζική εξαγωγή, εισαγωγή και ενημέρωση των όρων τόσο μέσω λειτουργιών που έχουν αναπτυχθεί σε κώδικα PHP όσο και με χρήση εντολών MySQL για περαιτέρω επεξεργασία στο Excel.

0. Introduction

Translators' websites with terminology modules, as well as translation fora, have been important tools of the trade for translators.

Early web-based forums date back as far as 1994[1]. However, they really took off from 2000 onwards. **VBulletin**, **YaBB**[2] and **phpBB** appeared in the year 2000. The first **SMF** version, SMF 1.0 Beta 1a, was released on September 30, 2003[3]. When it comes to translators' websites with terminology modules, **ProZ.com** was launched in 1999 and **TranslatorsCafe** in August 1, 2002.

It is to be noted, that a *terminology module* is substantially different to a *forum*; therefore, any forum script, if not heavily modified, is hardly the optimum medium for asking, answering and searching terminology.

To that end, a Greek forum based on SMF[4] (Simple Machines Forum), was developed and modified over the years specifically to serve the above purpose.

What follows is a brief exploration of the difficulties presented and the steps taken, as well as what was done differently in comparison to an ordinary forum.

1.0 Forum rules

Almost every forum has a page with rules prominently displayed. Rules ensure the smooth functioning of the forum and describe what is allowed and what is not.

[1] https://en.wikipedia.org/wiki/Internet_forum
[2] The precursor of SMF which was written in Perl.
[3] https://en.wikipedia.org/wiki/Simple_Machines_Forum
[4] http://www.simplemachines.org

Apart from the typical forum rules[5], like the ones dealing with advertising, spam, double posting, overquoting (i.e. not quoting a full post in the reply immediately after it), using the modify button to edit a post rather than posting below and thus cluttering the forum, flaming and staying on topic; there are rules which are specific to a terminology forum. For example:

- limiting a translation query to 12 words
- creating new topics for different terms
- avoiding greeklish
- using proper accents, punctuation marks (i.e. do not use English question mark in Greek text) and capitalization
- the query should be entered in the subject field in lowercase, unless grammar demands a capital (i.e. "Greek" and not "greek").
- a query should not be in quotation marks or be preceded with notes like "please help with a translation", etc.

In addition to the above, there is a number of other "rules" which are more aptly expounded in the "Style guide" section.

1.1 Moderators

Responsible for enforcing the forum rules are the (per board and global) moderators. *Moderators* can edit posts only in the boards they moderate. *Global moderators* can edit any post throughout the forum. The role of the moderators is extremely important as they are the final "authorities" who decide on the appropriate translations by adding and removing them. The role of the moderator at times can verge on the role of a wiki editor, as there is an understanding between moderators who are actively engaged in posting terminology, that other moderators may edit

[5] http://www.translatum.gr/forum/index.php?topic=35.0

their posts and correct mistakes and typos or even add and remove translations.

2.0 Modification of SMF forum software

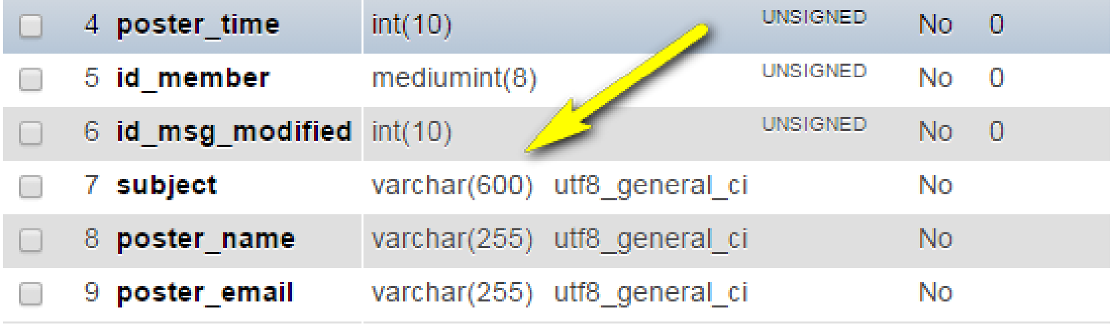
2.1 Php modifications

Php modifications are the most numerous. One of the most important modifications is increasing the subject field length (number of characters allowed). In order to ensure that, multiple changes had to be implemented in various files as well as in the **subject** field of the **smf_messages** MySQL table (see “MySQL modifications” below).

2.2 MySQL modifications

The use of UTF-8 is a very important issue in a multilingual forum. Although widely adopted in modern fora and CMS systems, during the first days of Translatum forum (i.e. SMF 1), UTF-8 was not officially supported. The disadvantage of this was that multibyte characters were saved and displayed as html mumerical entities, thus increasing database size and providing poorer search results. In order to remedy this, the database (in ASCII format) was exported and converted to UTF-8, and then reimported. Some issues appeared as the result of this, i.e. broken characters in certain parts of the forum. Full compatibility with UTF-8 was achieved with SMF 2.

The length of the subject field is extremely important as it affects the length of the terminological entry. The default is **subject varchar(255)** and it was increased to **varchar(600)**. The limit of **varchar** is 65,535 bytes but it is not advisable to have a very high value as it could affect database performance. In previous SMF versions (1 and 1.1) this field was **tinytext** with a limit of 255 bytes.



□	4	**poster_time**	int(10)		UNSIGNED	No	0
□	5	**id_member**	mediumint(8)		UNSIGNED	No	0
□	6	**id_msg_modified**	int(10)		UNSIGNED	No	0
□	7	**subject**	varchar(600)	utf8_general_ci		No	
□	8	**poster_name**	varchar(255)	utf8_general_ci		No	
□	9	**poster_email**	varchar(255)	utf8_general_ci		No	

2.3 Mods

Modifications or “mods” as they are known, are code packages which can extent or alter forum functionality and can be downloaded from the SMF Modifications list[6]. They change/add/remove the original source code by means of find/replace or add and remove operations. These are some of the mods installed:

1. Recent Topics On Board Index (the standard functionality is to display recent posts, thus possibly cluttering the forum “Info Center” with the same topic and its replies).
2. Share This Topic (a functionality for Facebook, Twitter and Google+ sharing)
3. Simple Audio Video Embedder (in order to autoembed youtube and other videos)
4. Smart Pagination
5. SMF4Mobile Mod (providing mobile functionality)
6. SMFPacks: SEO Pro Mod
7. Stop Spammer
8. TopicRenamer
9. WYSIWYG Quick Reply

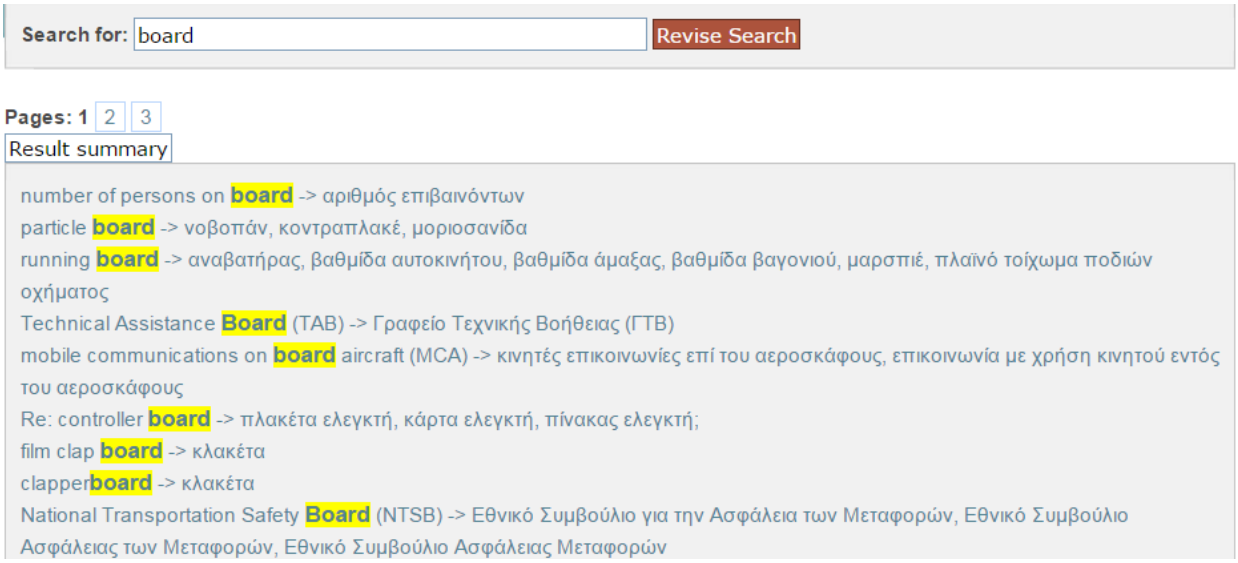
Some of the modifications though were custom-made specifically for Translatum forum.

[6] http://custom.simplemachines.org/mods/

2.3.1 Search enhancement

There is *Search enhancement* mod, which has some important bugs, like not functioning correctly when at least one of the returned results contains an apostrophe[7], and a custom mod which was developed to party emulate its behaviour.

Usually, search results appear in an expanded style in a forum, containing text from the subject as well as the body of the post. This is hardly ideal for a terminological search, where one needs to see relevant results quickly and succinctly. To this end, a “result summary” is displayed with the topic subjects only and the search term in yellow highlight. As subjects contain the term and its translation(s), the required level of succinctness is achieved.

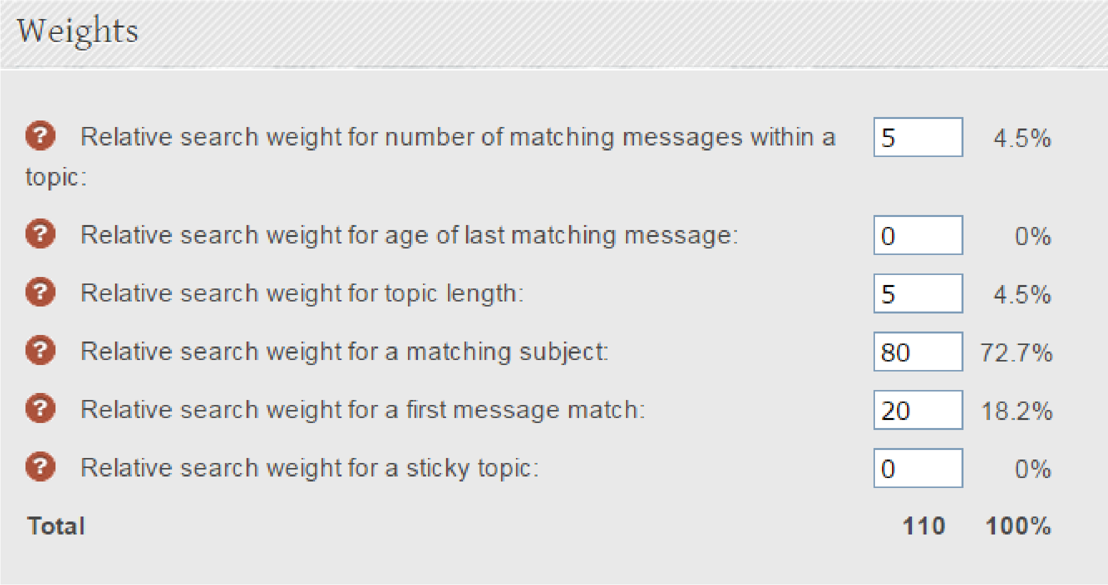


Search for: board

number of persons on **board** -> αριθμός επιβαινόντων
particle **board** -> νοβοπάν, κοντραπλακέ, μοριοσανίδα
running **board** -> αναβατήρας, βαθμίδα αυτοκινήτου, βαθμίδα άμαξας, βαθμίδα βαγονιού, μαρσπιέ, πλαϊνό τοίχωμα ποδιών οχήματος
Technical Assistance **Board** (TAB) -> Γραφείο Τεχνικής Βοήθειας (ΓΤΒ)
mobile communications on **board** aircraft (MCA) -> κινητές επικοινωνίες επί του αεροσκάφους, επικοινωνία με χρήση κινητού εντός του αεροσκάφους
Re: controller **board** -> πλακέτα ελεγκτή, κάρτα ελεγκτή, πίνακας ελεγκτή;
film clap **board** -> κλακέτα
clapper**board** -> κλακέτα
National Transportation Safety **Board** (NTSB) -> Εθνικό Συμβούλιο για την Ασφάλεια των Μεταφορών, Εθνικό Συμβούλιο Ασφάλειας των Μεταφορών, Εθνικό Συμβούλιο Ασφάλειας Μεταφορών

Another point is the way search results are ordered. By default, in fora, recent posts have a much higher “weight”. However, since this parameter is hardly of any importance when it comes to terminological data, the weights have been adapted with emphasis on the weight for a matching subject.

[7] http://www.simplemachines.org/community/index.php?topic=188513.msg3457985#msg3457985



Weights

Relative search weight for number of matching messages within a topic:	5	4.5%
Relative search weight for age of last matching message:	0	0%
Relative search weight for topic length:	5	4.5%
Relative search weight for a matching subject:	80	72.7%
Relative search weight for a first message match:	20	18.2%
Relative search weight for a sticky topic:	0	0%
Total	**110**	**100%**

2.3.2 Google button

One of the commonest ways to verify a linguistic hypothesis, especially when it is in a language which is not our mother tongue, is to use a Google search with that specific phrase. Normally, one would have to perform the Google search (using quotation marks if it is a phrase search), and then paste the (long) URL in the forum. However, this is a copious procedure given its high frequency.

A typical feature of a forum is BBC buttons (Bulletin Board Code[8]) which is a way to easily add formatting tags in a non-WYSIWYG environment. A custom mod was developed which adds a Google button to the available BBC buttons when a user posts. The functionality is simple: select the word or phrase and click the button. This will automatically create a **[google]**Google search**[/google]** for that word or phrase (in bold the actual tags). To enforce phrase search, one has to add first quotation marks to that phrase.

[8] https://en.wikipedia.org/wiki/BBCode

2.3.3 URL trimming

Given that long URLs are often pasted in the forum, if no trimming was applied, the readability of many posts would decrease. To that end, a CSS-only mod was developed that would trim long URLs. The benefit of this approach, instead of affecting the actual code that produces the output, is that it does not add any delay and when URLs are copied from the forum, the full URL is pasted.

2.4 Server modifications

Due to the inherent limitations and performance and speed of MySQL search, a server-based approach is recommended for large fora (i.e. fora with more than 750,000 posts). Sphinx[9] search engine is supported by SMF, provided that the forum is hosted at the very least in a virtual private server[10] (VPS), and it is being considered for a future implementation on Translatum forum[11].

3.0 Handy SMF features

As I mentioned in the **Introduction**, a forum is hardly the ideal medium for terminology management. However, there are certain SMF features that often missing from such dedicated terminology management environments.

One of the most important feature is *inline editing*[12]. Inline editing, using Ajax[13] technology, allows the user to click on editable element on the page and edit it in-place (that is without having to wait for a new page to load). This functionality is implemented in

[9] http://sphinxsearch.com

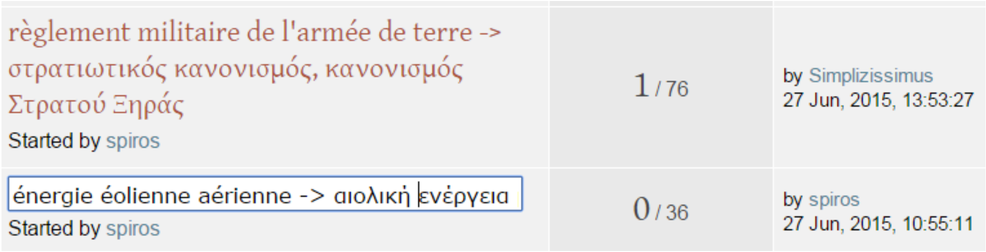
[10] https://en.wikipedia.org/wiki/Virtual_private_server

[11] A similar search tool, Lucene, is already in use for Translatum's wikis. Refer to the paper "Όταν το LSJ γνώρισε τη Βίκυ", 9th Hellenic Language and Terminology Conference, 2013 (http://www.eleto.gr/download/Conferences/9th%20Conference/Papers-and-speakers/9th_23-05-01_DoikasSpyros_Paper_V05.pdf).

[12] The following videos demonstrate this function: https://youtu.be/l7F-HELDGQQ and https://youtu.be/RfgZe1A79r4

[13] https://en.wikipedia.org/wiki/Ajax_(programming)

SMF’s board index where all the topics are listed one below the other. By double clicking in the box that contains the text (not in the text itself), the text becomes editable. Then one can make the desired changes, and save by clicking outside the box. This feature allows the quick editing of multiple entries and the workflow is extremely fast, i.e. if compared to a wiki workflow. If inline editing was implemented in search results, then it would have been even more useful.



règlement militaire de l'armée de terre -> στρατιωτικός κανονισμός, κανονισμός Στρατού Ξηράς Started by spiros	1 / 76	by Simplizissimus 27 Jun, 2015, 13:53:27
énergie éolienne aérienne -> αιολική ενέργεια Started by spiros	0 / 36	by spiros 27 Jun, 2015, 10:55:11

4.0 Style guide

Below is the style guide governing the submission of terminology questions and their editing by a moderator/user once the translation has been finalised.

4.1 Delimiters

Delimiters are important for future processing of terns.

Normally, the source/target delimiter is “->” (hyphen plus greater-than sign) and the translations delimiter is the comma “,”. However, there are some exceptions.

Multiple translations are normally separated by a comma. In case one of the proposed translations includes a comma, then the vertical bar is used in all translations (“|”) as translation delimiter.

(Not necessary in case, for example, of translations including "ό,τι", as there is no space after the comma).

Another way (complimentary to the aforementioned or to be used when there is only one translation and hence, not possible to use the pipe translation delimiter) to filter this out is to use an n-dash plus greater-than sign (–>) instead of hyphen plus greater-than sign as source/target separator. This is a way of pointing out some non-standard usage to the person who will do the term filtering. I.e. in the following example entering a vertical bar is irrelevant (one source term, one translation):

day in, day out –> μέρα μπαίνει, μέρα βγαίνει

Terms (source text) which include a comma is another way of filtering entries which do not abide to the comma separator rule. Hence, there should never be two terms in the source segment, but either broken in two topics of the specific language pair, or reverse the language pair, if the equivalent translation is only one term. For example, instead of:

foot, leg -> πόδι [in English-Greek pair]

use

πόδι -> foot, leg [in Greek-English pair]

When the poster has an idea about possible translation(s), then it should be entered in the subject, after the term. I.e.

βαθμολογική εξέλιξη -> promotion in grade, grade advancement;

Full translations of the terms should be used rather than parts of the term in parenthesis. I.e. instead of:

βαθμολογική εξέλιξη -> (grade) advancement use

βαθμολογική εξέλιξη -> advancement, grade advancement

4.2 Special use of Greek question mark

When the subject ends with a Greek question mark (;) then it is an indication that the translations given are not reliable or that they are tentative. The Greek question mark is also used to indicate a question, when one enters one or more possible translation equivalents, to assist the answerers. The Greek question mark is not necessary when only the source term is entered. Similarly, use only the Greek question mark to indicate this and not the English one.

Do not use a question mark (Greek or English) for finalised translations (resolved questions), even if the phrase is a question, as the question mark will be used to filter out all inconclusive translations.

Also, appropriate spacing should be used. For example, there is no space before comma and there is always ONE space after the comma (unless the comma is a decimal separator).

4.3 Initialisms

Initialisms:[14]

1. Are used only when they are accepted initialisms of the term language

[14] http://www.translatum.gr/forum/index.php?topic=218935.msg415559#msg415559

2. May contain a combination of upper and lower case letters (i.e. mRNA)
3. Fullstops are to be avoided (i.e. “OTE” rather than “O.T.E.”)
4. When initialisms contain (capital) letters which can be typed in Greek or English with the same visual effect, we make sure we use the term language
5. When there are two or more initialisms, they are separated with a comma, i.e. Reverend (Revd, Rev.)
6. When initialisms are not representative of the full term, but only of a part thereof, then they better be avoided. For example in “radiografia toracica postero-anteriore (PA) -> οπισθοπρόσθια ακτινογραφία θώρακα”, “PA” is an initialisms for “postero-anteriore” and not the full term.
7. The initialisms are listed with their expansions. The expansion comes first and then the initialisms follow in parenthesis.
8. The use of parenthesis is strictly for initialisms and other uses are discouraged (i.e. part of speech, register, linguistic variant, etc).

5.0 Import, export, update and maintain entries

5.1 Import

Entries can be edited and prepared offline and then posted one by one manually. However, this is time-consuming. To attend to this need, an import script has been developed which required a 4 or 5 column tab-delimited format. The columns are 1) username 2) subject 3) body 4) user ID and, optionally, 5) board ID. In other words each term can be imported on a different board on the basis of the board ID contained in column 5.

To run this script, a cron job (scheduled task) has to be created and configured to run on set times and intervals. “Sleep” functionality has been added to the script to make it look as if it was posted by a human by randomly delaying the posting time of each term based on a given delay range.

5.2 Export and mass editing

As the number of terms grew in the forum, a number of issues started to affect it. One of the most important being duplicates. Mass editing involves exporting the database to a tab-delimited format and then performing a number of maintenance tasks like removing duplicates, adding translations to entries, case fixing (different rules for different languages, i.e. in French in names of organisations only the first letter should be capital).

To export the terms a MySQL query is used that provides a csv export with poster ID, subject, topic ID and board ID.

5.3 Excel manipulations

In order to process the terms in Excel, a number of tasks has to be performed.

1. import the csv export (mentioned in previous section) into Excel[15]
2. normalise apostrophe variants (like **&#39;** and **&#039;**) to a standard symbol
3. replace the "->" and "–>" delimiter with a tab character
4. delete duplicate and leading/trailing spaces (using ASAP Utilities)
5. create an extra column with source text (optionally, separate initialisms found in parentheses in an extra column in order to locate duplicates with or without the initialism)
6. change all text of that column to lowercase (and, in languages like French for example, remove all accents in order to identify errors in accent usage)
7. use a duplicate manager like Duplicate Master[16] to colour duplicate cells of that column

[15] Since Excel fails to convert UTF-8 as exported from phpMyAdmin, I had to use the following regex pairs to manually convert to tsv: **Find:** ";" **Replace:** \t **Find:** "\n" **Replace:**\n

8. use Excel filter functionality to sort on coloured cells

At this stage the duplicates have been identified and the second stage deals with actions directly on the forum.

5.4 Forum manipulations

The following courses of action can be taken for forum duplicates:

1. Reverse the entry (i.e. change "car -> αυτοκίνητο" to "αυτοκίνητο -> car") and reverse the language pair (by moving to the reverse language pair board). This implies that it has been confirmed that there is no such entry in the other board (i.e. "αυτοκίνητο").
2. If there is already an entry in the reverse language pair board, then change one of the pairs and move to a different board altogether. I.e. if "αυτοκίνητο" already exists in Greek>English then move to Italian>Greek as "auto -> αυτοκίνητο". An easier workflow would be to first reverse the language pairs of topics, *en masse*, and when that is done, sort on Subject and, if the languages use different alphabets, then it would be easy to select them via check boxes in board index and move them *en masse* to the reverse pair board.
3. When two or more duplicate entries have a lot of replies, then it is a good idea to simply merge them, so that people in the future get more out of the topic. When merging, the subject should be amended in order to contain all the translations available in both topics. The down side of merging is that the newest topic ID(s), when accessed via the Internet, will display an error page (since the topic no longer exists). Merging can be done either by selecting the check boxes after a search, or by adding the topic ID of the topic to be merged.
4. Another possibility is to change from singular to plural or vice versa.

[16] http://www.translatum.gr/forum/index.php?topic=49093.0

5.5 Mass updating

Sometimes one needs to merge translations from an offline glossary with the forum contents. The first step in order to do that is to follow the procedure described in “Excel Manipulations”. Then:

1. replace the translations delimiter with “|”
2. add in the same spreadsheet the offline glossary, following the same delimiter conventions
3. Sort the spreadsheet on source column
4. run a bespoke macro which will merge duplicates, add extra translations if they exist, while maintaining the topic ID in the last column
5. merge source and target column[17] ending up with a spreadsheet with a subject column and a topic ID column.
6. save as UTF-8 tab-delimited text
7. upload to server
8. run a bespoke php script which batch updates the subjects of the topics listed in the txt file.

6.0 Conclusion

We have seen from the foregoing that a great deal of custom work has gone into making SMF more amenable to terminological use. That does not mean that there is no space for improvement. Much more can be done. For example, implement server-level Sphinx search to increase search speed, provide autocomplete and correction suggestions functionality[18]; enable inline editing in search results; overhaul the way subjects are stored in database level to allow for source and target fields and thus a higher degree

[17] I.e. by using ASAP Utilities’ “Merge column data” option.

[18] See: http://sphinxsearch.com/blog/2013/05/21/simple-autocomplete-and-correction-suggestion/. Autocomplete functionality is provided on the subject when creating/editing a post with the use of a the “Related Topics” and “Similar Topics” mods (Nos 189 and 3473 respectively). However, these mods are not compatible with Sphinx and they were not used in the forum given the plans to implement Sphinx in the future.

of flexibility in search output; provide a user interface for the export of terminology subsets in csv, xml or MultiTerm xml format.

Ultimately, the terminological data contained in the forum will be exported, revised and merged with other data and be made available in a terminological platform like mediawiki. However, the numerous discussions that have taken place in the forum should by no means be considered as a useless trouble; they play a major role in highlighting potential pitfalls and determining the preferred translations. In that sense, a terminological forum can be considered a workshop in applied translation, providing valuable insights into the hidden depths of the translation process and opportunities for creative synthesis of apparently discordant views.

Spiros Doikas
http://translatum.gr/cv.htm

Bio

Spiros Doikas read English at Manchester Metropolitan University. His postgraduate studies include Machine Translation at UMIST and IoL's Diploma in Translation (EN>EL). He has been working as a translator since 1995, initially as a literary translator and then as a technical translator specializing in software and IT. He has a keen interest in translation technologies and has been teaching translation tools and localization in meta|φραση School of Translation Studies since 2003. His research interests and skills include multilingual web site development, online terminology management systems, wiki and forum software. In 2001 he created Translatum.gr, a Greek translation portal providing, among other things, terminological assistance in a customized version of an open source forum platform. He is a member of the Administrative Board of the Hellenic Society of Terminology.

Βιογραφικό

Ο Σπύρος Δόικας είναι κάτοχος του πτυχίου Αγγλικής Φιλολογίας του Manchester Metropolitan University και του Diploma in Translation του Chartered Institute of Linguists. Έχει κάνει μεταπτυχιακές σπουδές στη Μηχανική Μετάφραση στο UMIST και έχει μετεκπαιδευτεί σε θέματα μεταφραστικών τεχνολογιών στο Πανεπιστήμιο του Βοσπόρου. Εργάζεται ως επαγγελματίας μεταφραστής από το 1995, αρχικά με εκδοτικούς οίκους και στη συνέχεια ως ελεύθερος επαγγελματίας μεταφράζοντας τεχνικά κείμενα με ειδίκευση στην πληροφορική και στις εφαρμογές μεταφραστικής μνήμης. Διδάσκει εργαλεία μεταφραστικής τεχνολογίας στη σχολή meta|φραση. Ασχολείται επίσης με την ανάπτυξη ιστότοπων (με έμφαση στο λογισμικό ανοιχτού κώδικα, το δυναμικό περιεχόμενο και την πολυγλωσσικότητα). Το 2001 δημιούργησε το Translatum.gr, την ελληνική πύλη για τη Μετάφραση. Είναι μέλος του Δ.Σ. της Ελληνικής Εταιρείας Ορολογίας.