Special punctuation handling

Since Tiki 2 (?) punctuation at the end of translatable strings is treated as a special case. For example, the string "Login:" is split into two strings, "Login" and ":", so that translators can translate the punctuation mark only once and we can have only one translation for "Login:" and "Login", thus reducing the amount of work.

This page is to discuss the pros and cons of this system and eventually decide if we should keep or remove it.

Statistics

Trunk, as of September 2011, has a total of 13000 strings. From those, 3018 end with a punctuation mark and are displayed to translators without it. Only 308 are duplicated (a version with and without punctuation mark at the end exist).

Problems of the current system

- For strings with more than one sentence we end up having a string that looks wrong in the language files because they miss a period at the end. For example: "Database `%0` creation failed. You need to create the database". Translators need to remember to skip the last period when working or the translation will be displayed with two ending periods.
- Sometimes the lack of the last punctuation mark makes it harder for translators to understand the meaning of the sentence. For example, the sentence "URL already added to the directory. Duplicate site?" appears for translators without the question mark.
- For Spanish (and maybe other languages as well) it is specially hard to translate the string shown above because translators have to guess it is a question, add an inverted question mark in the beginning of the question and leave to the software to add the question mark at the end. The guessing if a sentence is a question or not is for German translators a problem, too.
- Strings ending in ellipsis appear to translators with only two dots. "Saving..." becomes "Saving.."
  - Unicode has a character for the ellipsis: ...
    Using that character instead of three full stops would solve this problem. Chealer
- It messes statistics because translated strings with the ending punctuation are kept resulting in a different number of total strings for each language. For example, in the German language file both "Required for podcasts." and "Required for podcasts" are present while in the Brazilian Portuguese language file only "Required for podcasts" is present.
- What else?

Advantages of the current system

- Reduces the number of strings that need to be translated, by about 2% according to the above statistics.
- In French language, it reduced the number of "& nbsp;" in the language file.
  These are used because a lot of punctuation marks are preceded by a non-breaking space - in French (?) - real utf8 non-breaking spaces are difficult to insert, depending of the text editor, so translators...
would write "&nbsp;" which only works in web pages, not mails. As a "solution", extracting the punctuation from the strings and having one correct translation for each punctuation (using the correct utf8 non-breaking space) seemed easier than educating translators.
I do not know if this benefits any other language than French 😒 Any insight?

- This distinction is also important in JavaScript, not just in mails.
- Probably only a minority of translators would even put an HTML non-breaking space. Most would probably put a normal space, or even no space at all... Chealer

**Conclusion**

It was an interesting and promising idea, but now that we have experimented it a lot, we see that the problems are more numerous and more important than the advantages.

We need to remove this system.

How do we do that?