Word count analysis

Word count analysis depend on the definition of “word”, which for the purposes of word count analyses of source texts in Latin alphabet is considered to be a unit of text between word boundaries (e.g. spacing and punctuation symbols). However, different tools will use different parameters to count words, so it's not realistic to expect reports from different sources to be always identical. If you add tags (e.g. <b>) or escaped tags (e.g. <b>) to the equation, this gets more complex. For example:

- **memoQ** has a simple approach: it simply counts as words chunks of text between spaces (\b|\s|^|$) and does not consider escaped tags as unitary entities.
- **OmegaT** counts escaped tags as unitary entities but for example considers the Saxon genitives as independent words.
- **Rainbow** ignores escaped tags and considers Saxon genitives as a suffix, thus as part of the preceding word.

Therefore, the following string:

```
wall’s&lt;br /&gt;
```

will yield totally different counts in the three tools considered above:

memoQ	**OmegaT**	**Rainbow**
2 words	3 words	1 word
wall’s 	wall ’s 	wall’s

Other characters might differently estimated too. The example above does not mean to be comprehensive.