You Can't Fix What You Can't Measure

By Hans Fenstermacher

In the localization business, we live and die by metrics. Every step of our work is carefully measured, whether it's price, word counts, engineering hours, number of pages, or percentages of leverage from translation memory. Costs and resources are allocated precisely to match those metrics, and we are regularly challenged to substantiate the business case for localization services through objective and verifiable metrics.

Whether they know it or not, technical writers and content authors are in the same boat. The fewer metrics technical writers have, the leakier the boat will be—and no amount of vigorous rowing or protesting against offshoring will keep the boat from sinking.

Only by playing the game according to the financial rules can writers expect to win any of the daily battles over budgets. Corporate financial decision makers know little (and care nothing) about usability, user experience, and other "intangibles"; their thumbs-up or thumbs-down is based on things they can measure, like money. Therefore, writers must adopt content metrics that objectively measure content on a more granular level, so they can make smarter decisions about content globalization (optimizing content for the most cost-efficient localization).

Localization's most meaningful metric

As noted earlier, in contrast to the content authors, we in localization collect lots of metrics. Most are static, like cost, time, and volume (word count, page count, graphics count). Also there are comparative metrics, like translation memory leverage from project to project, and normative metrics, like budget and turnaround time. But the value that localization business cases typically must establish rests mostly on the ratio between two sets of well-known metrics: costs versus benefits. To optimize value, we can either reduce costs or increase benefits, or both.

No other metrics have proven so influential. While the benefits of improved, fully globalized content are obvious to those of us who produce it, it's tough to come up with hard metrics on usability, user experience, and the effects of having localized content (versus unlocalized). Even if we manage to assemble metrics in these areas, they likely fall on the deaf ears of corporate decision-makers. Let's face it, the most convincing metric we have at our disposal is cost—and the less of it the better.

As a result, in the localization services business, cost reduction is a non-stop objective and usually the top priority. Strategies to reduce multilingual costs abound, and they always represent a tradeoff between the cost reduction

Hans founded ArchiText® in 1994 after a successful 15-year career in technical communication and translation, including experience with some of today's industry leaders.

He started a different kind of company because he realized that localizers were really in the same business as their clients: building good communications with customers all over the world. Since then, ArchiText has grown into a well-established provider of localization and multilingual services to Fortune 2000 companies.

ArchiText developed ABREVE® in response to an expressed need for new ways to reduce localization costs while preserving the integrity of content. ABREVE® addresses these issues directly while improving content usability.

ABREVE® is a patent-pending methodology that enables low-risk step changes that can provide drastic localization cost reductions.

itself and its likelihood of success, which applies on several levels: ability to actually achieve the reduction, quality of the outcome, and process performance. The best strategies score highly on both fronts, cost reduction and performance, but these are rare.

One success story in our business is translation memory (TM) technology. It has demonstrated over time both high performance and cost reductions. Localizers have been using TM for years to provide static and comparative metrics, and these, in turn, have been driving decision-making on the buyer side.

But the vast majority of buyers have yet to take real advantage of TM as a normative and probative metric for multilingual content development. By internalizing TM metrics on the source side of content, technical writers could significantly enhance the business cases for their content globalization efforts.

Content building blocks

Translation memory tools filter source content through a digital prism that parses content for localization. When content is broken down into granular building blocks, it can be measured more effectively. The building blocks then become powerfully prescriptive and normative for establishing both cost and cost savings in localization.

Each TM building block is associated with a different level of localization cost, with new words being most expensive and "100% matches" being least expensive. A *100% match* is a building block that exactly matches other content that is already written and translated. It matches 100 percent. Other levels of match percentages also exist: *90% matches*, *75% matches*, etc. These are referred to as "fuzzy matches" because they are not exact matches. But they still offer a degree of leverage and, thus, a lower cost than a new word.

The cost of a *100% match* is relentlessly being driven down, but there is almost always still some cost associated. Only one type of content truly costs nothing: content that isn't there. Eliminating content is the most effective cost-savings strategy of all.

Thus, from a cost-savings perspective, the most effective content-globalization strategies are these:

- Create content entirely from *100% matches*.
- Have no content at all (don't laugh, it's possible).

Lowering costs through content globalization

With these extremes in mind, technical writers can actually build a sound multilingual-content business case by using the content building blocks. A case study using real data demonstrates how this works.

Suppose a software maker is about to release a long-awaited version 2.0 in multiple languages. Like most documentation sets, this release is a combination of legacy content (from version 1.0), new documentation (added features), and edits (corrections, changes, arbitrary rewrites). Translation memory analysis on the English documentation (comparing version 1.0 content to version 2.0 content) shows the content building blocks for the version 2.0 content to be as follows:

Original			
Cost	**Building Block**	**Words**	**Perc.**
$$$$	No Matches (new words)	47,800	41%
$$$	Low Fuzzy Matches	11,785	10%
$$	High Fuzzy Matches	17,677	15%
$	Exact Matches (100%)	23,450	20%
$-$$	Internal Repetitions	17,138	14%
	Total	**117,850**	

These building blocks now form the basis of a content globalization strategy with three goals:

- Increase the amount of *reused* words in the English source.
- Decrease the number of *new* words in the English source.
- Maximize the consistency of the words in the English source.

To increase reuse, the main focus is on exact matches (*100% matches*). Where can more of them come from? The most likely places are the "High Fuzzy matches" because, by definition, they are very close to exact matches. By strategically editing the High Fuzzy segments, they can be shifted to *100% matches*.

Next, the aim is to reduce the amount of "no matches" (new words) as much as possible. Since there is no TM to be concerned about here, these words can be freely rewritten and reduced. Some of the reduction in new words will inevitably reduce "internal repetitions" (exact matches within the new words), which are less expensive than new words. And remember, the only words that cost less than reused words are words that aren't there at all.

Finally, the content globalization strategy aims to maximize the consistency of the remaining content by increasing the number of internal repetitions and shifting Low Fuzzy matches to High Fuzzy matches wherever possible. The prime motivation is always to turn higher-cost content building blocks into lower-cost ones. Here are the metrics post-globalization:

Post-Globalization			
Cost	**Building Block**	**Words**	**Perc.**
$$$$	No Matches (new words)	23,937	27%
$$$	Low Fuzzy Matches	11,525	13%
$$	High Fuzzy Matches	2,660	3%
$	Exact Matches (100%)	37,236	42%
$-$$	Internal Repetitions	13,299	15%
	Total	**88,657**	

Did the strategy work? And how! Exact matches have more than doubled from 20 to 42 percent; no matches are down from 41 to 27 percent; internal repetitions and Low Fuzzy matches are up slightly, from 14 to 15 percent and from 10 to 13 percent, respectively.

Still, the proof of the pudding, as they say, is in the eating. Let's examine the effect of this globalization strategy on cost reduction, which was the overall goal to begin with. Using identical tiered per-word pricing for the content building blocks, the following table compares localization costs for an average language between the original and post-globalization contents:

Building Block	Original Cost	Post-Globalization Cost
No Matches	$11,950	$5,984
Low Fuzzy Matches	$1,768	$1,729
High Fuzzy Matches	$1,768	$266
Exact Matches	$586	$930
Internal Repetitions	$1,714	$1,330
Total	**$17,786**	**$10,239**

This comparison shows a reduction in cost not only overall, but in virtually every building block category. Why? Because the globalization strategy used detailed

Hans is also founding Chairman of the Board of the Globalization and Localization Association (GALA).

Founded in 2002, GALA is a fully representative, non-profit, international industry association for the translation, internationalization, localization, and globalization community.

metrics to target both content reuse and volume reduction, simultaneously. This approach also creates a useful aggregate metric we can call the "cost per leveraged word." That is, the overall cost per word, taking into account the discounted localization pricing for various types of building blocks as well as the total number of words. In this case the cost per leveraged word dropped from 15.09 cents to 11.55 cents.

Toward better content metrics

Many localization cost-savings strategies are based on one or two isolated metrics. Such approaches are flawed, because they ignore the effects of changing one or more building blocks in content. For example, a singular focus on overall word count assumes that all content types are equal. However, we have just seen that some types are, well, more equal than others. Cutting the number of words indiscriminately—whether it's removing certain content subsets in the localized versions or chopping text—runs the risk of destroying significant value that can be easily afforded with low-cost, reused words. On the other hand, a pathological fear of "damaging" translation memories and the consequent unwillingness even to touch legacy or partial-legacy content can also leave money on the table (as our case study showed in terms of the High Fuzzy matches).

A content globalization strategy based on content building blocks and on granular metrics represents a better approach. In terms of the case study here, indiscriminate cutting might have left most of the highest-cost new words intact, while removing only the lowest-cost exact matches. And leaving the content untouched would have meant overspending for localization. Instead, even though the original content contained a substantial amount of leveraged words in translation (over one-third was either exact matches or High Fuzzy matches), careful globalization work on the building blocks produced the optimal results—and, incidentally, a much higher-value TM for future releases.

Every piece of content is different, and results will vary, but a content globalization approach based on these granular building blocks and metrics promises a far better localization value than virtually all other strategies.

About the Author

Hans Fenstermacher is President and founder of ArchiText, a language service provider. He has developed ABREVE, a methodology for reducing volume and globalizing content. Contact Hans at hansf@architext-usa.com.

Reduce costs by 20% or more right now.
Ask us about:

ABREVE®

- Consulting
- Training
- Services

only from

ArchiText®

800.688.7205

www.architext-usa.com