



Understanding the Complexities of CEEC Languages:

Translation After 1 May 2004

By Joe DiDamo

May Day 2004—Accession of the Central and Eastern European Countries (CEECs) to the European Union (EU) has arrived, and so has the need to adapt to the changes that are no longer nigh but here to stay. Over the past several years, these 10 new members of the EU, eight former Soviet-bloc countries and two Mediterranean islands—Cyprus, Czech Republic, Estonia, Hungary, Latvia, Lithuania, Malta, Poland, Slovakia and Slovenia—have been transposing a large number of EU directives for the medical device and pharmaceutical industries.

Organizations such as the Pan European Regulatory Forum (PERF) were established to provide to the countries' competent authorities and stakeholders a central forum to help establish and manage the seemingly overwhelming processes and complexities that accompany EU accession.

Manufacturers in the highly regulated life sciences industries,¹ whether headquartered close by in neighboring countries or more remotely in the Americas or Asia, are compelled to manage fundamental shifts in their approach to the "new," expanded Europe.

In this article, this complexity is addressed by providing information about the individual countries and market conditions, as well as some suggestions (good practices) on how to manage the translation of content for regulatory requirements and market acceptance. Also discussed are models that manufacturers have set up to manage a quality localization process.

Acquis Communautaire

The BBC News Euro-glossary defines *acquis communautaire* as follows: "The entire body of European Laws is known as the *acquis communautaire*. This includes all treaties, regulations and directives passed by the European institutions as well as judgments laid down by the Court of Justice. They [candidate countries] must adopt, implement and enforce all the *acquis* to be allowed to join the EU."

So what has this meant to the once-candidate, now new-member states? A huge task. In the years leading up to 1 May 2004, the new member states have had to adapt, adopt and harmonize the regulatory requirements to be in compliance. They have then had to be prepared to enforce EU regulations such as *Conformité Européenne* (CE) Marking and directives like the Medical Device Directive and the In Vitro Device Directive. Certain countries were more prepared for this tran-

sition than others were. Some countries have taken local language requirements to a new level of intensity after years of seeming indifference or inability. The existing member states, having blazed this trail years earlier, have helped ease the sometimes-painful transition.

So what does this mean to the life sciences industry? Time and money. Although most of the CEECs have had some level of local language requirements in place in the past, the industry as a whole had not been strongly compelled by local governments to deliver local language content (labeling, prescribing information, user manuals, etc.), nor had it been sanctioned even when required by pre-accession local laws. As of 1 May 2004, all this has changed. The new member states have a new foundation, a massive EU political vehicle and greater technical and organizational infrastructure to enforce language compliance.

Table 1. The 10 New EU Countries

Country	Language	Language Family	Population (hundreds of thousands) ^a	GDP (billion USD) ^b	GNI per capita (USD) ^{b,c}
Cyprus	Greek, Turkish	Indo-European, Greek, Attic (Greek)	772	10.1	12,320 ^(d)
Czech Republic	Czech	Indo-European, Slavic, West	10,249	69.5	5,480
Estonia	Estonian (official)	Uralic, Finno-Ugric	1,409	6.5	4,190
Hungary	Hungarian	Uralic, Finno-Ugric	10,045	65.8	5,290
Latvia	Latvian	Indo-European, Baltic, Eastern	2,349	8.4	3,480
Lithuania	Lithuanian	Indo-European, Baltic, Eastern	3,593	13.8	3,670
Malta	Maltese, English	Afro-Asiatic, Semitic (Maltese)	400	3.9	9,260
Poland	Polish	Indo-European, Slavic, West	38,623	189.0	4,570
Slovakia	Slovak	Indo-European, Slavic, West	5,430	23.7	3,970
Slovenia	Slovenian	Indo-European, Slavic, South, Western	1,936	22.0	10,370

Notes:

^a Population data from CIA World Factbook. ^b GDP/GNI data from 2002 World Bank estimates, except where noted.

^c GNI is Gross National Income, as calculated in accordance with the World Bank Atlas method. ^d GNI 2001 World Bank estimate.

What Are the CEECs?

There are many views and definitions of what constitutes the region of Central and Eastern Europe, just as there are many views about where Europe actually starts and ends. According to the most accepted definition, which is based on geographical and political sentiments, the CEECs include all European countries east of Germany and south to the Balkans. The region is made up of a richly diverse population, encompassing over 20 national states and a population of more than 345 million people speaking a mixture of Indo-European and Uralic languages in addition to English (Malta) and Greek and Turkish (Cyprus²). The future of the EU may include additional accession countries.

Table 1 provides a quick snapshot of the 10 new countries that joined the EU on 1 May 2004.

Although these countries have joined the EU, integration efforts are far from finished. Full integration of the 10 new countries is likely to take several years. Meanwhile, the EU has already started negotiations with additional countries. Bulgaria and Romania are likely to join the EU before 2010, the EU is in talks about potential accession with Turkey and it has already indicated that it will accept the Balkan countries. The accession of Belarus, Russia and Ukraine is not currently under consideration, but, as in many other cases, what is unthinkable today can become quite feasible tomorrow.

The people of the CEEC region

speak Slavic languages (Czech, Polish, Slovak, and Slovenian), Baltic languages (Latvian and Lithuanian) and Finno-Ugric languages (Estonian and Hungarian), as well as Greek (Attic) and Turkish (Altaic). Although the Slavic languages have common roots and share a number of similarities (and are still mutually understandable to some extent), each language possesses specific characters and diacritics (accent marks). Most languages use Latin scripts, but there are new issues and levels of complexities for corporate content publishing in traditional media and the Web.

Publishing for the CEEC Languages

The CEECs bring rich and varied languages and scripts to the European

Table 2. Other Key CEEC Economies

Country	Language	Language Family	Population (hundreds of thousands) ^a	GDP (billion USD) ^b	GNI per capita (USD) ^{b,c}
Albania	Albanian (Tosk is the official dialect)	Indo-European, Albanian	3,582	4.8	1,450
Bulgaria	Bulgarian	Indo-European, Slavic, South, Eastern	7,538	15.5	1,770
Bosnia and Herzegovina	Croatian, Serbian, Bosnian	Indo-European, Slavic, South, Western	3,989	5.6	1,310
Belarus	Belarusian, Russian	Indo-European, Slavic, East	10,322	14.3	1,360
Croatia	Croatian	Indo-European, Slavic, South, Western	4,422	22.4	4,540
Macedonia, FYR	Macedonian, Albanian	Indo-European, Slavic, South, Eastern (Macedonian)	2,063	3.8	1,710
Moldova	Romanian (Moldovan)	Indo-European, Italic, Romance	4,400	1.6	460
Romania	Romanian	Indo-European, Italic, Romance	22,272	45.7	1,870
Russia	Russian	Indo-European, Slavic, East	144,526	346.5	2,130
Serbia/Montenegro	Serbian, Albanian	Indo-European, Slavic, South, Western (Serbian)	10,656	15.7	1,400
Turkey	Turkish (official)	Altaic, Turkic, Southern	68,110	183.7	2,490
Ukraine	Ukrainian, Russian	Indo-European, Slavic, East	48,055	41.5	780

Notes:

^a Population data from CIA World Factbook. ^b GDP/GNI data from 2002 World Bank estimates, except where noted.

^c GNI is Gross National Income, as calculated in accordance with the World Bank Atlas method.

community (see **Table 2**). Many producers will have little experience with these new languages and scant legacy materials on which to draw. They will be relieved to know, though, that the tools and processes to support translation into these languages are by now already fairly streamlined. This is because, although these countries are new to the EU in 2004, over the last 14 years they have become important consumer and information technology

markets that have long required localized³ products and content.

As a result, for instance, all the major computer operating systems, such as Microsoft Windows, Apple Macintosh, Sun Solaris and Linux, already support CEEC languages. For many of the CEEC languages, the user interface and support documentation of these vital operating systems have been localized into the local language, including complete regional and lan-

guage settings, font, and keyboard and display support. In a similar vein, major word-processing and desktop publishing software applications work easily with CEEC languages. These software programs are either localized in local-language versions or released with an English user interface as “CE editions” with extended support for CEEC languages for fonts, import filters, hyphenation, spell-checking modules and so forth (e.g., Adobe Photoshop CE, Adobe Acrobat CE).

At the same time, however, potential marketers for the region would be ill advised to ignore the many complexities that accompany support for the Central European or Baltic languages. Fonts are one example. Most fonts that are used in product documentation will work correctly, but some will not, simply because they are not available in the CEEC encoding, and so font substitutions may be required. It would also be wrong to assume that there is one font encoding that covers all the CEEC languages; there are many:

- Central European—for Czech, Polish, Hungarian, Slovak, Slovenian, Croatian, Romanian, Serbian (Latin) and Albanian
- Cyrillic—for Russian, Ukrainian, Bulgarian, Serbian (Cyrillic), Macedonian and Belarusian
- Baltic—for Estonian, Latvian and Lithuanian
- Greek—for Greek
- Turkish—for Turkish

Additional or supplemental fonts, such as Arial CE and Arial Cyrillic, may be required by publishers. The good news, however, is that these encodings work as an extension of the existing Western European characters sets, meaning that they contain all the routinely used non-CE characters.

Standard desktop publishing processes may fail when working with CEEC languages. There are, for instance, frequent problems with printing CEEC characters in PostScript or Portable



Document Format (PDF) files, and a myriad of modifications are needed to set up the printing environment, printer drivers or operating system. These problems are due to the varied levels of character display (Unicode or CE) support. In some cases, there may be a need to revert to previous versions of desktop publishing applications or operating systems to produce a CEEC document with a correct display of extended characters. Another known issue concerns the use of bookmarks in PDF files, which may not correctly display CEEC characters. Manual corrections or a work-around may be required.

Notwithstanding any technical issues, there is the need to cater to specific language typographical rules. Whereas each language has its own rules and best practices, the CEEC languages have many typographical differences and exceptions—more than can be found, for instance, between English and German or French. **Table 3** provides examples of how quotation marks and numeric thousands and decimal separators are used for CEEC languages. Users of US English are accustomed to open and closed quotation marks (“like this”) and to using a comma for a thousands separator and a decimal point for a decimal separator (1,234.56). In CEEC countries, each language has its own combination of rules.

Although the process of publishing for CEEC languages is by now a fairly well-trodden path, it is not without its issues. A number of aspects should be considered to help save money and time and to prevent embarrassing mistakes.

Corporate Compliance—or, What to Translate

Deciding what and how to localize in terms of corporate content continues to be an important and potentially complex process. Some organizations consider the exercise a costly nuisance, whereas others consider it an investment. Ideally, the strategy to localize

should be driven from the top down, originating at the highest levels of the company. Industry leaders and CEOs who think and act for a global market will generally have a leg up in enabling the departments that are most directly affected in their organizations to adopt a smooth and efficient process. Most often, the decision-making process has many touch points but lands squarely on the shoulders of product marketing or regulatory affairs. The actual process of executing translation and localization is then often established in a centralized or decentralized fashion. Internal and external resources to carry out the deliverables are identified.

Managing Translation

Global producers in the life sciences industries have approached localization of content in a variety of ways over the years. In all of the world markets, large multinational corporations, not unlike other industries, have followed a model of:

1. Central control of all content and process,
2. Decentralized creation of content and process, or
3. Establishment of a central control with local empowerment and inclusion to process and acceptance.

These three approaches have many nuances. The third option—central con-

Table 3. Examples of Typographical Rules in CEEC Languages

Language	Quotation Marks	Thousands and Decimal Separators
English	“x” (Windows Keystrokes Alt+0147, Alt+0148)	1,234.56
Greek (Cyprus)	“x” (Alt+0171, Alt+0187)	1.234,56
Turkish (Cyprus)	“x” (Alt+0147, Alt+0147; Alt+0171, Alt+0187 also applicable)	1.234,56
Czech	“x” (Alt+0132, Alt+0147)	1 234,56
Estonian	“x” (Alt+0132, Alt+0147)	1 234,56
Hungarian	“x” (Alt+0132, Alt+0148)	1 234,56
Latvian	“x” (Alt+0147, Alt+0148)	1 234,56
Lithuanian	“x” (Alt+0132, Alt+0147)	1.234,56 <small>Note: use of no thousands separator may also be acceptable, e.g., 1234,56</small>
Polish	“x” (Alt+0132, Alt+0147)	1 234,56
Slovak	“x” (Alt+0132, Alt+0147)	1 234,56
Slovenian	“x” (Alt+0171, Alt+0187; Alt+0132, Alt+0147 also applicable)	1.234,56

trol with local empowerment—enables the corporate entity to “re-purpose” content to a much greater degree, establish good practices common to all localization issues and provide acceptance and input from the local stakeholders who are closest to the consumer or end user. Naturally, one of the key differentiators between, say, a producer of a blood analyzer and a producer of an inkjet printer is academic. In either case, there is much at risk should poor content sneak out through the chain of review and acceptance. At the end of the day, however, the required result for the user and consumer is accuracy.

Because of this need for high accuracy and tight control, and before the proliferation of and access to information on the Web, “Big Pharma” and other large multinationals have historically decentralized almost everything at the national or regional level. Historically, this model has seemed to have served its purpose and to have satisfied local regulators. However, regulators are not necessarily concerned with cost to the producer and may or may not have an interest in the speed with which content can be produced at any given stage of a trial.

Build or Buy?

Not all producers have approached localization the way just described. Some have set up in-house translation or localization departments to manage the central control, incorporating local company professionals for the review and acceptance process. Some have outsourced this model and have invested in a relationship with a vendor partner to manage this process. Now, with the prospect of managing another 10 or more languages, companies are faced with the decision of how best to proceed while managing quality, cost, turnaround time and permission to market in the form of a Market

Authorization Holder, or CE Mark.

Small and medium-sized enterprises can more easily adopt the model that best fits with fewer legacy issues to resolve. In either case, those who want to sell globally should develop an integrated, cost-effective strategy.

Whether an organization chooses to build or buy a center for content localization, it all begins with the source and a well-defined method of managing one of their most vital assets: corporate content. Typically, but not always, the source is in English. If this is not the case, then English is often used as a “pivot” language. Ideally, the content source is controlled, identified, and packaged; written; and created in the source with a global audience in mind.

Translating Content for the CEECs

Today in Europe, the EU is one of the largest employers of in-house and freelance translators, interpreters and technical publishing specialists. This is especially true for the pharmaceutical and medical device industries, where there is a great deal of content, and scrutiny of that content.

According to Angus Roxburgh of *BBC News Online* in Brussels, in his article “Translating Is EU’s New Boom Industry,” this is what the “before” and “after” will look like for the European Commission, the policy and legislation guiding body of the EU, as of 1 May 2004:

Before 1 May 2004:

- The European Commission had 1,300 translators.
- They processed 1.5 million pages per year.
- They cost the EU 550 million euros.

After 1 May 2004:

- Staff will almost double in size.
- They will translate 2.5 million pages per year.

- Their budget will be over 800 million euros.

What does this mean to the producers? For one thing, this sudden demand will greatly limit the availability of translators, especially during the early surge to get content produced. Companies that have had the foresight to begin content translation and localization ahead of deadlines will be less affected.

Further, it will greatly exacerbate any review and acceptance cycle sign-offs, both internal or with competent authorities, including the European Agency for the Evaluation of Medicinal Products (EMA). Human medicinal medicine products approved by the EMA must be delivered in all official EU languages by the present or prospective Market Authorization Holder. The EMA controls final output for all content needed for the technical file through their Centralized Procedure and Quality Review Document group.

In theory, final acceptance can be achieved within 90 days of submittal, but companies must be prepared for longer periods due to fixes to content earmarked by the Quality Review Document group and sent back to the producer for correction and, of course, backlogs.

Manufacturers of medical devices needing to comply with the Medical Device Directive, the In Vitro Device Directive, Implantable Directives, and other directives can expect much of the same scenario, perhaps with added scrutiny and potential roadblocks.

EU and local regulations for medical software user interfaces are still unclear, yet most manufacturers with a software user interface component are preparing to localize for those languages in which they will market and sell product. Localization for user manuals, for the most part, must be translated into the local language, but there are some exceptions, at least on a regulatory

basis. Dutch is one example in which regulators and users will typically accept an English-language version.

Details, Details...

A number of translation-specific technology matters must be addressed, each with a potential cost impact. Publishers need to be aware of the region's scripts, diacritics and fonts (special CEEC font sets may need to be acquired), and publishing computer systems and software programs must be capable of viewing, displaying and printing CEEC languages. Similarly, if translation memory technology for re-purposing of already-translated text has been used for other languages, it must support CEEC languages. The same support is required for vital translation support materials such as glossaries and style guides.

Organizations should be confident with their translators' subject matter expertise and specific linguistic experience—whether part of an in-house staff or externally sourced. However, multiple translation production details must be attended to. It starts, as always, with cost. The CEEC countries are generally known as a cost-attractive region, but there are increased costs associated with translating and localizing content for this region. In addition, there is usually a range of costs associated with translation for these new languages. Some languages will be “expensive”—on a par with the costs associated with translation for Western Europe or Asia—depending upon local cost structures or due to supply-demand considerations.

In-locale reviewers also must be technologically prepared. If an organi-



zation has historically relied on the exchange of PDF versions of a document during the review of translated content, will their in-country CEEC reviewers have the up-to-date versions of Adobe Acrobat applications so that they can participate in the review process? Are reviewers familiar with using PDF annotation features? If files will be exchanged electronically, do reviewers have cost-effective Internet access with sufficient bandwidth?

The addition of 10 or more languages to the translation process also brings additional project logistics concerns, and the importance of effective translation project management should not be overlooked. Finally, just because key dates have come and gone does not mean that the translation work ends; the translation of newly developed content and updates to existing materials guarantees an almost never-ending process.

NOTES

1. In this article, the term life sciences refers primarily to the biotechnology, medical device and pharmaceutical industries, and less so to cosmetics, veterinary products and nutritional supplements.
2. At press time, Greek Cypriots had rejected a United Nations-backed plan to unite Cyprus. As a result of this vote, only the internationally recognized Greek Cypriots, and not the Turkish Cypriots, have joined the EU.
3. Localization is the process of adapting a product or service to make it linguistically and culturally appropriate to the target locale (country/region and language) where it will be used and sold. Typically, localization involves more than just translation and can require elements of publishing or engineering.

Joe DiDamo is vice president of Global Business Services for Moravia Worldwide. He is a veteran of the translation and localization business with a focus on the life sciences industry. He has held several leadership positions in the industry during the last 14 years, mainly in sales and marketing, but also in operations. DiDamo is based in Moravia's US office in Thousand Oaks, CA, and leads the global sales and marketing team. He can be reached at joed@moraviaworldwide.com.

Copyright © 2004 by the Regulatory Affairs Professionals Society (RAPS). Posted with permission on www.moraviaworldwide.com. Reprinted from the August 2004 issue of *Regulatory Affairs Focus*. This article may not be published, reposted or redistributed without express permission from RAPS and payment of appropriate fees when applicable. To obtain such permission, send a message to reprints@raps.org.