



Localization: On its nature, virtues and dangers

Anthony Pym

Universitat Rovira i Virgili, Tarragona, Spain

Abstract

Discourse on the localization of software and websites now underlies one of the major language industries of our time. Its basic concepts also address issues that are of interest to sociolinguistics, text linguistics, translation theory, and terminology. A survey of the main areas in which localization discourse might interact with linguistic theory and indeed with languages themselves reveals both positive and negative aspects. On the positive side, there is much to be learnt from interlingua architectures as opposed to transfer approaches. Similarly positive may be the way commercial language technology can strengthen the diversity of selected languages and cultures, rather than necessarily impose just one international *lingua franca*. However, the concepts of localization are rather more worrying with respect to the loss of linearity in text production and reception, since the predominance of chunking threatens to reduce the person-to-person aspects of communication. Also worrying is the use of leveraging strategies to restrict the role of translation to natural-string replacement, resulting in a hierarchy of work processes whereby IT technicians have political and financial priority over the uses of language for communication. These issues can only be addressed once the locale of localizers themselves is identified as an interculture, and a call for ethics is made in humanistic terms.

1. Basic terms

‘Localization’ is here understood as a general mode of thought informing cross-cultural text adaptation in the fields of software, product documentation, web technology, and some international news services. It is part of a series of concepts that are steadily making their way into Translation Studies, although they rarely appear in mainstream Linguistics.

Standard definitions of ‘localization’ and associated terms are offered by the Education Initiative Taskforce of the Localization Industry Standards Association (at <http://www.ttt.org/leit/terminology.html>):

Localization involves taking a product and making it linguistically and culturally appropriate to the target locale (country/region and language) where it will be used and sold.

Internationalization is the process of generalizing a product so that it can handle multiple languages and cultural conventions without the need for re-design. Internationalization takes place at the level of program design and document development.

Globalization addresses the business issues associated with taking a product global. In the globalization of high-tech products this involves integrating localization throughout a company, after proper internationalization and product design, as well as marketing, sales, and support in the world market.

We might thus say that there is one general process called ‘globalization’, of which ‘internationalization’ and ‘localization’ are parts. In order to globalize, you first make your product general in some way general (‘internationalization’). Then you adapt (‘localize’) to specific target markets (‘locales’). The terms are by no means as standard as they may appear (Microsoft uses them differently). Yet they encapsulate a very rudimentary mode of thought – the basis of many fragmentary discourses – that is of interest to us here.

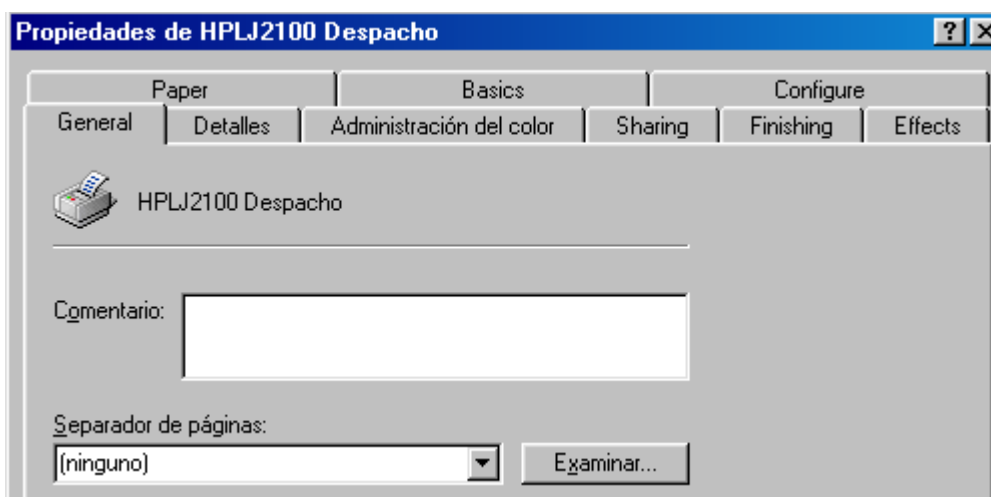
Several initial points should be made.

First, these terms appear to come from marketing. They primarily concern what is done with products, not just with texts as linguistic constructs. The concepts would seem to have been brought to bear on language problems in the late 1980s, to handle the skyrocketing costs of marketing computer software in many different languages and cultures (cf. Brooks 2000).

Second, the simple lesson is that high costs and numerous cultural problems result from just taking a home product (usually software produced for the US market) and translating its natural-language strings into many languages. Overall costs are lower if the home product is first prepared for later localizations. Perhaps the most significant example of this is double-digit encoding of all natural-language fonts (instead of the single-digit encoding used for English) so that the source codes can then be transferred into Oriental languages. More generally, specific local content is removed and the points that require translation or adaptation are in some way separated out (this is variously called 'leveraging'). Such localization-sensitive engineering means much of the hard work is put into producing just the one generic product, which can then be localized into any number of languages and cultures.

Third point: The term 'locale' designates both a language variety and a set of cultural preferences. To see simple examples in most Microsoft operating programs, go to Configuration in the Start Menu, select Control Panel, select Regional Configuration, and you should see a long list of locales. Try it. For English we have Australia, Canada, the Caribbean, the United States, Ireland, Jamaica, New Zealand, the United Kingdom and South Africa. For Spanish there are 20 such locales listed. Microsoft also gives us the same locales for the actual languages concerned (in Word, go to the Tools menu, select Language, select Set Language, and you should find the same list). Of course, you have to buy the dictionaries and thesauruses to make all those locales work. But it is possible to do so. I use Word to correct my Catalan.

Fourth: Like a great deal of work on language, localization is most visible in its absences or failings. We are all unfortunately used to computer menus like the following (from my own bilingual computer):



Here only some parts of the menu have been localized into Spanish, leaving the users to read English for most of the information actually from Hewlett Packard. Such mixes are usually

not accidental, since partial localization may be reserved for the interfaces that the non-specialist user is most likely to need.

Brooks (2000: 48-50) discusses three levels of localization, ranging from ‘enabled’ software, where the interface remains in English, to ‘adapted’ versions, where all content and examples are from the new locale. These levels would then correspond more or less to the nature of the markets being targeted: a large international market like French would have everything ‘adapted’; smaller markets (Thai, Romanian) would have everything enabled but only the most popular programs are localized; and then there are ‘emerging markets of limited potential’, for which only core products like Windows are localized. The theory would seem not to apply to our example above, unless Spanish is regarded as a mysteriously limited market. One suspects the reason is more likely to concern passive knowledge of English among technically-minded users across the planet. Whatever the case, the point to be made is that the degree of localization corresponds roughly to different user profiles.

A less innocent example of a bilingual interface would be the following, which similarly haunts my computer:



Not only do we have the two languages (the user has no choice but to ‘Aceptar’ in Spanish) but the English syntax has stayed in some kind of half-way house (I would personally feel much happier with ‘An error has occurred’). And then there is the famous although possibly apocryphal localization of ‘Status: Active’ (another HP printer menu) as ‘Estado civil: Activo’ (suggesting a matrimonially active printer). No matter: Having learned to live with such things, we click *Aceptar* and proceed regardless.

That is about all we really need to know about localization for the moment (more detailed information is readily available in Esselink 2000 and the web sites listed in our References). The rest of what we have to say will be based on the few concepts presented above.

2. Localization theory as a linguistics

There is something willfully perverse in dragging ideas away from the realm of commercial practice and making them act like a scholarly discipline. They will inevitably fail, or at least appear terribly naïve. Yet we suggest that localization discourse, even in the very simple terms in which we have explained it, might be operating like an industry-based linguistics. Recall that Benveniste (1966: 23) described the phonetic alphabet as the first descriptive linguistics. Localization is similarly using analytical concepts in order to divide up language. Further, its skeletal theory addresses several key issues in linguistics. Consider:

- ‘Internationalization’ or ‘globalization’, as the writing of an elaborate source text that can then go into all languages, uses amplified memory space to construct what become deceptively decontextualized texts, in something approaching universal technical languages. If you like, the combinatory logic of a Ramon Llull finds its

ultimate fulfillment in the ideally internationalized Microsoft operating system, big enough to embrace the user needs of all possible languages and cultures. Our irony is probably too easy, as if the academic armchair were by nature sure that cultural difference will conquer the largest of commercial imperialisms. On the contrary, universalism is indeed possible within the narrow fields of human experience touched by localization, with the proviso that the texts actively *form* those fields of experience (here, interaction with computers). In such circumstances, the dreams of universality may indeed be taken seriously.

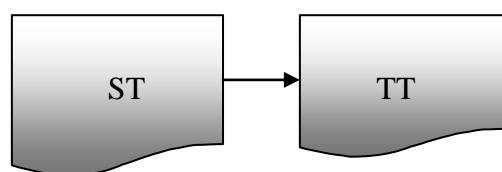
- ‘Locale’, as a specific union of language variety and cultural norms, effectively divides the space of varieties into interdisciplinary units, using market criteria to resolve contradictions between sociolinguistic levels. This is a potentially useful contribution to sociolinguistics, and the term ‘locale’ is short enough to catch on. Further, on the side of text linguistics, one could opine for hours about how many angels fit into a linguistic ‘context’, but the concept of ‘locale’ has to be defined and implemented in the daily struggle for market shares. It is an economically testable reality, as indeed are all the features of the localization process.
- ‘Localization’, as the adaptation of an internationalized text to a particular locale, covers everything that theorists of translation have been trying to legitimize for several decades, incorporating extensive rewriting as part of standard cross-language transfer. As such, it should be viewed as a potentially useful addition to the vocabulary of Translation Studies, if and when it is not given to belittling the concept of ‘translation’ itself.

The concepts are thus attached to an active empiricism, coming to grips with the languages of the world and doing much to shape their technological future. Mainstream linguistics and academic translation theory could surely benefit from such empiricism, and indeed from the funds invested in research projects along these lines.

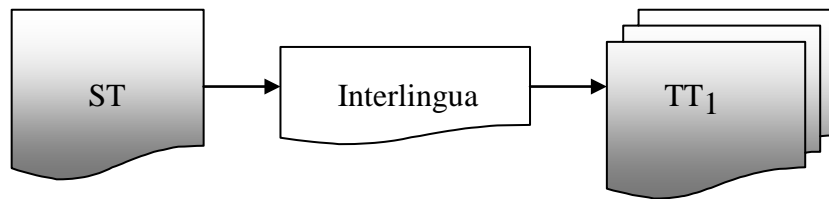
If it is a linguistics of some kind, localization is also a mode of thought that actually produces texts, modifying its own object and bringing about far-reaching effects on many levels. Our purpose now is to survey what the discourse of localization might actually be doing on this level.

3. Transfer vs. interlingua

One of the fundamental lessons to be learnt from localization is that language pairs form an inefficient model. That is, rather than compare Language A with Language B, we should be thinking about constructs that can be manifested in all (or a good number of) languages and cultures. In terms of translation theory, we might reduce the message to the following: Do not think in terms of directional transfer between the source text and the target text:



Think instead in terms of an interlingua version, derived from the ST, from which many target versions can be produced:



This is a very simple lesson. But it is one being overlooked by the Translation Service of the European Commission, for example, which insists on feeding glossaries into its transfer-based Systran architecture, which works quite well between Romance languages. However, attention to an interlingua architecture, hopefully with a lot of statistics and some of the discourses of localization, would do far more to meet the linguistic challenges of a fully enlarged European Union.

The consequences of the above are numerous. In terms of work on language, it means that a lot more needs to be invested in the controlled language of source texts, in field-specific concept mapping, in multilingual translation memories, and in modular text production. Some of this is happening in the European Union bureaucracy, where English is coming closer to the interlingua and English sections are thus becoming official scribes and revisers, rather than translation departments. With or without the concepts, some inklings of localization are changing the very nature of our jobs.

4. Recycling information

Electronic texts, especially the kind most apt for localization, are given to non-linearity. Think of the online Help file, or the web site we read with the Find function. Such texts are structured in terms of repetition rather than deictics, hyperlinks rather than hypotaxis, and text length has no socially necessary restriction. All this, of course, just when linguistics has established disciplines of text linguistics and discourse analysis given to analyzing such things. All this just when we have learnt to teach students the fundamentals of text construction. Now we are concerned not with the dynamics of real or feigned interaction, nor with any narrativity of beginning-middle-end. All texts are 'information objects' (Hofmann and Mehnert 2000), 'information elements' (Lockwood 2000), fragments of an unseen whole. Such is non-linearity. And it might be a further feature of localization.

Non-linearity on the side of text production more readily goes by names such as 'leveraging' (which we have met) and 'chunking' (which is understandable enough). It ensues from localization in two ways.

First, leveraging concerns the breaking up of text into parts that have to be worked on at different levels, usually by different people. When this concerns the separation of natural-language strings from source code, the process is actually increasing discursive linearity, since the localizer need then only work on what the reader will actually see. However, the notion of leveraging may also apply to the extraction of terminology from a set of strings, the separation of new from old elements, the isolation of updates, or the tagging of fragments in accordance with reception levels (some elements are for the home user, others are for

professionals). In these latter cases, the localizer is invited to work only on the fragments thus separated out, with significantly less regard for discursive linearity.

Chunking may be seen as another level of the same process. Texts are broken down into minor units that can be mixed and matched to form new texts. If you like, this is the step that makes concept-based leveraging possible. Yet it has far-reaching consequences as a general approach to language production. It encapsulates not only the general logic by which paragraph entities are tagged for XLM, but also the innocent alignments achieved by translation-memory programs. The software increasingly invites language producers to work only on the fragments, the new, the different, and to forsake linearity, or to trust that the user (no longer ‘reader’) will create their own lines of use.

Is this loss of discursive linearity a good or a bad thing? One can find arguments that nothing essentially new is happening here. For example, O’Donnell (1998) claims that linearity was lost with the historical move from papyrus scrolls to the bound manuscript, since the latter allows a text to be ‘thumbed through’ and consulted in terms of an index. The great libraries of antiquity would share the same ambition as the Internet as virtual library. There would be nothing fundamentally different between surfing the web and using canon tables to jump between sites in the Gospels. Indeed, on this broad historical view, the illusory fixed linearity of the printed book has been a moment of exception rather than the general rule: both before and after print, the written text was constantly subject to updating and adaptation, localization *avant la lettre* (cf. Pym 2000: 80-89). On this basis, arguments could be made that leveraging and chunking are even returning us to some kind of more fragmentary experience of the world, as if we were moving in and out of so many oral conversations.

Esselink (2000) has a vision in which all localization projects will be based on what he terms ‘multilingual database publishing’. This means that all texts are broken down into information elements, stored in a database in many languages, and recombined and modified in order to create new texts. Thus, all documentation will be ‘created, managed and published using database technology’ (478). This should radically reduce workloads, since ‘only new or changed information is extracted from the database and processed by translation memory’ (479). Localizers working at the level of natural-language strings would only have to translate or edit the new pieces of information, making this level of localization an on-going update process. Localization projects would then have no beginning or end (another loss of narrative linearity); workloads can be distributed more evenly, and ‘translators would probably end up working on several small projects each day’ (480).

This vision is offered as some kind of brave new world, ruled by criteria of efficiency. Language would be produced and consumed with little reference to textual wholes, with few ideas about who is actually producing or using the text, and with minimal reference to any kind of overarching discursive purpose. On all those levels, we could claim that text production risks losing the fundamental values of human communication. Esselink’s ideal world is full of things, not people.

5. Localization and translation theory

The dehumanizing tendency of localization discourse is of particular importance to translators (who risk getting lousy jobs) and translation theory (whose recent advances are threatened with immediate extinction).

Perhaps surprisingly, the term ‘translation’ does not generally appear among the key concepts of localization theory (it is not mentioned at the head of this paper). ‘Translation’ is not listed in the glossary of terms in Esselink (2000); it has no prominent place in the LISA literature. Where we do find ‘translation’ is in Sprung (2000), which is a volume produced for the American Translators Association and thus could hardly exclude the term. Yet even there we learn that ‘companies that specialize in software localization are quick to distinguish themselves from ‘translation companies’’ (xi). Why? Because, according to the same source, ‘localization’ means ‘taking a product [...] and tailoring it to an individual local market’. And ‘tailoring’ involves both ‘translation’ (‘converting text from one language to another’) and ‘adaptation’ (apparently everything else that has to be done) (Sprung 2000: x).

So what is ‘translation’ when seen from the perspective of localization? Basically, it is the minimalist replacement of natural-language strings. And this is certainly the least interesting part of localization. Sprung (2000) consistently demonstrates that the real costs are in internationalization, leveraging, structuring hierarchies of target languages in terms of market priorities, organizing complex language-service teams, drawing up schedules, testing localized products, evaluating translations, creating cooperative working relations between specialized service companies, using or developing appropriate software for localization, and working with controlled writing. In short, the replacement of natural-language strings (‘translation’) is shown to be a minor part of localization. The breakdowns of budgets rate ‘translation’ at about a third of the total costs, with the remaining two thirds split between ‘product re-engineering’ and ‘project management’. Small wonder that many translators and translator-training programs are switching over to the more lucrative mysteries of localization.

Perhaps the major irony in this process is that translation theory has been moving in precisely the opposite direction for the past twenty years or so, in tune with developments toward text linguistics, discourse analysis and greater attention to cultural determinants. No longer do we regard translation as a sentence-level language replacement exercise. Interpersonal dynamics and cultural specificity are now seen as playing a major role in the solving of translation problems, thanks to German-language *Skopostheorie* and the development of system-based Descriptive Translation Studies. And now, precisely when some theorists are on the point of affirming that all cross-cultural relations are translational, that translators should be experts in the management of cultural difference and the like, right at this point of maximum expansion, translation theory is being outflanked by the discourse of localization, and translation itself has been returned to the narrow linguistic exercise it was in the 1960s. To little avail might we point out that software, websites and news services are just a handful of the many genres undergoing moves from culture to culture. Of little use the arguments that localization should really be a small part of translation. The technicians have developed their discourse with disdain for all the academic waffle about translation. Theirs are the terms that have attracted serious investment in the commercial sphere; theirs is the discourse that has built the major companies in the field. Indeed, the only thing that ‘translation’ can show for itself at the large institutional level is the Translation Service of the European Commission and perhaps the fading aura of the UN, neither of which organizations is likely to score highly on ranks of efficiency. There is ground for real conflict here.

One might argue, of course, that such conflict only exists on the level of theory. In practice, translators and localizers are perhaps indistinguishable from each other, and many of the people working in localization have a background in translation anyway. Perhaps so. Yet one suspects that the terms used in theories are not entirely without correspondence to linguistic

processes on the ground. For example, click around the various web pages of an international site like Ikea (www.ikea.com). You will soon find a clear difference between pages that are ‘translated’ and those that are ‘localized’ in a broader sense. Ikea has exactly the same format in Australia, Greece and Malaysia; the only difference is natural-language string replacement. However, other countries (try Sweden, Germany and France) have clearly localized sites, reconceptualized for the target market. And Ikea Hong Kong has completely different colors and presentation, in accordance with its different script and need for a bilingual site.

The message of such examples is that translation is indeed used as a part of localization, and that there is a fairly clear cut-off point between the two. That is, there is a *discontinuity* between translation and non-translation (here, between translation and the non-translational aspects of localization). This might be a surprise for theorists like Halverson (1999), who argues at length that translation is a prototype concept but entirely ignores the discourse of localization (and indeed the discontinuities internal to the reception process). On this interface between translation and non-translation, there is more evidence of discontinuity than of prototypicality.

The electronic tools of our age are designed to exploit this discontinuity. For instance, an advertisement for Atril’s DéjàVu translation-memory software asks translators ‘Would you rather translate this... [image of complex source code] ... or *just this* [computer screen with just those two words]’. The translators of the world are apparently supposed to clamor for liberation from all the technical codes; they ostensibly have to buy the software so their natural-language strings can be automatically leveraged away from the rest. The technology condemns translators to remain ‘just translators’, while the big money and more interesting work flows the way of marketing specialists and information engineers.

6. For and against localization

There is good and bad in the discourses of localization. To summarize: As a massive intervention into the future of our languages, localization might also be denounced as a progressive dehumanization of communication. As an applied language policy, however, localization may yet provide a workable platform for a European Union with 50 to 60 languages. It may induce a return to interlingua architectures and centralized domains, rather than compare natural languages. And it may make us aware that power in this field is accruing in technical interculturalures, which need to be understood and addressed directly. Then might localization enhance rather than manipulate democracy.

References

- Benveniste, Émile (1966) *Problèmes de linguistique générale*. I. Paris: Gallimard.
- Brooks, David (2000) What Price Globalization? Managing Costs at Microsoft. In: Sprung, Robert C. (ed.) *Translating Into Success. Cutting-edge strategies for going multilingual in a global age*. American Translators Association Scholarly Monograph Series XI. Amsterdam/Philadelphia: Benjamins. 43-57.
- Esselink, Bert (2000) *A Practical Guide to Localization*. Amsterdam/Philadelphia: Benjamins.
- Halverson, Sandra (1999) Conceptual Work and the ‘Translation’ Concept. In: *Target* 11 (1). 1-31.
- Hofmann, Cornelia, and Thorsten Mehnert (2000) Multilingual Information Management at Schneider Automation. In: Sprung, Robert C. (ed.) *Translating Into Success. Cutting-edge strategies for going multilingual in a global age*. American Translators Association Scholarly Monograph Series XI. Amsterdam/Philadelphia: Benjamins. 59-79.

- Lockwood, Rose (2000) Machine Translation and Controlled Authoring at Caterpillar. In: Sprung, Robert C. (ed.) *Translating Into Success. Cutting-edge strategies for going multilingual in a global age*. American Translators Association Scholarly Monograph Series XI. Amsterdam/Philadelphia: Benjamins. 187-202.
- O'Donnell, James (1998) *Avatars of the Word. From Papyrus to Cyberspace*. Cambridge Mass. - London: Harvard University Press.
- Pym, Anthony (2000) *Negotiating the Frontier. Translators and Intercultures in Hispanic History*. Manchester: St Jerome.
- Pym, Anthony (2004) *The Moving Text: Localization, Translation and Distribution*. Amsterdam/Philadelphia: Benjamins.
- Sprung, Robert C. (ed.) (2000) *Translating Into Success. Cutting-edge strategies for going multilingual in a global age*. American Translators Association Scholarly Monograph Series XI. Amsterdam/Philadelphia: Benjamins.