

Localization and the Changing Role of Linguistics

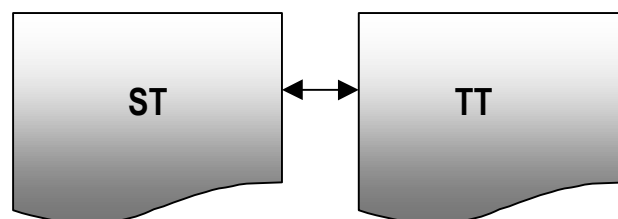
Anthony Pym
Universitat Rovira i Virgili
Tarragona, Spain

Abstract: The discourse of localization, now generalized beyond the domains of software and product documentation, has successfully outflanked restricted linguistic notions of translation. This new situation has certain consequences for the traditional role of linguistics, notably with respect to the enhanced need for projects based on interlingua architectures. This includes the use of multilingual terminology databases, controlled writing, disambiguity tagging, and the integration of MT, such that most linguistic intervention is effectively prior to the moment of actual translation. These features call into question the usefulness of comparative or contrastive linguistics, especially when limited to language pairs. At the same time, what is urgently needed is empirical study of the ultimate commercial and cultural effects of the various localization strategies.

We begin with a bad joke, based on the originally American form of the ‘dumb blonde’ joke. That form is now operative in many languages and many cultures (we will say, many ‘locales’), and may be adapted (let’s say, ‘localized’) to many purposes. Here the archetypically vapid blondes have transcended mere sexism and, with equal lack of justice, have turned into comparative linguists. This is unfair. But here goes anyway:

A comparative linguist is walking along and comes to a river. They look across and, seeing another comparative linguist on the other side, call out ‘How do I get to the other side?’. This, after all, is the fundamental problem of translation, to get to the other side, to reach that culture B from the known culture A. And the second comparative linguist replies, of course, ‘But you are on the other side!’

The moral of the story might be that the problem of translation (getting across the river) is not the problem of comparative linguistics (comparing the two sides of the river). Or more philosophically, *arriver chez l’autre* (translation) is not quite the same thing as *analyser l’altérité* (in linguistics). This is a profound difference. One might continue to think in terms of the following:



But then, that arrow in the middle might as well go both ways, from source text to target text and back again. All the work that allows one side to understand the other also allows the reverse comprehension. However, translation—and by extension translation studies—is condemned to deal with movement, with asymmetries in time and place, and with the consequent changing of cultural relations, hopefully for the better. Linguistics of the comparative kind avoids such asymmetries by presupposing confronted systems, be they of tongues, cultures or fields, in such a way that what is essential to either side should remain that way and ideally survive the process of translation. That is essentialist thought *par excellence*. It is also part of the noble

struggle to preserve the diversity of human languages and cultures. And this is so even when the theories and models of linguists cannot grasp the challenge they confront. Without substantial directionality, a strong arrow going one way and not the other, such theories cannot see movement as anything more than the illusory threat of sameness, of a globalized culture, of the non-academic future as disaster. The noble struggle is thus fought blind.

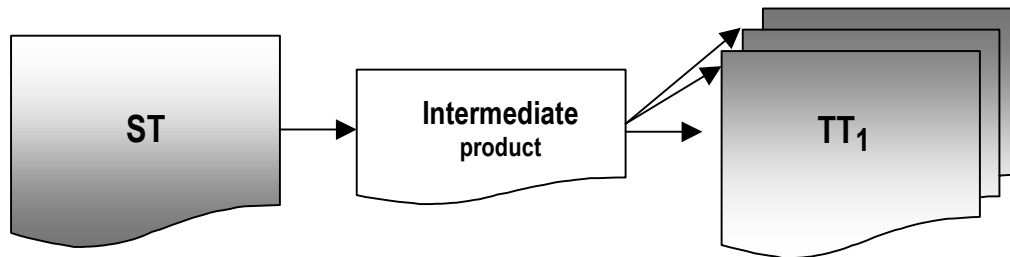
Here I would like to consider something between those two poles. The discourse of localization has not developed from translation theory, nor is it an invention of academic linguistics, as far as I know. It comes from practice, perhaps from the best-paid mediating practice of our time. The term 'localization' originally referred to the production of software for many languages and cultures, 'locale' being a convenient term for a particular configuration of language and culture. Localization would thus mean taking the (usually) American software and rewriting it so that it can work in a different locale. This would involve both translation (in the restricted sense of replacing user-visible natural language strings) and adaptation (adjusting to the local conventions of numeric representation, currency, dates and so on). A rough idea of this can be grasped from your computer. Here I am working with Microsoft Windows98: I go to Configuration in the Start Menu, select Control Panel, select Regional Configuration, and I find 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 Word97, 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 Word97 to correct my Catalan.

That much seems simple enough: there are localized versions of the best-selling software; the multinational can help develop the local. When people in IT talk about localization, however, they are most concerned with what we might call the non-translational aspects. Given the complexity of software programmes, the replacement of user-visible natural-language strings costs far less than the necessary re-engineering of the programme, the changing of dialogue boxes, the re-allocation of hot-keys, the coordination of all the various changes, the testing of the localized version, and the project management necessary to organize all the teams involved with strict time constraints. By most accounts, translation would account for about a third of total localization costs. In other words, to put things bluntly, the kinds of linguistic problems that our traditional comparative linguists deal with are no more than a third of the problems that the fastest-growing language industry has to deal with.

Confronted by extremely high localization costs in the 1990s, Microsoft successfully reduced those costs by 'internationalizing' their products prior to the moment of translation (Brooks 2000). This involved taking the American product and deleting all elements that were in some way specific to the American market. The most significant part of this was double-digit encoding of all natural-language fonts (instead of the single-digit encoding used for English) so that the source codes could then be transferred into the Oriental languages. Such localization-sensitive software engineering meant that all the hard work was put into producing just the one generic product, which could then be localized into any number of languages and cultures.

This would seem to be the main message of software localization: investments in a generic product mean savings in multiple locales. Admittedly, 'internationalization' is

not a happy term for this generic product—nations have little to do with this—and one might talk more fashionably about ‘delocalization’ as the step to the middle stone. Whatever the term, the basic model would be something like the following:



In recent years the discourse of localization has extended its range of application beyond software. It is now common enough to talk about the localization of websites, of product documentation, or of news items (see Sprung 2000 for a series of case studies). This is proving to be a very successful theory, commercially outflanking anything mainstream translation theorists have done in recent decades. Why the success? Well, people can be made to understand these terms; these are words and concepts that are easily explained to clients or managers; they bring the aura of things happening now, the *Jetztzeit* of the New Economy. Second, the fundamental model is constantly tested in practice, in ways that seek and need little from university locations or academic prestige. Third, the process can be shown to save money, and to do so in the sectors where the most money is moving. Fourth, as a consequence of the above, it is the basis of a high-pay job market that is currently hungry for competent language professionals. There are thus quite a few very good reasons why a university translator-training school might pay close attention to the discourse of localization, and why we should be training localizers and not just translators (for further reasons, see Pym 1999). Localization theory is achieving virtually everything that traditional translation theory has ceased to achieve.

Then again, you will object, the concept of localization is bringing us nothing essentially new. And that is quite right. Translation theorists have been talking about ‘language and culture’ or ‘the cultural component’ for at least 30 years. If we should now have a convenient word for the thing (‘locale’ is at least shorter), the conceptual advance is not of overwhelming proportions. More important, the basic idea of producing an intermediary generic text has long been applied in many fields, and in many facets of standard translation procedures. We might name the following as examples of the same underlying idea:

- Certain theorists (perhaps starting with Gouadec 1989) insist that the first stage of the translation process is to revise and verify (‘établir’) the source text so as to remove any potential typographical errors or otherwise disturbing ambiguities prior to the moment of translation as such.
- When we emphasise the need for documentation and terminology research prior to the moment of translation, we are similarly investing effort in the ‘before’ so as to save effort in the ‘after’.
- When the Bible translators of the SIL work into all kinds of exotic languages, their immediate source is not the Hebrew or Aramaic texts but skilfully annotated English versions, giving all the glosses and explanations needed for the actual translation process.

- When screen dialogues are dubbed or subtitled (from English, at least), the source text is increasingly marked with instructions for the translation, glossing culture-specific words, tagging ambiguities, and in some cases indicating the importance of certain implicatures.
- By extension, our work on standardized terminologies of all kinds might be seen as effort invested prior to translation so as to make actual translation easier.
- The same could be said of the text alignment at the basis of translation-memory software, where the distribution of major linguistic effort is similarly focused on the preparatory phase.
- And from that perspective, the integration of post-edited machine translation is yet another application of the same principle: the more effort we invest in the MT databases, the better the outputs and the less the effort needed for post-editing.
- The production of source texts in a controlled language (with a limited vocabulary and syntactic repertoire), necessary for the successful use of MT, can be seen in terms of the same logic. This is by no means new: the idea of Basic English has been with us from at least the 1920s.

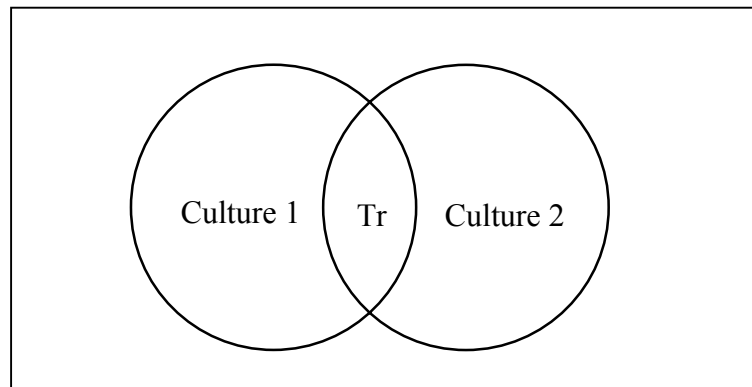
Translation theorists might thus rightfully claim that the discourse of localization is a simple case of new polish for old shoes. The idea of an intermediary version has indeed been around for a long time. So should we refrain from paying attention to these overpaid commercial upstarts?

There does remain one fundamental difference that might yet justify the term ‘localization’.

If we look closely at the above list of long-established and otherwise mainstream procedures, we usually find a mix of two architectures: the principles are considered to be the same no matter whether we are moving from one locale to another (i.e. mediated but paired transfer) or from a generic version to many locales (i.e. one-to-many transfer). What we need for moves from French to Arabic might thus be more or less what we need for moves from French to the ten other official languages of the European Union; our research may serve both purposes. In localization, on the other hand, the emphasis is consistently on the one-to-many, or on what MT knows as an ‘interlingua’ (rather than ‘transfer’) architecture. This, I believe, is a very fundamental difference. There are at least three reasons for stressing the point:

- The very notion of localization, in denominating a process of *working into* a locale, incorporates a degree of directionality and thus asymmetry that is lacking models where cultures simply face cultures. Admittedly, this is slightly paradoxical, given that the notion of internationalization or human interlinguas incorporate enough naïve aspirations to crown a modernist Tower of Babel. Yet the underlying geometry of movement remains unavoidable.
- In emphasising the many non-translational operations that language professionals are called upon to perform, the discourse of localization recognizes that our graduates will have to be able to do things other than translate. They will have to compile terminology bases, leverage natural-language strings, revise controlled inputs, post-edit MT outputs, hopefully manage projects, and so on. A narrow view of translation would leave students without training in this wide range of marketable skills.
- The basic geometry of the ‘intermediate product’ model further dovetails into a mode of thought that can no longer place translators in the falsely homogeneous space of a native or mother language. The professionals carrying out localization are active in the overlaps of cultures, in places inhabited by teams of complementary competencies whose job it is to work on relations between

cultures. Instead of Culture 1 (C1) facing Culture2 (C2), we would have to think in terms of an Interculture (IC), which would be the more or less narrow intersection of the two larger entities:



- Finally, the 'intermediate product' model necessarily emphasizes the need to work into many target languages, and not just the few major languages that can afford major extensive linguistic defence. We are made to think about entire sets of languages, placed in a hierarchy of market priorities (medieval language hierarchies once measured distance from divine utterance; now they quantify market potential). We would then defend not the particular case of Catalan, for example, but of all similarly stateless European languages of limited extension. This last point requires some elaboration. For example, the Systran machine translation system used by the Translation Service of the European Commissions works very well for transfers between French and Spanish. But this is because the primitive transfer architecture has been enhanced by extensive databases and many years of pair-specific linguistic parsing. The political result is undoubtedly an indefinite extension of French as a viable source or intermediate language in the EU context, along with a certain privilege for the Romance languages that share many of the syntactic features of French. But the non-Romance languages consequently sink toward the bottom of the language-effort hierarchy, and there is little perspective of Europe's many minor languages entering the Systran epiphany. Would the years of terminological work and syntactic rule-writing also be invested for Catalan or Welsh? On the other hand, some such effort might become worthwhile if an interlingua architecture were adopted for controlled-language inputs in highly specified official domains. To that extent, there might yet be political ideals to be attained through thinking about localization.

What is to be done?

The basic concept of localization calls for a departure from some of the classical linguistic concerns of translation studies. It implicitly calls into question the usefulness of pair-restricted comparative or contrastive linguistics. Yet that critical potential, at best appropriate to only some parts of linguistic studies, is no saving grace. We cannot assume that adopting the language of Microsoft will lead us to some kind of earthly paradise. On the contrary, the globalizing companies using localization are implanting globalizing culture under the guise of linguistic difference. Using Microsoft in French is not the same as using Microsoft in English, but in both cases the mode of operation is still Microsoft more than anything else. It is through

localization, not in spite of it, that professional users of technology become the same across the globe.

That might be why, even in the age of an international tongue, there is a growing demand for translations, mainly from English to the larger languages of consumption. Localization ultimately takes place to keep producers separate from readers, agents from end-points, centre from periphery, internationalization from translation, sometimes in the name of protecting cultures, always in the name of identifying and expanding markets. Thus are maintained the technology gaps that have long been the secret of economic imperialism. Localization may not be a good thing (the spread of non-native English might even be preferable). But the beast must be named before we can attempt to assess its ultimate effects.

References

- Brooks, David (2000). 'What Price Globalization? Managing Costs at Microsoft'. In Robert C. Sprung, ed. *Translating into Success. Cutting-edge strategies for going multilingual in a global age*. Amsterdam/Philadelphia: John Benjamins. 43-57.
- Gouadec, Daniel (1989). *Le traducteur, la traduction et l'entreprise*. Paris: AFNOR Gestion.
- Pym, Anthony (1999). 'Localizing Localization in Translator-Training Curricula', *Linguistica Antwerpiensa* 33. 127-137.
- Sprung, Robert C., ed. (2000). *Translating into Success. Cutting-edge strategies for going multilingual in a global age*. Amsterdam/Philadelphia: John Benjamins.