[*Machine Translation Today: Translating and the Computer 15*. Papers at a conference…18-19 November 1993 (London: Aslib)]

Multilingual Drafting of Instructional Texts

Donia R. Scott
Information Technology Research Institute
University of Brighton
Brighton, UK
email: donia.scott@itri.bton.ac.uk

Abstract

This paper presents a new approach to computer assisted production of parallel instructions in multiple languages: *multilingual generation.* At the IT Research Institute at the University of Brighton, we are working on two major projects on multilingual generation of instructions, funded by the CEC and SERC/DTI and involving collaboration with groups of professional technical authors and translators from commerce and government. These projects are directed towards the development of two types of software tools to assist in the production of multilingual instructions: graphical tools for capturing the procedures to be expressed in the text and drafting tools for producing a 'first draft' of the text itself in selected languages.

Multilingual Drafting presents a new approach being developed at the University of Brighton aimed at providing support for two basic needs of technical authors and translators: (a) access to the necessary knowledge of the product or process to be described in the document and (b) guidance in structuring the information in a form appropriate to the target linguistic audience and the genre. This work is being carried out by a team of linguists, computer scientists, and specialists in human-computer interaction, working closely with groups of professional translators and technical writers to develop two types of computational tools: a (non-linguistic) tool for capturing the necessary technical information from the domain expert and a (linguistic) tool for automatically generating draft outlines of instructions in multiple languages.

The chosen genre for our current work is that of instructions. Instructional texts provide an interesting domain for the development of multilingual drafting tools, faced as we are with an ever increasing demand for congruent multilingual instructional documents (especially within the European Community) and the notoriously low quality of commercial instructions. Clearly, writing good quality instructional texts is not a trivial affair.

Current practice for the production of commercial multilingual documentation hinders the efficient writing of effective instructions insofar as product/process development, authoring and translating tend to be three discrete

and sequential processes, making use of different information and involving different skills. Technical experts, while having a thorough understanding of the subject for documentation, rarely have the required linguistic skills for producing appropriate multilingual instructions. Technical authors and translators, on the other hand, have the appropriate linguistic skills but often do not have sufficient access to the appropriate technical information. This combination of factors lead to instructional texts which are all to often badly written or suffer from incomplete or misleading information, problems which are compounded with translations.

One possible solution is the use of controlled languages for the original instructions, combined with human or machine translation. This approach is a popular one, and has achieved a certain measure of success. However, it does not overcome the knowledge gap between the domain expert and technical authors/translators. Our approach provides a bridge for this gap in the form of a tool to capture the knowledge of the domain expert in a form that does not require linguistic expertise but which nonetheless is able to represent the domain knowledge necessary for producing good instructions in a number of languages (which at present include English, French, German and Italian).

A further motivation for multilingual drafting comes from the observation that texts are necessarily selective, embodying choices of expression and structure appropriate only to the requirements of that one language. The task of producing congruent documents in multiple languages is thus necessarily disadvantaged when the source for multilingual documents is textual (as it is with any translation scenario), especially when the translator has no access to the primary source (i.e. the domain expert). A clear example of this disadvantage can be seen in cases where the writer of the source language instruction set is able to exploit the resources of the language at hand to leave implicit one of the steps in a plan of action that the user must carry out. Take, for instance, the following instruction in English

```
Unscrew the lid to expose the box.
```

where our knowledge of the meaning of "expose" in the *given context* allows us to infer the missing, but crucial, intervening step of removing the lid. Note that this step is not left out in the utterance; rather, it is left implicit and is recoverable from the semantics of the second verb. Translating to another language where the equivalent verb does not carry the same presuppositions does not provide the reader with the missing step. In Italian, for example, a straight translation would produce

```
Svitare il coperchio per esporre la scatola.
```

Not only is the relation between "svitare" and "esporre" lost, but the intervening step is not recoverable from the utterance. A more suitable rendition would be:

```
Svitare il coperchio per aprire la scatola.
```

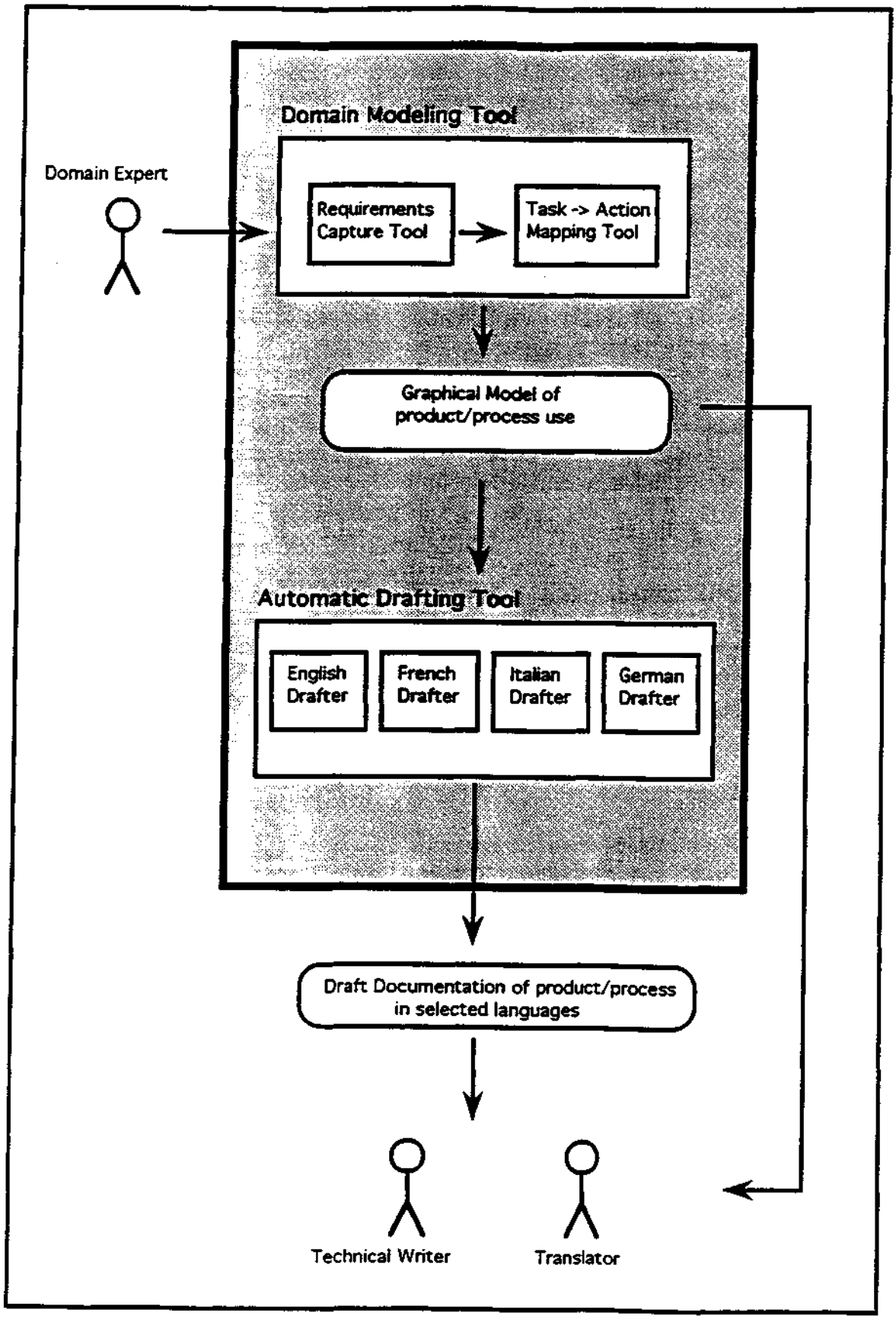
but this can only be reached if the translator has access the knowledge that led the English author to choose "expose", i.e. the domain goal of making the contents of the box visible.

For the automatic generation of equivalent instructions, a linguistically predetermined representation clearly cannot be the primary input, since the choices made for one language can often not be undone for the purposes of another.

Similarly, choices of global and local discourse structure made for one language often require mapping onto different structures in the target language. Different languages often present instructions from different perspectives, and this difference is reflected in discourse. A typical example of this can be seen in the following excerpts (taken from instructions for a step aerobics machine) which have the same underlying goal of getting the reader to balance the apparatus. Note that in the English version, the reader is informed of the *causal relation* between the actions of turning the wheel and balancing the machine; in the German version the reader is informed *how* to balance the apparatus; the French reader is told *why* the wheel needs to be turned.

> The apparatus can be balanced on uneven floors by turning the milled wheel.
>
> Durch Drehen der Rändelmutter kann das Gerät unebenen Bodenverhältnissen angeglichen werden.
> *By turning of the milled wheel can the apparatus (to) uneven floor conditions adapted become.*
>
> Tourner l'écrou molleté pour adapter l'appareil aux inégalités du plancher.
> *Turn the milled wheel to adapt the apparatus to the unevenness of the floor.*

These different perspectives lead to different choices at the level of discourse structure, which in turn impact on choices at the clausal level. Such decisions can be manifest at any level of the structure of a document, and can therefore be expensive to undo in translation. Multilingual drafting helps to avoid the problem by providing (a) a source of knowledge that is not language bound or language biased and (b) guidance to the writer/translator to produce appropriate discourse and clausal structures.


Domain Modeling Tool
Domain Expert
Requirements Capture Tool
Task -> Action Mapping Tool
Graphical Model of product/process use
Automatic Drafting Tool
English Drafter
French Drafter
Italian Drafter
German Drafter
Draft Documentation of product/process in selected languages
Technical Writer
Translator