ISSN: 2456-1134 www.isjcresm.com
Vol-10 Issue-01 Mar 2025

https://doi.org/10.5281/zenodo.15230666

MULTILINGUAL LANGUAGE TRANSLATOR

K.Anjaneyulu[1], Jukanti Ruchitha[2], Goshika Manohar[3], Duvvuri Sainath[4], Bandi Akanksha[5]

[1] Assistant Professor, Dept. of AI-ML, Sri Indu College of Engineering and Technology, Hyderabad,
[2 3 4] Research Student, Dept. of AI-ML, Sri Indu College of Engineering and Technology, Hyderabad

Abstract

In this paper, we look at the current scenario in multilingual documentation generation, identi- fying the reasons for the introduction of compu- tational tools into the process and the types of tools currently provided in support of the trans- lation tasks. We note that certain requirements of the multilingual documentation task cannot be satisfied by the tools currently available and identify what the problems are. We examine trends now emerging in the document indus- try, observing a reorganisation of the workflow. Based on these observations, we hypothesize that, as a radically different approach to sup- porting the multilingual documentation effort, multilingual generation can offer assistance in situations where current translation-based tools are very limited. We believe that, in fact, the type of support afforded by multilingual generation is one that fits well with the new trends in the doc- ument industry and that, ultimately, translation- oriented systems and generation-based systems can be employed in complementary ways to pro- vide a suite of tools to support the documentation process.

This work is partially supported by the Engineering and Physical Sciences Research Council (EPSRC) Grant J19221, by the British Council grant BC/DAAD ARC Project 293, and the Commission of the European Union Grant LRE-62009.

Anthony Hartley is also a member of the Language Centre of the University of Brighton.

1 Introduction

It is well known that the need for multilingual documentation is growing as organisations are motivated by incentives such as access to global markets and the need to provide information about their activities and products in many lan- guages – see, for example, (Language Technol- ogy Symposium, 1993). Currently, the usual method of producing multilingual documenta- tion is to finalise a document in one language and then *translate* it into the other languages re- quired. As the volume of documents produced and the number of languages of publication in- crease, accomplishing the translation task within a commercially acceptable time-frame becomes

ISSN: 2456-1134 www.isjcresm.com
Vol-10 Issue-01 Mar 2025

impractical without computer support. As a result, many tools for translation have been developed. These range from translation aids (e.g., on-line dictionaries, terminology management packages, translation memories) to automatic translation systems (using both rule-based and empirical approaches).

Recently, multilingual generation has been proposed as an alternative to machine translation, e.g., (Rösner and Stede, 1991; Scott, 1993; Bateman et al., 1993; Kosseim and Lapalme, 1994; Reiter et al., 1995; Paris et al., 1995). Instead of supporting the *translation* of a document into a variety of languages, its aim is to support the *generation* in parallel of the documents in the required languages, thus eliminating the translation process entirely.

In this paper, we look at the current scenario in multilingual documentation generation, identifying the reasons for the introduction of computational tools into the process and the types of tools currently provided in support of the translation tasks. We note that certain requirements of the multilingual documentation task cannot be satisfied by the tools currently available, and identify what the problems are. We also examine trends now emerging in the documentation industry, observing a reorganisation of the workflow. Based on these observations, we hypothesize that, as a radically different approach to supporting the multilingual documentation effort, multilingual generation can offer assistance in situations where current translation-based tools are very limited. We believe that, in fact, the type of support afforded by multilingual generation is one that fits well with the new trends in the documentation industry and that, ultimately, translation-oriented systems and generation-based systems can be employed in complementary ways to provide a suite of tools to support the documentation process.

2 Multilingual document generation: current practices

In examining the current scenario in multilingual document generation, we will restrict our attention to applications typical in an industrial setting, that is, non-literary texts largely independent of a particular language or culture. Thus, we consider neither publicity materials, which exploit the linguistic devices of a particular linguistic code, nor legal texts, which are culturally specific. A good example of what we do have in mind is the instructions and reference materials accompanying products marketed internationally.

The prevailing scenario in such applications is to write a text in a given language and translate it into the other languages required once the final version of the text has been agreed.[1] The authoring process typically overlaps with the design of the product itself. It is not unusual for there to be changes to the design specification and functionality which entail corresponding revisions of the draft documentation (Power et al., 1994). Translation can take place only when a final document has been produced, and the product launch can take place only when the translation, in its turn, has been completed. In such an application, the relation between the status of the source language text and the target language texts is generally one of *interdependence* (Sager, 1994), whereby content and intention are kept constant across the different language versions. This may result in a parallel document containing all the language versions, or in separate documents.

Increasingly, and for a variety of reasons, companies have found the need to introduce lan-

[1] In some cases, companies translate a pre-final version of the original text into one other language, in order to highlight and eliminate translation difficulties that would otherwise be compounded by translation into multiple target languages. In this case, however, the translation is not

ISSN: 2456-1134 www.isjcresm.com
Vol-10 Issue-01 Mar 2025

done to obtain the text in a new language but has instead a quality control function.

ISSN: 2456-1134 www.isjcresm.com
Vol-10 Issue-01 Mar 2025

guage engineering tools into this process (Freibott, 1992). First, the amount of documentation to be produced can be enormous, and the cost it involves is correspondingly high; automatic tools can help reduce unit costs. Second, translation tools can also reduce the time it takes to bring the product and its related documentation to the market and the end users. Third, automatic tools are currently being used to compensate for gaps in some areas of competence of human translators, such as lexical knowledge. Finally, one of the main perceived benefits of these tools is that they promote, if not enforce, consistency of terminology and, potentially, syntax.

In the next section, we briefly review the type of tools employed and their limitations in term of the support they provide. This sets the background against which to view both significant new trends emerging in the language industry and multilingual generation.

3 Translation Tools

The most widely employed tools are translation aids such as on-line dictionaries and automatic term-replacement, and translation systems that attempt sentence processing, either by rule-based or by empirical methods. In the latter case, the output is revised by the translator insofar as its quality does not meet the reader's requirements.

The first class of tools addresses an important but small part of the translation process. The other class takes over more of the process and, potentially, more of the effort. Understandably, however, it may also present more problems. The decision to use such a tool depends on the amount and kind of effort required on the part of the revisor and on an assessment whether the particular division of labour between human and machine results in an overall increase in productivity. The user of these kinds of support tools is the *translator*, who has sufficient do-

ISSN: 2456-1134 www.isjcresm.com
Vol-10 Issue-01 Mar 2025

main knowledge to understand the input text, together with the bilingual knowledge necessary to check that the output is consistent with the in- put. Thus, there is a duplication of competence across system and human user, both possess- ing source-language, target-language and con- trastive knowledge.

4 Current tools and practice: limitations

There are limitations to the support that current translation-oriented tools offer. Clearly, term re- placement systems, for example, are limited by design. However, even translation tools have se- rious limitations, one of which is the potentially poor quality of the output. Another is that the systems themselves are geared to producing in- terdependent texts; this relationship between the status of source and target texts is maintained by default by a strategy of minimal revision aimed only to enforce the well-formedness of the out- put.

In rule-based systems, these shortcomings stem either from difficulties in understanding the input or from the fact that such systems tend to be structure preserving.[2] Problems in the analy- sis stage arise when the coverage of the analysis component is insufficient. These are especially acute when the input text is itself of 'poor' qual- ity, that is, either ambiguous or deviating from the expected patterns of the sublanguage. In general, during analysis, incomplete representa- tions of the input text are obtained, and these are not sufficient to control transfer and genera- tion. Furthermore, the representations obtained are representations at the sentence level, which take into account only limited information about discourse phenomena and the intentions or pur- poses of the input text.[3] This entails that the

[2]For an exception to this trend, see (Mitamura et al., 1993).

[3]Some recent research in Knowledge-Based Machine

ISSN: 2456-1134 www.isjcresm.com
Vol-10 Issue-01 Mar 2025

system predisposes to the production of interdependent texts, which additionally have the same (discourse) structure as the input texts. This in turn entails that (1) the tools are best used only when it is acceptable to produce such interdependent texts and that (2) a relationship of interdependence is imposed on the texts when, in fact, a relation of *congruence* would have been preferable. By congruence, we mean that the texts have the same purpose (or function), but that the realisation of that purpose can differ at various levels of the linguistic system in order better to meet the readers' expectations of the text type. For example, the distribution of the information may differ, or the texts may be different in their discourse structure or their phrasing.

It is not only the current tools that are limited. The very way the task is organised currently engenders inefficiencies and is itself being called into question. First, the serial phasing of the authoring and translating processes makes it hard to reduce time-to-market. Second, the organisation of the task, the stricture to actually use the tools available in the workplace, and the inherent limitations of these tools (as described above) all conspire to produce documents in a relation of interdependence, when a relation of congruence might have been better. Finally, new documents are often being translated, even when they are largely the same as previous versions already translated (Clarke, 1994). For example, new versions of a manual may be produced which contain essentially the same information as the previous versions, but expressed differently. In some documentation departments, the new manual is finalised and re-translated in its entirety, although modular document design allowing targetted updating is intended to reduce the volume of text submitted for re-translation.

In summary, the main limitations of current practices are the poor quality of the output, due essentially to either the analysis or to structure

Translation (KBMT) aims at capturing discourse phenomena, such as theme (Nirenburg et al., 1992).

ISSN: 2456-1134 www.isjcresm.com
Vol-10 Issue-01 Mar 2025

preserving translation, a predisposition to interdependent texts and the need to re-translate. These limitations have been noted, and two new avenues are being explored. On the one hand, other tools have been developed to try to allevi- ate some of these problems; on the other hand, the document production itself is being changed.

5 Addressing the limitations

5.1 New tools

Two approaches have been adopted to remedy- ing problems at the linguistic analysis stage. One is simply to side-step the issue by using statistical string matching techniques, which in their pure form exploit no linguistic knowledge. Another approach is the use of controlled languages to constrain the input explicitly, and thus ease the burden of analysis and, in theory, ensure that analysis will not fail. In an integrated environ- ment – e.g., SECC (Adriaens, 1994) – offering both a controlled language critiquing module and a machine-translation system, it is intended that the principal user will be not a translator but a *monolingual technical author*. These are implementations of the idea of machine translation for monolinguals proposed by Johnson and Whitelock (1987). In this scenario, the human user brings to the task source-language knowl- edge and domain expertise, while the system possesses the contrastive linguistic and target- language knowledge. Thus, there is complemen- tarity rather than overlap of expertise. Having such tools available offers the opportunity for radically re-thinking the multilingual documen- tation task and the division of labour between human and machine.

To address the problem of re-translation, tools to promote 'text re-use' are being developed. In particular, *translation memory* is an empirical approach related to the statistical approaches described above, e.g., (IBM, 1994). The motivation for translation memory is to reduce the

ISSN: 2456-1134 www.isjcresm.com
Vol-10 Issue-01 Mar 2025

amount of text that needs to be translated by storing aligned pairs of previously translated sentence tokens and retrieving the 'matched' sentence in the target language.

5.2 New trends

A new conception of the task itself is also emerging, with the realisation that the goal of having multilingual documentation does not entail translation. Some companies are now cutting out the translation process and, instead, have introduced a working practice which they term 'parallel technical documentation': technical writers of different native languages are briefed at the same time, and write the documents more or less independently in the different languages, conferring with each other when necessary. This is designed to avoid delays inherent in the translation scenario, and also to ensure that from the outset all documents are biased to the expectations of their respective readership. This practice is a recognition that texts written for the same purpose in different languages can vary in content, structure and pragmatics. It represents a move away from interdependent translations to texts which, had they been translated, would have been said to stand in a *derived* relation to one another. A derived text is one which has been modified in purpose and/or content with respect to the source text (Sager, 1994).

This shift from translation to localisation, with its emphasis on reader-oriented writing, is also reflected in the new currency in the language industry of designations like 'Sprachvermittler', 'langagier' and 'language mediator' as opposed to ' Ü bersetzer', 'traducteur' and 'translator', and this quote from the director of a leading UK translation company:

> In the future, information technology is set to change language services dramatically … In the year 2010 … the job will not

ISSN: 2456-1134 www.isjcresm.com
Vol-10 Issue-01 Mar 2025

be translation but language process- ing. You might be a technician who operates between two languages, rather than a translator.

G. Kingscott, Praetorius Limited,
in (The Guardian, 1995)

6 Multilingual Generation

The novel dimension of multilingual genera- tion is the fact that text is constructed not from an existing text but from an underlying knowl- edge base, which must either pre-exist or be constructed. Multilingual generation was orig- inally viewed as a radical alternative to multi- lingual document production assisted by trans- lation tools. In fact, in the light of these new developments in industry, its radicality lies less in its philosophy than in the fact that it offers the prospect of technical support for the task as newly conceived.

Multilingual generation does not suffer from the same limitations as translation tools. First, since the texts are generated from an underlying knowledge base, problems of analysis do not arise. Second, there is no source text to con- strain the 'target' texts, and there is no neces- sary imposition of a relation of interdependence between the texts generated. Third, the texts are generated in parallel, thus reducing time-to- market. Finally, multilingual generation permits extensive knowledge re-use.

Knowledge re-use an important factor in doc- umentation production, as is evidenced by the fact that up to 85% of a new document may consist of unchanged or 'familiar' text (Birch, 1993).[4] Furthermore, many companies gener- ate large volumes of documentation which re- express the same knowledge in a variety of text

[4]Familiar text here refers to text which exactly or approximately matches previously translated text held in translation memory.

ISSN: 2456-1134 www.isjcresm.com
Vol-10 Issue-01 Mar 2025

types. For example, the aerospace industry is required to produce a crew manual, an operating manual and a workshop manual for every aircraft; automobile manufacturers produce owner's manuals and workshop manuals; software houses provide tutorials, reference manuals and user guides. A multilingual generation system with the appropriate linguistic knowledge could in principle generate these ranges of texts in a given domain from the *same* underlying knowledge base.

Note that multilingual generation is consistent with the most innovative practices in industry. Any company that has already introduced parallel technical authoring could integrate a multilingual generation tool into its workflow without disruption. Such a tool could be used by the technical authors, since, like the drafting tools for monolinguals, multilingual generation eliminates the requirement for users to have bilingual knowledge. However, it is equally possible to imagine its use by technical specialists (e.g., engineers) who are not trained as authors, because the system is endowed with the necessary linguistic knowledge (strategies and realisation). Thus multilingual generation avoids the need for duplicating knowledge between human and machine.

The reorganisation of the workflow and the availability of multilingual generation techniques open up the possibility for a company to:

produce a range of documentation types (including context-sensitive, on-line documentation) using no more resources than are required to produce a single type, since all texts are generated from the same underlying knowledge base;

ensure consistency at all stages between all the documents and the actual product;

generate textual representation of data not currently made available as texts.

These benefits hinge on the availability of a knowledge base. If such a knowledge base al-

ready exists or can be constructed automatically, then multilingual generation is a viable option for producing multilingual documents. This op- tion becomes even more attractive as the call for knowledge re-use increases, either because the knowledge base changes often, or because there is a need to produce congruent texts in the dif- ferent languages or different types of documents. If no knowledge base is readily available, then it must be constructed by technical specialists. In this case, the cost of building this knowledge base must be offset by savings on authoring and translation. Thus, the scope for knowledge re- use becomes the decisive factor in determining the appropriateness of multilingual generation. Obviously, there will be situations where ma- chine translation remains the preferred option: to enable browsing of texts which have been exter- nally authored and to produce multilingual ver- sions of documents which have previously been authored internally. In the future, however, there is no reason why a single platform to support multilingual documentation should not integrate translation-oriented tools and generation-based tools to be employed as appropriate by different types of users in different circumstances.

References

Adriaens, G. (1994). Simplified English Gram- mar and Style Correction in an MT Frame- work: the LRE SECC Project. In *Proceed- ings of Translating and the Computer 16*, pages 78 – 88, London, UK.

Bateman, J. A., Degand, L., and Teich, E. (1993). Multilingual textuality: Some ex- periences from multilingual text generation. In *Proceedings of the Fourth European Workshop on Natural Language Genera- tion, Pisa, Italy, 28-30 April 1993*, pages 5 –
17. Also available as technical report from GMD/Institut fü r Integrierte Publikations- und Informationssysteme, Darmstadt, Ger- many.

Birch, R. (1993). Future Translation Workbenches: Some Essential Requirements. In *Proceedings of Translating and the Computer 15*, pages 181 – 192, London, UK.

Clarke, B. (1994). Director of Praetorius Limited, Personal Communication.

Freibott, G. P. (1992). Computer Aided Translation in an Integrated Document Production Process: Tools and Applications. In *Proceedings of Translating and the Computer 14*, pages 45 – 66, London, UK.

IBM (1994). IBM Translation Manager (available commercially).

Johnson, R. L. and Whitelock, P. (1987). Machine Translation as an Expert Task. In Nirenburg, S., editor, *Machine Translation: Theoretical and Methodological Issues*. Cambridge University Press, Cambridge, UK.

Kosseim, L. and Lapalme, G. (1994). Content and Rhetorical Status Selection in Instructional Texts. In *Proceedings of the Seventh International Workshop on Natural Language Generation*, pages 53–60, Kennebunkport, Maine.

Language Technology Symposium (1993). Technology and Language in Europe 2000, London.

Mitamura, T., Nyberg, E. H., and Carbonell, J. G. (1993). An Efficient Interlingua Translation System for Multi-Lingual Document Production. In Nirenburg, S., editor, *Progress in Machine Translation*. IOS Press, Amsterdam, NL.

Nirenburg, S., Carbonell, J., Tomita, M., and Goodman, K. (1992). *Machine Translation: A Knowledge-Based Approach*. Morgan Kaufmann, San Mateo, CA.

Paris, C., Vander Linden, K., Fischer, M., Hartley, A., Pemberton, L., Power, R., and Scott, D. (1995). A Support Tool for Writing Multilingual Instructions. Submitted for Publication.

Power, R., Pemberton, L., Hartley, A., and Gorman, L. (1994). User Requirements Analysis. WP2 Deliverable, Drafter Project IED4/1/5827, financed by the Engineering and Physical Sciences Research Council (EPSRC) Grant J19221.

Reiter, E., Mellish, C., and Levine, J. (1995). Automatic Generation of Technical Documentation. *Applied Artificial Intelligence*, 9.

Rösner, D. and Stede, M. (1991). Towards the Automatic Production of Multilingual Technical Documents. Technical Report FAW-R-91022, Research Institute for Applied Knowledge Processing (FAW), Ulm, Germany.

Sager, J. C. (1994). *Language Engineering and Translation: Consequences of automation*, volume 1, Benjamins Translation Library. John Benjamins, Amsterdam.

Scott, D. R. (1993). Generating Multilingual Instructions. Contribution to the panel on "Instructions: Language and Behavior". In *Proceedings of the International Joint Conference on Artificial Intelligence (IJCAI-93)*, Chambery, France.

The Guardian (1995). Gift of Tongues. 18 February 1995.