

Scalability in MT Systems

Elisabeth Maier*, Anthony Clarke⁺

*Canoo Engineering AG, Kirschgartenstr. 7, CH-4051 Basel, Switzerland

elisabeth.maier@canoo.com

CLS Corporate Language Services AG, Elisabethenanlage 11, CH-4051 Basel, Switzerland

anthony.clarke@cls.ch

Abstract

In this paper we show why scalability is one of the most important aspects for the evaluation of Machine Translation (MT) systems and what scalability entails in the framework of MT. We illustrate the issue of scalability by reporting about an MT solution, which has been chosen in the course of a thorough hands-on evaluation and which in the meantime has been developed from a pilot system to a MT turnkey solution for mid- to large-scale enterprises.

1 Introduction

Many evaluations published in the literature focus on the translation quality of MT. With more and more systems entering the market and being integrated into large networks -- be it the Internet or Intranets and Extranets of large enterprises -- questions of operability and maintainability of systems come into focus. In this paper we elaborate on this evaluation aspect, explain in detail, for whom this aspect is important, what it entails and how a scalable MT architecture can be flexibly adapted to a changing situation. It is shown that the MT solution provided by CLS Corporate Language Services AG fulfils the scalability criterion and what this entails for the everyday operations of the system.

2 Why is MT system scalability an important issue?

To be able to respond to changing user and business requirements an MT system must be easily adaptable to new situations. In the following subsection we discuss various dimensions which influence the scalability of a MT system.

2.1 (Expected) machine translation volume

The introduction and operation of MT systems in the abovementioned settings bears a high degree of uncertainty: in the literature, only little information is available about the average volume of machine

translated texts in large enterprises (see [Petrits, 2001], [Flanagan, 2002]). Also, the throughput of such systems cannot be reliably predicted from the order volume of human translation as machine translation offers access to a new level of service, i.e. to the translation of emails, memos, etc. which are usually too costly and too time critical for a translation by humans. Predictions by the future users about their prospective use of the MT service are usually not reliable either as the usability of MT can only be estimated once the system is in place. In order to respond to such a situation, the system has to be scalable to be adapted to changing demands.

2.2 Number of users and sites

Over the years, companies often undergo substantial changes -- new departments are created, new branches in various locations are opened, new markets are explored, all of which must be readily equipped with Machine Translation services. Therefore, it must be taken into account that the MT infrastructure can be easily and efficiently extended to serve a growing number of sites and customers. Also, within a company of constant size, it often happens that an MT service is accepted gradually and over a long period of time, as word-of-mouth spreads and positive feedback is made available on corporate intranets.

2.3 Responsiveness to customer requirements

With the structure of corporations changing continuously and with new uses for MT being

defined once the system has been successfully introduced, new requirements have to be met which have to be assessed and possibly foreseen at evaluation time. Such new requirements concern:

- **Introduction of new languages.** Mergers and acquisitions of new companies as well as the extension of the market segment often require the addition of additional language pairs to an existing MT system. For example, the expansion of a company to also cover the American market often makes the introduction of English <-> Spanish and English <-> <Asian Languages> necessary.
- **Treatment of new document formats.** As MT gets to be an important part of the workflow in large enterprises the need arises to automatically translate a full variety of document formats; while MT systems typically translated .html and .txt formats in the past, the capabilities of MT systems have now been extended to also cover Office formats as .rtf, .doc, .ppt. Recently, efforts are under way to also handle .pdf files and, most importantly, .xml Files as XML is the basis of many document and exchange standards.
- **Integration with new applications and processes.** With MT becoming an essential part of the everyday work processes in a company, the wish arises to fully embed the system into a company's infrastructure. This entails a full integration into the company's accounting system, document production workflow as well as document storage and retrieval. It is therefore necessary that the system dispose of an open architecture with well-documented APIs, which can be employed for system integration and adaptation.

3 What are the Benefits of Choosing a Scalable MT System?

Scalability is of importance with respect to any mission-critical software system that is operated by large enterprises. If such a piece of software is unable to be adapted to changing or growing needs, it has to be either replaced by a new and better system or it has to undergo massive re-engineering. For the following reasons, both

possible solutions are unacceptable from the customers' point of view:

3.1 Costs

System changes that are made very late in the lifetime of software systems are among the most costly; moreover, if a system has to be replaced entirely, license costs are doubled as are costs of education, system introduction and rollout. In the phase of system replacement, the MT software packages have to be operated redundantly (both the old and the new system) in order to allow for front-up testing of the new system in a realistic environment and to ensure a smooth transition.

3.2 Continuity of Service

Regardless of extensive system testing, introduction of new software very often leads to interrupts in the service. Even under the best of circumstances, changes in the look and feel of the system and differences in the usage of the system lead to a (even if tiny) drop-back in the system use. Any discontinuities in the performance of an MT system may result in reduced acceptance of the overall system, which is a critical factor for any provider of MT services.

3.3 Time to Market

Problems in the adaptation of a system to new requirements result in a long time period, which elapses from the specification of the needs to the rollout of a new system or a new system version.

4 What are the Aspects of Scalability in Machine Translation Systems?

To ensure that a system is scalable, the following aspects have to be taken into account in the MT system evaluation:

4.1 Distributed modular architecture (vs. monolithic architecture)

A modular and possibly distributed MT system is typically composed of the following components:

- **Linguistic engines**, responsible for a specific translation pair or translation direction;
- **Lexicons**, both delivered with by the MT provider and user-generated lexicons;

- **Translation memories**, if available for the specific customer;
- **Communication infrastructure**, responsible for the communication with the end user, distribution of the translation jobs among the linguistic engines and delivery of the results.

4.2 Openness of architecture and accessibility of interfaces (vs. closed systems)

To be able to flexibly adapt an MT system to various environments the following possibilities should be available:

- **Adaptive system behaviour.** Depending on the user profile, the system must be open to minor manipulations of the system, e.g. to display translation alternatives, to prefer specific grammatical variants, etc. Such features are typically coded in one or a few centrally available parameter files, which are open for changes by the local admin staff.
- **Security.** As data security is one of the main reasons for large corporations to employ corporate MT, MT systems must provide interfaces and/or components for various types of security solutions, ranging from encryption to the use of various security protocols.
- **Logging and Accounting.** To offer flexibility with regard to different license and cost models the usage of such systems needs to be transparent. This implies that reporting functionality needs to be available for processing by various types of reporting tools. At the same time, APIs need to be available to attach accounting systems to the usage data. Depending on the cost model chosen, user authentication as e.g. based on an LDAP-based service, must be available to allow for billing of users.
- **Monitoring.** MT systems operated in large corporations are subject to high demands concerning availability. In multinational companies a 7x24 hour service with > 95% availability is a necessary requirement. This can only be achieved if the software is stable and contains mechanisms for automatic error detection and analysis, restart and fail-over. Hooks for monitoring software have to be provided.

- **Integration of 3rd party MT Knowledge Sources.** Last but not least, every system needs to be open for extension on the linguistic level: systems must be open for the integration of new dictionary data and translation memories. If these data are not available in the format required by the system, tools need to be available to support the import of 3rd party products (see below).

4.3 Compliance with standards (vs. proprietary solutions)

To allow for the easy exchange of data between various systems and various data providers, every MT system evaluation must take care that MT system developers participate in the development of exchange standards and implement these standards in new system releases. Minimally, standards concerning Translation Memory and Lexicon Formats, i.e. TMX [TMX] and OLIF [OLIF] should be supported. The implementation of standards also helps companies, which employ MT systems, to change to a different provider.

5 An example for a scalable MT architecture – the MT solution at CLS Corporate Language Services AG

We present an example for a scalable MT architecture as employed by CLS Corporate Language Services AG. The system, Compendium's Machine Translations solution (see [Thurmaier & Alonso, 2003]), has been chosen in a thorough evaluation of various MT systems, which has been published in [Maier et al., 2001]. The system was chosen both for

- its translation quality, which was assessed by both lay persons and linguistic experts, and
- its technical features, which have been evaluated during an installation and test phase. The scalability of the MT systems was one of the major technical evaluation criteria next to maintenance, support and reliability.

Another result of the evaluation phase was the development of a plan for the ramp-up of the MT solution based on the chosen system.

In the remainder of this section we show how the system evolved from a small pilot system into a full-scale enterprise solution, which is in operation worldwide and which can be easily and quickly adapted to serve new customers and new domains.

Over two years, the development of the MT suite went through various stages, which are described in the following subsections.

5.1 Pilot and productive use

Pilot Phase 1: 20 users. After evaluation, the chosen MT system was installed on a single server. Six language pairs were provided; the look and feel of the system was adapted to the needs of the prospective users. At the same time, the import of terminological data, available in the company's Multiterm database, was initiated. Load tests were carried out to assess both the potential maximum translation volume of the installation and to get an impression of the system's stability.

In parallel, a test system was set up, on which new software and lexicon versions are tested before their rollout in the productive system.

Pilot Phase 2: ca. 200 users. Based on the customer's feedback, the system's user interface was slightly changed; the stability was massively improved by adding monitoring and fail-over services. 30'000 financial terms per were imported semi-automatically for all language pairs. Mechanisms were added to log the translation volume with respect to the business units they originate from. To prepare for the first productive phase, the system was distributed over a server farm consisting of four servers with redundant linguistic engines for 6 translation directions. Again, load tests were carried out on the new extended architecture to assess the maximum translation volume.

After the second pilot phase a survey among all pilot users was made. The feedback was almost exclusively positive; a majority of the users said that their daily work would suffer if the system were no longer available.

Detailed results of the pilot phases described in [Clarke et al., 2002].

Productive Phase 1, 30'000 users. The possibility for single-user login and reporting was added to the system. 3'000 - 9'000 telecommunication terms per language were imported semi-automatically. Stability was improved once more and system availability is now well over 99%. Translation quality benchmarks are carried out on a regular basis. Extensive reporting facilities were set in place and are produced regularly.

Productive Phase 2, over 70'000 users. A dedicated server for overseas users with special IT requirements was set up. Additional translation pairs were introduced on all servers; lexicon data were imported accordingly.

The pilot and productive phases described here have been carried out for UBS AG, Switzerland's largest bank, which is operating worldwide. The system is currently being rolled out at additional customer sites following the principle of stepwise introduction as described above.

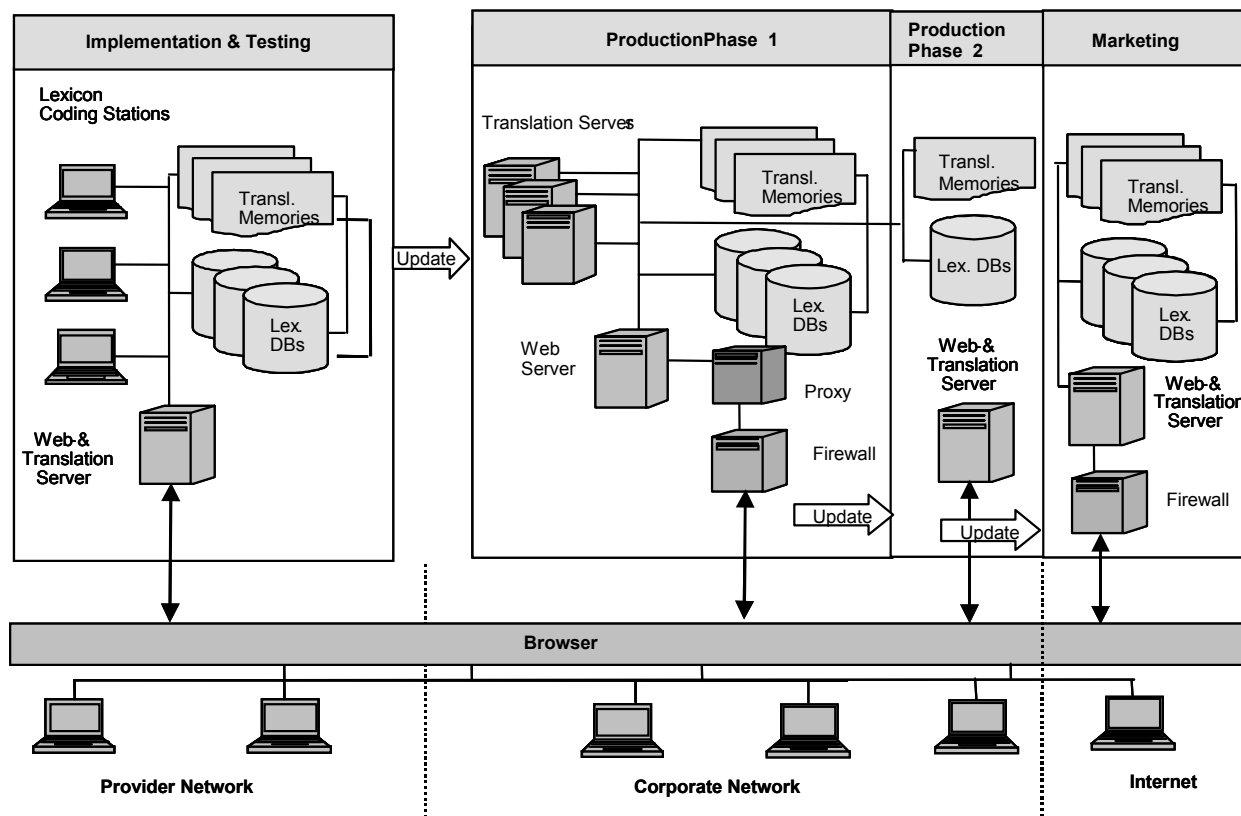
5.2 In-house use

Postediting: As an additional service level CLS Corporate Language Services AG recently started to offer postediting services on the basis of MT output (see [Hyland, 2003]). To this end, the MT test environment is used. It is planned to install a dedicated environment specifically for postediting purposes.

5.3 Marketing

Demonstration and evaluation: In order to allow test access to potential customers a demo-server was put in place at a very early stage. This demo-server is accessible from the Internet via user-specific accounts. As access can be provided to new customers within minutes this option is an important marketing instrument.

In the following figure the current MT architecture is displayed in graphical form.



thank the anonymous reviewers of this paper for their valuable suggestions.

6 Conclusion

In this paper, we have shown, why scalability is one of the key evaluation criteria for MT system evaluation. We also illustrated how a system that fulfils these criteria has been built up incrementally. The infrastructure, which is currently in place, allows

- New customers to test the system under realistic conditions,
- The development of fast and exact offers for potential customers,
- The rapid development and rollout of dedicated MT solutions to new customers, including value-added services as e.g. hosting, support and postediting.

7 Acknowledgements

The authors would like to thank all members of the Machine Translation Team at CLS Corporate Language Services AG. We would also like to

8 Bibliographical References

- Anthony Clarke, Elisabeth Maier, Hans-Udo Stadler. 2002. A Report on the Experiences of Implementing an MT System for Use in a Commercial Environment. In *Machine Translation: From Research to Real Users, 5th Conference of the Association for Machine Translation in the Americas, AMTA 2002*, Stephen D. Richardson, ed., Lecture Notes in Computer Science 2499 Springer 2002, pp. 187-194.
- Mary Flanagan. 2000. *Daimler Chrysler's Language Engine*.
<http://www.systransoft.com/Technology/DaimlerChrysler.pdf>.
- FEMTI. 2003. The FEMTI Classification of Machine Translation Systems,
<http://www.issco.unige.ch/projects/isle/taxonomy3/>,
 mirror site will be published as
<http://www.isi.edu/natural-language/mteval/>.
- Catherine Hyland. Testing "Prompt": The Development of a Rapid Post-Editing Service at CLS Corporate Language Services, *this volume*.
- Elisabeth Maier, Anthony Clarke, and Hans-Udo Stadler. 2001. Evaluation of Machine Translation

Systems at CLS Corporate Language Services AG. In *Proceedings of MT Summit VIII, Machine Translation in the Information Age*, Bente Maegaard, ed., pp. 223-228.

OLIF. Open Lexicon Interchange Format, <http://www.olif.net>.

Angeliki Petrakis. 2001. EC Systran: The Commission's Machine Translation System, see http://europa.eu.int/comm/translation/reading/articles/pdf/2001_mt_mtfullen.pdf.

TMX, Translation Memory eXchange standard, see <http://www.lisa.org/tmx/>.

G. Thurmair and J. Alonso. 2003. The Compendium Translation System, *this volume*.