Introduction

The output quality of Machine Translation systems has been always the central issue in MT research and development. However, it is not equally handled in different research paradigms. In recent decades Machine Translation became a commercially viable technology with a growing number of industrial applications. This was motivated by increasing usefulness of state-of-the-art MT systems in the workflow of professional translators due to the following developments: much wider lexical and grammatical coverage, integration of pre/post-editing and translation memory tools, the use of controlled language approaches, domain-specific terminological databases and disambiguation strategies, user dictionaries, etc. The "usefulness" of MT is no longer associated exclusively with the "quality" of raw MT output: there is recognition that even imperfect text produced by the systems can find its applications.

As a result there appeared two separate directions in MT research. The first one – the "perfectionist" direction – is treating MT as a "*venerable scientific enterprise*" and/or a "*technological challenge*" (Nirenburg and Wilks, 2000); to a large extent this direction is motivated by the idea of achieving "fully automatic high quality translation" (FAHQT). The second direction is "pragmatic"; it views MT as an "*economic necessity*" and is concerned primarily with the usefulness of existing systems and techniques, having conceded that very limited progress in MT quality is achievable. Still there is a gradual progress in text quality, and the advances in the "pragmatic" direction build up on extended capabilities of MT provided by the "perfectionist" route. However, the MT quality comparable to the quality of professional human translation (HT) has not been reached. In terms of MT evaluation scores there is still a huge gap between the quality of HT and MT. The disagreement between the "perfectionist" and "pragmatic" directions concerns the question whether MT quality can be made comparable to HT in a foreseeable future, or whether there is a ceiling for output quality, which MT may have already reached or is about to reach. There is an intuitive recognition that such limits exist, so new conceptual models are needed:

> The *prima face* case against operational machine translation from the linguistic point of view will be to the effect that there is unlikely to be adequate engineering where we know there is no adequate science. A parallel case can be made from the point of view of computer science, especially that part of it called *artificial intelligence*. (Kay, 1980).

> "To produce "good enough" output, an MT system has to be based on linguistically sound principles, at least in the long run. To get the linguistically sound ideas, we need to study language more intensively. […] If we are doing something we understand weakly, we cannot hope for good results. And language, including translation, is still rather weakly understood. Therefore, I think it is rather deceptive to give an impression that currently existing MT programs resemble closely, or even are equal to, FAHQT" (Kettunen, 1986: 37).

This very general criticism of MT doesn't identify any productive lines of research in linguistics, artificial intelligence (AI) or translation studies, which may yield improved MT quality. However, constructive attempts to identify limits of current approaches to MT often inspire new research directions: many technological developments in MT may be viewed as responses to such limits (Wilks, 2003: 204), e.g., recognition of the role of knowledge in Ontological Semantics (Nirenburg and Raskin, 2004) responds to Bar-Hillel's "demonstration of the nonfeasibility of FAHQT" (Bar-Hillel, 1960), i.e., accessing ontological knowledge can successfully disambiguate word senses in Bar-Hillel's example: *Little John was looking for his toy box…The box was in a pen*. Similarly, development of data-driven approaches such as statistical MT (SMT) and Example-Based MT (EBMT) try to overcome the data-acquisition bottleneck in MT technology. Therefore, systematic identification of current limits on MT is essential for achieving progress in output quality.

However, most demonstrations of MT limits are based on abstract theoretical lines of reasoning, on isolated and artificially constructed examples rather than on empirical corpus-based data. Little effort has been made to rank the importance of the problems and to identify which difficulties are most typical for the state-of-the-art MT systems. Inventories of possible improvements of MT often look like unstructured wish lists. Claims have been made that solution to the following problems are crucial and even play a vital role for MT: anaphora resolution (Mitkov, 2002: xii; 2003: 257), disambiguation, including word sense and syntactic disambiguation (e.g., McEnery, 2003: 459), term extraction (Jacquemin and Bourigault, 2003: 604), representations of the rhetorical structure of texts, of common-sense knowledge, etc. (Wilks, 2003: 203). However, yet there is no bird's-eye view on the problem which limits on MT are more serious and which are less serious; there is no empirical assessment of what impact particular aspects of NLP could have on MT quality. Without such assessments it is hard to identify scientific problems that are likely to provide best engineering solutions for MT: we just don't know what linguistic or AI issues need to be resolved in the first place to ensure considerable improvements in MT quality. There even is a suggestion by M. Kay that "even if all problems of syntax, morphology, and computational semantics had

been individually solved, it might not improve MT" (quoted in Wilks, 2003: 203). Such suggestion calls for empirical verification on corpus data, so there is a need:

- to systematically asses the exact impact of different NLP and AI technologies on MT;
- to identify theoretical and technological problems whose solutions will be the most important for improving MT quality, i.e., those problems that need to be solved in the first place.

These two tasks are related: corpus-based evaluation of the impact, which has a particular NLP technology on MT, also highlights its limitations and makes it comparable within a bigger picture, where it becomes possible to identify the relative importance of individual solutions and discover some missing, perhaps previously unknown technologies, which might appear essential for improving MT quality.

This thesis concentrates on the first of the two tasks mentioned above. I have chosen to investigate the impact of some aspects of *Information Extraction* (IE) technology on MT quality, particularly those aspects which were found relevant for MT: Named Entity Recognition (NER) and identification of salient terms (which typically are Named Entities (NEs), names of events, etc.) in text. The thesis aims at setting an example of how a systematic corpus-based evaluation of the impact on MT quality can be carried out for other NLP and AI technologies.

There are two aspects how a particular NLP/AI technology can be useful for MT. On the one hand it may be integrated into analysis and transfer modules of some MT system and *directly* contribute to improvements in MT quality. On the other hand it may become a part of some evaluation metric for MT, and highlight different problems which need to be solved by MT researchers. I will show that this *indirect* contribution of an NLP/AI technology to MT quality even more important than its direct impact on MT quality, because in this case we may discover some new, previously unknown facts and arrive at some unexpected empirical results, while with the direct impact we just measure the effect of some known approach. The thesis:

- proposes an MT evaluation framework based on ideas from IE;
- identifies technological limits for MT that can be revealed by the IE-based MT evaluation;
- suggests ways of improving MT quality with IE techniques, such as Named Entity Recognition, and outlines a wider IE-guided MT architecture.

To conclude, the thesis is an attempt to systematically identify some typical needs and perspective lines of improvement for the state-of-the-art MT systems, and to evaluate the effect of the proposed solutions on MT quality, which can be done using IE technology and corpus-based MT evaluation techniques. The thesis tries to show how it is possible to arrive at a bigger picture of limitations of MT quality from the perspective of IE and to empirically assess the effect of the suggested technological improvements using corpus data.