Dissertationen aus der
Philosophischen Fakultät der Universität des Saarlandes

Lexical Ambiguity in Machine Translation and its Impact on the Evaluation of Output by Users

Gabriel Armand Djiako

universaar
Universitätsverlag des Saarlandes
Saarland University Press
Presses Universitaires de la Sarre

Universitätsverlag des Saarlandes
Saarland University Press
Presses Universitaires de la Sarre
Postfach 151141, 66041 Saarbrücken

D 291
Dissertation zur Erlangung des akademischen Grades eines Doktors der Philosophie der Philosophischen Fakultäten I und II der Universität des Saarlandes
Dekan: Prof. Dr. Roland Marti
Berichterstatter: PD Dr. Peter Godglück
Prof. Dr. Ingo Reich
Tag der letzten Prüfungsleistung: 12.02.2016

ISBN 978-3-86223-231-4 gedruckte Ausgabe
ISBN 978-3-86223-232-1 Onlineausgabe
URN urn:nbn:de:bsz:291-universaar-1764

Projektbetreuung *universaar*: Susanne Alt, Verena Wohlleben

Satz: Gabriel Armand Djiako
Umschlaggestaltung: Julian Wichert

Gedruckt auf FSC-zertifiziertem Papier durch readbox unipress

Bibliografische Information der Deutschen Nationalbibliothek:
Die Deutsche Nationalbibliothek verzeichnet diese Publikation in der Deutschen Nationalbibliografie; detaillierte bibliografische Daten sind im Internet über <http://dnb.d-nb.de> abrufbar.

ABSTRACT

The artful use of ambiguity in politics and marketing presents many advantages, as ambiguity can help, both politicians and advertisers, navigate between several possible meanings while communicating effectively enough to not cause any controversy; however, ambiguity has often been described as a "key bottleneck for progress in Machine Translation" (Dale, Moisl and Somers 2000). While several studies have been centered on the description of ambiguity as a linguistic and philosophical phenomenon; there has been no study, to the best of our knowledge, that has measured its impact on machine translation output. This quantitative case study highlights the influence of ambiguity on machine translation's output quality in the judgement of users thus providing concrete data to measure this impact. In an empirical study, selected corpora containing lexical ambiguities were translated using different Machine Translation (MT) systems such as Google Translate and Personal Translator. A comparative, and contrastive, analysis using human judgement helped to measure users' judgement of quality and, by extension, the influence of ambiguity on MT output quality. To achieve this objective, the most common fluency and adequacy metrics were employed. Two sets of corpora were compared: firstly, corpora containing ambiguous sections; and secondly, corpora without ambiguous lexemes. Based on 10 users' judgement of MT output quality, we show that both fluency and adequacy metrics are negatively influenced by the presence of ambiguous words. Our experiments, also, shed light on the fact that no absolute correlation exists between fluency- and adequacy scores. Lastly, this dissertation also includes a comprehensive survey of different forms of ambiguities and ambiguity resolution techniques in MT.

ZUSAMMENFASSUNG

Der kunstvolle Einsatz von Mehrdeutigkeit in der Politik und im Marketing hat viele Vorteile, weil Mehrdeutigkeit Politikern und Werbeagenturen dabei helfen kann, zwischen mehreren möglichen Bedeutungen zu navigieren und gleichzeitig effektiv zu kommunizieren, ohne Kontroverse auszulösen. Jedoch ist Mehrdeutigkeit in der maschinellen Übersetzung häufig als ein „wichtiger Engpass für Fortschritt" beschrieben worden (Dale, Moisl and Somers 2000). Obwohl viele Studien sich auf die Beschreibung der Mehrdeutigkeit als linguistisches und philosophisches Phänomen konzentrieren, hat bislang keine Studie die Auswirkung von Mehrdeutigkeit auf maschinell übersetzte Korpora gemessen. Diese quantitative Fallstudie setzt es sich zum Ziel, den Einfluss von Mehrdeutigkeit auf die Qualität von Übersetzungsergebnissen zu untersuchen und liefert somit konkrete Daten, um diese Auswirkungen zu messen. In einer empirischen Studie wurden ausgewählte Korpora, welche lexikalische Mehrdeutigkeit enthalten, mit unterschiedlichen M.Ü.-Systemen wie Google Translate und Personal Translator übersetzt. Dank einer komparativen bzw. kontrastiven Analyse menschlicher Urteile konnte die Qualitätsbewertung gemessen werden und demnach auch der Einfluss der Mehrdeutigkeit auf maschinell übersetzten Output. Zu diesem Zweck wurden Metriken wie Flüssigkeit und Adäquanz, die in den meisten Studien verwendet werden, eingesetzt. Zwei Reihen Korpora wurden verglichen: Korpora, die mehrdeutige Abschnitte enthalten und Korpora ohne mehrdeutige Lexeme. Basierend auf 10 Benutzer-Urteilen zur M.Ü.-Qualität zeigen wir, wie die Metriken der Flüssigkeit und Adäquanz durch Mehrdeutigkeit negativ beeinflusst werden. Unsere Untersuchungen zeigen, dass keine absolute Wechselbeziehung zwischen den für Flüssigkeit und Adäquanz gemessenen Werten besteht. Abschließend umfasst diese Arbeit zudem eine Übersicht unterschiedlicher Formen von Mehrdeutigkeit sowie Disambiguierungsansätze in der maschinellen Übersetzung.

To my son Aaron

ACKNOWLEDGMENTS

I thank Prof. Dr. Johann Haller and Prof. Dr. Erich Steiner who have provided me with a wealth of insight and expertise that greatly assisted me with this research.

I thank the "*Universitätsgesellschaft des Saarlandes*", particularly Prof. Dr. Torsten Stein for their allowance that helped me attend the Machine Translation Summit XII in Ottawa, Canada. Attending this summit gave me the perfect opportunity to get an overview and gain some insight into the recent developments in the machine translation field.

I am indebted to my wife, Sandrine, for her unwavering support throughout this research endeavor.

Finally, to my family and friends. Thank you for your constant encouragement and support.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

1 General Introduction

The past three decades have borne witness to significant transformations in the scientific, economic, political and cultural domains. Globalization, has been the main driving force behind these changes. One of the main paradigms of globalization is, undoubtedly, the Internet. The Internet has transformed our planet into a vast global village, in which everything is interconnected. As a consequence, hereof, access to information has never been so easy and so fast, at any other point of time in history. Furthermore, the volume of human knowledge is increasing exponentially and more often, than not, this new information is only available in a select few languages. For instance, most scientific reviews are only published in the English language. Those who need this information for their own research must either: learn the language, presently English; or have the English documents translated into their first language. There is no need here, to stress the importance of translation in the dissemination of knowledge. As a matter of fact, the linguistic landscape of our planet is a very diversified and complex one. There are more than 6000 languages (Crystal 2003) and most of the essential information only available in English. Making this information available to users of other languages is a major challenge. Moreover, with the development of science and the increased expansion of human knowledge, the volume of documents to be translated has increased tremendously. International organizations, such as the United Nations and the European Union, often implement a language policy whereby all official documents must be made available in the languages of the constituent members, (which would make 23 languages for the European Union). The traditional human translation alone, however important it is, cannot arguably meet all the translation needs that have emerged as a result of globalization, given how fast new information is released and how quickly this information has to be made available to an ever-increasing number of people. In this information age, in which people rely extensively on technology to solve all kinds of problems, the scientific community has yet to successfully design tools that can help break down all of the language barriers.

With this in mind, translators -but also the scientific community as a whole- are faced with at least two major challenges: firstly, how to handle the ever-growing need for translation; and secondly, how to keep pace with the pressure to make all information readily accessible. During the industrial revolution, the mechanization of tasks helped meet the overgrowing demand for goods. Can the 'mechanization of translation' help meet these translation demands? At least, for some, this premise has been the underlying justification for research in machine translation (MT). For more than fifty years, the scientific community has been trying to develop tools that can either accelerate the translation process, such as translators' workstation, or function as an alternative to the human translator, to achieve Fully Automatic High-Quality Machine Translation (FAHQMT), see (Bar-Hillel 1964). The present study will, therefore, take an in-depth look at machine translation and review some of the major developments that have occurred in this field in recent years. Furthermore, this research will also report different evolutions in the field of machine translation. To this end, different approaches, to machine translation, shall be reviewed namely the linguistic approach to MT which helped design early machine translation systems, but also the statistical approach to machine translation which has revolutionized the body of knowledge. Recent trends such as the hybrid approach shall also be dealt with, albeit only briefly.

Since the first public demonstration of the Russian-German prototype machine translation models in 1954 to the latest translation software, major technical and methodological advances in the field of MT have been made. There has been a shift from the rule-based methodology which dominated research in MT during the early days, to the statistical approach which represents the new generation in MT methodology (Stein 2009). The above the question of methodology, users are usually more interested in a translation tool that works for them; professional translators are interested in getting effective help in terms of cost and quality. A fast review of most commercial and experimental MT systems will reveal that the linguistic aspects of machine translation still constitute a big challenge. The linguistic aspects of MT should not be understood here in terms of the approach implemented; but rather, it refers, in this context, to the grammaticality and acceptability of the machine translation output.

Saving time can be identified as one of the main advantages of machine translation. Translation software is able to translate large amounts of documents in a relatively short period of time. For multinational companies, or international institutions, such as the United Nation or the European Union, which require many documents to be translated in various languages in a very short time, any MT system that can help reduce translation costs and the meeting of deadlines is also another valuable asset. To date, machine translation systems are faster than human translators, still human post-editing is usually inevitable, if the machine translation output is to be published. Speed is therefore not the only criterion that enters into play when choosing an MT system.

If a translation system is fast, yet produces results that cannot be readily exploited, further additional time, and cost, has to be spent in editing the output, thus the usefulness of such a tool becomes questionable. An important part of the present research endeavor shall, therefore, be devoted to the quality of MT output.

1.1 Objectives of the work

It would be an understatement to assert that most machine translation systems currently available on the internet often produce unintelligible or unidiomatic output. In fact, most fail to successfully handle ambiguous sentences, thus producing translations that sometimes cannot be used. Ambiguity resolution has been one of the most difficult issues to solve in MT since the beginning of research in this field, (Bar-Hillel 1964). It would, therefore, be interesting to see how this linguistic phenomenon affects MT quality and what solutions can be implemented to resolve ambiguity in MT. This research endeavor therefore: firstly, sets out to explore how source text ambiguity affects the understandability of MT output; and secondly, what disambiguation approaches can be successfully implemented. Using a survey, some current machine translation systems shall be evaluated in terms of their efficiency with concern to the fluency and adequacy metrics.

As has been mentioned earlier, there are two main approaches to machine translation, namely the rule-based approach and the statistical approach. Early research on MT largely implemented the rule-based approach which is based on a large linguistic description following the assumption that the more linguistic information is available on a language, the better the MT system would perform. On the other hand, Statistical Machine Translation (SMT) which emerged as a response to the complexity of the rule-based approaches does not rely on grammatical rules to generate translation; rather, it relies on probability of co-occurrence. Statistical approaches operate through statistical models or algorithms that enable the production of translations. Linguistic knowledge here plays only a subsidiary role if at all. Taking these two approaches as a backbone to the current study, a contrastive analysis shall be carried out to help showcase the many evolutions that have shaped the field of machine translation. In recent years, the so-called hybrid approaches have enriched the machine translation paradigm offering new prospects for the future, this too shall be briefly discussed.

1.2 Research problem

Six decades ago, interest for machine translation was diminishing due to the disappointment that followed from the first experiments. Some prominent researchers, Bar-Hillel among others (op. cit.), who had support for machine translation early on, were now providing arguments as to why machine translation would not succeed. Failure to develop MT systems able to produce a *fully automatic high-quality translation* provided Bar-Hillel with arguments for casting doubt on the viability of machine translation. In his now famous example, "*The box was in the pen*", (Bar-Hillel 1964) argued that no machine translation would be able to determine the sense of *pen* in the above-mentioned sentence as taken from the following context:

Example 1[1]

Little John was looking for his toy box. Finally, he found it. The box was in the pen. John was very happy.

Following is a table showing how some current MT systems handle this ambiguous sentence more than five decades later.

Source Text	Google Translate	Personal Translator	Human Translator
Little John was looking for his toy box.	Little John war für seine Spielzeugkiste suchen.	Kleiner John suchte seine Spielzeugkiste.	Little John suchte seine Spielzeugkiste.
Finally, he found it.	Schließlich fand er es.	Schließlich fand er es.	Schließlich fand er sie.
*The box was in the **pen**.*	Die Box wurde in die ***Feder***.	Der Kasten war im ***Stift***.	Die Schachtel war im ***Laufgitter***.
John was very happy.	John war sehr glücklich.	John war sehr glücklich.	John war sehr glücklich.

Table 1: Disambiguation of "pen" 50 years later

From Table 1, it can be observed that despite new methodologies and five decades of research in MT, most systems still cannot successfully disambiguate

[1] Borrowed from (Bar-Hillel 1964)

"pen" as it appears in the above-mentioned context. In fact, ambiguity resolution has been one of the trickiest linguistic aspects of MT since research began in this area. Since the ALPAC report (ALPAC 1966), which some argue, was the main reason for defunding most research on machine translation (Ide and Véronis 1998), a lot of technological progress has been made. New methodologies have been developed and to date, most systems generate translations that are far better than the translations that could have been obtained five or even three decades ago. Sadly, as the example above illustrates, some aspects of machine translation seem to have failed to catch the attention of research; therefore, little advancement seems to have been made in these areas. The present study intends to review this development.

1.3 Scope and structure of the work

The present work falls within the framework of research in Applied Linguistics. Machine Translation is an interdisciplinary field that extends from Applied Linguistics, Artificial Intelligence up to such fields as Computational Linguistics. As the name suggests, Machine Translation also has a lot to do with translation *per se*. The present work will be divided in three parts. The first part will be theoretical and focus on the scope of the research shall be determined, in relation to translation studies and computational linguistics. This is where the addition to machine translation's body of knowledge will take place. As such, a brief historical review of MT shall be conducted including, the different methodological and technological developments that have shaped this field, over the past six decades, and how this shaping has occurred. Also discussed in the first part, are the differences and commonalities between human and MT.

In the second part, the notion of ambiguity shall be tackled from a linguistic perspective and, furthermore, a typology of ambiguities shall be established. Additionally, CAT2, an experimental rule-based MT system shall be used to illustrate some of the intricacies of ambiguity resolution in a rule-based environment. CAT2 is a transfer-based machine translation system that has been developed at Saarland University. It is mostly being used as an experimental system and is also a teaching tool for students of machine translation. This section will also serve as the setting to begin the discussion of ambiguity as a "translation problem". The strategies implemented in resolving ambiguity in both human and machine translation shall be examined. To this extent, an experimental and evaluative study shall be carried out in what will constitute the empirical section.

In the abovementioned empirical section, selected corpora containing lexical ambiguities shall be translated using different MT systems such as Google Translate and Personal Translator. These MT systems shall be presented succinctly in subsequent chapters. The results of the experiments shall be subjected to human evaluators and a comparative and contrastive analysis thereof shall be conducted. The aim here will be to measure users' judgement of quality, and by extension, the influence of ambiguity on MT output quality.

When machine translation is mentioned, most people cannot accurately tell what it refers to. A further point being the fact that more often, than not, people see Computer Assisted Translation (CAT) and machine translation as the same thing. The present research endeavor does not deal with localization; nor with CAT, notwithstanding the fact that these concepts will be referred to at several stages. This study does not set out to develop a new methodology for MT; nor how to develop a new MT system. As was pointed out above, the research sets out to mainly explore the question of: how source text ambiguity affects the understandability of MT output; and what disambiguation approaches can be successfully implemented. MT output evaluation also constitutes an important part of this study. The next section briefly deals with the methodology implemented for this study.

1.4 Methodology

The present paper will follow the research methodology in the humanities, that is, the results will be interpretative and analytical. First, and foremost, it is important to emphasize that this work falls within the framework of basic research, since its stated purpose is simply the exploration of a field and the acquisition of new knowledge. It is hoped, that the results of the present research endeavor will not only contribute to answering some questions in the field of MT, but more importantly, that it shall help lay some groundwork for further research in this area. Based on the method of case study, a comparative and contrastive analysis of two specific approaches to MT will be carried out. This research will, as has already been mentioned in our introductory remarks, evaluate translation output. To do this, selected ambiguous corpora shall be submitted to machine translation. Since this research has a comparative and contrastive component, MT systems implementing various approaches, namely: rule-based and statistical approaches shall be selected for the experiment. This comparative and contrastive study shall help determine whether, or not, the statistical approach upholds its promise of improving the fluency of translation.

Google Translate and Personal Translator are the software selected for the experiment. Google Translate and Personal Translator are two well-known translation software systems and have been extensively studied in the literature ((Barreiro, et al. 2014); (Koletnik Korošec 2011); (Zuo 2010); (Aleksić and Thurmair 2011) and (Thurmair 2005)).

Personal Translator has been chosen partly because it implements a linguistic approach, but also because of its success as a commercial MT system. Linguatech Personal Translator has experimented a hybrid approach that combines both RBMT and neural network, to be discussed in detail later. The second MT tool Google Translate, adopts a purely statistical approach to machine translation. It is freely available on the Internet and can be downloaded as a free application on smartphones. It has been chosen because of its growing popularity among internet and smartphone users.

To perform the experiments, selected corpora shall be submitted to English-French and German-French translations, using the aforementioned translation software. The corpora shall be made up of various texts displaying various levels of ambiguity, especially lexical ambiguity. Once the corpora are translated, a qualitative analysis of the results shall be carried out, that is, the translation shall be evaluated in terms of fluency and adequacy by human annotators.

The present work shall be arranged in two major parts, namely: a theoretical part, where the theoretical implications of machine translation shall be dealt with; and an empirical part, which focuses on users' assessment of machine translated output.

The theoretical part shall consist of a further two sub-sections: The first being the related literature shall be reviewed in chapter two; and the second being a brief historical overview of the MT field shall be carried out with a focus on the current state of research and the various approaches to machine translation. In chapter three, general linguistic problems in machine translation shall be discussed, such as lexical and structural mismatches, multiword units and compound words. These linguistic problems seem to pause the same kinds of difficulties to Machine Translation as does lexical ambiguity. A large section shall be devoted to ambiguity and its different forms in 3.6. Not only shall the general theoretical aspects of ambiguity be examined, but a test suite shall be used to illustrate some of the difficulties rule-based systems such as CAT2 has handling ambiguous segments.

In chapter four, the evaluation of machine translation output will build the cornerstone of this empirical and evaluative part. Evaluation is becoming an increasingly integral part of machine translation. In recent years, the focus seems to have shifted to developing a standard for MT evaluation, especially, as far as automatic evaluation is concerned. The focus in the present study shall be on human evaluation and various schools of thought shall be presented. Chapter four will also, be where the main experiment will be carried out to this end, machine translated output will be submitted to human evaluation. The results will be analyzed in terms of their fluency and adequacy as these metrics help determine machine translation quality. Chapter five will then summarize the outcome of the human evaluation and section 5.6 shall be devoted to some of the solutions that can be implemented to disambiguate and better translate an MT input. This chapter will end with a comparative analysis of both human and machine translation. Chapter 6 will act as the conclusion where recommendations shall be made and thereafter move to the suggestions for further future research.

2 State of Research

In this chapter, the development of research in Machine Translation will be presented with a focus on the relationship between Machine and Human Translation. Some theoretical aspects of Machine Translation will also be discussed. The last two sections of this chapter address the current state of research in Machine Translation, especially hybridization which reconciles two approaches which have long been opposed. Some related researches shall also be mentioned.

2.1 Human vs. Machine in Translation

Human and machine translation can be examined contrastively at various levels ranging from the theoretical to the practical aspects. The relationship between human translation and machine translation is marked by a continuum which is characterized by a more, or less, pronounced human/machine intervention (Hutchins and Somers 1992). On the one side of the continuum, there's a human translator whose linguistic skills seem to be the determining factor in the translation process; and on the other side, there is a fully automated translation with no human intervention needed. Human translation falls within the framework of Translation Studies while Machine Translation on the other hand falls within the framework of Computational Linguistics and Artificial Intelligence. These two disciplines are, however, intertwined in terms of the goals and in terms of their approaches.

Translation Studies primarily deal with theories of translation, that is, the "appropriate translation methods for the widest possible range of texts or text-categories". (Newmark 1981, 19). These "appropriate methods" can be summed up into theories of translation. The practices and theory of translation have evolved dramatically throughout the years and have generated various approaches and/or schools of thought. Historically speaking, translation was performed only in regard to official and religious documents. Early known translations followed the so-called word for word approach. Translation Studies in the mid-twentieth century have long been dominated by linguistic theories, which (Cheung 2013) refers to as the "linguistic era". One very interesting theory of this period comes from (Nida and Taber 1969), with their theory on functional or dynamic equivalence, which itself was inspired by Chomsky's generative-transformational grammar theory (Chomsky 1957).

Nida and Taber's linguistic theory identifies at least two dimensions in translation, referred to as "formal equivalence" and "dynamic equivalence" The first refers to a translation that respects the grammatical structure (surface structure) of the SL. The "dynamic equivalence" meanwhile will focus on the "deep structure" i.e., the underlying meaning; therefore, the translator is released from the formal structure of the source language and can recreate an equivalent structure in the target language.

The field of translation has experienced an important shift in the latter half of the twentieth century when scholars such as (Holmes 1988/2004) began to advocate the emancipation of translation from the field of linguistics which hitherto was the framework in which translation was studied. Holmes is often referred to as the father of translation studies. Following Holmes' early view on translation studies, modern translation theories tend to distance from the linguistic approach and adopt a more functionalist approach with (Gentzler 2001) as one of the main representatives, *Cf.* (Cheung 2013). The functionalist approach advocates a move from "source-text oriented theories" to "target-text oriented theories" in other words, the translation process is primarily considered in terms of its functionality. As a consequence, cultural and geographical elements of the target text play a vital role in the translator's shaping of the output. This approach is widely used in the localization industry where the aim is to offer contents that are tailored to the target audience. The functionalist approach to translation was further developed within what is commonly referred to as the "*skopos* theory" (*Skopostheorie*, in German) by (Reiss and Vermeer 1984) who state that: "Geglückt ist eine Interaktion, wenn sie vom Rezipienten als hinreichend kohärent mit seiner Situation interpretiert wird und kein Protest, in welcher Form auch immer, zu Übermittlung, Sprache und deren Sinn (‚Gemeintem') folgt" (Reiss and Vermeer 1984, 112).

A successful translation is, therefore, one that fulfills its function as expected from the commissioner. The focus is no longer placed on the "formal equivalence", which refers to the intricacies of the source language, but on the receiver whose primary objective may be to be informed of the general content of the source text.

The functionalist approach to translation adapts very well to applications of machine translation. There is, therefore, no definitive answer as to whether machine translation is useful, or not, because once the "*skopos*" or objective of translation is determined (dissemination, assimilation, interchange, information access) a fully automatic machine translation may well be adequate even if the surface structure of the output does not match the SL surface structure. In some cases, machine translation, even when it contains mistakes, is sufficiently clear to enable the reader to draw out the essential information. This scenario occurs when information is quickly needed and should be readily available especially, in the context of a press review. Here, the essence of the communication/information is more important than any grammatical or stylistic errors that may have occurred during translation; therefore, investment in human translation may be completely unnecessary. To put it in Reiss' words, "if the target text fulfills the *skopos* outlined by the commissioner, it is functionally and communicatively adequate". (Reiss and Vermeer 1984) cited in (Munday 2008, 80).

Even though modern practices of human translation have moved away from linguistics theories, there continues to be an interaction between machine translation and Linguistics especially, as far as the rule-based approach to machine translation is concerned.

While human translation has been performed for centuries, machine translation is a relatively new field and has primarily to do with the technological aspects of translation. Technological, in the sense that software is integrated into the translation process. Consequently, a great deal of programming is needed to help the computer emulate the human translation process. Early models of machine translation implemented the word-for-word approach which relied on a morpho-lexical analysis of the SL. Early translation systems were equipped with a bilingual dictionary and translating actually meant establishing a lexical equivalence between source language and target language, which was basically the same strategy implemented in literal human translation. This resulted in translations of very poor quality because the machine translation systems could not conduct syntactic analysis necessary for a better 'understanding' and therefore, a better rendering of the underlying meaning of the source text. As a matter of fact, relying only on the morpho-lexical analysis of the SL text may help understand the lexeme of the source text taken one by one and out of context, but obviously, translating is far more than replacing a SL word with its TL equivalence.

Thus, human translation and machine translation have both been largely influenced by linguistic theories, at least in the early stages of their development; however, these two disciplines have evolved over time to adopt a more pragmatic approach. As far as human translation is concerned, this theoretical evolution gave rise to the functional approach which is free from the linguistic constraints imposed by the "source-text oriented approaches" to focus on the "*skopos*", that is, the goal set forth by the receiver of the translation. Machine translation has also witnessed a shift from the purely linguistic approaches, adopted in early experiments, in machine translation; to an approach based on a detailed description of linguistic rules, modern MT has moved to a pragmatic approach, which heavily relies on data collection. The quality of translation will therefore depend on the size of the available databases as well as the robustness of the translation system. It should however be pointed out that the break between machine translation and Linguistics is not completely consummate since current research in MT are mostly oriented toward the hybrid approach, which combines the benefits of the statistical approach with the benefits of linguistic approach.

This section has highlighted some of the differences between machine translation and human translation. Depending on the purpose of translation, either machine or human translation can be used separately or conjointly. Despite the difference in approach and goal between MT and Human Translation (HT), these two disciplines should not be seen as opposing each other. Human translation would gain much in paying more attention to machine translation since the latter, is becoming more and more prominent among translators. A 2008 study shows that over 70% of translators and translation agencies report using CAT (Computer Aided Translation) technology, (LT2013 2013).

Advances in MT research in recent years have certainly changed the approach to translation altogether. These advances have also helped translators to be more productive and certainly more effective. New strategies for solving translation problems could be achieved as a consequence of the integration of translation software in the translation process. Ignoring all these changes equates to the refusing to recognize the developments in the science of translation. Inversely, machine translation should learn from the many theories developed in translation studies because machine translation cannot be thought of outside any theory of translation. (Hauenschild and Heizmann 1997) provide a complete overview of the relationship between human translation, machine translation and interpreting.

2.2 Theoretical aspects of machine translation

In this section, the focus will be on the theoretical aspects of machine translation. The two dominant approaches in MT architectures shall be discussed, namely: the knowledge-based and the data-based approaches. Machine translation architectures can be classified into two main categories: 'knowledge-driven'; and 'data-driven' systems, which also refer to rule-based systems and empirical systems respectively. Knowledge-driven systems implement the classical MT architectures and they correspond to the first and second-generation MT, namely direct translation and rule-based MT.

Data-driven systems on the other hand refer to a more recent trend in machine translation and they correspond to the so-called third generation machine translation. (Quah 2006, 68) categorizes MT architectures in terms of generations as illustrated in the following figure:

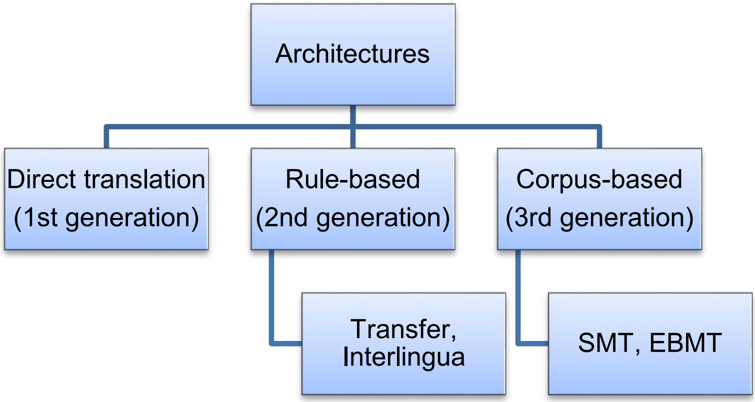

Figure 1: Machine translation architectures

Machine translation is the first non-numerical application of computers (W. J. Hutchins 1986, 16) and although it was mainly researched in the field of Computer Science, the linguistic theories that dominated the field of translation have shaped the early architectures in machine translation. As a consequence, much thought and effort were put into developing rule-based approaches that were thought to emulate the human brain ((Trujillo 1999); (Bhattacharyya 2015)). The rule-based machine translation paradigm can be subdivided into three architectures, namely transfer-based machine translation, ((Sharp 1994)); (Nirenburg and Somers 2003)), Interlingua machine translation (Zhang 1993) and dictionary-based machine translation or the direct approach ((Wheeler 1984); (B. J. Dorr 1993)). The late 1980s however has witnessed the development of more empirical methods that starkly contrast with the early linguistic approaches. Contrary to rule-based approaches, empirical methods to machine translation are data-driven, meaning that translation is not directly generated from a set of rules; instead, it is modeled on an existing corpus that is used as a basis for determining which translation is most plausible. Example-based machine translation (EBMT) ((Carl and Way 2003); (Dietzel 2009)) and Statistical machine translation (SMT) ((Manning and Schuetze 1999); (P. Koehn 2010); (Petrzelka 2011)) are representative of this approach. SMT has gained much interest in recent years following rather promising results. (W. J. Hutchins 1986) and (Hutchins and Somers 1992) provide a historical overview of machine translation as well as the linguistic implications of machine translation. (Arnold, et al. 1994) also break down the linguistic implications of machine translation with a focus on meaning representation and processing.

2.3 Recent developments and current status of MT

As the different approaches to machine translation were reviewed, it was pointed out that this field has undergone various evolutions in the methodological approach. The knowledge-based approaches have dominated early research in machine translation and therefore the first systems adopted rule-based architectures. Very soon however, the knowledge-based approaches started to expose their limitations as far as development time and cost are concerned. The lexical "bottleneck" also seems to be one major problem in this approach (Arranz 1997). In addition, knowledge-based approaches usually score poorly in adequacy measures compared to SMT (Eisele, et al. 2008). The desire to circumvent the obstacles of the rule-based approaches gave rise to the corpus-based approaches which require less development time since no linguistic expertise is required to build these systems; the main focus being the necessity to build a robust bilingual aligned corpus to generate translation. However, corpus-based MT approaches also have some disadvantages such as a strong dependency on large-scale bilingual corpus which are not available (as for yet) for many languages. Example-based MT for instance largely depends on the size of the example database. The following table from (Eisele, et al. 2008) summarizes strengths and weaknesses of both Rule-based and Statistical machine translation.

	RBMT	SMT
Syntax, Morphology	++	--
Structural semantics	+	--
Lexical semantics	-	+
Lexical adaptivity	--	+
Lexical reliability	+	-

Table 2: Advantages and disadvantages of RBMT and SMT

Given the aforementioned limitations of both the corpus-based and the knowledge-based approaches, various attempts have been made in recent years to capitalize on the advantages of both approaches. System combination or hybridization seems to offer this possibility. Hybridization in machine translation refers to the latest generation of machine translation. This approach results from the premise that both knowledge-based and corpus-based approaches may not be perfect, yet they seem to be complementary in various ways. The aim of the hybrid approach is therefore to combine the advantages of both approaches in order to maximize their benefits.

The starting point in hybrid architecture is, generally, a baseline system that can be either knowledge-based or corpus-based. Hybridization, therefore, consists of integrating a different architecture to a baseline system. Two main methods seem to characterize the hybrid approach: Hybridization guided by RBMT and Hybridization guided by corpus-based MT. See (Costa-jussàa and Fonollosa 2015).

2.3.1 Hybridization guided by RBMT

Hybridization guided by RBMT refers to hybrid or combination approaches with a baseline system that is rule-based. One of the most common ways of combining RBMT and SMT is done by using RBMT output to constitute the parallel corpus to train the statistical model for SMT, ((Hu, Wang and Wu 2007); (Dugast, Senellart and Koehn 2008)). Linguatech Personal Translator is an example of a system that is originally rule-based, but in recent years, much research and effort have been put into integrating statistical approaches especially when it comes to erroneous input correction, subject area recognition, and word disambiguation (Aleksić and Thurmair 2011).

(Haugereid and Bond 2012) in their approach to hybridization use lemmatized and aligned parallel corpus to extract semantic transfer rules for a rule-based system. (Labaka, et al. 2014) implement a statistical method to calculate candidate translations and enrich the tree-based representation of the RBMT with more translation alternatives. Translation generation is the result of the most probable combination among the available fragments following the RBMT reordering. Also see (Streiter, Carl and Haller 2000) for a complete overview of first hybridization attempts, especially the integration of translation memories and termbases in Example-based machine translation systems.

2.3.2 Hybridization guided by corpus-based MT

Hybridization guided by corpus-based MT refers to a hybrid approach where the baseline system is corpus-based (Example-based or Statistical Machine translation). Hybridization can work either through system combination or a rule-based translation module can be added to the baseline system. In SMT architecture, the availability of a large corpus is instrumental in building a system that covers a wide range of areas and such a corpus is not always readily available, especially for domain-specific translation or languages with low resources. (Hu, Wang and Wu 2007) for instance, propose an architecture that combines an SMT with a RBMT whereby the RBMT is used to produce a synthetic bilingual corpus. The obtained RBMT output is subsequently used as training data for the SMT for alignment.

Some corpus may be difficult to translate just because of the inherent source language syntactic structures (languages with different word order) or just because of the long complex sentences. While RBMT usually handles syntax well, this has been shown to be SMT's weakest point (Eisele, et al. 2008). Many combination efforts are therefore focused on developing strategies that can solve the syntactic bottleneck in SMT.

(Chen, et al. 2007) propose an approach to syntactic reordering. This approach consists in a pre-processing stage for SMT whereby systematic differences between source and target language are established which enables a reordering of the source sentence words following the target language syntax prior to the alignment. In an English-German translation for instance, this reordering technique allows the placement of SL verbal elements in the positions within the clause they will have in the target language. Also see ((Li, Kim and Lee 2010); (Sangodkar and Damani 2012); (Gojun and Fraser 2012) and

(Huang and Pendus 2013)). Another technique consists in splitting long sentences into clauses and then translating the clauses separately followed by a reordering of the clauses as per the TL syntactic structure ((Jiang, Du and Way 2010); (Goh, Onishi and Sumita 2011)). Another area of syntactic difficulty is the possible long distance between a preposition and the head it modifies. (Shilon, Fadida and Wintner 2012) incorporate linguistic knowledge in SMT especially by pre-determining verb-preposition agreement. Implementing the aforementioned techniques usually yields better fluency and adequacy results as opposed to the baseline system. Finally, (Xiong and Zhang 2015) produce a seminal work on some of the linguistic challenges of SMT. The book also presents case studies on linguistically motivated statistical machine translation.

2.4 Related Research

Ambiguity has long been studied in the field of Semantics as it appears to be primarily a linguistic problem. However, early on, it also became a problem in machine translation as well and was even considered a "key bottleneck for progress in MT" (Dale, Moisl and Somers 2000). Given the dominance of linguistic driven research in the early days, early works on ambiguity in machine translation were, therefore, mostly descriptive and were tackled from a morpho-syntactic stand point ((Arnold, et al. 1994); (Hutchins and Somers 1992)). The description of ambiguity will be further developed in the next chapter 3.7.

Word Sense Disambiguation (WSD) has been the focus of research in recent years as non-linguistic approaches to Natural language processing (NLP) have been developed especially with the development of statistical techniques to translation. (Prescher, Riezler and Rooth 2000) investigate the use of probabilistic class-based lexica for disambiguation in target word selection with the use of contextual information. (Hearne and Way 2006) investigate ambiguity resolution within a hybrid MT environment. Numerous works have also been dedicated to specific types of ambiguities using statistical techniques. (Lefever and Hoste 2011) evaluate the translation quality of ambiguous nouns and show that integrating a WSD module improves quality. (Schäfer 2002) evaluates the machine translation of economic texts and investigates the suitability of domain specific texts for MT and the linguistic implications thereof.

(Ceccato, et al. 2004) present a large description of various types of ambiguities as the authors consider ambiguity identification to be instrumental in determining text quality. The authors also present a tool for ambiguity identification and measurement. This approach is primarily useful in pre-editing and the authoring process as the tool for ambiguity identification can better help author text that is meant for machine translation.

Given the wide variety of works which focus on ambiguity or word sense disambiguation, it is widely assumed that ambiguity influences machine translation output; however, there is little empirical evidence to support and characterize this influence. This contribution investigates the impact of ambiguity on both rule-based and statistical machine translation outputs. The next chapter shall describe general linguistic problems in machine translation with a focus on lexical ambiguity in the source text.

3 Some General Linguistic Problems in MT

This research was undertaken, recognizing from the outset that the linguistic aspects of machine translation would build the backdrop of this work. Machine translation, it has been pointed out, is an interdisciplinary field straddling between linguistics and computer science. As stated above, the focus of the present study lies more on the linguistics aspects of machine translation, even though some of the technical aspects have been described in chapter two. In this section, some linguistic issues in human and machine translation are examined in contrast to chapter two. A first distinction is therefore made between translation problems and then translation difficulties.

3.1 Translation problems and translation difficulties

One of the first to address this topic is (Krings 1986, 158) who categorized translation problems into two types: on the one hand, there are translation problems related to the understanding of the source text; and on the other, there are issues related to the translation of what was understood from the original text. Krings' underlying idea is to understand the background processes that translators use to make their decisions during translation. Simply put, what are the elements that inform and motivate the translator's choices? (Krings 1986), therefore, differentiates between reception problems (*Rezeptionsproblem*) and production problems (*Wiedergabeproblem*).

3.1.1 Reception problems

Reception problems deal with the understanding of the source text. A text may not be fully "received" or understood due to the translator's L1 deficiency. A word or a sentence, which is not well understood, may not be translated properly.

3.1.2 Restitution problems

Restitution problems refer to the difficulties a translator may encounter while trying to "restitute" in the target language what he has understood from the source language. In other words, his/her own interpretation of the source text. As an example, if the source text contains words that the translation is not familiar with, due for example to the latter's deficient L1 proficiency, this may end up influencing the way he/she ultimately translates the part of the text that has not been well understood. Restitution problems may then also be caused by a deficient L2 proficiency if the translator is unable to find the equivalent L2 formulation for a given L1 utterance. While this kind of difficulty may appear in any kind of text, technical texts seem to display such difficulties more often, than not.

Following Krings, (Nord 1990, 30) first introduced a distinction between problems of translation and translation difficulties. Unlike (Krings 1986), however, (Nord 1990) considered issues related to competence of the translator as translation difficulties. Translation problems therefore refer to language inherent problems, regardless of the person who undertakes the translation task. (Nord 1990, 32) further, categorizes translation problems into four subcategories: source text-specific translation problems; pragmatic translation problems; culture-pair specific translation problems and language pair specific translation problems.

3.1.3 Source text-specific translation problems

Source text-specific translation problems are a result of the individual style of a text. A poem, for instance, should not be translated the same way as prose would. Some texts are full of stylistic devices, while others may be littered with illustrations; therefore, a text can be categorized according to its content, its nonverbal elements if any, its lexis and its syntax.

3.1.4 Pragmatic translation problems

Pragmatic translation problems as its name indicates refer to the problems that may result as a consequence of the fact that a text may be produced in a specific context and for a specific purpose yet be translated for distinctly different purpose. This is usually the case with scientific literature. A scientific article usually targets specialists in its domain. If the article is to be translated for dissemination, an effort has to be made as to how the content would fit to the general public. The transmitter-receiver balance has to be respected in order for the text to function as it was intended.

3.1.5 Culture-pair specific translation problems

Culture-pair specific translation problems refer to text type conventions, which can be different in the source and target text, depending on the specific requirements of the source and the target language. While translating property texts or a simple recipe, it is necessary to compile the relevant conventions of the text type.

3.1.6 Language pair specific translation problems

These are the consequence of the differences in the way languages operate. This includes lexical differences, and structural mismatches. They may be language specific. The comparative and contrastive study between languages offer a great insight into how to tackle this kind of translation problems. The translation problems that have been reviewed above summarize the main issues in Translation Studies. Some of these problems are relevant in machine translation as well, especially language pair specific problems. In the following section, translation problems that are more specific to machine translation shall be reviewed.

3.2 Translation problems in machine translation

The translation problems that have been reviewed above summarize the main issues in Translation Studies. In the following section, translation problems that are more specific to machine translation are reviewed. (Diekema 2003) provides an in-depth look at these phenomenon as summarized in the following table:

Translation problems from the translation literature	
lexical ambiguity *The fisherman walks along the bank.*	- words having multiple meanings *Bank, Ufer, Sessel, Damm, Deich*
lexical mismatches *Brown (color)*	- differing conceptual structures between language communities *Brun; châtain (of hair); marron (of shoes/leather)*
lexical holes *J'ignore le contenu du colis.*	- unlexicalized concepts across languages *Ich weiß nicht was das Packet enthält.*
figures of speech *La Dame de fer.*	- words that should not be taken literally - are used to create a certain literary effect *Die eiserne Lady.*
multiword lexemes *Pomme de terre,* *le savoir-faire*	- idioms, phrasal verbs, and collocations *Kartoffel,* *das Können*
false cognates *Das Gift # a gift*	- words that seem to be the same across languages but are not *poison # a present*

Table 3: MT translation problems

The following sections will discuss problems related to structural and lexical differences between languages and the problem of collocations and multiword units drawing from (Arnold, et al. 1994). Since ambiguity constitutes the main focus of the present study, a whole section shall be devoted to its description under paragraph 3.4.

3.2.1 Lexical and structural mismatches

Lexical mismatches occur on two levels: first, at the semantic level, because the source language word may have several possible equivalents in the target language. As a result, the multiplicity in interpretation leads to a multiple lexical rendition of the source language word. In addition, mismatches occur because each language has different ways of viewing, organizing and expressing the same reality, as shown in the following examples borrowed from (Arnold, et al. 1994, 109):

Example 2

(a) know (V) savoir (a fact)
(b) connaitre (a thing)
(c) leg (N) patte (of an animal)
(d) jambe (of a human)
(e) pied (of a table)
(f) brown (adj) brun
(g) châtain (of hair)
(h) marron (of shoes/leather)

Seen above, some cases of lexical mismatches between English and French. As the name indicates, lexical mismatches occur at the lexical level and this notion describes an instance where one single word, in the source language, has multiple target language equivalences, each of the latter having a specific context in which they are used. It has been shown above that while English uses the single word “leg” to designate the lower limb of a biped from the knee to the ankle (human being or animal), the French language makes a further distinction by differentiating between human beings, animals and things, thus:

- *patte* fits for animals and may only figuratively be used for human
- *jambe* fits for humans
- *pied* is usually used in reference to tables.

The difficulty in establishing the lexical mapping in such cases may make MT architectures more complex, especially in the context of a rule-based machine translation.

Structural mismatches on the other hand, refer to an instance where a single SL word will be used to express a target language phrase, and *vice versa*;

Example[2]:
a) These are the letters which I have already replied to.
b) *Ce sont les lettres lesquelles j'ai déjà répondu à.
c) These are the letters *to which* I have already replied.
d) Ce sont les lettres *auxquelles* j'ai déjà répondu.
In the above example, the English verbal phrase "to which" has been translated with a French single word "*auxquelles*" Structural mismatch also refers to a change in the syntactic structure during translation from the source to the target language. This notion corresponds to (Vinay and Darbelnet 1995)'s "transposition". Structural and lexical mismatches can potentially complicate machine translation.

3.3 Multiword units

This category entails all the semantic units that are materialized at the graphemic level by several tokens or words such as collocations and idioms.

3.3.1 Idioms

Idioms are expressions whose meaning depends on all the constituent units. Taken separately, the lexical units that make up the idiom do, usually, not make any sense whatsoever or, at the very least, mean something completely different than the intended meaning. Following are a few examples of idioms:
Example 3
(a) To kick the bucket
(b) To run errands
(c) It is raining cats and dogs
(d) Be in the spotlight
(e) Rub someone the wrong way
(f) Jump the gun
(g) Pay the piper
(h) To buy a farm
These idioms cannot be understood by simply combining the meaning of the various lexical elements that compose them, because the result would be nonsensical; instead, they are to be construed as one single semantic unit, as shown in the following example:

[2] Borrowed from (Arnold, et al. 1994, 133)

Example 4[3]

(a) If Sam ***mends the bucket***, her children will be rich.

(b) If Sam ***kicks the bucket***, her children will be rich.

The above two instances where 'bucket' appears within two different structures and meanings. In the first instance, 'bucket' can be construed as what it literally means, that is, a container with an open top and a handle. In the second instance, 'bucket' can only be construed in relation to 'kick' and these two elements put together convey a meaning that is far from the original single meanings taken separately. In this case, 'kick the bucket' would mean 'to die'. A similar expression in French would be "*casser sa pipe*".

In machine translation, idioms can be translated at least in two ways. In certain cases, the target language also has an equivalent idiom.

Example 5

(a) It's pouring. (En)

(b) Es regnet Bindfäden. (Ge)

(c) Il pleut des cordes. (Fr)

The second technique for translating idioms would be to render them as a single word in the other language as shown in the following:

Example 6

He bought the farm:

(a) He died. (En)

(b) Er ist tot. (Ge)

(c) Il est décédé. (Fr)

Translation problems such as multiword units or lexical and structural mismatches constitute a real problem in rule-based machine translation during the semantic transfer. The question is how to represent such a structure in a machine translation system. (Arnold, et al. 1994) suggest that idioms should be treated as a "single unit". In order to figure out how this works, we tried to experiment this technique in the CAT2 MT system (Sharp 1994). The compound noun "customs officer" was used as an example and was codified as follows:

Example 7

(a) 'customs_officer'={role=gov,cat=n,lex='customs_officer',semf=per
s,gen=masc,v=no}.[].

(b) atom = {lex='customs_officer'}.[] <=> {lex=douanier}.[].

Example 7 (a) represents the codification of "customs officer" in the syntactic grammar of English during the grammar compilation of CAT2; whereas

---

[3] Borrowed from (Arnold, et al. 1994)

example 7 (b) shows the bilingual English-French dictionary. Here, "customs officer" is treated as one lexical unit. Its equivalent in French, however, does not have the same structure and is rendered with a single token "*douanier*". The use of the hyphen allows the compiler grammar to merge both lexical items into one. This merger then facilitates the semantic transfer from L1 to L2, which produces "*douanier*" in L2. This is relatively easy since in this case, a compound word in L1 produces a single word in L2. This operation is, however, more complicated when it comes to the transfer of idioms and collocations from L1 into L2. The idiom "it is raining cats and dogs" was submitted to Google Translate which yielded the following result:

Example 8

SL: I'm not going out in that storm. <u>It's raining cats and dogs.</u>

Google Translate: Je ne vais pas dans cette tempête. <u>Il pleut des chats et des chiens.</u>

This result is all the more surprising, because this type of problem is usually easily solved in statistical machine translation architecture. Statistical MT systems function on the basis of aligned corpora, which are used to determine the probability of a given sentence in the target language being the translation of a given source language sentence. The advantage of this system lies in the fact that the equivalence is established on the basis of translation units. A translation unit is a distinctive grammatical or semantic unit that can be translated separately without the loss of the meaning.

Turning to the idiom "pay the piper" which cannot be understood by considering the lexical units taken separately. This expression is a semantic unit that translates as such. It translates in French as "*payer le prix*". Handling idioms in statistical machine translation architecture is usually easy as the translation units used for alignment goes beyond the lexical level. The only instances in which a particular idiom may not be well translated, in a SMT system, are cases where the trained-data does not take into account a particular phrase or idioms. This seems to be the case with "it is raining cats and dogs", which to our surprise is not part of the trained data from Google. Consequently, it shouldn't be a surprise if on a later date, this very idiom is translated properly by Google Translate, since the system undergoes constant updates.

3.3.2 Compound words

Composition is a concept in morphology and syntax that designates one of the many ways in which words are coined within a language. While it is true that compounds can be found in almost all languages, the way they come into existence is specific to each language. For instance, German and French are two languages of different origin; both have different ways of coining their words. German for example, allows a greater number of compound words. In addition to their greater frequency, compound words in the German language are fairly intuitive since the meaning of compound words can be fairly determined from the different elements that make them up.

Example 9

Staplerfahrer

In this example, there is Stapler + Fahrer.

Intuitively, we understand that "*Staplerfahrer*" is someone who drives engines. This compound word displays the morphology of a single word since the various elements which constitute the compound word underwent a morphological fusion. This means that the compound word is no longer distinguishable from other words in the language at the morphological level.

In French, compound words are of different kinds. They may consist of at least two words, separated either by a blank or a hyphen, like in: "*pomme de terre*" or "*après-midi*". "*Pomme de terre*" is made up of the three elements (*pomme* + *de* + *terre*). When hyphenated, compounds are easily identifiable as such. Their mapping may however be difficult in MT if no distinctive sign marks their singularity. A difference has to be made between a compound word and a phrase. Compound words are identifiable by their idiomatic nature. A phrase is a sequence of words forming a syntactic unit, while a compound word is a sequence of words forming a lexical and semantic unit. Compound words are semantically and lexically fixed, while a phrase is not. Phrases usually have a greater freedom in the paradigmatic and the syntagmatic axis while compound words do not always allow for such flexibility. For instance, if "*pomme de terre*" is transformed into "*pomme de mer*" the notion of idiomaticity is broken down and, additionally, the French language does not identify "*pomme de mer*" as the result of a morphological transformation, therefore the criteria for "compoundness" is not being fulfilled. To better illustrate the difference between phrase and compound words, consider the following sentences:

Example 10

(a) *La pomme de terre* est un élément essentiel de la cuisine allemande.
(b) *La pomme de courge* est un élément essentiel de la cuisine allemande.
(c) *La femme de terre* est un élément essentiel de la cuisine allemande.
(d) *La charcuterie* est un élément essentiel de la cuisine allemande.

Example 11

(a) *Les personnes qualifiées* n'ont pas de mal à se trouver un emploi stable.
(b) *Les femmes qualifiées* n'ont pas de mal à trouver un emploi stable.

(c) *Les ouvriers qualifiés* n'ont pas de mal à trouver un emploi stable.
(d) *Les personnes en mal d'orientation* n'ont pas de mal à trouver un emploi stable.
(e) *Les cadres* n'ont pas de mal à trouver un emploi stable.

In the first series of sentences (example 11), any change in the paradigmatic axis within the compound word "*pomme de terre*" produces semantically incorrect sentences. Except for the last sentence, example 11 (d), in which not just one element of the compound word has been replaced, but all of them. This replacement shows that the compound word "*pomme de terre*" may be replaced by another word; however, it does not allow changes in its internal structure. Thus, "*pomme de courge*" appears as a surprising and erroneous construction.

Conversely, the second set of sentences (example 11) admits some modularity. Thus making it possible to make changes in the paradigmatic axis without this affecting the syntactic structure of the noun phrase that is the phrase "*Les personnes qualifiées*". It is also worth noting that compound words can be of different types according to their morphological and graphemic realization. Thus, there are:

- Unified compound words whose elements are clustered (e.g., a "*portmanteau*");
- The compound words apostrophe, whose elements are separated by one or more quotation marks (e.g., "*aujourd'hui*");
- Hyphenated compound words, whose elements are separated by one or more hyphens (e.g., "*après-midi*"); in the compounds, the elements are separated by at least one space (e.g., "*pomme de terre*").

In German, composition is a way of word formation used very often and its functioning follows a different logic than the French composition. While it is true that the phenomenon of composition fulfills the same function in both languages, they differ in their implementation. Below, we will look at the phenomenon of composition in the German language. As in French, there are well-defined German compound words typologies. So, there are a range of compositions that can be defined according to their types.

The difficulty with compounds in German lies in the fact that a compound is not only made up of lexical elements, that constitute the meaning, but sometimes a melting element can be added. For example, "*Strahlenbündel*" consists of the elements "*Strahl*" and "*Bündel*". Yet "*en*" is inserted between these two clearly identifiable components.

In machine translation, German compound word are generally marked as "unknown" and not translated because there is usually no corresponding target word matching the source language compound. One very successful approach is the splitting of the compound. ((Koehn and Knight 2003); (Niessen and Ney 2000); (Popović, Stein and Ney 2006)) investigate and compare strategies for splitting compounds when German is the source language.

The splitting of compound words is an approach in which each lexical unit consisting of two or more words is considered as a compound word and is matched against words in the training corpus to determine the constituent elements of that word. For example, if the compound is "*Aktionsplan*" the word-splitting module will ensure that "*Aktion*" and "*Plan*" are found in the system's dictionary to help recover the sense of the desired word (action plan). The number of unknown source word (usually German) is reduced therefore enabling the translation of the compound by the translation of its parts. (Niessen and Ney 2000) report an improvement of the translation results in German-English corpora. (Koehn and Knight 2003) reports a 0.039 points quality improvement in the BLEU score.

Failure to develop MT systems able to produce a *fully automatic high-quality translation* provided Bar-Hillel with an argument for casting doubt on the viability of machine translation. Since the ALPAC report, (ALPAC 1966) which was the direct cause for defunding most research efforts on machine translation (Ide and Véronis 1998), there has been a lot of progress. New methodologies have been developed, some of which have been reviewed in chapter two. Nevertheless, as has been shown previously, many linguistic aspects of MT still need to be thoroughly researched. The impact of ambiguity in machine translation output quality constitutes one of these research areas.

In the previous sections, some of these linguistic problems have been reviewed, namely the translation of multiword units, metaphors and compounds. In this section, the focus is shifted to establishing the difference between the concepts of ambiguity and vagueness. Various definitions of ambiguity, found in the literature, are discussed further below. Using CAT2, we also showcase how some forms of ambiguity can be solved in a transfer-based machine translation system.

3.4 Typology of ambiguities

In linguistics, the term ambiguity refers to the fact that a sentence or a word allows for two or more interpretations. Ambiguity can be observed within a language or when translating from one language to another, referred to as "transfer ambiguity". Some linguists consider ambiguity to be a generic term that encompasses other linguistic phenomena such as vagueness and equivocation. (Fries 1980) takes this stance and defines vagueness as an ambiguity, which is not clear in context. He defines ambiguity further, as the possibility to attribute different interpretations or meanings to a morpheme, a word, a sequence or a sentence.

(Gillon 2004) on the other hand, makes a clear distinction between the concepts of vagueness, ambiguity, generality and indeterminacy. He argues that an expression is said to be vague if three types of cases apply:

i. there are instances where the term clearly applies;
ii. there are instances where it is clear that the term cannot apply; and
iii. there are instances where the speaker cannot decide, given the sense of the expression, whether it applies or not (failure to decide doesn't lie on the speaker's lack of knowledge).

Example (Kennedy 2007):

(a) Mercury Mbah Gotho is old.
(b) Prince George of Cambridge is old.
(c) Prince Williams is old.

(a) is clearly true since Mbah Gotho is reported to be the oldest man alive; (b) is clearly false since Prince George is only 3 years old; (c) is borderline because this assertion might be true if Prince William is compared to his younger brother Prince Harry and would not be true if compared to Prince Charles, his father who is obviously older. See further, (Kennedy 2007).

Indeterminacy refers to an instance where there is some property of an expression "which neither is included in the expression's connotation nor is it a species of any property included in its connotation" (Gillon 2004, 394). The word 'house' is indeterminate, since its connotation does not include, or exclude, being any particular type of housing.

Generality refers to an expression which is used as a generic term to describe some general group or species. E.g.: metal: gold, copper, silver, iron, mercury, etc.

(Andersen 2002) on the other hand, considers ambiguity to be a subcategory of indeterminacy and summarizes the relationship between indeterminacy, ambiguity and vagueness as illustrated in the following figure:

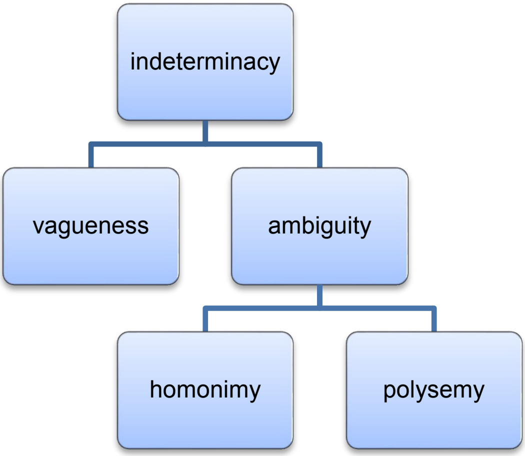

Figure 2: Types of indeterminacy borrowed from Andersen (Andersen 2002, 137).

In (Andersen 2002)'s distinction (drawing on (Pinkal 1985)), ambiguities can be of various types, depending on the linguistic level that enters into play. For instance, two words that display the same morphological structure, ("bank" and "bank") are said to be ambiguous and they fall under the linguistic category homograph. When there's some degree of semantic proximity between two morphologically identical words, they are said to be polysemous. The spelling may be identical, but the words differ in their meanings and as a consequence, are entered into different entries in dictionaries. Sometimes, these words with the same spelling can also be classified under different grammatical categories. For instance, "address" could either fall under category verb or noun, as illustrated in the following examples:

Example 12

(a) The speaker *addressed* the crowd with contempt.

(b) It took the firefighters 15 minutes to get to the right *address*.

At the phonetic level, two or more words may be pronounced the same way, yet having different meanings. These are called homophones.

Example 13

(a) *straight* # *strait* /streIt/

(b) *sole* # *soul* /səʊl/ (BrE) /sōl/ (AmE)

It should be pointed out, though, that homophones are rarely problematic during human translation nor are they problematic during automatic translation; rather, they might be of particular interest in speech recognition or automatic interpreting research.

At a sentence level, the syntactic structure of an utterance may give way to more than one interpretation depending on how the elements of the sentence are put together. The more domain-specific we get, the less likely it is that MT systems are going to encounter ambiguities. In a financial text, for instance, one would expect the word “bank” to have the meaning “financial institution”. But as we get more general, it becomes problematic to determine its specific meaning.

3.4.1 Lexical ambiguity

3.4.1.1 Homographs and polysemes

As was pointed out earlier, ambiguity may arise at various levels depending on the criteria chosen for analysis. “Bank” is a perfect example of lexical ambiguity because it has many different definitions that are not related as shown in the following dictionary entry from the Collins dictionary of English (Collinsdictionary 2015).

Bank[1] /bænk/

Definitions

noun

(1) an institution offering certain financial services, such as the safekeeping of money, conversion of domestic into and from foreign currencies, lending of money at interest, and acceptance of bills of exchange

(2) the building used by such an institution

(3) a small container used at home for keeping money

(4) the funds held by a gaming house or a banker or dealer in some gambling games

(5) (*in various games*)

(6) the stock, as of money, pieces, tokens, etc, on which players may draw

(7) the player holding this stock
(8) any supply, store, or reserve, for future use. A data bank, a blood bank

verb

(1) *tr* to deposit (cash, cheques, etc) in a bank
(2) *intr* to transact business with a bank
(3) *intr* to engage in the business of banking
(4) *intr* to hold the bank in some gambling games

bank^2 /bænk/

Definitions

noun

1. a long-raised mass, esp of earth; mound; ridge
2. a slope, as of a hill
3. the sloping side of any hollow in the ground, esp when bordering river
4. an elevated section, rising to near the surface, of the bed of a sea, lake, or river
5. the area around the mouth of the shaft of a mine
6. the face of a body of ore
7. the lateral inclination of an aircraft about its longitudinal axis during a turn

verb

1. *when tr,* often foll by *up* to form into a bank or mound
2. *tr* to border or enclose (a road, etc) with a bank
3. *tr, sometimes foll by up* to cover (a fire) with ashes, fresh fuel, etc, so that it will burn slowly
4. to cause (an aircraft) to tip laterally about its longitudinal axis or (of an aircraft) to tip in this way, esp while turning
5. to travel round a bank, esp at high speed
6. *tr* (*billiards*) to drive (a ball) into the cushion

bank^3/bæŋk /

Definitions

noun

1. an arrangement of objects, esp similar objects, in a row or in tiers ⇒ a bank of dials
2. a. a tier of oars in a galley
   b. a bench for the rowers in a galley.
3. a grade of lightweight writing and printing paper used for airmail letters, etc
4. telephony (in automatic switching) an assembly of fixed electrical contacts forming a rigid unit in a selector or similar device

verb

5. (transitive) to arrange in a bank

Figure 3 : English dictionary entry for “bank”

The (Collinsdictionary 2015) has three main entries for “bank”: financial institution, a landside and an arrangement of objects. Further distinctions can be added to these three senses, which results into twenty-four different definitions of bank. An ambiguous lexeme with no related meanings is called homograph. Conversely, when there is a semantic proximity between the lexemes, they're said to be polysemous as shown in the following example.

Example 14

(a) *I went for a walk.*

(b) *I walk the dog.*

“Walk” in example 14 (a) and “walk” in example 14 (b) are semantically related, because the two actions relate to a pedestrian activity that is practiced by living beings, yet the two occurrences have different meanings, and grammatical function. The meaning of the homographic lexemes can usually only be derived from the context in which they arise. Consider the following example:

Example 15

(a) *The fisherman walks along the bank.*

As taken in the following context:

(b) *After a long day of fishing, the fisherman walked along the bank of the river.*

Sentence 15 (a) may be subject to more than one interpretation:

i the fisherman is walking along the bank which is the *riverside*, or

ii the fisherman is walking along a [*financial institution*.]

In the above example, the word *fisherman* gives us a hint that the first interpretation (example 15 (a)) is most likely, since it can be intuitively assumed that fishermen are most likely to be found around riversides; however, this at first glance conclusion might prove tricky, insofar as the second interpretation (example 15 (b)) cannot be totally ruled out, unless the context gives further information that will help disambiguate to interpretation. Sentence 15 (a) has been used as a test for both our illustrative system (CAT2) and Google Translate. Most of the sentences were translated from English into French and German. Google Translate produced the following result:

Example 16

(a) **Le pêcheur promenades le long de la banque.*

The result of the Google translation was a syntactically erroneous sentence. Indeed, Google Translate seemed unable to properly identify "walked" as a verb in the present tense. Therefore, "walks" was rendered as "*promenades*" which is a substantive in French. The semantic proximity can be noticed between the English "walk along" and the French "*promenade*", but because of the poor syntactic structure of the above Google translation, the above rendition cannot be identified as a successful transposition. A syntactically more acceptable translation would have been:

(b) Le pêcheur fait une promenade le long de la banque.

The Google Translate system selected the "financial institution" sense of "bank" over the "riverside" sense. Our assumption is that the "financial institution" sense of "bank" more often than not occurs in the trained corpus used to generate Google translations. Thus, the system generates translations that are arguably erroneous. Even after providing more context to sentence 16 (a), no amelioration was observed in the translation, hence:

Source Text 1

(a) *After a long day of fishing, the fisherman walked along the bank of the river.*

Google Translate

(b) *Après une longue journée de pêche, les pêcheurs marchant le long de la <u>banque</u>.*

While the above interpretation of source text 1 is fully plausible, (obviously putting aside the poor syntactic and stylistic quality) one could still argue that the probability of it being the "correct" translation of our sentence is rather low, especially if one considers "fisherman" and "fishing" as a contextual reference.

Using the CAT2 translation system, a word sense disambiguation was attempted by first providing the CAT2 glossary with the various definitions of *bank*, attributing the following possible translations:

```
atom = {lex=bank}.[]       <=> {lex=(banque; rive; groupe; rangée)}.[].
         atom = {lex=bank, cat=v}.[]  <=> {lex=surhausser}.[].
```

Figure 4: Entry for "bank" in the CAT2 En-Fr dictionary.

Following this distinction, the CAT2 system was not able to produce the expected result during the first attempt, since it wasn't provided with further selection restriction for the various meanings of "bank". Thus, the first attempt to translate sentence 16 (a) produced the following results:

Example 17

(a) *Le pêcheur marche le long de la banque.*
(b) *Le pêcheur marche le long du groupe.*
(c) *Le pêcheur marche le long de la rangée.*
(d) *Le pêcheur marche le long de la rive.*

Resolving lexical ambiguities involving homographs of the same grammatical category requires the use of semantic information. Borrowing from the rules of inclusion-based approach to lexical ambiguity resolution (Fass 1988), we specified which features are compatible in given syntactic constructions, *via* the satisfaction and violation of selection restrictions. We specified that the following semantic features can be attributed to "bank":

- Location
- Finance
- Abstract
- Collection

bank = {role=gov, cat=n, lex=bank, semf=(loc; river; coll; abs),v=no}.[].

Figure 5: Selection restriction of 'bank' in the CAT2 system dictionary.

These details provide the system with specific information as to how to interpret 'bank' if it is encountered in a fishing context. As a consequence, CAT2 produced the following translation:

(d) *Le pêcheur marche le long de la rive.*

The inclusion-based approach to lexical ambiguity resolution proved to be effective in this case; it would, however, be cumbersome to disambiguate all ambiguous English lexical items using this technique.

3.4.2 Transfer ambiguity

Transfer ambiguities occur when a non-ambiguous word in a source language produces several target language interpretations as in example 18 (a).

Example 18

(a) The catholic priest married my sister.

None of the constituents in the above sentence are ambiguous in English. However, when translating into French, "marry" may refer to two different actions depending on whether the emphasis is on the subject (*marier quelqu'un*) or the object (*épouser quelqu'un*). Google Translate produced the following translation:

(b) *Le prêtre catholique a épousé ma sœur*

If we compare the original text, and its translation into French, it becomes obvious that a semantic shift occurred during translation. Indeed, "married" is rendered in French by "*épouser*". While it is true that "marier" and "*épouser*" are semantically linked (they refer to the act by which two people are united by the bond of marriage), it is important to note that "marry" may refer to two different actions depending on whether the emphasis is on the subject (the performer of the marriage: civil authority, priest etc...) or the object (the benefactors of the marriage, usually husband and wife).

Considering the above examples, our world knowledge reminds us that catholic priests do not get married. The first attempt with CAT2 produced the following:

(c) *Le prêtre catholique* ***épousa*** *ma soeur.*

(d) *Le prêtre catholique* ***maria*** *ma soeur.*

The semantic relation between "*prêtre*" and "*épouser*" in example 18 (c) is anomalous because "*prêtre*" is not a preferred agent of "*épouser*". This brought us to allow a selection restriction on "*épouser*" which would prevent it from unifying with "*prêtre*" in a sentence where "*épouser*" is the governor and "*prêtre*" is agent, Hence:

épouser={role=gov,cat=v,lex=épouser,stem=épous,end=er,flex=reg,vpre=yes,
frame= {arg1={role=agent, cat=n, lex **~=prêtre**},
arg2={role=goal,cat=n}}}.[].

Figure 6: Selection restriction for 'épouser' in the CAT2 system.

As a result, we obtained *Le prêtre maria ma soeur* as the sole translation to sentence 18 (a).

3.4.3 Metaphors

A metaphor is a figure of speech which uses analogy to designate an object or an idea with a word or another object or another idea not related to a previous. A comparative element is therefore introduced that merges the two distinct ideas/words into one.

Example 19
"Britain bids farewell to Margaret Thatcher, the Iron Lady"
((Washintonpost.com April 17, 2013); (Faiola and Mackintosh 2013))
In this classical example, "The Iron Lady" does not designate a person made of iron, but a lady whose fearless and unyielding character recalls the characteristics and/or attributes that we attribute to iron.

Metaphors can be classified into two categories: the first category, being metaphors that one could easily translate, especially when the languages involved belong to the same cultural family as shown in the French and German translations below:
French:
Grande-Bretagne fait ses adieux à Margaret Thatcher, <u>la Dame de fer</u>.
German:
Großbritannien nimmt Abschied von Margaret Thatcher, <u>die eiserne Lady.</u>
Whereas the second category of metaphors, however, are more specific to their culture of origin. These pose a much more substantial, especially when translation takes place between two unrelated cultures. Machine translation has yet to find a way to deal with this second category.

3.4.4 Category ambiguity

Category ambiguity is another type of lexical ambiguity. It occurs when a word may be assigned more than one grammatical or syntactic category (e.g. noun, verb or adjective) as illustrated in the following examples:
Example 20
(a) The insurance is *invalid* for the *invalid*.
(b) The carpenter must *polish* the *Polish* furniture.
(c) Ich bereite das *Essen* vor, kann aber nicht *essen*.
In sentence 20 (a), the word "invalid" occurs twice, yet the two occurrences do not play the same syntactic function because they belong to two different grammatical categories; they also have two different meanings. The first occurrence of "invalid", is an adjective whereas the second occurrence is a noun. Likewise, in sentence 20 (b), "polish" first occurs as a verb, the second occurrence is an adjective.

3.4.5 Syntactic ambiguity

When the structure of the sentence generates several possible interpretations, this is referred to as structure or syntactic ambiguity.

Example 21

(a) John hid the photo in the drawer[4].
(b) I saw the woman with the telescope.
(c) Pregnant women and children are vulnerable.

Turning towards the following sentence: (a) John hid the photo in the drawer. More than one interpretation can be attributed to this sentence depending on its underlying structure, hence:

- The *photo* is being hidden *in the drawer*.
- The *photo in the drawer* is being hidden somewhere else.

In contrast to lexical ambiguity, syntactic ambiguity does not arise from the meaning of the individual elements of a sentence, instead it results from the relationship between those elements, the parts of the sentence and the sentence structure. Syntactic ambiguities are typically difficult to resolve, even by human translators, unless further situational context is provided to help disambiguate. CAT2 generated the following apparently redundant translations with the same surface structure:

i *John cacha la photo dans le tiroir.*
ii *John cacha la photo dans le tiroir.*

In contrast, Google Translate tended to be more selective when generating translations involving syntactic ambiguities, thus:

iii *John caché la photo dans le tiroir.*

In this particular instance, source language and target language ambiguity coincide. Putting aside the past tense form "hid", which has been mistranslated into the French past participle "*caché*", Google Translate actually keeps the ambiguity intact, which (Emele and Michael 1998) refer to as the ambiguity preserving approach whereby the ambiguous segment is maintained as is during translation. If the surface structures of examples 21 (i) and (ii) look identical, syntactic parsing revealed quite different underlying structures as shown in Figures 7 and 8 below.

[4] This example is discussed in (Trujillo 1999, 230)

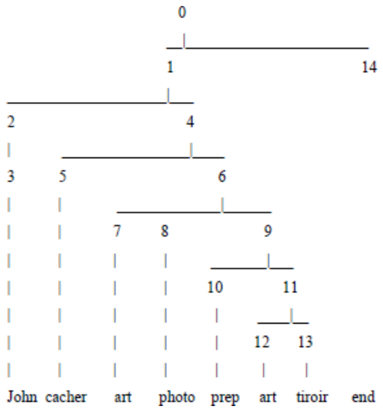

Figure 7: Constituent structure 21 (i)

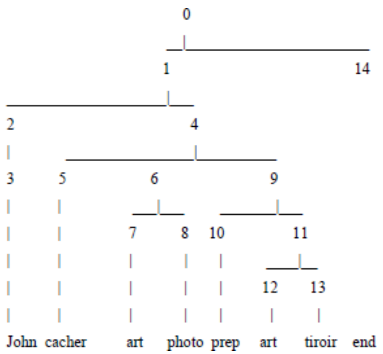

Figure 8: Constituent structure 21 (ii)

Parsing these sentences will produce more than one syntactic structure. (Ceccato, et al. 2004, 3) identify four different types of syntactic ambiguity, namely coordination, analytical, attachment, and elliptical ambiguities.

3.4.5.1 Coordination ambiguity

Coordinating conjunctions are language elements that syntactically link words of similar grammatical nature such as nouns, verbs, adverbs and adjectives. Coordinating conjunctions are also used to juxtapose larger constituents such as, phrases. Their role is, therefore, to establish a semantic relation between the "coordinated" elements. A careful use of coordinating conjunctions ensures a logical discourse. Some of these conjunctions include: "and, but, or, yet, for, nor, so". A careless use, as opposed to a coordinated use, of conjunctions can break a statement's logic and introduce a variety of varying interpretations that may complicate the translation task significantly. Coordination ambiguity, therefore, refers to an instance where a sentence is subject to multiple interpretations due to a careless use of coordinating conjunctions.

3.4.5.2 Analytical ambiguity

Analytical ambiguity occurs when the role of the constituents within a phrase or sentence is ambiguous:

Example 22

Porcelain egg container.

The above phrase can be understood in at least two ways, depending on the syntactic distribution that is made of the components of the sentence:

(a) [The porcelain egg] container.

(b) The porcelain [egg container].

In the former, example (a), the substantive "egg" is attached to "porcelain", and in this case, porcelain is understood as modifying "egg". With this configuration, the final interpretation would be: "a container in which you can put porcelain eggs."

In the latter, example (b), "porcelain" modifies "container", and this time, it is not "egg" which is made of porcelain, but it's "container" that displays this feature because the substantive "egg" is attached to container, prompting a quite different interpretation of this phrase.

3.4.5.3 Attachment ambiguity

(Ceccato, et al. 2004) define attachment ambiguity as an instance where a particular syntactic constituent of a sentence can be attributed to at least two constituents of that sentence.

Example 23

He was looking for a woman with money.

The PP "with money" can be read as attached to the subject "he". In this case, the sentence should be read as meaning "he used money to entice women … to do something". The second interpretation arises when the PP "with money" is attached to "women." In this case, the subject "He" would be looking for women who have money.

3.4.5.4 Elliptical ambiguity

Ellipsis consists in retracting one or more words in a sentence, which in principle, are necessary to understanding the overall meaning of the utterance. Ellipsis are usually used: either to avoid repetition; or for stylistic purpose. A statement that employs ellipsis is almost always ambiguous if no further context is given, to fill the gap that is created by the omitted segments of the sentence.

Example 24

Paul likes cheeseburger more than Mary.

The above example is a typical case of elliptical ambiguity as this sentence can be understood as least two ways:

- Paul likes cheeseburger more than Mary *does*.
- Paul likes cheeseburger *more than he likes* Mary.

The omission can, therefore, lead to ambiguity or possibly be misinterpretation.

3.4.6 Scope ambiguity

A scope ambiguity is a sentence in which the scope of a verb, noun, adjective, or adverb is unclear, therefore leading to multiple interpretations.

Example 25

(a) All pregnant women and men are vulnerable.

(b) Every man loves a woman.

Usually, the quantifier or negation operators, such as "every, each, all, some, several, a, not etc…", can be interpreted as referring to one or several constituents of a sentence at a time as in:

"all pregnant women and men are vulnerable" where the question is whether "all" can be said to refer only to "women" or "women and men". Sentence 25 (b) has two distinct readings as well:

- for each man there is "his" woman, and he loves her.

or, alternatively:

- there is a particular woman who is loved by all men.

For scope ambiguity, (Maienborn, von Heusinger and Portner 2013) advocate ambiguity preserving translation since text understanding doesn't presuppose that all ambiguities have to be resolved and scope ambiguity is typical for this category.

3.4.7 Referential ambiguity

Referential ambiguity arises when in a discourse, an element of reference (personal pronoun, grammatical deictic, anaphora) can be interpreted as referring to more than one antecedent. Referents are elements of discourse that evoke a referent without making an abusive repetition. As an example:

Example 26

Paul is a brilliant student. He is diligent in his work.

By using the personal pronoun "he", repetition is avoided. In the above example, the discourse is organized so that the pronoun-antecedent relationship is sufficiently clear. There is no doubt, considering the statement, that the personal pronoun "it" refers directly to Paul. Sometimes, however, the referent-antecedent relationship is not obvious as in the following examples:

Example 27

(a) He drove the car over the lawn mower, but it wasn't hurt.
(b) Bob waved to Jim in the hallway between class. He smiled.

In example 27 (a), one could argue that "it" refers both to "lawn mower" and to "car". In machine translation, the referent is usually preserved in the target language when no specific context for disambiguation is provided (Nirenburg 1993, 56); however, preserving the referent in the target language might be confusing, especially when the target language marks the gender of pronouns. In French, for instance, the pronoun always agrees with its antecedent.

(Mitkov, Choi and Sharp 1995) propose a more specific anaphora resolution model implemented in the CAT2 system that is based on syntactic and semantic constraints and preferences, such as the pronoun-antecedent agreement, for antecedents displaying the same syntactic and semantic role as the pronoun. The most recent techniques, (Hardmeier, Tiedemann and Nivre 2013) implement the pronoun prediction model which uses a neural network (contextual analysis) to determine possible target language pronouns in statistical machine translation.

The previous steps have helped to make an inventory of machine translation and a review of some linguistic aspects such as multiword units, metaphors and compound words. A linguistic description of ambiguities has also been undertaken. In the following chapter, an evaluation will be conducted in order to measure the impact of ambiguity on the output of automatically translated corpora.

4 Experimental and Evaluative Part

The previous chapter has been the venue to explore the concept of ambiguity in language studies. Several categories of ambiguities have therefore been identified. In this chapter an experiment will be conducted using Personal Translator, a commercial machine translation system based on rules, and Google Translate, a free online machine translation system that implements the statistical method for MT. A brief description of these tools will be given below, in section one. The present research endeavor has been undertaken under the premise that ambiguity in the source text may influence machine translation output. As a consequence, a critical part of this chapter is devoted to evaluating how much of an influence lexical ambiguity represents in machine translated output. To this end, a review of various evaluation approaches shall be carried out, followed by a description of the methodology, which was implemented in collecting the data and conducting the evaluation. *Costa MT*, the evaluation tool which assisted annotators during the evaluation task, shall also be presented. Sections 4.5 and 4.6 shall be devoted, respectively, to the experiment which was carried out in two steps. The first step consisted of a machine translation of the corpus; thereafter followed by a raw analysis of the machine translation output. During the second step, the human evaluation of the same output was carried out. Both experiments aimed at assessing how ambiguity influences the understandability of a machine translated output. Bar-Hillel (Bar-Hillel 1964) pointed out that ambiguity would prevent any MT system from attaining Fully Automatic High Quality Machine Translation. On the basis of this postulate, the present experiment shall help validate or refute this assessment. The final section of this chapter shall be devoted to presenting the results of the experiments. Further analysis of these results shall be dealt with in the subsequent chapter.

4.1 Description of the MT systems used

In the first part of the present research endeavor, the CAT2 machine translation system was used as a laboratory that helped illustrate various types of ambiguities. As this study moved forward, it became obvious that the CAT2 system could not handle forms of ambiguity that go beyond the sentence level. A comprehensive evaluation of ambiguity cannot be made without contextualization being a key element of sense determination. Moreover, since the present study has a comparative component, which is, comparing rule-based approaches to statistical approaches, it was therefore deemed necessary to alleviate the deficiencies of the CAT2 system by introducing more robust commercial systems such as Personal Translator and Google Translate both of which are well established translation tools. Under the present section, these tools shall be briefly presented.

4.1.1 Personal Translator

Personal Translator (PT) is a commercial machine translation system. It belongs to the category of rule-based machine translation systems, but comes additionally with a statistical component and can therefore, be labeled as a hybrid machine translation system. Hybrid machine translation systems are the latest generation of translation systems and function by integrating two different types of machine architecture. There are two types of system architecture: firstly, the statistical translation systems that integrate linguistic knowledge; and then there are rule-based machine translation architectures that incorporate a statistical module. PT belongs to the latter category, with a disambiguation module which makes it an ideal candidate for our experiments. Like most machine translation systems based on rules, the PT system solves lexical ambiguities by conducting a morphological and syntactic parsing which is necessary for the subsequent categorization and translation of the ambiguous words. The limitations of the rule-based approach have previously been discussed, especially the fact that rule-based architectures usually conduct a morpho-syntactic analysis that does not transcend the sentence level; meanwhile an efficient disambiguation requires an analysis that not only takes into account the immediate constituents of a sentence, but most importantly the overall context in which the utterances appear, which furthermore reflects the way human beings operate when they want to understand a word that might seem unclear at first glance. Indeed, with his now famous example “The box was in the pen”, (Bar-Hillel 1964) demonstrated that the best way to understand the meaning of a word is to “contextualize” it, as without non-verbal cues added to the context, it would almost be impossible even for a person to decipher certain utterances. Following this observation, PT has developed a module that performs a contextual analysis in order to identify the semantic network to which the ambiguous word pertains. This, contextual analysis is called “neural transfer” which can be defined as a model designed “to single out the best translation for a word by identifying its semantic network.” (Aleksić and Thurmair 2011, 306).

(Aleksić and Thurmair 2011) *op. cit.* moreover, note that contextual analysis should not be limited to the sentence, as is often the case in some translation systems. If the human behavior has to be taken as an example, it should be pointed out that in addition to context, the "world knowledge" is a tool which human beings usually refer to, in order to establish the connections between words and their meanings. Most systems that emulate this process generally achieve greater reliability through their *contextual analyzer*. PT, for instance, reports improvement of the translation quality by about 40%, for texts containing the affected concepts. Borrowing, from (Aleksić and Thurmair 2011), the German word "*Gericht*", ("court" or a "dish", depending on the present context) shall be used as an example to illustrate the functioning of the PT disambiguation module. In the following examples where "*Gericht*" appears under its various meanings, hints for the context are provided through subsequent sentences:

Source Text 2[5]

Ich kann mich noch an dieses Gericht erinnern. Es hat die Klage meiner Firma auf Entschädigung abgewiesen.

Target Text 3

I can still remember this court. It has rejected the complaint of my company on reimbursement.

Versus:

Source Text 3[6]

Ich kann mich noch an dieses Gericht erinnern. Es war eines dieser Gerichte aus der Küche der Balkanländer, mit Gemüse und Knoblauch.

Target Text 4

I can still remember this dish. It was one of these dishes from the kitchen of the Balkan States with vegetables and garlic.

In the above examples, the *neural transfer module* enables a connection of the dots between:

Gericht + Klage + Entschädigung = court

Gericht + Küche + Gemüse = dish

When considering the above, some contextual elements militate in favor of one translation at the expense of the other. In the first example, words such as "*Klage*" and "*Entschädigung*" are cues that "court" could be the best translation for "*Gericht*", given the thematic proximity that exists between "*Klage*"; "*Entschädigung*" and "*Gericht*". The same analysis applies for the second

[5] (Thurmair 2005)

[6] (Thurmair 2005)

example, where "dish" appears as the best candidate translation for "*Gericht*". The co-occurrence of "*Küche*"; "*Gemüse*" and "*Knoblauch*" are cues that enable the disambiguation module to connect the dots at the semantic level. Indeed, the semantic proximity between these words seems evident.

Obviously, the above experiment which has been carried out with a very limited corpus shows that, in addition to the morpho-syntactic analysis, a semantic analysis can significantly reduce the probability of error. The question, however, is now: how to determine whether this achievement can be reiterated where a larger corpus is involved.

4.1.2 Google Translate

Google Translate is a web-based translation system operated by the internet giant Google. This translation system implements the statistical approach to MT which is based on the use of aligned bilingual corpus. The corpus used for the alignment is selected from pre-translated high-quality corpus by professional translators. The main corpora used by Google Translate are the United Nations documents which are translated and available in all six official United Nations languages (Arabic, Chinese, English, French, Russian, Spanish). The *Acquis communautaire* or the Europarl database also serves as a corpus for other statistical machine translation systems. The corpora used for generating translations are obtained by aligning each segment of a source corpus with its translation in the target corpus. Thus, when the German word *Europarat* appears in a text, the system will search the corresponding translation (*Council of Europe* in English). These two segments are aligned *via* an alignment module and statistical learning techniques that are then applied to build a translation model. Using a statistical method, the translation model calculates the most probable translation of a source phrase. The trained data provide a corpus for generating translations. Google Translate, which is available in 52 languages, and other translation software online such as Babel Fish, AOL, and Yahoo have become very useful tools for immediate access to information on the internet.

Google Translate recently introduced GTT which is a tool that enables users of this software to be the architects of their machine translation. The GTT allows the user to build up and integrate a glossary and translation memory. Once the glossary is integrated to Google Translate, users can customize their translations. Part of the current experiment will therefore be to assess if indeed a customization of the translation system by the user can yield better results as far as ambiguity is concerned.

4.2 Evaluation Methods in MT

The subsequent sections will be devoted to the evaluation of machine translation output, but before moving on to that point, the current section provides a framework where the concept of evaluation is explored in-depth. While it is true that a good translation *par excellence* does not exist, it is necessary to differentiate a successful translation from a less successful one. In the context of machine translation, MT output evaluation is even more important because it offers a metric that can account for the efficiency of a given translation system. Even if, today no single methodology exists, as to how to carry out MT evaluation, the practice of assessing machine translation systems obeys a number of standards, in particular the ISO standards for software quality which define evaluation as “the measure of the quality of a system in a given context, as stated by the definition of quality as the totality of features and characteristics of a product or service that bear on its ability to satisfy stated or implied needs” (ISO/IEC9126, 1991, p. 2). The ISO standard for software defines 3 quality model frameworks for software evaluation:

4.2.1 Quality model for external quality

Also called “black box” evaluation, this quality model measures the external qualities while the software is running and the focus lies on the system output.

4.2.2 Quality model for internal quality

Also called “glass box” evaluation, these quality models are measured in a context of use and aim at assessing software’s effectiveness, productivity, safety and user satisfaction.

4.2.3 Quality model for quality in use

The quality in use metric is also measured in a context of use and aims at assessing software’s usability, functionality, reliability and efficiency. It is user oriented and provides a framework to establish whether the designed goals of the software are met.

Using the ISO standards, as a general guideline for MT evaluation, quality measurement has to obey more specific aims which may vary depending on the purpose of the evaluation. The evaluation is particularly important to at least two categories of users: the first being, designers; and then, users of MT devices. Evaluation allows designers of MT to identify weaknesses in their system and find adequate solutions. MT evaluation also allows potential buyers to get an idea of what should be expected from the purchased system. Thirdly, MT evaluation also allows Language Service Providers to assess the need to invest in a particular machine translation system. The question of how much post-edition is needed to attain a translation of publishable quality is also decisive in the purchasing process. The following section presents both automatic and human evaluation.

4.3 Automatic evaluation of machine translation

As the name suggests, automatic evaluation methods perform an evaluation that is bereft of, any, human intervention. Three main parameters govern the automatic evaluation of machine translation:
- the proximity to human translation;
- adequacy; and
- fluency.

NIST, BLEU (Papineni, et al. 2002) and METEOR are some of the most prominent automatic evaluation systems. Automatic evaluation methods are mostly based on a metric which, using specific criteria such as understandability, fluency and adequacy, computes the proximity between machine translation and translations from professional human translators. The count of error rate at the word level (word error rate) helps measure this degree of proximity.

Automatic evaluation methods appeared as a response to human evaluation methods that are deemed too slow and unnecessarily expensive, even if, human evaluation remains the most reliable of all forms of evaluations. As (Papineni, et al. 2002, 311) point out, “the closer a machine translation is to a professional human translation, the better it is”. To date, no standard automatic evaluation metric exists. The BLEU metric however appears very often in the literature.

BLEU (BiLingual Evaluation Understudy) is an MT evaluation system that compares a candidate machine translated output with a reference translation based on word sequences (n-gram) translations. This metric measures the lexical coverage of the translated sentence with that of reference translations (from professional translators). Following various evaluation campaigns, the Bleu metric is said to correlate highly with human judgement (Papineni, et al. 2002), (G. Doddington 2002). The BLEU score ranges from 0 to 1, and the higher the score, the better the translation.

4.4 Human evaluation

Human evaluation methods are the earliest and most reliable evaluation methods, since they mostly rely on trained human evaluators. The main difficulty in describing human evaluation of MT is of methodological nature because to date, various MT evaluation methods have been developed, but none of these have unanimously been awarded a standard status. If there seems to be a real difference in the methods proposed in the literature, there is one point on which everyone seems to agree: the evaluation of machine translation depends on the aims of the evaluation. Consequently, a user of an MT system will evaluate MT output for a different aim than an MT system developer. Both would understandably use different methods to evaluate MT output, as the focus would lie on different aspects. In the following paragraphs, some of the most prominent approaches to MT evaluation are reviewed.

4.5 Approaches to MT Evaluation

Numerous works have been devoted to evaluation in machine translation. While enumerating all of them is not within the scope of this work, some of the most prominent approaches are presented here. ((Van Slype, et al. 1984); (Hutchins and Somers 1992)) and (Lehrberger and Bourbeau 1988) shall be discussed in turn. See further, (Krenz and Ramlow 2008).

4.5.1 Van Slype's approach to MT Evaluation

In Van Slype's description (Van Slype, et al. 1984), MT evaluation can be categorized in terms of the aim of the evaluation and the evaluation doer. When the evaluation is described in terms of who carries out the evaluation, then a distinction can be made between upstream evaluation and downstream evaluation. On the one hand, an upstream evaluation is carried out by the manufacturer before the product is released for commercialization; whereas, a downstream evaluation will be carried out by the end-user who wants to make sure the product meets its description. When the evaluation is described in terms of the aim of the evaluation, then the macro-evaluation is opposed to the micro-evaluation.

4.5.1.1 Macro-evaluation

The macro-evaluation refers to an assessment of the productivity or efficiency of a translation system. The aim here is to determine to what extent a machine translation system meets users' requirements. Van Slype distinguishes at least four different levels for macro-evaluation:

Cognitive level:
- Intelligibility;
- Fidelity;
- Coherence;
- Usability; and
- Acceptability.

Economic level:
- reading time;
- correction time; and
- translation time.

Linguistic level:
- reconstruction of semantic relationships;
- syntactic and semantic coherence;
- "absolute" quality;
- lexical evaluation;
- syntactic evaluation; and
- analysis of errors.

Operational level:
- automatic language identification; and
- verification of the claims of the manufacturer.

4.5.1.2 Micro-evaluation

In a micro-evaluation, a detailed list of translation errors is established followed by an analysis of the origin of these errors. The error analysis aims at determining the improvability and the subsequent improvement of a system. For Van Slype, micro-evaluation is divided into 5 sequences or levels. These range from the detection of "symptoms" to the therapeutic measures that can be taken to improve a given system:

Grammatical symptomatic level:

In a first sequence, "symptoms" are identified through the analysis of the grammatical errors detected in the translated texts.

Formal symptomatic level:
This comes after the revision and post-editing task has been performed. During a formal symptomatic evaluation, deletions, additions, modifications, shifts and replacement of words by the revisers and post-editors are counted and rated.
Diagnosis level:
This level of evaluation mostly refers to the output text and various sources of error that they may contain. Indeed, a source text that contains typing errors or words that are unknown to the dictionary may lead to a wrong translation. Therefore, analyzing the input can help identify the types of structures that are problematic to a given translation system.
Forecast level:
After a diagnosis is established, proposals can be made on how to improve a given translation system.
Therapeutic level:
Evaluation belongs to the normal life cycle of any software. Even when improvements are made and the system has been upgraded, a "therapeutic evaluation" is still necessary to make sure that the implemented solutions are effective.

4.5.2 Lehrberger & Bourbeau's approach to MT Evaluation

Lehrberger and Bourbeau (Lehrberger and Bourbeau 1988) propose a methodology that distinguishes three levels of machine translation evaluation: an evaluation by the manufacturer; a cost-benefit evaluation by the user and a linguistic evaluation.

4.5.2.1 Evaluation by the manufacturer

This form of evaluation refers to the review of the performance of a system. Here, the origins of the errors are determined in order to optimize the system. This step can be equated to Van Slype's "formal symptomatic and diagnosis levels" of evaluation.

4.5.2.2 Cost-benefit evaluation by the user

This type of evaluation aims at assessing the profitability of a system. It is done in form of a comparative study in which the revision time and cost of human translation is compared to the revision time and cost of machine translation. Compared to human translation, machine translation is undoubtedly much faster; however, the time spent editing machine translation could prove to be longer than the whole human translation process (translation and editing combined). A profitable machine translation system is, therefore, one that requires few or no revision at all. In this regard, a machine translation system that is combined with a translation memory may be the best way to reduce time and cost.

4.5.2.3 Linguistic evaluation by the user

This type of evaluation refers to the evaluation of the performance of an MT system in terms of the linguistic quality of the MT output. While some MT systems may provide a translation of publishable quality, others, on the other hand, produces translations that are nearly undecipherable. The performance of a system may, also, depend on the quality of the input and may not be consistent from one text type to another. Compare the following translations, provided by the same MT system:

Source Text 4[7]

Die Europäische Kommission

Die Europäische Kommission ist das Exekutivorgan der EU und vertritt die Interessen der gesamten EU (im Gegensatz zu den Interessen einzelner Länder).

Der Begriff „Kommission" bezeichnet sowohl das Kollegium der Kommissare als auch das europäische Organ selbst, dessen Hauptsitz sich in Brüssel (Belgien) befindet, mit Büros in Luxemburg. Die Kommission hat überdies so genannte Vertretungen in allen EU-Mitgliedstaaten.

Target Text 4

La Commission européenne

La Commission européenne est l'organe exécutif de l'UE et représente les intérêts de l'UE dans son ensemble (par opposition aux intérêts de chaque pays).

Le terme «Commission» désigne à la fois le collège des commissaires et l'institution européenne elle-même, dont le siège est situé à Bruxelles (Belgique), avec des bureaux à Luxembourg. La Commission a également demandé des représentations dans tous les États membres de l'UE.

Source Text 5

Angeschlossene Druckleitungen bzw. -schläuche

An das Gerät angeschlossene Druckleitungen bzw. -schläuche, Fittings und Verschraubungen müssen dem max. Betriebsdruck von 100 bar/1450 PSI mit

[7] (Kommission 1997)

dem Sicherheitsfaktor 2,5 entsprechen, d.h. die Komponenten müssen für einen Mindestbetriebsdruck von 250 bar/ 3625 PSI ausgelegt sein.

Target Text 5

Conduites et tuyaux de pression affiliés

Pour les périphériques connectés conduites de pression ou de tuyaux, raccords et accouplements ont le max. Pression de service de 100 bar / 1450 PSI avec un facteur de 2,5 correspondent, à savoir la sécurité les composants doivent être conçus pour une pression de service minimale de 250 bar / 3625 PSI.

Although both texts were translated using the same MT system, they performed differently on the fluency measure. Text 4 obtained 0.7 while text 5 obtained 0.3, in the fluency measure, which clearly indicates the fact that users' evaluation of the linguistic quality of an MT system is highly subjective and that it, furthermore depends on the text type that is submitted to translation. This type of simple evaluation by the user may help them decide whether a given system is suited for the type of task they plan to undertake.

4.5.3 D. Arnold's approach to MT Evaluation

Taking a user's perspective (Arnold, et al. 1994), discusses the most common evaluation methods. These methods describe some of the ways a potential buyer of an MT system can determine whether purchasing such a system might be profitable to their business. In Arnold's (op. cit.) description, two main forms of evaluation are highlighted: performance and operative evaluation.

4.5.3.1 Performance evaluation

An MT system's performance can be assessed in terms of the intelligibility score, the accuracy of the output, error analysis and test suite.

Intelligibility

Under this criterion, an MT system is evaluated in terms of how intelligible the output sentences are. An intelligible target language sentence is one that displays lower scores of grammatical errors, mistranslations and untranslated words, because these can affect the way a sentence is construed.

Accuracy

An accurate translation is one that renders the information in the source language faithfully, i.e. so that no piece of information is diluted in the process of translation. The form and style may differ depending on the aim of translation, but the main information must be accurate.

Error analysis
Error analysis helps make the account of the main errors that an output may display and, therefore, provides an insight into how much post-editing work is needed. The higher the error rate, the poorer the MT system is. Error analysis can also be a valuable tool for system developers, as it allows them to identify areas of the system that need to be improved.
Test suite
A test suite consists of a number of sentences, displaying various types of linguistic difficulties that are submitted to an MT system to see how they handle these phenomena. In the context of the present research endeavor, a test suite would consist of various sentences with varying degrees of ambiguity. This helps the developer pinpoint areas where improvement is necessary. To complete this range of evaluation possibilities, (Arnold, et al. 1994) propose that an operative evaluation also be carried out.

4.5.3.2 Operative Evaluation

This type of evaluation is a more pragmatic one which is carried out within the context of use. The aim here is to allow the prospective buyer of the MT system to conduct a testing of the engine he/she intends to purchase. By doing so, they can precisely assess how the introduction of such a system impact the workflow as well as the cost devoted to translation. For this to work, a comparison has to be made between the time and cost of human translation and the time and cost of introducing an MT device into the workflow. The time and cost of post-editing is a determining factor because it determines the profitability of the whole process.

4.5.4 Hutchins' approach to MT Evaluation

From the outset, (Hutchins and Somers 1992) identify five different stages at which evaluation may be carried out. They correspond to the different stages of an MT system's life: during the development of a prototype; during the actual development of the system; while the system is being operated; and finally, an evaluation by translators and recipients of the translation.

4.5.4.1 Prototype evaluation

This is carried out by the system developer during the development phase as they want to ensure that the system functions properly. In a rule-based system, for instance, the system developer would make sure that the changing in the translation rules of the system doesn't negatively affect the proper functioning of the system. The system developer may, furthermore want to ensure that changes made, in the dictionary, positively affect the translation output.

4.5.4.2 Development evaluation

The development phase includes "the design of facilities for inputting text, compiling and updating dictionaries, for revising output, for interacting with the computer etc" (Hutchins 1992). At this stage, the system is tested against its ability to adapt to specific working environments and operating systems.

4.5.4.3 Operational evaluation

End users are essential in any evaluation process as they are the most affected (positively or negatively) by the product; therefore, they must at a certain point figure out whether the MT system under development fits their specific needs and whether the claims of the manufacturer are true or not. In this regard, the potential user may want to know what kind of proficiency is required of the user.

4.5.4.4 Translator evaluation

As was pointed out earlier, cost effectiveness is key. Potential users of a system are primarily interested in knowing whether an investment in a given MT device is time and cost effective compared to HT. Human translators are best equipped to carry out this task. They master the translation workflow and they know best how to be time and cost effective and, therefore, the translator evaluation may want to know how introducing an MT system affects the translator's workflow.

4.5.4.5 Recipient evaluation

Just like translators, recipient of a translation, are interested in knowing how cost-effective a translation system may be. In translation practice, certain Language Service Providers, and clients, expressly forbid the use of machine translation in performing the task because they claim this may affect the translator's creative ability. Other clients, on the other hand, make extensive use of machine translation and only require post-editing from the translator. The recipient evaluation is concerned with the linguistic quality of the translation output.

The different approaches reviewed above reflect the multiplicity of conceptions regarding MT evaluation; however, when taking a closer look at these different approaches one finds that they may seem very different, at first, but they have some similarities which are summarized in the recap chart below. The similarities are symbolized in colors.

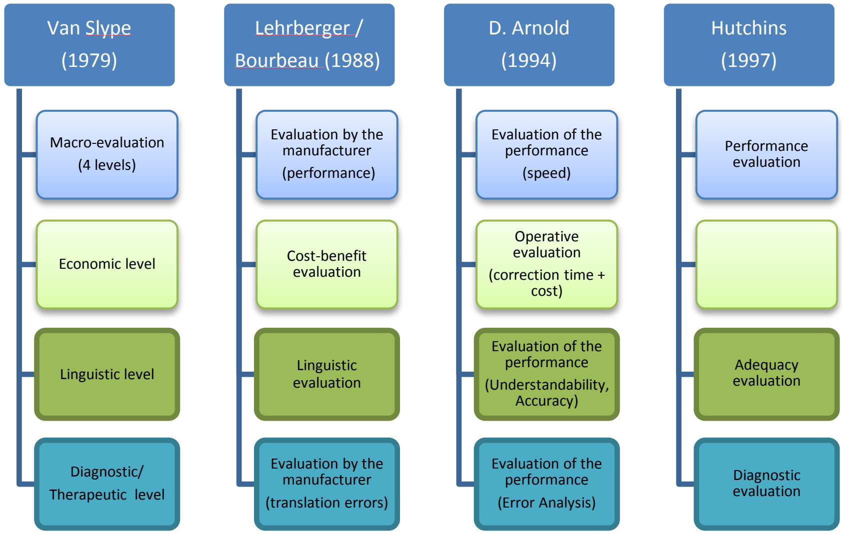

Figure 9: Comparative review of some evaluation methods.

4.6 Pro and cons of human and automatic evaluation

Evaluation plays a crucial role in machine translation. Once the different evaluation approaches have been presented, it is useful to highlight the advantages and disadvantages of both human and automatic approaches.

4.6.1 Speed

While human evaluation might be tedious and time consuming, automatic evaluation takes little time and is easy to perform. Preparing the data for automatic evaluation, however, is a laborious, arguably even more so, task as well. Even though the alignment of source and target text may be done automatically, much time is spent "cleaning up" the aligned data to ensure consistency, and therefore quality. Moreover, automatic evaluation requires a reference translation that needs time to be performed, but at the end of the day, human evaluation proves more time consuming because annotators need to go through long corpora and grade the translation sentence by sentence. This may take days, if not weeks. Annotators also need to be familiarized with the testing scheme, which may take additional time to be completed.

4.6.2 The cost

Carrying out human evaluation comes at a cost. It is practically impossible to find human annotators who are willing to spend hours, or days, carrying out evaluations without any financial compensation. If one considers, that a reliable evaluation is one that relies on multiple annotators, this multiplies the cost that is associated with using human evaluators.

The disadvantages of human evaluation are well known: it takes time; and it is expensive. While human judgment may be subjective, they usually display higher degrees of reliability than automatic MT as automatic evaluation systems do not provide details about the nature of translation errors. In (Sun 2010, 1726) terms, "human judgement is the benchmark to assess the usefulness of automatic evaluation metrics." Based on this assumption, a human evaluation of machine translated corpora was conducted. The next section presents the methodology that was employed for the experiments.

4.7 Empirical Part: Experiments with GT and PT

The present evaluation section has two parts: a test and an evaluation. The test consists of translations obtained through Google Translate and Personal Translator. A first analysis of the MT outputs is performed with a focus on how ambiguous segments in the source text are handled by both systems. The evaluation consists of a collection of human judgments on the quality of the machine translated output. The evaluation will help determine if MT outputs of source text containing ambiguous segments are consistently evaluated in a different way than the MT output of source text without ambiguity.

4.7.1 Methodology

Under this section, the methodology employed to arrive at the results is discussed. In the absence of a standard method in evaluating machine translation, this study was faced with the question of choosing the most appropriate method to attain its purpose. It has previously been mentioned that there are two major approaches to MT evaluation namely human and automatic evaluation. Another big issue, facing MT evaluation, is whether human or automatic evaluation is the most reliable method of evaluation. What is the amount of corpus to be considered for the evaluation to be reliable? When considering human evaluation, for instance, how large should the panel of judges be to guarantee reliability? It is often argued that the larger the evaluation corpus, and/or the larger the number of subjects, the more reliable the MT evaluation results would be. See (Elliott, et al. 2003) and (Kulesza and Shieber 2004).

Other studies have focused on the specific question of the volume of the corpus that is needed for a reliable MT evaluation. (Estrella, Hamon and Popescu-Belis 2007) have evaluated through a bootstrapping method, the reliability of MT evaluation on a scale. For this study, the authors evaluated several machine translation systems and focused on a series of corpora numbered from one to fifteen. The MT outputs were first evaluated based on one single corpus; gradually, more corpora were added until the fifteenth text. The aim was to measure how increasing the volume of the corpus affects the outcome of the evaluation. This experience covered both human evaluation methods and automatic evaluation metrics. The texts used in this study are those of the CESTA campaign, which assesses the automatic translation of texts from English into French. Thus, the corpus consisted of 15 documents from the Official Journal of the European Community, made up of 790 segments in total, with an average of 25 words per segment. The results are displayed in the following table borrowed from (Estrella, Hamon and Popescu-Belis 2007):

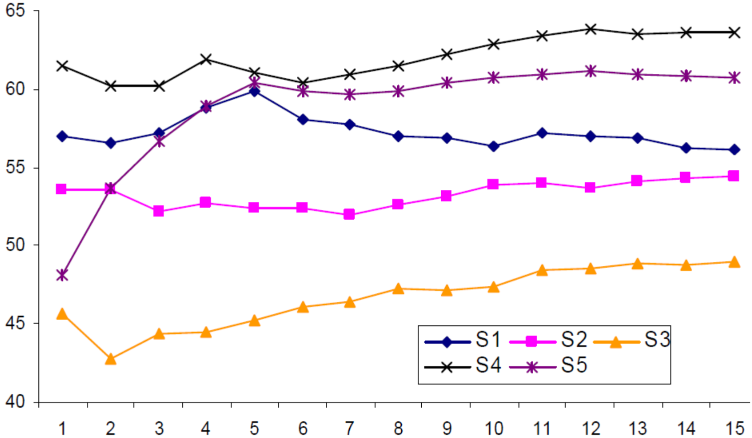

Figure 10[8]: Average adequacy values (on a 0-100% scale) for systems S1 to S5, computed on 15 documents.

The stated goal of the experiments was to "observe how the average scores obtained by human judges and automatic metrics evolve, as more documents are incrementally added to the evaluation corpus." (Estrella, Hamon and Popescu-Belis 2007, 168). The observation that can be made is that the values appear to be stabilizing at ten texts. Between the tenth and fifteenth text, the values appear to change marginally. From this, it can be concluded that huge amounts of data are not absolutely necessary to obtain reliable evaluation results, as is usually put forward. This study indicates that a significant volume of corpus does not necessarily affect the evaluation of the results. These findings are decisive in any evaluation endeavor because less test data also means less cost and, therefore, less time dedicated to the evaluation. Another important finding of this study is that: "for human or automatic evaluation about five documents from the same domain -with ca. 250 segments or 6,000 words- seem sufficient to establish the ranking of the systems and about ten documents are sufficient to obtain reliable scores" (Estrella, Hamon and Popescu-Belis 2007, 174).

[8] Borrowed from (Estrella, Hamon and Popescu-Belis 2007)

Borrowing from these results of the aforementioned study, the corpus for the present evaluation will be limited to ten texts which will be machine translated and submitted to human evaluation. The corpora used for the current evaluation consisted of a subset of six documents from various internet sources and displaying various forms of ambiguity, especially lexical ambiguity. This subset of German texts was machine translated into French. The second subset consisted of four English texts, also retrieved from the internet, with the same characteristics as the first subset, the only difference being the translation direction, English-to-French. The documents were segmented and aligned into sentences of variable lengths *via* the *Bitext2tmx*[9], an automatic sentence alignment software. Each test corpus was of a variable length ranging from 145 up to 695 words amounting to a total of 2940 words, divided in to a total of 187 segments. The corpora were UTF-8 encoded following the tagging format defined by the COSTA MT Evaluation Tool and the texts were organized in three parallel text files.

Translations were obtained from two different origins. Firstly, the MT translations performed with Google Translate and Personal Translator; no human processing of these translations was done. The second set of translations (hereafter, simply referred to as "reference translation") consisted of expert human translations of the same material from professional translators. This served as a reference for the evaluation method described under this section. These data were aligned and stored in three parallel files namely one source file (made up of all the source texts), two machine translation files, where the machine-translated versions were stored, and finally the reference file, which consisted of the human expert translations. Based on a 1-5 scale, the evaluators were instructed to rate the translations on their fluency and adequacy following the metrics defined during the DARPA 1994 campaign: (White, O'Connell and O'Mara 1994).

4.7.1.1 The evaluation tool

The human judgments were obtained using the Web-based interface COSTA MT Evaluation Tool[10] (Chatzitheodorou and Chatzistamatis 2013) which displays translated segments to the evaluator. The COSTA MT evaluation tool is a human evaluation system based on the evaluation criteria commonly used

[9] http://bitext2tmx.sourceforge.net/ (accessed March 15, 2014)

[10] The software is available for download at: https://code.google.com/p/costa-mt-evaluation-tool/ (accessed March 15, 2014)

including fluency, adequacy, and the translation error classification. Figure 11 below illustrates fluency evaluation with the COSTA MT evaluation tool.

Figure 11: Costa MT evaluation Tool.

The fluency and adequacy measures were the metrics retained to assess the degree of understandability of the MT output. Twenty-five judges initially volunteered for the evaluation and fourteen eventually returned their copies. Only ten were exploitable, which remains within the frame that was previously defined for this study. Fluency and adequacy were assessed separately.

4.7.1.2 Fluency

The objective of the fluency evaluation is to determine how much like "good French" a translation appears to be, without taking into account the correctness of the information. In this evaluation, evaluators who were native speakers of French made intuitive judgments on a sentence by sentence basis, for each translation (on a 1-5 scale), without access to any reference text. The source text, however, was displayed (*cf* Figure 11). The annotators were given the following definitions of fluency (Chris Callison-Burch 2007):

Fluency

5. Flawless language;

4. Good language;
3. Non-native language;
2. Disfluent language; and
1. Incomprehensible.

4.7.1.3 Adequacy

The objective of the adequacy is to determine the extent to which all of the content of a text is conveyed, regardless of the quality of the French in the output. The annotators were given the following definitions of adequacy also borrowed from (Chris Callison-Burch 2007):

Adequacy

5. All meaning;
4. Most meaning;
3. Much meaning;
2. Little meaning; and
1. None.

Unlike the fluency evaluation, the annotators had access to the reference translation, which is human performed translation. This helped determine whether the machine translated output could be deemed adequate compared to the reference text.

4.7.2 Aims of the experiment

With respect to the above review on evaluation in machine translation, the following evaluation falls within the frame of a macro-evaluation and a micro-evaluation. It is a macro-evaluation because it compares the quality of two different MT systems, but it is also a micro-evaluation because the detailed evaluation of fluency and adequacy provides some insight into the improvability of those systems. This evaluation can also be labelled a linguistic evaluation because it assesses the performance of MT systems with respect to ambiguity. Lastly, this will be a diagnostic evaluation because the set goal is to figure out how ambiguity affects the intelligibility of MT outputs.

4.7.3 Corpus analysis

In chapter three, above, various forms of ambiguity were presented; however, only lexical ambiguity will be assessed under the present evaluation because of the difficulty to assess other forms of ambiguities in a larger context. While it is easy to determine how a lexically ambiguous word influences the understandability of a text, it is a much more complicated venture to perform the same task for other forms of ambiguities. To find out how an ambiguous word in the source text influences the understandability of a text, it has to frequently appear in the corpus for it to be considered relevant, in the overall meaning. If the ambiguous unit appears only once then, its influence on the overall understandability may be considered minimal. Lexical ambiguities tend to be more recurrent in a given corpus than syntactic or elliptical ambiguities for instance. Evaluating the influence of syntactic or elliptical ambiguity on the understandability of a given corpus would generally be carried out through a test suite, which is an artificial construction. A test suite generally consists of a string of sentences which are meant to illustrate a given phenomenon. The present study, however, is aimed at evaluating ambiguities as they might appear in natural situations.

Below, is a table of the ambiguous words that appear in the corpora that were selected for the current experiment and their possible translations in the target language:

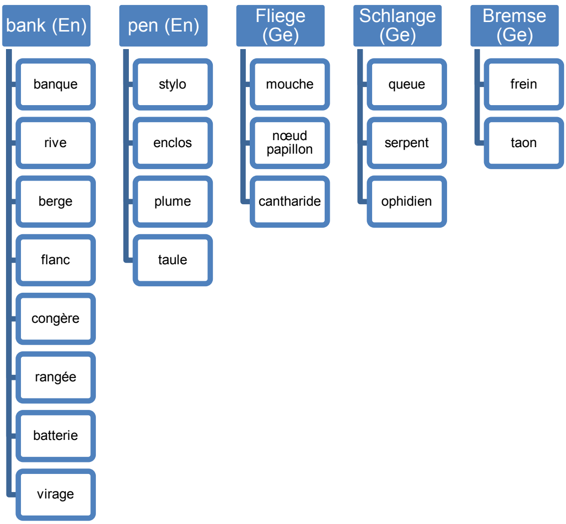

Figure 12: List of ambiguous words and their translations[11] into French.

[11] Definitions are obtained from the Cambridge English-French Dictionary

4.7.4 Analysis of the MT Output

For the purpose of the present experiment, two sets of source texts were chosen. In the first set, the above mentioned lexical items appear in their most common meaning, that is, the first definition that appears in most dictionaries. In the second set of texts, the same lexical items appear in a different context with a different meaning. For instance, the German word "*Fliege*" first appears in a context where its "insect" sense is highlighted. In the second context however, it bears the meaning "bow tie". The current experiment therefore only used the first and second definitions as presented in the first and second rows of Figure 11 above. The aim was to understand how the MT systems react to context change.

The following two tables display how the same German word "*Fliege*" is used within two different contexts and how Personal Translator, and Google Translate, react to this change of context: In the first text[12], "*Fliege*" means a "fly" while in the second corpus[13], it means "bow tie".

[12] http://www.augsburger-allgemeine.de/bayern/Fliege-laesst-Motorradfahrer-stuerzen-id27154727.html (accessed July 15, 2015)

[13] http://www.krawatte-binden.com/fliege.html (accessed July 15, 2015)

Source Text	Personal Translator	Google Translate
Fliege lässt Motorrad-fahrer stürzen.	**Mouche** laisse ren-verser motocyclistes.	**Fly** permet de renver-ser les motocyclistes.
Weil er eine **Fliege** ins Auge bekommen hatte, verunglückte ein Motorradfahrer im Landkreis München.	Parce qu'il avait reçu une **mouche** dans l'œil, un motocycliste a eu un accident dans le district Munich.	Parce qu'il avait une **mouche** dans l'œil, un motocycliste a été tué dans le quartier de Munich.
Kleine **Fliege**, großer Schaden: Ein 55-jähri-ger Motorradfahrer ist am Mittwochmittag auf der Kreisstraße zwischen Dingharting und Beigarten gestürzt, weil ihm eine **Fliege** ins Auge geflogen war. Als er das Insekt bemerkte, bremste er das Motorrad stark und fiel auf die Straße.	Petite **mouche**, dom-mage important : Un motocycliste de 55 ans est tombé sur la rue de cercle entre Dinghar-ting et de l'à-jardin le midi de mercredi parce qu'une **mouche** avait volé pour lui dans l'œil. Quand il remar-quait l'insecte, il frei-nait fortement la moto et il tombait dans la rue.	Petite **mouche** grand mal : A 55-year-old motocycliste s'est écrasé mercredi après-midi sur la route de comté entre Dinghar-ting et Beigarten parce que c'était une **mouche** vole dans l'œil. Quand il a vu l'insecte, il a freiné le vélo solide et est tombé sur la route.

Table 4: Translating “Fliege” in the insect context.

Source Text	Personal Translator	Google Translate
Die **Fliege**	Les **Mouches**	The **Fly**
Fliegen sind in denselben Farben und Mustern wie die klassische Krawatte erhältlich.	Les **mouches** sont disponibles dans les mêmes couleurs et dans les mêmes modèles que la cravate classique.	Les **mouches** sont disponibles dans les mêmes couleurs et les motifs comme la cravate classique.
Eine **Fliege** ist meistens schwarz und wird zu einem Abendanzug und einem weißen Hemd mit Frackkragen getragen. Sie kann auch etwas zwangloser zu einem Anzug und einem Hemd mit breitem Kragen getragen werden.	Un **nœud papillon** Est la plupart du temps noir et il est porté à un smoking et une chemise blanche avec Frackkragen. Il peut être porté aussi un peu plus sans contrainte à un costume et une chemise avec un col large.	Une **mouche** est généralement de couleur noire et est porté à un costume de soirée et une chemise blanche à col smoking. Il peut aussi être quelque chose de moins portés avec un costume et une chemise avec un col large.

Table 5: Translating “Fliege” in “a-bow tie” context.

From the above Table 4 and Table 5, Google Translate rendered “*Fliege*” as “*mouche*” in most instances, despite the textual indications providing the context of use. In fact, in the first text, “*Fliege*” appears in its most common sense which is “*mouche*”. Textual elements such as “*geflogen*” or “*Insekt*” corroborate the choice of “*mouche*” as the most appropriate translation; on the other hand, the choice of “*mouche*” as a translation for “*Fliege*” in the second text is not justified as textual elements such as “*Krawatte*; *Hemd*; *Anzug*; *Kragen*; *Knote*; *binden*” are all indicators that “*nœud papillon*” is the appropriate translation. We would assume that the alignment “*nœud papillon-Fliege*” does not exist in the Google corpus. To confirm this assumption, a text was translated from French into German, which included the term “*nœud papillon*”. As a result, “*Fliege*” was the preferred translation. This leads us to conclude that, indeed, the combination "*Fliege-nœud papillon*" is part of the aligned Google corpus.

It must also be pointed out that there are two instances in which Google Translate repeatedly translates "*Fliege*" as a "Fly", which is a surprising rendition of "*mouche*" because this is not a French word, but an English word. No clear explanation could be found for this phenomena that accounted for this mistranslation. In both rule-based and SMT, unknown words in the SL are usually rendered in their SL form, that would be "*Fliege*", in this instance.

The same two texts were translated with Personal Translator and produced unsurprisingly similar results, with one crucial exception, however: in this exceptional case, PT translated "*Fliege*" with "*nœud papillon*" which is the expected translation. It is nevertheless surprising to note, the lack of consistency in the translation of this word in the rest of the corpus. Indeed, "*Fliege*" was translated to "*mouche*" in all other instances of this word. We will try to remedy this by producing a glossary that defines "*nœud papillon*" as the preferred translation for "*Fliege*" when it is used in a clothing context. For the description of the complete corpora, please refer to the appendix.

4.8 Presentation of the findings

This section will be devoted to the presentation of the results of the evaluation. In chapter five, further analysis of these results shall be carried out. The ten texts selected for evaluation were subjected to a panel of ten annotators. Originally, twenty-five evaluators had volunteered to take part in the evaluation, but in the end, only fourteen of them were able to submit their copy (which represents a ratio of 82%) and ten were exploitable. The first criterion for selecting the evaluators was target language. The evaluators were required to be French native speakers. This criterion was important in judging the fluency measure. Annotators were also required to have written knowledge of German and English. Most evaluators were studying or had studied translation and were more or less familiar with the subject. However, specific instructions were given as to the conduct of the evaluation, without specifying the objective, to avoid any bias in favor or against the set goal. The results will be presented in figures. As a reminder, the lexical ambiguities appeared recurrently in the following forms:

Bank: financial institution	**Bank**: riverside
Pen: writing tool	**Pen**: enclosure for animals
Fliege: insect	**Fliege**: bow tie
Schlange: reptile	**Schlange**: queue
Bremsen: brakes	**Bremsen**: horsefly

The results presented below are each an average of the evaluation of the different ten texts.

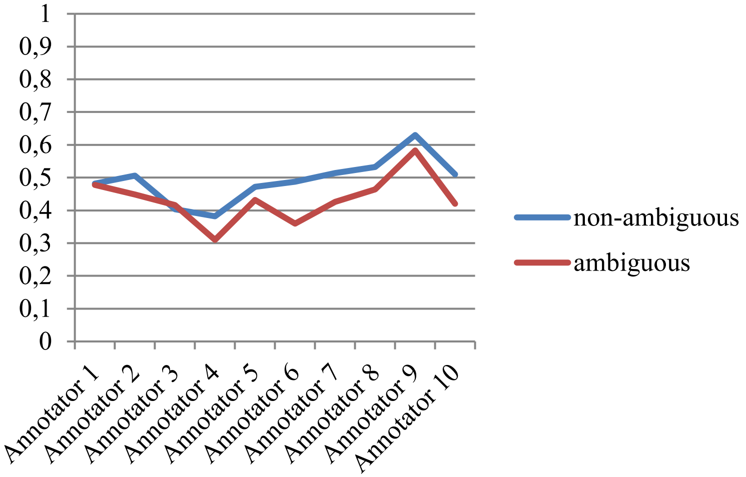

Figure 13: Results of human evaluation of fluency for Google Translate with and without ambiguous utterances.

4.8.1 Fluency results for Google Translate

From these results, it appears that the values range from 0.31 to 0.63, with a majority of judgments below 0.5. It also follows from this that in a majority of cases (33/17), non-ambiguous texts get better values than texts containing ambiguities. This seems to confirm our early hypothesis on the fact that the quality of machine translation is even worse when the source text contains ambiguities.

In general, the judgment on the fluency of the translations is quite severe for both texts with and without ambiguity. The values barely reached 0.5.

4.8.2 Fluency results for Personal Translator

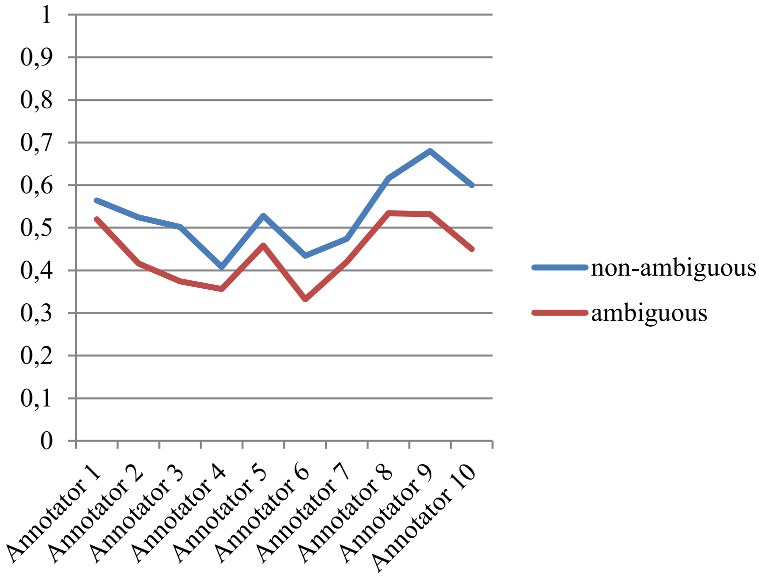

Figure 14: Results of human evaluation of fluency for Personal Translator with and without ambiguous utterances in the ST.

Fluency values for Personal Translator are clearly distinctive depending on whether the texts contain ambiguities or not. Thus, non-ambiguous texts outweigh the ambiguous texts in positive value, but it is still to be observed that the different judgments do not display a large variation-maximum of 0.1 points between ambiguous and non-ambiguous corpora respectively, even if the scores range from 0.332, on average, and for the most severe 0.68 on average for more generous.

4.8.3 Adequacy results for Google Translate

Figure 15 summarizes the results of human evaluation of adequacy for Google Translate with and without ambiguous utterances.

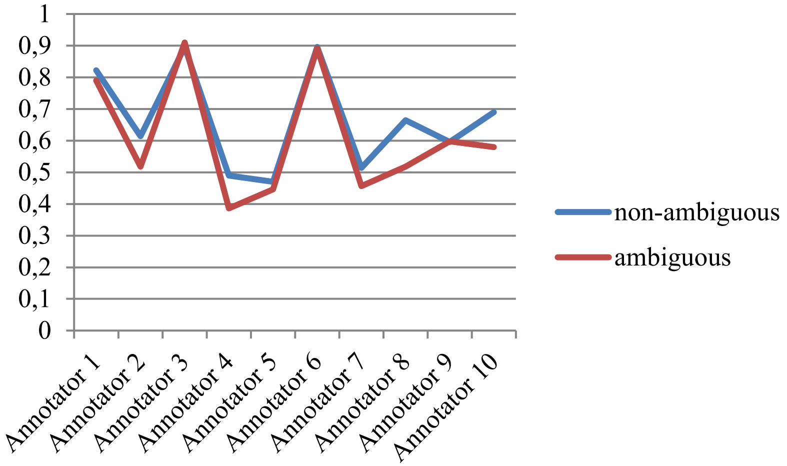

Figure 15: Results of human evaluation of adequacy for Google Translate with and without ambiguous utterances in the ST.

The adequacy scores exhibit significant variations from one annotator to another (e.g. 0.4 vs 0.9 for annotators 4 and 6 respectively). The unambiguous texts certainly emerge with a majority of favorable judgments, but the difference seems very small. Unlike judgments about the fluency, the scores assigned to the adequacy are, in a majority of cases, above 0.5 i.e. (15/20). Some values were even closer to 1, which is the maximum value.

4.8.4 Adequacy results for Personal Translator

Figure 16 below summarizes the results of human evaluation of adequacy for Personal Translator with and without ambiguous utterances.

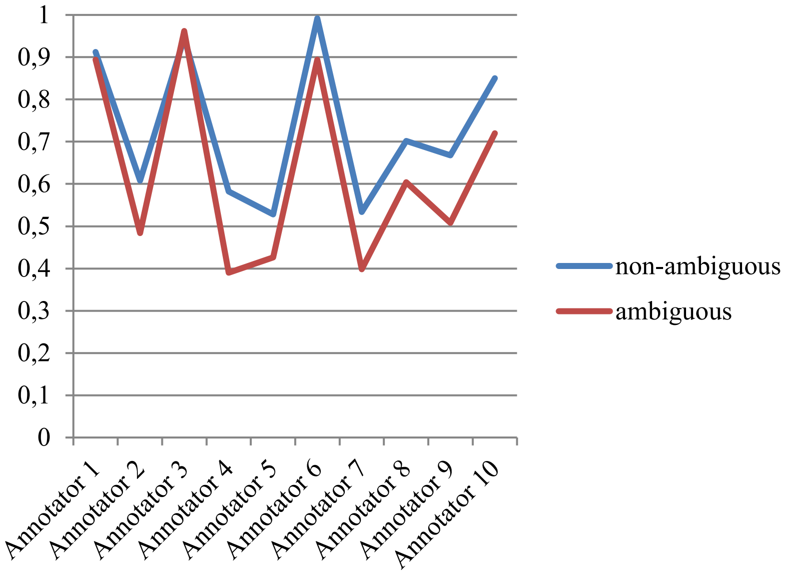

Figure 16: Results of human evaluation of adequacy for Personal Translator.

The above figure presents the adequacy values of ten texts for Personal Translator. Again, the values differ widely from one annotator to another. The values range from 0.47, for texts with ambiguity, to 0.96, for texts without ambiguity. There is a majority of positive values on the adequacy of unambiguous texts, which again seems to confirm our hypothesis that the presence of ambiguity in a SL corpus influences the quality of its machine translation.

5 Summary of Results (Analysis and Synthesis)

In the beginning of this work, it was found that machine translation is still struggling to solve the problems of ambiguity and that this could influence the understanding or the quality of MT output. Given this situation, we decided to explore the concept of ambiguity, to better understand its implications in machine translation. This study was aimed at quantifying the influence that lexical ambiguity in the ST may have on the quality of automatically translated texts. In pursuance of completing this goal, the criteria of adequacy and fluency were identified as key indicators of the quality of automatically translated text. In the following sections there will be an analysis of the results of the human evaluation that was carried out in the course of the present study. This data analysis will enable the research team to respond to the research question: how does the presence of lexical ambiguity in the source text influence MT quality?

This chapter will be divided into three sections. The first, a comparative analysis of the results of the texts with and without ambiguity. The objective of the first section is to test whether, the presence of ambiguity in the source text influences the perception of fluency and adequacy. During the second part, the analysis focuses on a comparison between the two translation systems. One will then observe that there might be a significant difference between the two systems in terms of translation quality, again, based on the criteria of fluency and adequacy that were chosen as the basis for analysis. In the final step, the significance and importance of these findings shall be discussed.

5.1 Fluency Google Translate

In translation, fluency relates to the extent to which a translation reads naturally, as if written by a native speaker. It is useful to note, however, that the fluency metric alone is not sufficient to determine the quality of an MT system. Some MT systems may produce translations that are fluent, albeit not properly rendering the content of the message. The adequacy metric, therefore comes into play as a complementary step which accounts for the accuracy of the message transfer from the source language to the target language.

Fluency is one of the criteria most often used in different evaluation campaigns. It reflects the quality of a text beyond the veracity of the information contained in the text. The fluency measure is an important indicator of a translation software's ability to reproduce grammatically correct and semantically intelligible sentences. This criterion also has the considerable advantage that it is not biased by the presence of a reference translation. Fluency is only measured through the annotator's objective assessment, regardless of the actual content of the source text.

It appears from the various figures presented in the previous chapter that in a majority of cases (78%), the values assigned to the translations whose ST don't contain ambiguous segments clearly override the MT output whose ST contain ambiguities. This difference in values appears most clearly in Figure 13 (fluency results for PT). In this figure, there seems to be a relative consistency in the evaluation of the ten annotators who penalize more severely texts with ambiguous terms. This severity in the judgment could be attributed to the presence of ambiguous words in the ST that are translated incorrectly. If a key word is not translated properly, the understandability of the whole text may very well be affected. This is what happened in Table 4, where "*Fliege*" is mostly translated as "*mouche*" by both translation software. As a consequence, hereof, a mistranslation is produced in the target text. It, therefore, seems only logical that most evaluators barely attribute 0.5 points to the translation; however, it can be observed in this particular case that Personal Translator scores better than Google Translate. This is probably due to the fact that in one instance, Personal Translator has succeeded in translating "*Fliege*" as bowtie. The Google translation provides no clear clue which indicates that the text refers to a bowtie and this causes confusion.

In some cases, however, the score of the MT output containing ambiguous segments exceeds that of non-ambiguous texts. This is particularly true in Figure 12 (fluency measure GT) where most of the time, evaluator 3 awards better values to ambiguous texts than non-ambiguous texts. Further comparative look at the figures shows that this case goes against the general trend and remains peripheral.

Finally, it can be observed that in a majority of cases (72%), the combined fluency and adequacy value for Google Translate (non-ambiguous corpora) score an average above 0.5 points, even reaching 0.68. This indicates a rather good performance for Google Translate in terms of fluency and corroborates the claim that quality seems to improve when a text is non-ambiguous.

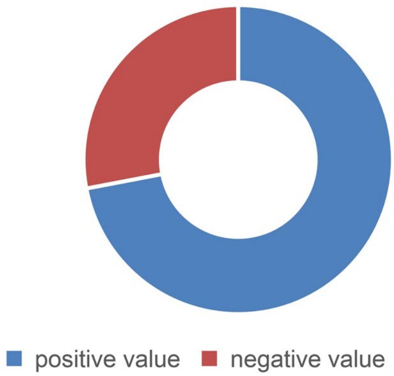

Figure 17: Combined fluency & adequacy value for Google Translate (non-ambiguous corpora).

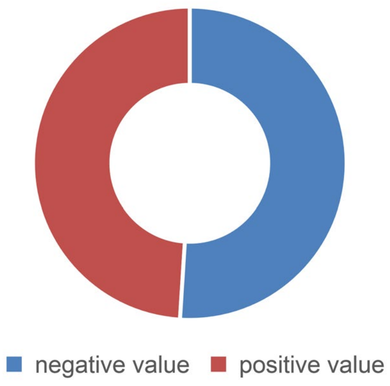

Figure 18: Combined fluency & adequacy value for Google Translate (ambiguous corpora).

5.2 Fluency Personal Translator

The fluency scores for PT are shown under Figure 13 above, where very distinctive features can be observed between the two sets of texts (with and without ambiguity). The scores form two nearly parallel lines, yet the positive scores awarded to non-ambiguous texts outweigh those of the ambiguous corpora: (70% vs 30%). Unlike the fluency score for Google Translate, evaluators are almost unanimous in their assessment. This could be explained by the fact that ambiguity was often less of a problem to PT. "Bow tie" has been mentioned as an example earlier on.

5.3 Adequacy Google Translate

Adequacy is a metric that helps evaluate the accuracy of the translation. As noted above, adequacy and fluency were considered separately as these metrics reflect two distinct phenomena. A text can indeed be fluent without being adequate and *vice versa*. Figure 14 (adequacy GT) gives us the adequacy values for Google Translate. This figure shows a large disparity in the evaluators' assessment of adequacy of translations. It should be noted that most of the scores lie above 0.5 and sometimes even reach 0.9 points. In general, the adequacy scores are higher (75%) than the fluency scores. It should be pointed out that adequacy is measured in the presence of the reference translation, that is, annotators are provided with a machine translation which they compare to a human translation before a judgement is made whether the machine translated segment conveys the same meaning as human translation. This comparative process could, in some cases, have influenced the judgment of the evaluators. Indeed, one can assume that in 25% of cases where the adequacy scores are lower than the fluency scores, the annotators may probably have been influenced by the reference translation that gives a hint on what a "good translation" should look like. Regarding judgments about ambiguous and non-ambiguous texts for Google Translate, no clear trend seems to emerge. In some cases, both ambiguous and non-ambiguous texts receive nearly identical scores. Again, it can be assumed that annotators are influenced by the reference translation. When the reference text comes into play, the focus is shifted to the reference text and the perceived influence of ambiguity decreases.

5.4 Adequacy PT

Unlike adequacy scores for Google Translate, a clearer trend seems to emerge in the adequacy measure of Personal Translator, that is, non-ambiguous corpora unequivocally score better than ambiguous ones. This confirms what has been previously observed concerning PT fluency scores. This hybrid machine translation system seems to better identify, and address, ambiguities than Google Translate, as was previously observed. It can also be observed that the adequacy scores of PT are very high, sometimes reaching 1. This is especially true for annotator 6 (Figure 15).

This study has, furthermore discovered that contrary to other studies, such as ((Carroll 1966); (Olive, Christianson and McCary 2011); (Wilks 2009)), there is no correlation between fluency and adequacy scores. Studies demonstrating the correlation between fluency and adequacy are often based on the argument that a fluent text is almost always adequate as far as the content is concerned. This assumption, it is argued, does not take into account the case where the text actually has a high degree of fluency. The correlation between fluency and adequacy is often less evident when the fluency scores are lower. It is quite conceivable that a text segment might score well in terms of adequacy while not being fluent at all. A text that is syntactically and grammatically calamitous would receive a very low score in terms of fluency, but if a text accurately conveys all the content of the source text, its adequacy score will certainly be much higher.

5.5 Fluency and adequacy correlation

Several studies have reported a correlation between fluency and adequacy. See ((Carroll 1966); (Krahmer und Theune 2010); (Depraetere 2011)). As part of the current study, we wanted to establish whether this claim could be deemed to be true. In terms of absolute value, it can be observed that adequacy tends to be rated far higher than fluency. There is an 85% positive score for adequacy while fluency gets positive scores only 15% of the time. This percentage variation suggests that fluency is, usually, more severely judged than adequacy.

This difference in positive scores could be explained by the way both metrics are evaluated. Fluency is evaluated in the absence of any reference translation. The evaluators' only tools are the source text and the target text, which in this case was a machine translated output. Even though annotators were provided with definitions of the different degrees of fluency, the MT output was to be held to the annotators' standard of "good language", which may differ depending on the latter's mastery of the target language. Another element that could very well justify the poor fluency scores is the fact that the fluency metric is judged on the basis of a system including standardized norms imposed by the language in terms of lexical combinations, syntactic construction and semantic coherence. Any breach of those standards is directly penalized by the evaluator.

The adequacy metric on the other hand is performed in the presence of a reference translation. The reference translation is a human translation that is supposed to reflect the meaning of the source text. The presence of the reference translation may have two effects: on the one hand, it can amplify the perception of poor quality that the annotator may already have of the translated text; on the other hand, the reference translation may instead mitigate that effect. The influence of the reference translation is not to be neglected since a person's perception of adequacy may vary from one individual to another, from one context to another. A translation intended for publication will be judged more severely than a translation intended for conveying non-essential information. A further point to note is that a translation may be deemed adequate for a specific purpose whilst it may not be fluent, therefore the severe judgement towards fluency. Fluency and adequacy correlation amounts to only 30% if one has to consider how the judgements match. The Pearson and Spearman's correlation coefficient confirmed this tendency of negative correlation between fluency and adequacy suggesting that a comprehensive evaluation has to consider both aspects, which contradicts views held in ((Olive, Christianson and McCary 2011); (Wilks 2009)).

This study shows that the fluency and adequacy metrics do not always correlate and, therefore, a comprehensive evaluation has to involve both metrics because they account for different aspects of a text's understandability. The data presented above, also show how a text can be well constructed on the lexical-semantic level but be a bad candidate for machine translation simply because it contains an ambiguous word that may be subject to various interpretations. Contextual analysis has often been advanced as a way to better handle ambiguity. While this may well be true in certain cases, the present study has shown that the presence of contextual elements is not always taken into account during translation.

In Section 4.1.2, it was stated that Google Translate Toolkit allows users to customize their translation by setting up a glossary which could be integrated to the translation process. A subsequent question then comes to mind. Does the integration of such a glossary impact the translation results and whether the sense-annotated data is taken into account during machine translation? Next is an illustration of the glossary:

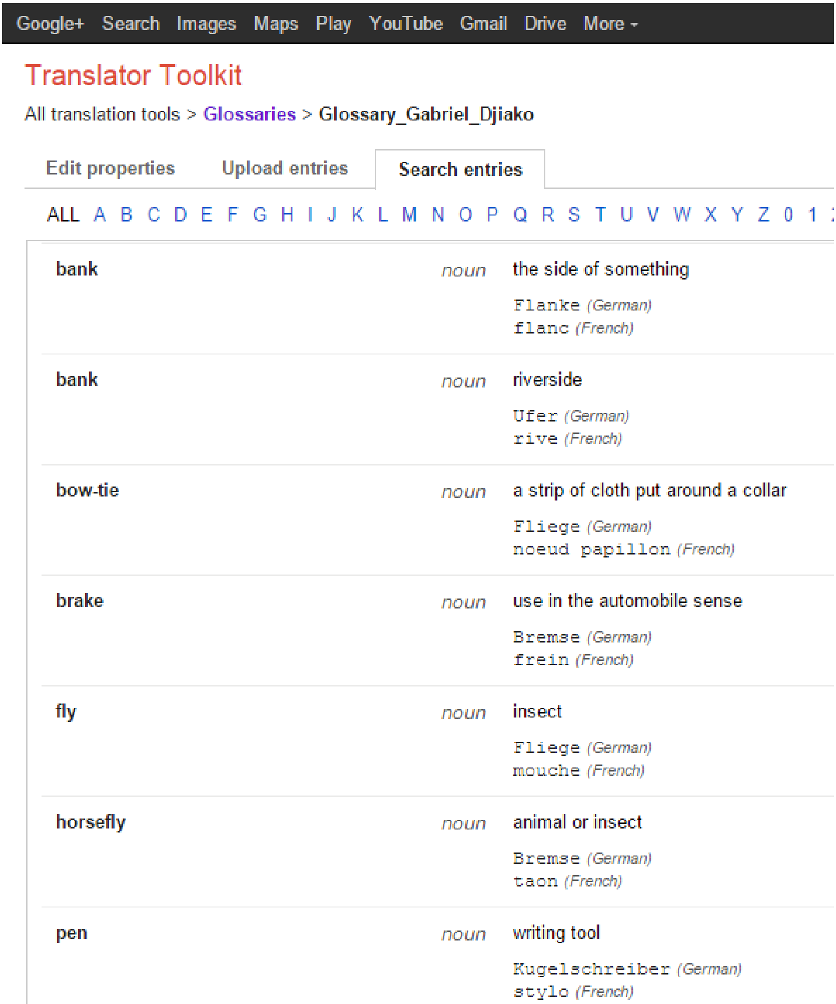

Figure 19: Example of a customized glossary in the Google Translate Toolkit.

The glossary primarily contained the ambiguous lexical items that have been selected in this study. Each candidate lexical item was entered with its various translations listed in Figure 12, as well as its POS, its translation and the context in which the lexical item can be used. After a CSV[14] glossary was created in UTF-8 format, it was uploaded and saved in the GTT. For a glossary to be used during translation, it has to be selected beforehand, just as one would select a translation memory. Once the glossary was customized and uploaded, the corpus was again subjected to translation.

Results showed that the introduction of a customized glossary did not change the output of the translation. Translations obtained were the same as the outputs obtained without glossary however, while trying to edit the MT output, the glossary was automatically activated when the segment containing any ambiguous lexical item appeared (see figure below):

[14] CSV or comma-separated values file stores tabular data in plain text.

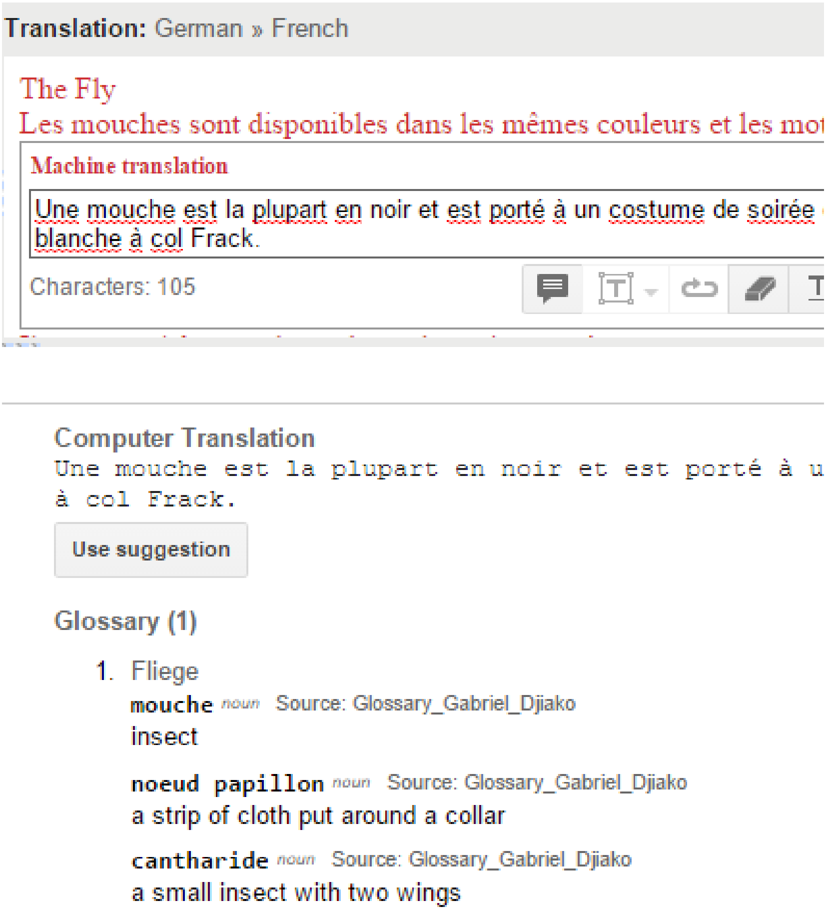

Figure 20: Glossary suggestions within the Google Translate Toolkit.

The activated box showed the source word, its various translations in the target language and annotations on the context of use. In conclusion, even though the customized glossary was not directly taken into account in the machine translation, it helped achieve a more accurate post-editing since fine-grained definitions were made available.

The possibility offered by GTT to customize the translation is obviously a step forward in the translation process since the translator has an additional tool - in addition to the translation memory - that helps save time and produces gains in accuracy. It should be noted however, that a better integration of the glossary during machine translation would be advisable. In other words, the user should be allowed to create domain-specific glossaries that are set up as "preferred translation" during the automatic translation process, and not during post-editing.

This study could, therefore, help translation system designers further explore this aspect of machine translation with an aim to refine the context analyzers. An example hereof would be where; one could imagine an option in a machine translation system that allows users to pre-define the meaning of keywords upstream of the translation. Key words are important because they define the themes represented in a document. An upstream "pre-determination" of the keyword would be a decisive step as it would avoid incorrect translations and improve translation fluency and accuracy. The next section further presents disambiguation approaches.

5.6 Ambiguity resolution in MT

Following the evaluation presented to this point, it clearly appears that improving the quality of translations goes hand in hand with resolving ambiguities. Several studies show that integrating a word sense disambiguation module to the translation system increases precision and therefore, the quality of the translation. For instance, (Dorr and Douglas 2000) discuss the use of a semantic filter for improving accuracy and verb meaning disambiguation. The evaluation following the introduction of the filter shows precision scores increasing from 62.5% to 85.3%, confirming the assumption that the resolution of the ambiguity in the text, whether lexical, semantic or syntactic, helps to improve the quality of machine translation. The question however, often lies in determining the best approach to achieve the best results. (Emele and Michael 1998), for instance, propose an ambiguity preserving approach whereby the ambiguous segment is maintained as is during translation by using underspecified representations. By so doing, the ambiguity of the source text is preserved and transferred into the target text therefore leaving it up to the reader to decide which meaning should be highlighted. This technique is unfortunately only possible with certain types of preservable ambiguities such as attachment ambiguities (see 3.6.4.3) as shown in the following example:

Example 28

(a) Wir treffen die Kollegen in Berlin.

(b) We meet the colleagues in Berlin.

From the above example 28 (a), it can be observed that syntactic attachment ambiguity can be preserved during translation without this influencing the quality of the translation output as neither fluency nor adequacy are affected. Lexical ambiguity is one such exception, because preserving lexical ambiguity can potentially lead to nonsense. In the following instances, some of the most prominent approaches to word-sense disambiguation are discussed.

5.6.1 Word-sense disambiguation

Word sense disambiguation (WSD) refers to the computational identification of the meaning of ambiguous words in context, (Navigli 2009). It is an important task, not only in machine translation, but also in other natural language processing tasks such as search engines, information retrieval, anaphora resolution etc. In the following, various approaches to WSD shall be reviewed.

5.6.2 Supervised WSD

Supervised WSD -(Navigli 2009)- is an approach that disambiguates by relying on manually sense-tagged /semantically annotated data. This can be done through a decision list or a decision tree.

Decision trees and decision lists refer to rules for assigning the appropriate sense to a target word. The decision tree, or list, identifies and classifies various semantic fields that a word can have. This initial classification enables a hierarchy of semantic proximity of words according to the context in which they are employed. In the case of a decision list, these are scores allocated to different features, depending on whether their meaning is similar to that of the term used. Selecting the appropriate word therefore depends on the score obtained, as shown in the table below. Each word with features that match those of the input word are attributed a score. The higher the feature scores, the closest it will match the input word.

Feature	**Prediction**	**Score**
account with bank	Bank/FINANCE	4.83
stand/v *on*/p … bank	Bank/FINANCE	3.35
bank *of blood*	Bank/SUPPLY	2.48
work/v bank	Bank/FINANCE	2.33
the left/J bank	Bank/RIVER	1.12
of the bank	-	0.01

Table 6: Decision list

In a decision tree, on the other hand, the selection of the target word will be based on the satisfaction of certain conditions or not. The conditions are expressed in terms of semantic or syntactic characteristics attributable to a word. The decision to select a word is taken only if at the end of the node, the requirement for a precise definition is met. Below is an example of a decision tree.

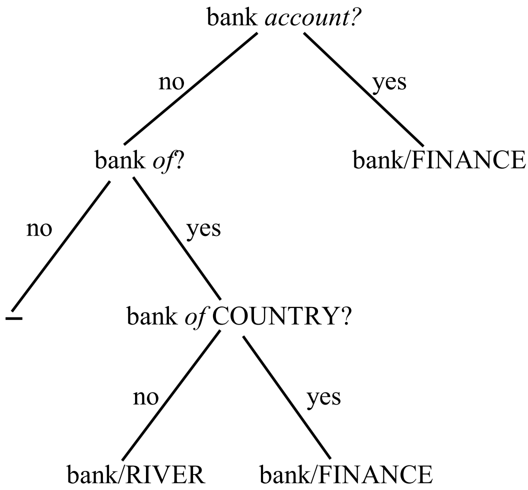

Figure 21: A decision tree

5.6.3 Unsupervised WSD

Unlike other supervised methods, the unsupervised WSD approaches do not rely on labeled training text and machine–readable resources to determine the meaning of words. Here, words are analyzed in context, under the assumption that a word that is used in similar contexts will "activate" the use of other words that are thematically or semantically related; therefore, by clustering word occurrences, the regularities that govern the use of a specific word are identified. Consider the following example:

Example 29[15]

(a) I went fishing for some sea bass.
(b) The bass line of the song is too weak.

Analyzing the word occurrence in the above example will yield the following:

- if *bass* has words *sea* or *fishing* nearby, it probably is in the fish sense;
- if *bass* has the words *music* or *song* nearby, it is probably in the music sense.

In a supervised approach, it will be easy to identify 'bass' in the dictionary and to determine its meaning. In the unsupervised approach however, all the words that tend to revolve around the target word will be clustered and the system will learn from these regularities.

5.6.4 Knowledge-based methods / dictionary based WSD

Knowledge-based systems exploit the information in a Lexical Knowledge Base to perform WSD, without using any corpus evidence, (Navigli 2009). This approach to word sense disambiguation exploits pre-defined knowledge resources (dictionaries, thesauri, ontologies, collocations, etc.) to infer the senses of words in context. The dictionary specifies the senses which are to be disambiguated through examining the overlap of sense definitions and selectional preferences/restriction. The overlap of sense definitions is a technique whereby the word overlap between the sense definitions of two or more target words is calculated (WordNet senses). Following this computation, the target word whose definition has the highest overlap is selected. Selectional preferences, or restrictions, on the other hand provide specific restrictions on the semantic category that the meaning of a given word imposes on the words with which it associates at the syntagmatic level. Selectional restrictions rule out senses that

[15] Borrowed from (Navigli 2009)

violate the constraint whereas selectional preferences select senses which better satisfy the requirements, (Asher 2014).

5.7 MT and human translation in perspective

The prospect of machine translation taking over the jobs of human translators has been a constant scenario played over and over by some critics of MT. It cannot be disputed that both MT and HT have one goal, at the very least, in common: they aim at producing the equivalent of a source language text in a target language, but it is important to note that the decision to use either HT or MT ultimately depends on the purpose of translation. Speed, cost and translation quality are some basic criteria that enter into account when deciding which approach is the best suited to one's needs.

5.7.1 Speed

In a time in which everything moves quickly, timeliness is an extremely important aspect if business growth and customer loyalty are the primary goals. In major international organizations such as the United Nations, the African Union and the European Union, meetings report as well as the various resolutions must be available in a timely manner in several languages. Faced with such a huge pressure, any human translator would have a hard time meeting both deadlines and quality requirements. In such circumstances, using a translation software can save much time and speed up the translation process. Human translators can barely translate 650 words per hour, whereas some translation software may translate more than 60.000 words per hour (Bass 1999). While it is true that most MT outputs are usually not of publishable quality, these outputs can be used as a first draft that the professional translator would have to edit before publication. The translator can also use a terminology database which avoids wasting time searching for words that best fit the context of translation. In any case, the use of translation software is very helpful and saves time.

5.7.2 The cost

A translation that needs a lot of time to be produced usually generates a high cost, and human translation for one, requires long working hours that must be remunerated accordingly. Just to name a figure, more than one billion Euro is spent, yearly, by the EU parliament for translation alone (Castle 2006). In a bid to reduce this cost, the European Union decided a few years ago to fund a project, Euromatrix, that would help substantially reduce translation costs. There is almost always a cost associated with hiring a human translator and the result might not always be satisfactory; when this happens, proof-reading and editing are additional steps that must be taken to ensure good quality, which implies additional costs. Conversely, many websites offer free online machine translation services that can help produce a gist of the text/message. In such instances, further costs are only involved when the client is not satisfied with the gist of the translation; in such cases the client would require a professional translation. In conclusion, whenever a translation device is involved, be it a CAT tool or an MT system, translation cost is likely to decrease substantially. Ideally, a fully automated high quality machine translation would solve all of the problems related to cost and quality. But to date, no such device exists.

5.7.3 Quality

Translation quality is a highly subjective concept. Firstly, because every translation can be further perfected, furthermore, translation quality can be measured through its efficiency. A translation may contain errors (grammatical or stylistic), but still fulfills its function in terms of information transfer. Secondly, a translation may be "void of errors" but fails to transfer information. Quality in translation can therefore relate to the efficiency of a given translation in a specific context. As a consequence, hereof, a translation can be graded "good" if the receiver of the message manages to draw the bulk of information s/he needs as can be illustrated in the following example:

Example 30

"I am new in the city and would like to know where to find a good restaurant."

<u>Personal Translator</u>

Ich bin neu in der Stadt und möchte wissen, wo ein gutes Restaurant zu finden ist.

By the standard of a tourist in a foreign country whose primary concern is not stylistic artistry, the above PT translation would be graded "good" if not "very good" because all basic information has been preserved during translation. Were this was something to be published, then the notion of quality would take a completely different meaning. A translation that is meant for publishing must be idiomatic and respect higher stylistic standards and while it is true that machine translation attempts to achieve this goal, only human translators are able to produce flawless translation to date.

5.7.4 Fluency and adequacy correlation

Correlation here refers to the fact that both adequacy and fluency are perceived equally because they are equally revealing about translation quality. Thus, the following two figures summarize the fluency and adequacy scores as perceived by the various evaluators. The figures show that for both Google Translate and PT, the same trend can be observed, namely, the adequacy scores are generally higher than the fluency scores. It can also be observed that the fluency scores never go beyond 0.6, while at the same time, the adequacy scores often get closer to 1 which is the highest score. To further establish a possible correlation between the two metrics, the correlation coefficient was calculated using both Pearson's and Spearman's correlation coefficients. The r value[16] on the Pearson correlation coefficient was -0,19 and the r value for the Spearman's correlation coefficient was -0,11. Both values display a negative correlation, therefore, no correlation was found between fluency and adequacy in evaluating Google Translate.

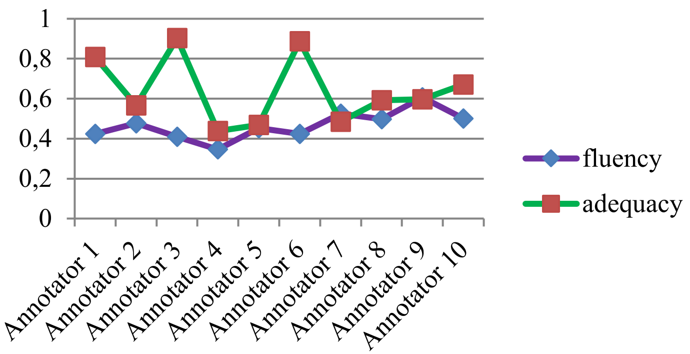

Figure 22: Fluency and adequacy correlation for Google Translate.

[16] The correlation coefficient value is expressed as *r*. the value r = 1 means a perfect positive correlation and the value r = -1 means a perfect negative correlation.

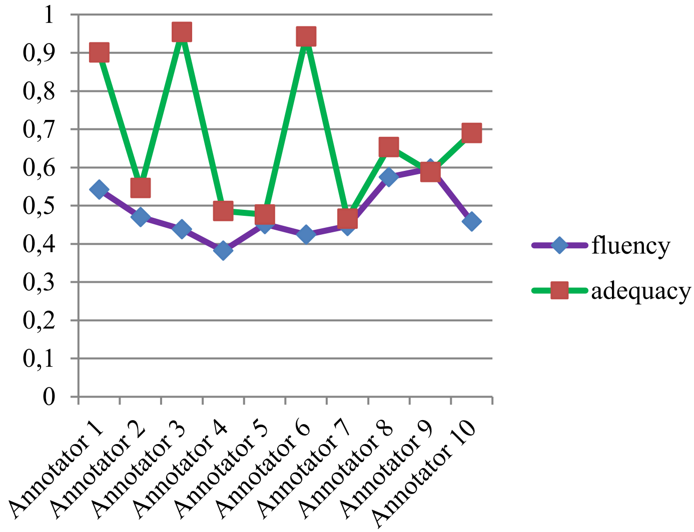

Figure 23: Fluency and adequacy correlation for Personal Translator.

The correlation coefficient was also calculated for PT (Figure 17) and the r value obtained in Spearman's scale was 0.00 while the Pearson correlation coefficient amounted to 0.05. Even if both values are positive, they are too small to be significant.

6 Conclusion

This quantitative case study has explored the influence of ambiguity on machine translation output quality. The theoretical framework proposed that ambiguity resolution has been one of the trickiest linguistic aspects of MT since research began in this area. The literature implied that ambiguity identification is instrumental in determining text quality and mistranslation of an ambiguous word may impact the understandability of a given sentence. According to the ten users' judgement of MT output quality presented in this study, both fluency and adequacy metrics are negatively influenced by the presence of ambiguous words in the source language. The underlying conclusion of this evaluation is that the presence of ambiguous words in the SL is partially accountable for poor quality MT output. The findings furthermore, suggested that there exists no absolute correlation between fluency and adequacy scores.

6.1 Summary of Contributions

This study was conducted with the aim to get an in-depth look at machine translation and review some of the major developments that have occurred in this field in recent years and additionally to study the concept of ambiguity in machine translation. Machine translation is a relatively new field, which has, however, seen quite dramatic changes over the past fifteen years, as a result of the development in statistical machine translation.

Despite these various methodological improvements witnessed in the field of machine translation however, ambiguity resolution has continuously constituted an important domain of research because of its complex nature. The starting point of this research was Bar Hillel's 1966 assertion that it would be impossible to reach fully automatic high quality machine translation. This admission of powerlessness came as a result of the impossibility to automatically translate a text containing ambiguous lexical items. Given the centrality of the issue of ambiguity in the evolution of research on machine translation, this study aimed to review the progress in this field since the famous ALPAC report. To this end, an evaluation of some current translation systems was carried out. In this evaluative study, a comparison was established between two sets of: corpora containing ambiguous sections; and corpora without ambiguous lexemes. The aim was to highlight the influence that ambiguity can have on the quality of machine translation output.

Previous studies on ambiguity focused either on the typology of ambiguities or on methodological approaches to ambiguity resolution. The literature does not abound with quantitative studies on the influence of ambiguity on the quality of machine translation output. Quite often it is put forward that ambiguity could be a problem in MT, but to the best of our knowledge, no study to date has measured this impact. The evaluative section of this thesis, therefore aimed to provide concrete data to measure this impact. To achieve this objective, the fluency and adequacy metrics used in most surveys were employed. The evaluation conducted as part of this study was essentially a human evaluation, as it is considered the most reliable, even though it is more expensive and difficult to implement.

It appears from the evaluation that, indeed, the higher the degree of ambiguity in a corpus, the poorer the translation of this corpus; conversely, a corpus that is less ambiguous usually yields better translation quality in terms of fluency and adequacy. It was also found that most machine translated corpora perform better in terms of adequacy than fluency. While fluency has to do with the grammaticality of a construction, adequacy merely measures the overall sense of an output, without taking grammaticality into consideration. As such, the machine translation systems evaluated seem to perform very well in terms of adequacy, meaning that they are suited for general translations that do not need much editing, for instance, for students who want to collect general information in a foreign language. As far as fluency is concerned, the MT systems that were evaluated have a long way to go to improve the grammaticality of the outputs. It came as no surprise that Personal Translator, a hybrid machine translation system with a strong rule-based component, performed better in terms of fluency than Google Translate, which implements the statistical approach to machine translation.

At least three lessons can be learnt from this observation: the first, that rule-based methodologies are not as obsolete as is often believed, as they succeed better at producing outputs that are somewhat closer to what is to be expected from a grammatically correct translation; whereas the second being that, statistical machine translation seems to perform very well in terms of adequacy, which can be equated to a higher degree of understandability. The SMT system, on the other hand, performed poorly in terms of fluency and required a substantial amount of editing; and lastly, the present study found absolutely no systematic correlation between fluency and adequacy which suggests that both metrics must be evaluated if a comprehensive evaluation of understandability is to be obtained. The evaluation results obtained in this study validate the hypothesis that the presence of ambiguous segments in a corpus is likely to influence its machine translation even though in exceptional cases, some ambiguous segments could get better scores than the non-ambiguous segments. An additional experiment was carried out with Google Translate to see whether integrating a customized glossary would impact the translation results and whether the sense-annotated data was taken into account during machine translation. It was found that the customized glossary was not fully integrated into the automated translation process. As a suggestion, MT systems should better integrate user's customized glossary into the automated process as this enables the user to tune the machine translation system to their specific needs.

6.2 Future research

The present study has highlighted the influence of ambiguity on machine translation output. In the review of literature, a plethora of ambiguities were identified. Studying the impact of all forms of ambiguities on MT output was, however, beyond the scope of this study. The focus in the present study has therefore been on evaluating texts containing lexical ambiguity. Future research might focus on the impact of other forms of ambiguity on the quality of automatically translated corpora, such as syntactic or semantic ambiguity. Such a study would complement the present study as it would help focus disambiguation efforts on aspects which are most pertinent.

7 Bibliography

Aikawa, Takako, Lee Schwartz, Ronit King, Mo Corston-Oliver, and Carmen Lozano. “Impact of controlled language on translation quality and post-editing in a statistical machine translation environment.” *MT Summit XI, Proceedings.* Copenhagen, Denmark, 2007. 1-7.

Aleksić, Vera, and Gregor Thurmair. “Personal Translator at WMT2011 - A rule-based MTsystem with hybrid components -.” *Proceedings of the Sixth Workshop on Statistical Machine Translation.* Edinburgh, Scotland, UK: Association for Computational Linguistics (ACL), 2011. 303-308.

ALPAC. *Language and machines: Computers in Translation and Linguistics.* A report by the Automatic Language Processing Advisory Committee, Division of Behavioral Sciences, National Academy of Sciences, National Research Council, Washington, D.C: National Academy of Sciences, National Research Council, 1966, 124.

Andersen, Øivin. “Terminology and Pragmatics.” Edited by Antero Niemkorpi. *Proceedings of the University of Vaasa Reports.* Vaasa, Finland, 2002. 136-147.

Arnold, Douglas, Lorna Balkan, Siety Meijer, Lee Humphreys, and Louisa Sadler. *Machine translation: an introductory guide.* Manchester: NCC Blackwell, 1994.

Arranz, Victoria. “Lexical Bottleneck in Machine Translation and Natural Language Processing: A Case Study.” *Transling and the Computer 19: Papers from the Aslib conference… 13 & 14 November 1997.* London: Aslib, 1997.

Asher, Nicholas. “Selectional restrictions, types and categories.” *Journal of Applied Logic* 122, no. 1 (2014): 75-87.

Austermühl, F. *Electronic Tools for Translators.* Manchester: St. Jerome Pub. Co., 2001.

Bar-Hillel, Y. *Language and information: selected essays on their theory and application.* Reading, Mass.: Addison-Wesley, 1964.

Barreiro, A., et al. “Linguistic evaluation of support verb constructions by OpenLogos and Google Translate.” *LREC 2014: Ninth International Conference on Language Resources and Evaluation.* Reykjavik, Iceland, 2014. 35-40.

Behnam, Biook, Farhad Azimi, and Peyman Hajarizadeh. "Journal of Language Teaching and Research." *Machine Translation and Skopos Theory: Post-modernist Approach to Interlingual Translation* 5, no. 3 (2014): 714-721.

Bhattacharyya, Pushpak. *Machine Translation.* CRC Press, 2015.

Carl, Michael, and Andy Way. *Recent Advances in Example-Based Machine Translation.* Edited by Michael Carl and Andy Way. Springer Science & Business Media, 2003.

Carpuat, Marine. "A semantic evaluation of machine translation lexical choice." *Proceedings of the Seventh Workshop on Syntax, Semantics and Structure in Statistical Translation, 13 June 2013.* Atlanta, Georgia, 2013. 1-10.

Carroll, John B. "An Experiment in Evaluating the Quality of Translations." *Mechanical Translation and Computational Linguistics* 9, no. 3 (1966): 55-66.

Castle, Stephen. "Cost in translation: EU spends €1bn on language services." *www.independent.co.uk.* July 15, 2006. http://www.independent.co.uk/news/world/europe/cost-in-translation-eu-spends-83641bn-on-language-services-407991.html (accessed July 26, 2015).

Ceccato, Mariano, Nadzeya, Zeni, Nicola Kiyavitskaya, Luisa Mich, and Daniel M. Berry. "Ambiguity Identification and Measurement in Natural Language Texts." Departmental Technical Report, 2004.

Chatzitheodorou, Konstantinos, and Stamatis Chatzistamatis. "COSTA MT Evaluation Tool: An Open Toolkit for Human Machine Translation Evaluation." *The Prague Bulletin of Mathematical Linguistics*, no. 100 (2013): 83–89.

Chen, Yu, Andreas Eisele, Christian Federmann, Eva Hasler, Michael Jellinghaus, and Silke Theison. "Multi-Engine Machine Translation with an Open-Source Decoder for Statistical Machine Translation." *ACL 2007: proceedings of the Second Workshop on Statistical Machine Translation.* Prague, Czech Republic, 2007. 193-196.

Cheung, Andy. "A History of Twentieth Century Translation Theory and Its Application for Bible Translation." *Journal of Translation* 9, no. 1 (2013): 1-15.

Chomsky, Noam. *Syntactic structures.* The Hague: Mouton, 1957.

Chris Callison-Burch, Cameron Fordyce, Philipp Koehn, Christof Monz, & Josh Schroeder: "(Meta-) evaluation of machine translation."

Proceedings of the Second Workshop on Statistical Machine Translation, 2007: 136–158.

Collinsdictionary. *www.collinsdictionary.com.* August 27, 2015. http://www.collinsdictionary.com/dictionary/english/bank (accessed August 27, 2015).

Costa-jussàa, Marta R., and José A.R. Fonollosa. "Latest trends in hybrid machine translation and its applications." *Computer Speech & Language* 32, no. 1 (2015): 3-10.

Crystal, David. *English as a global language.* Cambridge: Cambridge University Press, 2003.

Dale, Robert, Hermann Moisl, and Harold Somers. *Handbook of Natural Language Processing.* CRC Press, 2000.

Dandapat, Sandipan, Sara Morrissey, Andy Way, and Joseph van Genabith. "Combining EBMT, SMT, TM and IR Technologies for Quality and Scale." *EACL Joint Workshop on Exploiting Synergies between Information Retrieval and Machine Translation (ESIRMT) and Hybrid Approaches to Machine Translation (HyTra): Proceedings of the workshop* . Avignon, France; , 2012. 48-58.

Depraetere, Ilse. "Analyse contrastive de deux techniques d'évaluation de traduction automatique." Edited by Tatiana Boukreeva-Milliaressi. *De la linguistique à la traductologie.* Lille: Presses Universitaires du Septentrion, 2011. 309-324.

Diekema, Anne R. "Translation Events in Cross-Language Information Retrieval: Lexical Ambiguity, Lexical Holes, Vocabulary Mismatch, and Correct Translations." *http://works.bepress.com.* 2003. http://works.bepress.com/anne_diekema/3 (accessed 08 25, 2015).

Dietzel, Stefanie. *Example-based Machine Translation.* GRIN Verlag, 2009.

Doddington, G. "Automatic evaluation of machine translation quality using n-gram cooccurrence statistics." *Proc. ARPA Workshop on Human Language Technology.* 2002.

Doddington, George. "Automatic Evaluation of Machine Translation Quality Using N-gram Co-Occurrence Statistics." *HLT '02 Proceedings of the second international conference on Human Language Technology Research.* San Francisco, CA, USA: Morgan Kaufmann Publishers Inc, 2002. 138-145.

Dorr, Bonnie J., and Jones Douglas. "Acquisition of Semantic Lexicons: Using Word Sense Disambiguation to Improve Precision." *Breadth and Depth of Semantic Lexicons.* Norwell: Kluwer Academic Publishers, 2000. 79-98.

Dorr, Bonnie Jean. *Machine Translation: A View from the Lexicon.* MIT Press, 1993.

Dorr, Bonnie. "Parameterization of the Interlingua in Machine Translation." *Proceedings of the 14th International Conference on Computational Linguistics.* Nantes, France, 1992. 624--630.

Dugast, Loïc, Jean Senellart, and Philipp Koehn. "Can we relearn an RBMT system?." *ACL-08: HLT Third Workshop on Statistical Machine Translation, Proceedings, June 19, 2008.* The Ohio State University, Columbus, Ohio, USA : ACL WMT-08, 2008. 175-178.

Eisele, Andreas, et al. "Hybrid machine translation architectures within and beyond the EuroMatrix project." Edited by John Hutchins and Walther v. Hahn. *EAMT 2008: 12th annual conference of the European Association for Machine Translation.* Hamburg, Germany, 2008. 27-34.

Elliott, D, Hartley, A, and Atwell. "Rationale for a multilingual corpus for machine translation evaluation." *Proceedings of CL2003: International Conference on Corpus Linguistics.* 2003. 191-200.

Emele, Martin, and Dorna Michael. "Ambiguity Preserving Machine Translation using Packed Representations." *COLING-ACL.* Montréal: Morgan Kaufmann Publishers / ACL 1998, 1998. 365-371.

—. "Ambiguity Preserving Machine Translation using Packed Representations." *COLING-ACL.* Montréal: Morgan Kaufmann Publishers / ACL 1998, 1998. 365-371.

Estrella, Paula, Olivier Hamon, and Andrei Popescu-Belis. "How Much Data is Needed for Reliable MT Evaluation? Using Bootstrapping to Study Human and Automatic Metrics." *Proceedings of the Machine Translation Summit XI*, 2007: 167-174.

Faiola, Anthony, and Eliza Mackintosh. *www.washingtonpost.com* . April 17, 2013. https://www.washingtonpost.com/world/a-final-farewell-to-the-iron-lady/2013/04/17/0bd78582-a73e-11e2-9e1c-bb0fb0c2edd9_story.html (accessed April 17, 2013).

Fass, Dan C. "An account of coherence, semantic relations, metonymy, and lexical ambiguity resolution." In *Lexical Ambiguity Resolution: Perspectives from Psycholinguistics, Neuropsychology, and Artificial Intelligence*, edited by Steven L. Small, Garrison W. Cottrell, Michael K. Tanenhaus and Morgan Kaufmann, 151-178. 1988.

Feng, Minwei, Weiwei Sun, and Hermann Ney. "Semantic cohesion model for phrase-based SMT." *Proceedings of COLING 2012: Technical Papers, December 2012.* Mumbai, 2012. 867-878.

Fries, Norbert. “Ambiguität und Vagheit: Einführung und kommentierte Bibliographie.” Tübingen: Niemeyer, 1980.

Gaspari, Federico. “Embedding Free Online Machine Translation into Monolingual Websites for Multilingual Dissemination: a Case Study of Implementation.” *Proceedings of Translating and the Computer 27.* London, United Kingdom, 24-25 November 2005, 2005.

—. “Integrating On-line MT services into Monolingual Web-sites for Dissemination Purposes: an Evaluation Perspective.” *Proceedings of the Ninth EAMT Workshop “Broadening Horizons of Machine Translation and Its Applications”.* Valletta, Malta: EAMT / Foundation for International Studies (University of Malta), 2004. 62-72.

Gentzler, Edwin. *Contemporary translation theories.* Second edition. Clevedon, UK: Multilingual Matters, 2001.

Gillon, B. S. “Ambiguity, indeterminacy, deixis and vagueness.” *Semantics: a reader*, 2004: 157-187.

Goh, Chooi-Ling, Takashi Onishi, and Eiichiro Sumita. “Rule-based reordering constraints for phrase-based SMT.” Edited by Heidi Depraetere, Vincent Vandeghinste Mikel L.Forcada. *proceedings of the 15th conference of the European Association for Machine Translation, 30-31 May 2011.* Leuven, Belgium, 2011. 113-120.

Gojun, Anita, and Alexander Fraser. “Determining the placement of German verbs in English-to-German SMT.” *[EACL 2012] Proceedings of the 13th Conference of the European Chapter of the Association for Computational Linguistics, April 23-27, 2012.* Avignon, France, 2012. 726-735.

Goutte, Cyril, Marine Carpuat, and George Foster. “The impact of sentence alignment errors on phrase-based machine translation performance.” *AMTA-2012: the Tenth Biennial Conference of the Association for Machine Translation in the Americas. Proceedings.* San Diego, CA,, 2012. 10.

Haller, Johann. “Multilinguale technische Dokumentation: Sprachkontrolle für die Sicherheit.” *S. 547-556. TUEV Saarland Foundation: World Congress Safetyof Modern Technical Systems - Congress-Documentation.* Haller, Johann; Hellmund, S., 2001. 547-556.

Hardmeier, Christian, Jörg Tiedemann, and Joakim Nivre. “Latent anaphora resolution for crosslingual pronoun prediction.” *Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing.* Seattle, Washington, 2013. 380–391.

Hauenschild, Christa, and Susanne Heizmann, . *Machine Translation and Translation Theory.* Berlin: de Gruyter, 1997.

Haugereid, Petter, and Francis Bond. "Extracting Semantic Transfer Rules from Parallel Corpora with SMT Phrase Aligners." *SSST-6, Sixth Workshop on Syntax, Semantics and Structure in Statistical Translation, 12 July 2012.* Jeju, Republic of Korea, 2012. 67-75.

Hearne, Mary, and Andy Way. "Disambiguation Strategies for Data-Oriented Translation." *Proceedings of the 11th Conference of the European Association for Machine Translation, June 19-20, 2006.* Oslo, Norway, 2006. 59-68.

Holmes, James S. *The name and nature of translation studies.* Edited by Lawrence Venuti. 1988/2004.

Hu, Xiaoguang, Haifeng Wang, and Hua Wu. "Using RBMT Systems to Produce Bilingual Corpus for SMT." *Proceedings of the 2007 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning.* Prague, Czech Republic: EMNLP-CoNLL-2007, 2007. 287-295.

Huang, Fei, and Cezar Pendus. "Generalized reordering rules for improved SMT." *ACL-2013: Proceedings of the 51st Meeting of the Association for Computational Linguistics, Short papers, August 4-9 2013.* Sofia, Bulgaria, 2013. 387-392.

Hutchins, W. John. "Machine translation over fifty years." *Histoire, Epistemologie, Langage*, 2001: 7-31.

—. *Machine translation: past, present and future.* Chichester: Ellis Horwood, 1986.

—. "Machine translation: problems and issues." *Proceedings of machine translation conference.* London: Academic Press, 2007.

Hutchins, W. John., and L. Harold Somers. *An introduction to machine translation.* Academic Press: London, 1992.

Ide, Nancy, and Jean Véronis. "Introduction to the Special Issue on Word Sense Disambiguation: The State of the Art." *Computational Linguistics* 24, no. 1 (1998): 1-40.

Jiang, Jie, Jinhua Du, and Andy Way. "SMT with syntactic reordering patterns learned from lattice scoring." *AMTA 2010: the Ninth conference of the Association for Machine Translation in the Americas, October 31 – November 4, 2010.* Denver, Colorado, 2010. 1-10.

Kennedy, Christopher. "Vagueness and grammar: The semantics of relative and absolute gradable adjectives." *Linguistics and philosophy* 30, no. 1 (2007): 1-45.

Kirchhoff, Katrin, Anne M Turner, Amittai Axelrod, and Francisco Saavedra. “Application of statistical machine translation to public health information: a feasibility study.” *Kirchhoff K, Turner AM, Axelrod A, Saavedra F. Application of statistical machine translation to public health information: a feasibility study. Journal of the Journal of the American Medical Informatics Association* , 2001: 473-478.

Kit, Chunyu, Haihua Pan, and Jonathan J. Webster. “Example-Based Machine Translation: A New Paradigm.” *Translation and Information Technology* (Chinese U of HK Press), 2002: 57–78.

Koehn, P., and K Knight. “Empirical Methods for Compound Splitting.” *Proc. 10th Conf. of the European Chapter of the Association for Computational Linguistics.* Budapest, Hungary: EACL, 2003. 347–354.

Koehn, Philipp. *Statistical Machine Translation.* Cambridge University Press, 2010.

Koletnik Korošec, Melita. “The internet, Google Translate and Google Translator Toolkit – nuisance or necessity in translator training? .” *Tralogy.* Paris, 2011. 1-17.

Kommission, Die Europäische. “Die Europäische Kommission.” *www.chemie.de.* 1997. http://www.chemie.de/behoerden/70595/europaeische-kommission.html (accessed march 15, 2014).

Krahmer, Emiel, and Mariet Theune. *Empirical Methods in Natural Language Generation Data-oriented Methods and Empirical Evaluation.* Springer, 2010.

Krenz, Michael, and Markus Ramlow. *Maschinelle Übersetzung und XML im Übersetzungsprozess: Prozesse der Translation und Lokalisierung im Wandel : zwei Beiträge.* Edited by Uta Seewald-Heeg. Frank & Timme GmbH, 2008.

Krings, H.P. *Was in den Köpfen von Übersetzern vorgeht?* Tübingen: Gunter Narr, 1986.

Kulesza, Alex, and Stuart Shieber. “A learning approach to improving sentence-level MT evaluation.” *Proceedings of the 10th International Conference on Theoretical and Methodological Issues in Machine Translation (TMI).* Baltimore, 2004.

Labaka, Gorka, Cristina Espana-Bonet, Lluıs Marquez, and Kepa Sarasola. “A Hybrid Machine Translation Architecture Guided by Syntax.” *Machine Translation*, October 2014: 91-125.

Lefever, Els, and Véronique Hoste. “An evaluation and possible improvement path for current SMT behavior on ambiguous nouns.” *Proceedings of SSST-5, Fifth Workshop on Syntax, Semantics and Structure in Statistical Translation, June 2011.* Portland, Oregon, USA: ACL HLT 2011, 2011. 52-60.

Lehrberger, John, and Laurent Bourbeau. *Machine translation: linguistic characteristics of MT systems and general methodology of evaluation.* Amsterdam/Philadelphia: John Benjamins, 1988.

Li, Jin-Ji, Jungi Kim, and Jong-Hyeok Lee. “Chinese syntactic reordering through contrastive analysis of predicate-predicate patterns in Chinese-to-Korean SMT.” *AMTA 2010: the Ninth conference of the Association for Machine Translation in the Americas, October 31 – November 4, 2010.* Denver, Colorado, 2010. 10.

Ljudskanov, A. *Mensch und Maschine als Übersetzer.* Translated by Aus dem Bulgar. übers. von Gert Jäger und Hilmar Walter. München: Hueber, 1972.

LT2013. “Status and Potential of the European Language Technology Markets.” *Language Technologies*, 2013: 15.

Maienborn, Claudia, Klaus von Heusinger, and Paul Portner. *Semantics: Handbooks of Linguistics and Communication Science.* Edited by Walter de Gruyter. Vol. 3. 2013.

Manning, Christopher, and Hinrich Schuetze. *Foundations of Statistical Natural Language Processing.* MIT Press, 1999.

Martin, Bruce Thomas. “A Typology of Ambiguity as it Relates to Natural Language Processing.” ProQuest, 2007.

Miguel, José P., David Mauricio, and Glen Rodríguez. “A REVIEW OF SOFTWARE QUALITY MODELS FOR THE EVALUATION OF SOFTWARE PRODUCTS.” *International Journal of Software Engineering & Applications (IJSEA)* 5, no. 6 (November 2014): 31-54.

Mitkov, Ruslan, Sung-Kwon Choi, and Randall Sharp. “Anaphora resolution in machine translation.” *TMI-95: proceedings of the Sixth International Conference on Theoretical and Methodological Issues in Machine Translation, July 5-7, 1995, .* Katholieke Universiteit, Leuven, Belgium, 1995. 87-95.

Munday, J. *Introducing Translation Studies. Theories and Applications.* New York / London: Routledge, 2008.

Navigli, Roberto. “Word sense disambiguation: A survey.” *ACM Computing Surveys* 41, no. 2 (2009): 1-69.

Newmark, P. *Approaches to Translation.* Oxford and New York: Pergamon, 1981.

Nida, Eugene A., and Charles R. Taber. *The theory and practice of translation.* Leiden: Brill, 1969.

Niessen, S., and H. Ney. “Improving SMT quality with morpho-syntactic analysis.” *Proc. 18th Int. Conf. on Computational Linguistics (COLING).* Saarbrücken, Germany, 2000. 1081–1085.

Nirenburg, Sergei. *Progress in Machine Translation.* IOS Press, 1993.

Nirenburg, Sergei, and H. L. Somers. *Readings in Machine Translation.* MIT Press, 2003.

Nord, Christiane. *Übersetzen lernen-leicht gemacht. Ein Kurs zur Einführung in das professionelle Übersetzen aus dem Spanischen ins Deutsche.* Vols. 1,2. Heidelberg: Universitätsdruckerei, 1990.

Olive, Joseph, Caitlin Christianson, and John McCary. *Handbook of Natural Language Processing and Machine Translation: DARPA Global Autonomous Language Exploitation.* Springer Science & Business Media, 2011.

Papineni, Kishore, Salim Roukos, Todd Ward, and Wei-Jing Zhu. “BLEU: a Method for Automatic Evaluation of Machine Translation.” *Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics (ACL).* Philadelphia, 2002. 311-318.

Petrzelka, Jiri. *Statistical Machine Translation: How to maximize the quality of translation.* LAP Lambert Academic Publishing, 2011.

Pinkal, Manfred. *Logik und Lexikon: die Semantik des Unbestimmten.* Berlin: de Gruyter, 1985.

Popović, Maja, Daniel Stein, and Hermann Ney. “Statistical Machine Translation of German Compound Words.” *Advances in Natural Language Processing*, 2006: 616-624.

Porro, Victoria, Johanna Gerlach, Pierrette Bouillon, and Violeta Seretan. “Rule-based automatic post-processing of SMT output to reduce human post-editing effort.” *Translating and the Computer 36: proceedings.* Asling: International Society for Advancement in Language Technology, 2014. 66-76.

Prescher, Detlef, Stefan Riezler, and Mats Rooth. “Using a Probabilistic Class-Based Lexicon for Lexical Ambiguity Resolution.” *COLING 2000, 18th International Conference on Computational Linguistics, Proceedings of the Conference, 2 Volumes, July 31 - August 4, 2000.* Universität des Saarlandes, Saarbrücken, Germany. : Morgan Kaufmann 2000, 2000. 649-655.

Quah, Chiew Kin. *Translation and technology.* Basingstoke (UK): Palgrave Macmillan, 2006.

Reinke, Uwe. "State of the Art in Translation Memory Technology." *Translation: Computation, Corpora, Cognition,* 3, no. 1 (2013): 27-48.

Reiss, Katharina, and Hans J. Vermeer. *Grundlegung einer allgemeinen translationstheorie.* Tübingen: M Niemeyer, 1984.

Rösener, Christoph. "Computational Linguistics in the Translator's Workflow—Combining Authoring Tools and Translation Memory." *Proceedings of the NAACL HLT 2010 Workshop on Computational Linguistics and Writing.* 2010. 1-6.

Sager, C. Juan. *Language engineering and translation: Consequences of automation.* Amsterdam: John Benjamins Publishing Company., 1994.

Sangodkar, Amit, and Om P. Damani. "Re-ordering source sentences for SMT." *LREC 2012: Eighth international conference on Language Resources and Evaluation, 21-27 May 2012.* Istanbul, Turkey, 2012. 2164-2171.

Sato, S., and M. Nagao. "Toward memory-based translation." *Proceedings of COLING.* Helsinki, Finland, 1990. 247-252.

Schäfer, Falko. *Die maschinelle Übersetzung von Wirtschaftsfachtexten: Eine Evaluierung anhand des MÜ-Systems der EU-Kommission Systran im Sprachenpaar Französisch-Deutsch.* Frankfurt: Peter Lang, 2002.

Schwanke, Martina. *Maschinelle Übersetzung: ein Überblick über Theorie und Praxis.* Berlin, Heidelberg: Springer, 1991.

Sharp, Randy. *CAT2 Reference Manual Version 3.6.* 1994.

Shilon, Reshef, Hanna Fadida, and Shuly Wintner. " Incorporating linguistic knowledge in statistical machine translation: translating prepositions." *EACL Workshop on Innovative Hybrid Approaches to the Processing of Textual Data, 23 April 2012.* Avignon, France, 2012. 106-114.

Sloculn, Jonathan. "A survey of machine translation: its history, current status, and future prospects." *Computational Linguistics* , 1985.

Somers, Harold. "Review Article: Example-based Machine Translation." *Machine Translation, 14(2), 113--*, 1999: 113-157.

Stein, Daniel. "Maschinelle Übersetzung -- ein Überblick." *Journal for Language Technology and Computational Linguistics* 24, no. 3 (2009): 5-18.

Streiter, Oliver, Michael Carl, and Johann Haller. "Hybrid Approaches to Machine." *IAI Working Paper No. 36.* Edited by Oliver Streiter, Michael Carl and Johann Haller. 2000.

Sun, Yanli. "An Investigation into Automatic Translation of Prepositions in IT Technical Documentation from English to Chinese." 2010. http://doras.dcu.ie/16053/1/Thesis-YLSun-Final.pdf.

Täuschel, Alexander. *Linguistic Aspects in Machine Translation.* GRIN Verlag, 2008.

Thurmair, Gregor. "Improving machine translation quality." *MT Summit X, Phuket, Thailand, September 13-15, 2005, Conference Proceedings: the tenth Machine Translation Summit.* Phuket, Thailand, 2005.

Tomita, Masaru, and Jaime G. Carbonell. "Another Stride Towards Knowledge Based Machine Translation." *In Proceedings of 11th International Conference on Computational Linguistics (COLING-86).* Bonn: Tomita, M. and J.G. Carbonell, 1986. 633-639.

Trujillo, Arturo. *Translation engines: techniques for machine translation.* Edited by Arturo Trujillo. London: Springer, 1999.

van der Eijk, P., M. de Koning, and G. van der Steen. "Controlled Language Correction and Translation." *In Proceedings of the 1st International Workshop on Controlled Language Applications.* Leuven: CLAW, 1996. 64-73.

Van Slype, G., F. J. Guinet, F. Seitz, and E. Benejam. "Better translation for better communication." *Journal of the American Society for Information Science*, 1984: 373.

Vinay, J.P., and Darbelnet. *Stylistique comparée du francais et de l'anglais: Méthode de Traduction.* Paris: Didier, 1995.

Wheeler, Peter J. "SYSTRAN." Edited by Margaret King. *Machine translation today: the state of the art.* Lugano, Switzerland: Edinburgh University Press, 1984. 192-208.

White, John S., Theresa O'Connell, and Francis O'Mara. "The ARPA MT Evaluation Methodologies: Evolution, Lessons, and Future Approaches." *Proceedings of AMTA Conference*, 1994: 193-205.

Wilks, Yorick. *Machine Translation: Its Scope and Limits.* Springer Verlag US, 2009.

Xiong, Deyi, and Min Zhang. *Linguistically Motivated Statistical Machine Translation: Models and Algorithms.* Springer, 2015.

Zhang, You-Peng Paul. *A Practical Design for an Interlingua Machine Translation System Based on a Case Grammar Model.* Georgetown University, 1993.

Zuo, Li. “Machine translation from English to Chinese: a study of Google’s performance with the UN documents.” *AMTA 2010: the Ninth conference of the Association for Machine Translation in the Americas,.* Denver, Colorado, 2010.

APPENDICES

Full evaluation results:

Adequacy GT

	Non ambiguous	**Ambiguous**
Annotator 1	0,482	0,478
Annotator 2	0,506	0,448
Annotator 3	0,404	0,416
Annotator 4	0,382	0,31
Annotator 5	0,472	0,432
Annotator 6	0,488	0,36
Annotator 7	0,514	0,426
Annotator 8	0,532	0,464
Annotator 9	0,63	0,583
Annotator 10	0,51	0,42

Adequacy PT

	Non ambiguous	**Ambiguous**
Annotator 1	0,564	0,52
Annotator 2	0,524	0,416
Annotator 3	0,502	0,374
Annotator 4	0,408	0,356
Annotator 5	0,528	0,458
Annotator 6	0,434	0,332
Annotator 7	0,474	0,42
Annotator 8	0,616	0,534
Annotator 9	0,68	0,532
Annotator 10	0,6	0,45

Adequacy Google Translate and Personal Translator.

Fluency GT

	Non ambiguous	**Ambiguous**
Annotator 1	0,822	0,79
Annotator 2	0,614	0,518
Annotator 3	0,896	0,91
Annotator 4	0,49	0,386
Annotator 5	0,47	0,446
Annotator 6	0,896	0,892
Annotator 7	0,514	0,456
Annotator 8	0,664	0,518
Annotator 9	0,596	0,598
Annotator 10	0,69	0,58

Fluency PT

	Non ambiguous	**Ambiguous**
Annotator 1	0,912	0,894
Annotator 2	0,608	0,484
Annotator 3	0,946	0,962
Annotator 4	0,582	0,39
Annotator 5	0,528	0,426
Annotator 6	0,992	0,894
Annotator 7	0,534	0,398
Annotator 8	0,702	0,604
Annotator 9	0,668	0,508
Annotator 10	0,85	0,72

Fluency Google Translate and Personal Translator.

Evaluation Corpora

<u>Text 1: „Bremse" as brake</u>

Bremsen, die intakt sind, gelten als überlebenswichtig.
Damit Sie Ihre Bremsbeläge wechseln, bevor es zu spät ist, werden regelmäßige Inspektionen empfohlen.
Der Wechsel ist jedoch nicht frei von Kosten. An der Sicherheit sollte allerdings grundsätzlich nicht gespart werden.
Bremsen in Schuss halten
In der Regel werden bei Autos Trommelbremsen an den Hinterachsen verarbeitet und Scheibenbremsen im Vorderachsenbereich.
Trotz moderner Technologie und diverser Assistenzsysteme entsteht an den Bremsen Reibung.
In der Folge verschleißen die Bremsbeläge.
Die Sicherheit kann dann kaum noch gewährleistet und die Beläge müssen ausgetauscht werden.
Wenn Sie die Bremsbeläge wechseln wollen, geschieht dies in einer professionellen Kfz-Werkstatt förmlich im Handumdrehen.
Um eine gleichmäßige Bremswirkung zu erreichen, sollten Sie nicht nur an einem Rad die Bremsen erneuern.
Der Austausch der Trommelbremsen an den Hinterreifen kostet in der Regel 150 bis 300 Euro.
An der Vorderachse können Sie 100 bis 200 Euro einplanen.
Kosten: In Sicherheit investieren
Höher sind die Kosten, wenn Sie die Bremsscheibe austauschen müssen.
Denn auch an diesen Aggregaten kann ein gefährlicher Verschleiß auftreten.
Die Schäden treten meistens dann auf, wenn Sie zuvor mit defekten Bremsbelägen gefahren sind.
Die Folge: In den Bremsscheiben bilden sich tiefe Risse, welche die Bremswirkung verringern.
Zwar gibt es die Möglichkeit, die Scheiben zu reparieren, Kfz-Experten raten jedoch zu einem Austausch.
Für die Bremsscheiben fallen Kosten in Höhe von rund 500 Euro an.

<u>Text 2: „Bremse” as insect</u>

Bremsenstiche - Stiche von Bremsen

Bremsenstiche sind besonders unangenehm.
Sie lassen sich jedoch mit einigen Hausmitteln wirksam behandeln.
Die Bremse gehört zur Familie der Fliegen.
Da sie gleichzeitig ein blutsaugendes Insekt ist, sticht sie sowohl Tiere als auch Menschen.
In unseren Breitengraden sind die lästigen Parasiten vor allem zwischen April und August aktiv.
Je nach Region trägt die Bremse verschiedene Bezeichnungen.
So nennt man sie in Westdeutschland Blinder Kuckuck, in Norddeutschland Blinde Fliege oder Dase, und in Süddeutschland Brämer.
Insgesamt gibt es etwa 4000 verschiedene Bremsenarten auf der Welt.
Bei den meisten davon sind die Weibchen für das Blutsaugen verantwortlich.
Die Männchen begnügen sich dagegen mit Nektar von Blüten.
Bremsen erreichen eine Größe von etwa 3,5 Zentimetern und besitzen auffallend gefärbte Augen.
Ihre Stiche sind deutlich unangenehmer als die von Mücken.
Das liegt daran, dass die Bremse mit sägeförmigen Mundwerkzeugen ausgestattet ist, was beim Blutsaugen zu schmerzhaften Stichen führt.
Als besonders unangenehm für Menschen und Weidetiere gelten drei Arten.
Dies sind die Tabanus, die Regenbremsen (Haematopota), die man auch Blinde Fliegen nennt, sowie die Chrysops.
Mithilfe ihrer großen Komplexaugen sind die Bremsen in der Lage, Bewegungen von Menschen und Tieren zu erkennen und diese zu verfolgen.
Besonders häufig trifft man die hartnäckigen Parasiten in der Nähe von Pferdegehöften oder Kuhweiden an.
Angelockt werden die Bremsen von Schweiß.
Am liebsten fallen sie über Weidetiere her.
Menschen sowie Haustiere wie Hunde oder Katzen werden dagegen eher selten von den Stechfliegen attackiert.
Besonders gefährdet sind dann Körperstellen, bei denen nur eine schwache Behaarung besteht, wie das Gesicht oder die Ohren.
Symptome
Im Gegensatz zum Mückenstich wird ein Bremsenstich umgehend bemerkt.
Neben den unangenehmen Schmerzen treten dabei auch Juckreiz und Nachblutungen auf.
Außerdem kommt es meist zur Bildung einer Quaddel.

Selbst das Tragen von langer Kleidung bietet nicht immer ausreichend Schutz vor Bremsenstichen, da die Parasiten imstande sind, durch die Kleidungsstücke hindurch zu stechen.
Einige Arten, die vor allem in Afrika vorkommen, können auch gefährliche Krankheiten wie Lyme-Borelliose, Milzbrand, Tularämie oder die Weilsche Krankheit übertragen.
Hierzulande sind Bremsenstiche normalerweise nicht gefährlich und kein Grund zur Panik.
Manchmal kann sich der Stich jedoch entzünden und stark anschwellen.
Besonders unangenehm sind die Stiche für Allergiker.
So kann ein Bremsenstich bei ihnen dazu führen, dass die betroffene Stelle so stark anschwillt, dass sie regelrecht deformiert aussieht.
Mitunter kommt es auch zu Atemnot oder sogar zu einem Schock.
In diesem Fall muss sofort ein Arzt verständigt werden.

Text 3: „Fliege" as insect
LANDKREIS MÜNCHEN
Fliege lässt Motorradfahrer stürzen
Weil er eine Fliege ins Auge bekommen hatte, verunglückte ein Motorradfahrer im Landkreis München.
Kleine Fliege, großer Schaden: Ein 55-jähriger Motorradfahrer ist am Mittwochmittag auf der Kreisstraße zwischen Dingharting und Beigarten gestürzt, weil ihm eine Fliege ins Auge geflogen war.
Als er das Insekt bemerkte, bremste er das Motorrad stark und fiel auf die Straße.
Der Motorradfahrer musste in die Klinik
Dabei zog er sich eine Fraktur des Wadenbeines zu.
Ein Rettungshubschrauber brachte den 55-Jährigen zur stationären Behandlung in eine Klinik.
Der Schaden an der Honda beläuft sich auf ca. 1.500 Euro. AZ

Text 4: „Fliege" as bow-tie
Die Fliege
Fliegen sind in denselben Farben und Mustern wie die klassische Krawatte erhältlich.
Eine Fliege ist meistens schwarz und wird zu einem Abendanzug und einem weißen Hemd mit Frackkragen getragen.
Sie kann auch etwas zwangloser zu einem Anzug und einem Hemd mit breitem Kragen getragen werden.

In 8 Schritten die Fliege binden:
Schritt 1: Legen Sie die Fliege so an, dass ein Ende kürzer ist als das andere.
Schritt 2: Legen Sie das längere Ende über das kürzere.
Schritt 3: Ziehen Sie das längere Ende unter der Fliege nach oben.
Schritt 4 und 5: Bilden Sie die beiden Flügel der Fliege, indem Sie das kürzere Ende horizontal falten.
Schritt 6: Klappen Sie das längere Ende über den soeben geformten Knoten.
Schritt 7: Dann verstecken Sie das längere Ende unter dem soeben gefalteten Teil.
Schritt 8: Richten Sie die Fliege aus, indem Sie behutsam an beiden Flügeln ziehen.
Wenn Sie fertig sind, sollten die beiden Flügel theoretisch auf einer Linie mit ihren Augen sein.

Text 5: „Schlange" as reptile

Alarm am Bahnhof

Schlange an Bord stoppt Intercity

Wegen einer Schlange an Bord musste ein Intercity seine Fahrt stoppen.
Der Zug war gerade in den Bahnhof eingefahren, als ein Passagier eine Schlange entdeckte.
Nach einer Suchaktion konnte konnten Beamte die Schlange im Intercity einfangen.
Der Intercity war am Dienstagmorgen um 7.30 Uhr planmäßig auf dem Weg nach Basel (Schweiz) unterwegs, als am Bahnhof Bern alle Reisenden die Wagen verlassen mussten.
© AFP Wegen der Schlange musste der Zug anhalten.
Grund: Passagiere hatten laut Kantonspolizei eine Schlange im Zug gemeldet.
Nachdem der Intercity komplett evakuiert worden war, starteten Beamten eine Suchaktion nach dem Tier.
Und fanden eine 50 Zentimeter lange Natter im Lüftungsschacht eines Wagens.
Allerdings handelte es sich laut Polizeibericht um ein absolut ungiftiges Exemplar.
Für Passagiere und Zugpersonal bestand demnach keine Gefahr.
Wie die Schlange jedoch in den Intercity gelangen konnte, ist noch unklar.
Es ist wohl sehr unwahrscheinlich, dass die Natter allein in den Zug gekrochen war.
Rund 450 Bahnreisende waren von dem Vorfall betroffen.
Mit einer Verspätung von 30 Minuten konnten sie ihre Zugfahrt fortsetzen.

Text 6: „Schlange" as queue

Warteschlangen-Theorie: Wie wir alle schneller shoppen könnten

Jeder hasst sie, keiner kommt an ihnen vorbei - kann man Warteschlangen nicht irgendwie erträglicher machen?

Klar, sagen Mathematiker: Sie haben das Stau- und Anstehphänomen gründlich erforscht und verblüffende Lösungen gefunden.

Nur Supermärkte wollen nicht auf sie hören.

Endlich Feierabend! Jetzt noch einen Drink an der Bar, und die Party kann beginnen.

Nur gibt es da zwei Theken: An der einen stehen zehn Leute brav in einer Reihe - an der anderen drängeln sich zehn Leute in einer Traube.

Wohin gehen, um möglichst schnell an den Drink zu kommen?

Als Partygänger ist man ziemlich schnell mittendrin in der Mathematik, genauer: in der Warteschlangentheorie.

Wissenschaftler beschäftigen sich seit fast hundert Jahren damit - sie suchen die optimale Anstehstrategie.

Refael Hassin, einer von ihnen, lehrt an der Universität von Tel Aviv.

Er rät Barbesuchern, die es eilig haben, sich in die Menschentraube zu mischen.

"Das ist rational", sagt Hassin SPIEGEL ONLINE.

Denn in der Traube habe man gute Chancen, nicht erst als elfter bedient zu werden.

Man müsse dafür nur ein bisschen drängeln.

In der braven Schlange sei langes Warten dagegen garantiert - man werde auf jeden Fall erst als elfter bedient.

"Die Situation ist sehr speziell", das gibt der israelische Mathematiker zu.

Sie zeige aber, dass Leute strategisch denken: "Sie überlegen sich sehr genau, was sie tun." Gerade beim Warten.

Hassin untersucht das Problem spieltheoretisch - doch das ist nur einer von vielen Ansätzen.

Bei komplexeren Problemen wie der Logistik von Karosserieteilen in einer Autofabrik oder den Starts und Landungen auf einem Flughafen greifen Mathematiker zu ganz anderen Werkzeugen.

Um die bestmögliche Lösung für ein Warteschlangen-Problem zu finden, jonglieren sie in der sogenannten Queuing Theory mit Größen wie Ankunftsströmen, Bedienraten, der Größe des Warteraums und der Anzahl der Bediener.

Wenn nur der Zufall nicht wäre

Erschwert wird die Suche vor allem durch den Faktor Zufall.

Er ist es in der Regel, der Warteschlangen überhaupt produziert.

Beispiel Flugverkehr: Ein Flugzeug hat einen Defekt und verspätet sich, eine andere Maschine muss auf Anschlussreisende warten - schon nimmt das Chaos seinen Lauf.
Beispiel Supermarkt: Kunden kommen nicht in gleichmäßigen Abständen zur Kasse; es gibt unter anderem Stoßzeiten am Abend.
Außerdem kann es an jeder Schlange unvorhersehbare Verzögerungen geben, wenn die ec-Karte Schwierigkeiten macht oder eine Ware nicht ausgezeichnet ist.
Für Alexander Herzog, Mathematiker an der Technischen Universität Clausthal, sind Supermarktschlangen gleich in doppelter Hinsicht ein Ärgernis.
Zum einen wird er immer wieder gefragt, wie man sich denn nun am intelligentesten anstellt.
Seine wenig befriedigende Antwort: Bei zwei längeren Schlangen ist es praktisch egal, für welche man sich entscheidet, denn “Unregelmäßigkeiten im Bedienprozess sind viel wichtiger als geringe Längendifferenzen”.
Auch in der kürzeren Schlange werde man nur in gut 50 Prozent der Fälle wirklich schneller bedient.
Zum anderen ärgert den Mathematiker, dass es durchaus eine faire Lösung gäbe - Supermärkte sie aber kaum umsetzen. Ihr Name: amerikanische Warteschlange.
Sie ist das Ideal der Experten.
Bei ihr stellen sich Kunden nicht an mehreren einzelnen Schlangen an, sondern an einer einzigen großen - und werden dann von dort an die nächste freiwerdende Kasse verteilt (siehe Fotostrecke).
Das System ist auf Flughäfen, Bahnhöfen und in Postfilialen inzwischen auch in Deutschland üblich.
Es führt zu einer gerechteren Verteilung der Wartezeit über die Kunden.
Auch wenn die amerikanische Schlange länger aussieht und deshalb manchen abschreckt: Sie wird in der Regel schnell abgearbeitet.
Mit diesem System kann es einfach nicht passieren, dass sich ein Supermarktmitarbeiter an der einen Kasse langweilt, während an der anderen drei Leute darauf warten, dass ein Kunde ganz vorn ein 20-Cent-Stück aus seinem Portemonnaie gefingert hat.
“Es gibt mehr Bediengerechtigkeit, es wird weniger Arbeitszeit verschwendet”, sagt Herzog SPIEGEL ONLINE.

<u>Text 7: „Bank" as financial institution</u>

Bulgaria's central bank has said there has been a systematic attempt to destabilise the country through attacks on the banking system.

It said it would use all powers at its disposal to protect citizens' savings.

Shares in Bulgarian banks fell sharply for the second day in a row.

These is speculation that a run on deposits at the country's fourth-biggest bank, Corporate Commercial Bank, could spread to others.

The central bank took control of Corporate Commercial Bank last week and said its problems were isolated.

Economists and Fitch Ratings agency have also played down the risk of contagion, while foreign banks with subsidiaries in Bulgaria insist their operations are safe.

But comments by a deputy from the country's ruling party on Thursday that another bank may suffer a similar fate further hit confidence and left investors rushing to ditch Bulgarian bank stocks.

'Ill-intentioned rumours'

"In recent days there has been an attempt to destabilise the state through an organised attack against Bulgarian banks without any reason," the central bank said in a statement.

The central bank urged all state institutions to work together to protect financial stability and take legal action against those spreading "untrue and ill-intentioned rumours" about the health of Bulgaria's banks.

Shares in First Investment Bank plunged 23% on Friday. Other bank shares also declined.

"The whole banking sector is being sold off due to the problems around Corporate Bank. Investors are worried the problems can spread," said Boyan Gatsev, a trader with Varchev Finance.

First Investment Bank said on Friday it would close its branches at 12:00 GMT and remain closed until Monday after depositors withdrew 800m lev (£328m) of funds in a matter of hours.

Separately, Bulgaria set 5 October as the date for a snap parliamentary election.

Prime Minister Plamen Oresharski's minority government agreed to resign earlier this month after the biggest party in his coalition, the Socialists, performed badly in May's European elections.

<u>Text 8: "bank" as riverside</u>

Floodplains and levees are found in the lower course of a river.

After a heavy downpour, the volume of flow in a river may increase drastically and the river may no longer be able to hold this sudden increase in volume. The water overflows its banks and a flood occurs.

As a river overflows its banks, the speed of flow is reduced and it begins to deposit its load especially when the flood starts to subside.

With its energy reduced, the river deposits the heavier and coarser materials first usually on its immediate banks while the finer and lighter materials e.g. clay and silt are carried further away from the banks before they are deposited.

Over a series of floods, sediments are deposited layer upon layer forming a flood— Floodplain of River Wyre, England.

The accumulation of coarser materials on the banks of the river helps to raise the banks higher than the flood plain forming natural embankments called levees.

<u>Text 9: "pen" as animal enclosure</u>

Muddy pens cause decrease in cattle gains

Livestock producers may not be able to eliminate all the stress placed on herds by Mother Nature, but if they want to maximize animal performance they should make management decisions to minimize animal exposure to mud and provide protection from adverse weather conditions.

Chris Reinhardt, Extension feedlot specialist for Kansas State University, understands that part of raising cattle is dealing with the weather and encourages livestock producers to take precautions that reduce stress from muddy pen conditions.

"Rain, snow, ice and extreme temperatures are a part of life in Kansas," Reinhardt said in a news release.

"However, each of these factors can steal a measure of the animal's performance as that animal moves outside of its comfort zone, called the thermal neutral zone."

Thermal neutral zone for healthy cattle is 23 degrees to 77 degrees Fahrenheit.

When the temperature outside falls below or rises above the animal's comfort zone, the body needs to produce more energy to stay cool or keep warm.

Feedlots and winter-feeding sites can quickly become muddy after receiving moisture and animals are active.

If cattle are too tightly confined and feeding grounds are not sufficiently spread out, even calving pastures can become riskily muddy.

Reason for concern

Reinhardt explained that producers should be concerned with the effects of mud in their pens because of four main reasons.
Slogging through a muddy pen increases the amount of energy cattle expend, thus reducing the amount of energy left for gain.
Mud on the hide reduces the insulation effects of the hair coat, increasing cold stress, reducing energy left for gain.
Muddy lots in a feed yard make lying down to rest uncomfortable, resulting in more time spent standing, increasing energy expenditure, reducing energy left for gain.
"Under stress-free conditions, only about half of animals' normal daily energy intake goes toward gain," Reinhardt said.
"All these increases in energy expenditures dramatically cut into what is left over for gain."
The National Research Council reported that mud 4 to 8 inches deep can reduce feed intake of animals by five to 15 percent.
When the temperature drops between 21 and 39 degrees Fahrenheit, mud that is dewclaw deep has the potential for a 7 percent loss of gain and the percentage doubles when the mud reaches shin deep.
Prepare for mud
Reinhardt encourages producers to prepare for muddy conditions, even though they won't eliminate the costs proper planning can reduce them. He suggests the following:
Build and repair mounds within the pen. Cattle should have about 25 square ft. of mound space per animal in the pen.
Increase pen space per animal. Whereas 125 square ft. of pen space is sufficient during dry summer conditions, 350 square ft. may be not nearly sufficient during wet conditions. Adapt as conditions dictate.
Smooth pen surfaces whenever the weather allows. The longer muddy conditions persist, the worse the pen conditions become and cattle will have an even greater difficulty moving throughout the pen.

Text 10: "pen" as a writing tool

This Pen Can Draw Every Single Color In The World

Have you ever felt limited by the colors contained inside a box of Crayola?
Imagine, instead of being forced to resort to "Forest Green" for the grass in your next masterpiece, you could take Photoshop's "eyedropper" tool to extract the color from a single, blade of grass and turn that color into ink.
Scribble is a new device that lets you do just that.

The pen matches hues from the world around you and transfers them onto paper or a mobile device.
For the latter, the tool works in conjunction with a stylus and a mobile app to sync the colors that attract you onto your phone or tablet. Pretty cool.
The pen is armed with a 16-bit RGB color sensor that stores the colors you tell it to.
Hold the device up to your friend's gorgeous blonde hair, a vibrant flower or the pizza crust on your plate and Scribble will analyze the color and reproduce it with ink from its refillable cartridges.
Say, for example, you were enticed by the bright, pungent orange sitting on your countertop.
You'd start by simply holding your Scribble pen up to the fruit.
Then, after the pen analyzed the specific orange of this particular orange, you could take the tint to paper.
Both the ink pen and the stylus are a little more than six inches, rely on bluetooth wireless technology and have a rechargeable battery.
Until now, the closest you've even gotten to this magical resource of color concoction was probably through something similar to Bic's assorted ball point.
Scribble, of course, offers more options than Bic's royal blue for when you want to draw the sky. The only limitation here, it seems, is your imagination.

Ambiguity has often been described as a "key bottleneck for progress in Machine Translation" (Dale, Moisl and Somers 2000). While several studies have been centered on the description of ambiguity as a linguistic and philosophical phenomenon, there has been no study, to the best of our knowledge, that has measured its impact on machine translation output. This quantitative case study highlights the influence of ambiguity on machine translation's output quality in the judgement of users, thus providing concrete data to measure this impact. Our experiments also shed light on the fact that no absolute correlation exists between fluency- and adequacy scores. Lastly, this dissertation also includes a comprehensive survey of different forms of ambiguities and ambiguity resolution techniques in MT.