A Survey of Current Paradigms in Machine Translation

Bonnie J. Dorr
Computer Science Department
University of Maryland
A.V.. Williams Building
College Park, Maryland 20742
bonnie@umiacs.umd.edu
(301)-405-6768

Pamela W. Jordan
Intelligent Systems Program
University of Pittsburgh
901 Cathedral of Learning
Pittsburgh, PA 15260
jordan@isp.pitt.edu
(412)-624-7573

John W. Benoit
Mitre Corporation
Artificial Intelligence Center
7525 Colshire Drive
McLean, Virginia 22102
benoit@starbase.mitre.org
(703)-883-6685

Abstract

This paper is a survey of the current machine translation research in the US, Europe and Japan. A short history of machine translation is presented first, followed by an overview of the current research work. Representative examples of a wide range of different approaches adopted by machine translation researchers are presented. These are described in detail along with a discussion of the practicalities of scaling up these approaches for operational environments. In support of this discussion, issues in, and techniques for, evaluating machine translation systems are addressed.

1 Introduction

Machine translation (MT), i.e., translation from one natural language into another by means of a computerized system, has been a particularly difficult problem in the area of artificial intelligence (AI) for over four decades. Early approaches to translation failed in part because the interactive effects of complex phenomena made translation appear to be unmanageable. Later approaches to the problem have achieved varying degrees of success. In general, most MT systems do not attempt to achieve fully-automatic, high-quality translations, but instead strive for a level of translation that suits the basic needs of the user, perhaps requiring controlled input or revisions (post-editing) or both to arrive at the final result.

This paper is a survey of the current MT research in the US, Europe and Japan. While a number of MT surveys have been written (see, e.g., [99], [100], [118], [134], [138], [152], [154], [198], [200]), this one discusses a wide range of current research issues in light of results obtained from a survey and evaluation project conducted by Mitre [26], [27], [111]. During this project we evaluated 16 systems (10 operational and 6 under development) and also studied 7 U.S. research systems. Because a number of innovative MT approaches have come about since the completion of the Mitre study, we also include some discussion about more recent research paradigms. However, we do not attempt to describe all MT research in detail. Rather, we present approaches as representative examples of a wide range of different approaches adopted by MT researchers.

The next section provides a brief description of the history of MT. In section 3, we discuss the types of challenges (both linguistic and operational) that one must consider in developing an MT system. Section 4 describes three architectural designs that are used for MT. Following this, we compare translation systems along the axis of research paradigms (section 5); these include linguistic, non-linguistic, and hybrid approaches. We then discuss the difficult task of evaluating an MT system in section 6. Finally, we conclude with a discussion about what results we should expect to see in the future and where more effort needs to be placed.

2 The History of MT

Numerous attempts have been made in the past, both in the United States and Europe, to automate various steps in the translation process. These attempts range from simple on-line bilingual dictionaries, terminology data banks, and other translation aids to complete MT systems. Much work was done in the 1950s and 1960s toward achieving MT. However, the 1966 Automatic Language Processing Advisory Committee (ALPAC) report [5] condemned those efforts, citing poor-quality technology and the availability of inexpensive manual labor as negative-cost factors. These early efforts failed for several reasons, not the least of which was the unreasonably high expectation for perfect translation without having the basic theoretical foundation to achieve this. The ALPAC report caused a major reduction in U.S. research and development (R&D) efforts in the area of MT in favor of some related areas, such as computational linguistics and artificial intelligence, that subsequently provided a better theoretical foundation for current MT

R&D. Nevertheless, reduced but still significant MT research did continue at such places as the University of Texas/Austin, Brigham Young University, and Georgetown University. The ALPAC report also affected the R&D effort in Europe, but again, significant research continued in Western Europe and the USSR. An important side-effect of the reduced R&D funding for MT was the stimulation of a number of commercial MT endeavors by those displaced from the research centers. This resulted in most of our current operational MT systems, including Systran and Logos.

In the late 1960s, MT R&D was initiated in Canada, driven by the bilingual status of the country. In the late 1970s and the 1980s, two significant events occurred. The first was the formation of the EUROTRA project by the European Communities (EC) to provide MT of all the member nations' languages. The second was the realization of both Japanese government and industry that MT of Japanese to and from European languages first, and later to and from other Asian languages, was important to their economic progress. Thus far the EUROTRA project has failed to meet its goal of complete intertranslation of all the member languages; however, it has initiated important new research in MT and computational linguistics, and augmented existing MT research. Commercial MT systems supporting limited language pairs are now beginning to emerge from this effort. The EUROTRA project continues with somewhat narrowed goals in that a large single system is not being attempted. The Japanese efforts have produced a plethora of prototypes and commercially available operational systems, most based on established technology. Japanese research in MT, while never extensive, has been increasing both in quality and funding. A small effort is also under way in the former Soviet Union.

In the United States, research and commercial development have expanded considerably since the mid-1980s. In part, this expansion has been stimulated by the desire for more foreign markets. Government funding has increased, and MT research has evolved out of computational linguistics work at such places as New Mexico State University, Carnegie Mellon University, and University of Maryland. Several commercial systems have been developed, providing translation capabilities that are limited, but effective for some applications. Several small companies are developing and marketing more complete MT systems based on more recent technology. The U.S. Government, through its civil, military, and intelligence branches, is showing increased interest in using and supporting MT systems. A market for MT is developing among international and domestic corporations that are competing in the world market.

In summary, work on MT has been under way for over four decades, with various ups and downs. The ALPAC report of the mid-1960s was a serious, but by no means devastating, setback to the effort, and the current trend is toward increased support. This history is illustrated by Figure 1, adapted from a chart by Wilks [225].

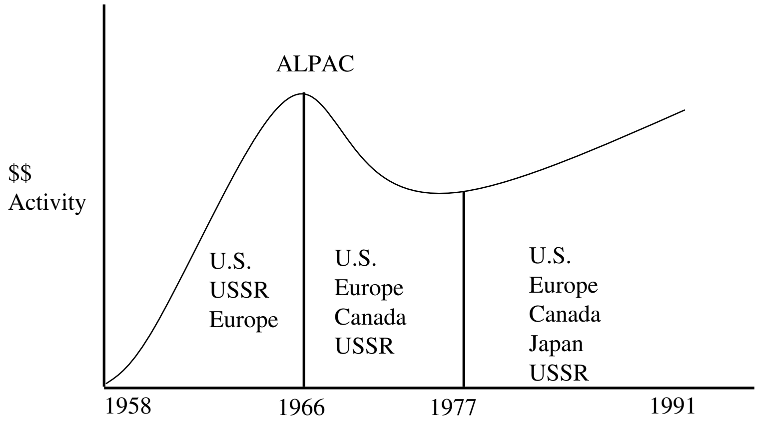

Figure 1: A History of MT

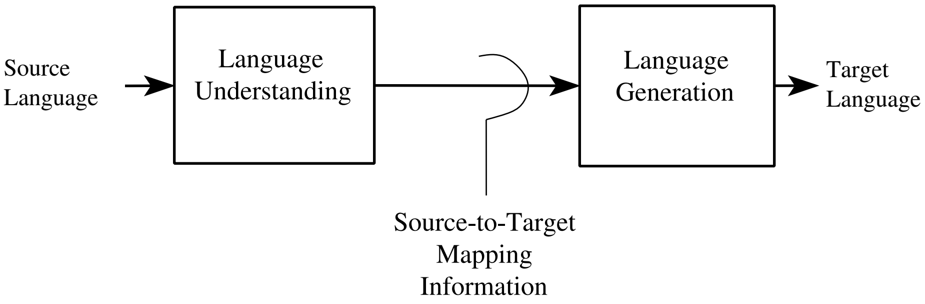

Figure 2: Three Categories of Linguistic Considerations

3 Translation Challenges

This section discusses the types of challenges that one must consider in developing an MT system. We examine these challenges along two dimensions, the first pertaining to different types of *linguistic considerations* (e.g., syntactic word order and semantic ambiguity) and the second pertaining to different types of *operational considerations* (e.g., extensibility, maintainability, and user interface).

3.1 Linguistic Considerations

There are three main categories into which the linguistic considerations fall; language understanding, language generation, and the mapping between language pairs. Roughly, these are related as shown in figure 2.

Regarding the first category, there have been many arguments in the past both for, and against, the idea that a complete understanding of the source text is necessary for adequate MT (see [18], [45], [133], [154], [156], among others). In more recent years, however, researchers have started to concentrate on the issue of whether it is possible to achieve a satisfactory translation with a minimal amount of understanding [25]. With respect to this issue, the areas to consider are: syntactic ambiguity, lexical ambiguity, semantic ambiguity, and contextual ambiguity. Each is addressed below.

In English, syntactic ambiguity arises in many contexts including the attachment of prepositional phrases, coordination, and noun compounding. For other languages the types of syntactic ambiguities will vary. The difficulty of prepositional phrase attachment is illustrated in the following English sentence:

(1) **Syntactic Ambiguity:**
I saw the man on the hill with the telescope

Here, we have no way to determine whether the telescope belongs to the man or the hill. However, in such cases, particularly for similar languages, it may not be necessary to resolve such an ambiguity since a particular source-language syntactic ambiguity may transfer to the target language and still be understandable to human readers.

In the case of lexical ambiguity, the choice between two possible meanings of a source-language lexical item is often easily resolved if enough syntactic context is available. Consider the following example:[1]

(2) **Lexical Ambiguity:**
E: book
S: libro, reservar

The English word *book* would be translated to the Spanish noun *libro* if it appeared after the word *the* or to the verb *reservar* if it appeared before the phrase *the flight*.

A more formidable problem is semantic ambiguity; the resolution of this type of ambiguity falls outside of the realm of syntactic and lexical knowledge as in the following examples:

(3) **Semantic Ambiguity:**

(i) Homography:
E: ball
S: pelota, baile

(ii) Polysemy:
E: kill
S: matar, acabar

Many words, such as *ball*, have distinctly different meanings (homography); MT systems are forced to choose the correct meaning of the source-language constituent in these cases (e.g., whether *ball* corresponds to a spherical object (pelota) or a formal dance (baile)). Other problems arise for words like *kill* which have subtle related meanings (polysemy) in different contexts (e.g.,

---

[1] Throughout this paper, the abbreviations C, D, E, G, and S will be used to stand for Chinese, Dutch, English, German, and Spanish, respectively. (Literal translations are included for the non-English cases.)

kill a man (matar) vs. kill a process (acabar)) and are frequently represented by distinct words in the target language.

Semantic ambiguity has often been considered an area in which it would be too difficult to provide an adequate translation without access to some form of "deeper" understanding, at least of the sentence, if not the entire context surrounding the sentence [18], [45]. The following well-known examples illustrate the difficulty of semantic ambiguity:

(4) **Complex Semantic Ambiguity:**

(i) Homography:
E: The box was in the pen.
S: La caja estaba en el corral / *la pluma
'The box was in the pen (enclosure) / *pen (writing)'

(ii) Metonymy:
E: While driving, John swerved and hit a tree.
S: Mientras que John estaba manejando, se desvió y golpeó con un arbol.
'While John was driving, (itself) swerved and hit with a tree'

In (4i), the system must determine that the *pen* is not a writing implement but some sort of enclosed space (i.e., a play pen or a pig pen) (homography resolution). In (4ii), the system must determine that it is John who is driving but John's car that hit the tree (metonymy resolution).

However, according to Bennett [25], the sort of ambiguity represented by these examples rarely arises in texts to which MT is typically applied; he argues that contextual ambiguity occurring in routinely translated texts (e.g., computer manuals) is often easily resolved by means of a simple feature-based approach. Consider the following examples:

(5) **Contextual Ambiguity:**

(i) E: The computer outputs the data; it is fast
S: La computadora imprime los datos; es rápida
'The computer outputs the data; (it) is rapid'

(ii) E: The computer outputs the data; it is stored in ascii
S: La computadora imprime los datos; están almacenados en ascii
'The computer outputs the data; (they) are stored in ascii'

In the context of a computer manual, determining the appropriate antecedent for the word *it* can be solved by distinguishing between storable objects and non-storable objects (storable ±) and between objects with a speed attribute and those without (speed fast/slow). Note that, although a computer is a storable object in other contexts, we can view it as a non-storable object in the limited domain of a computer manual.

More difficult ambiguities arise in translations that are truly ambiguous without extensive contextual cues, i.e., those that require discourse or pragmatic knowledge for correct interpretation. An effective discourse analysis would recognize themes and theme shifts in the text surrounding a sentence. As a simple example, consider the ambiguity in the following sentence:

(6) **Complex Contextual Ambiguity:**
E: John hit the dog with a stick
S: John golpeó el perro con el palo / que tenía el palo
'John hit the dog with the stick / that had the stick'

This ambiguity could be resolved by remembering from the earlier text that John was carrying a stick to protect himself (and not that there were several dogs, one of which had a stick). Pragmatic analysis deals with the intentions of the author in affecting the audience. This is as important for language generation (to be discussed next) as it is for language understanding. In particular, the author's intentions affect the choice of words and how they are realized (e.g., the use of active rather than passive voice to emphasize urgency). Together, discourse knowledge and pragmatic knowledge are useful in resolving many types of ambiguities.

A second type of linguistic problem for MT is that of language generation. Most MT researchers are of the opinion that generation of the target-language sentence does not require a full language generation capability, i.e., it is not necessary to fully plan the content and organization of the text; this is because the source-language text provides much of the information that will appear on the surface in the target language. However, generation for MT is a non-trivial exercise since it is often difficult to choose the words that adequately convey the conceptual knowledge behind the source-language sentence. This is the *lexical selection* problem.

Some simple examples for Spanish and English are given here:

(7) **Lexical Selection:**

(i) S: esperar
E: wait, hope

(ii) G: können
E: know, understand

(iii) E: be
S: ser, estar

(iv) E: fish
S: pez, pescado

Assume, for the sake of the current discussion, that an MT generator must select the appropriate target-language words from general notions such as *expect*, *have knowledge of*, *be*, and *fish*, respectively. In the above example, additional information is required for choosing the relevant term from each target-language pair. A possible scheme would be to use distinguishing features, e.g., ±desire, ±fact, ±permanent, and ±edible, respectively.

Further problems arise for MT generation in cases where linguistic information required in the target language is not explicit in the source language sentence. Consider the following example:

(8) **Tense Generation:**
C: Wǒ bèi Hángzhōu de fēngjǐng xīyǐnzhù le
E: I was captivated by the scenery of Hangchow
E: I am captivated by the scenery of Hangchow

In this example, two different English sentences might be generated from the Chinese. This is because tense information (past, present, future) is not overt in the source-language text. The information used to select the target-language tense depends entirely on the context of the utterance. For example, the second sentence would be generated if the speaker is looking at the scenery at the time of speech.[2].

The generation of tense is problematic in other languages as well. In Spanish there is a distinction made between simple past (preterit) and the ongoing past (imperfect). This type of distinction is not made explicitly in English. Consider the following example:

(9) **Tense Generation:**

(i) E: Mary went to Mexico. During her stay she learned Spanish.
S: Mary iba a Mexico. Durante su visita, aprendió español.

(ii) E: Mary went to Mexico. When she returned she started to speak Spanish.
S: Mary fue a Mexico. Cuando regresó, comenzó a hablar español.

In the first example, *went* is translated as the Spanish imperfect past since the sentence that follows is an elaboration, making *went* stative. In the second example, *went* is translated as a preterit past since the sentence that follows does not elaborate the visit to Mexico. (For a discussion about analogous examples in French, see [75].)

As we will see in section 3.2, the problems of understanding and generation in MT are often addressed by restricting the domain of the text so that the lexicon and grammar are constrained.

A third type of linguistic problem for MT concerns the mappings between source- and target-language representations. There are a number of dimensions along which source- and target-language representations may vary. These *divergences* make the straightforward mapping between languages impractical. Some examples of divergence types that MT researchers strive to address are *thematic*, *head-switching*, *structural*, *categorial*, and *conflational*.[3] Each of these will be discussed, in turn.

Thematic divergence involves a "swap" of the subject and object positions:

(10) **Thematic Divergence:**
E: I like Mary
S: Mary me gusta
'Mary (to) me pleases'

---

[2] This example is based on personal communication with Yan Qu

[3] Many sentences may fit into these divergence classes, not just the ones listed here. Also, a single sentence may exhibit any or all of these divergences.

Here, *Mary* appears in object position in English and in subject position in Spanish; analogously, the subject *I* appears as the object *me*.

Head-switching divergences occur commonly across language pairs. In such cases, a main verb in the source language is subordinated in the target language:

(11) **Head-Switching Divergence:**
E: I like to eat
G: Ich esse gern
'I eat likingly'

Observe that the word *like* is realized as a main verb in English but as an adverbial modifier (*gern*) in German.

In structural divergence, a verbal argument has a different syntactic realization in the target language:

(12) **Structural Divergence:**
E: John entered the house
S: Juan entró en la casa
'John entered in the house'

In this example, the verbal object is realized as a noun phrase (*the house*) in English and as a prepositional phrase (*en la casa*) in Spanish.

Categorial divergence involves the selection of a target-language word that is a categorial variant of the source-language equivalent. In such cases, the main verb often changes as well:

(13) **Categorial Divergence:**
E: I am hungry
G: Ich habe Hunger
'I have hunger'

In this example, the predicate is adjectival (*hungry*) in English but nominal (*Hunger*) in German. Note that this change in category forces the generator to select a different main verb.

Conflation is the incorporation of necessary participants (or arguments) of a given action. A *conflational* divergence arises when there is a difference in incorporation properties between the two languages:

(14) **Conflational Divergence:**
E: I stabbed John
S: Yo le di puñaladas a Juan
'I gave knife-wounds to John'

This example illustrates the conflation of a constituent in English that must be overtly realized in Spanish: the effect of the action (knife-wounds) is indicated by the word *puñaladas* whereas this information is incorporated into the main verb in the source language.

Resolution of cross-language divergences is an area where the differences in MT architecture are most crucial. Many MT approaches resolve such divergences by means of construction-specific rules that map from the predicate-argument structure of one language into that of another. More recent approaches use an intermediate, language-independent representation to describe the underlying meaning of the source language prior to generating the target language. The details of these contrasting approaches will be discussed further in section 4.

3.2 Operational Considerations

In addition to the above linguistic challenges, there are several operational challenges. These include: extension of the MT system to handle new domains and languages; handling a wide range of text styles; maintenance of a system once it has been developed; integration with other user software; and evaluation metrics for testing the effectiveness of the system.

Typically, to handle the linguistic challenges associated with understanding or generating a text, the text is restricted by domain so that the lexicon and grammar are more restricted. By doing so, the problems of lexical ambiguity, homography, polysemy, metonymy, contextual ambiguity, lexical selection, and tense generation are reduced. Then when building or extending an MT system to handle a particular domain and language, the designer must take on the smaller but still expensive task of acquiring and adapting the lexicon. To give an idea of the size of a domain lexicon, we have seen domain lexicons in commercial MT systems ranging from around 8000 to 12000 entries. The lexicon size varies according to the domain and whether an entry represents multiple senses or a single sense.

Although several researchers have developed tools to help with the acquisition of the lexicon (see, e.g., [28], [31], [32], [34], [36], [43], [53], [68], [73], [80], [88], [131], [142], [144], [164], [165], [166], [169], [171], [172], [212], [219], [227], [231]) these tools only help reduce the overall amount of work that is required by a small amount. The majority of the work still requires manual entry and fine-tuning by people with specialized expertise in linguistics and in the domain for which the system is being built. The words that should be included in the system can be extracted from a representative corpus and their possible parts of speech assigned. However, each of these entries must be reviewed to correct the part of speech assignments since the automated process is not 100% accurate. In addition, the entries must be manually modified so that other linguistic features can be added.

While some argue that one would want to manually review and fine-tune each entry anyway [141], the expense involved depends on the system architecture and research paradigm involved (i.e., statistical-based MT systems do not require detailed linguistic and domain knowledge). For systems that require large amounts of encoded knowledge, research is in progress to automatically extract other linguistic features from published bilingual and monolingual dictionaries and from parallel corpora [31], [40], [67], [78], [87], [148].

Another issue is how to cost effectively maintain a lexicon once it has been acquired. Most interfaces that have been built for users with no specialized linguistics training still look much like the first such interface created for the TEAM project [92]. (Other lexical interfaces are described in [17], [20], [70], [89], [91], [93], [124], [209], [208].) The maintainer is presented with various sentences utilizing the word being updated and asked to indicate which usages are correct and which are not. Each sentence represents a test to determine whether or not a particular linguistic feature applies.

One problem with these interfaces is that asking these types of questions does not work for all words. Someone with linguistic expertise will still have to review the results of the maintenance session. For example, once the Spanish verb *gustar* is entered into the lexicon as a psyche verb, someone with knowledge about linguistic structure must check that the argument structure (i.e., ordering of subject with respect to the object) is the reverse of the argument structure for the analogous English word *like*. (Example (10) given earlier illustrates this thematic divergence between Spanish and English.)

The extension of a system to handle additional languages also involves providing an analysis grammar for the source language and a generation grammar for the target language. Creating these grammars requires specialized linguistics knowledge as well as an understanding of the domain for which the system is being built since the grammar writer usually must understand the text in order to write a grammar that will produce an appropriate analysis. Grammar writing is the point at which many of the linguistic challenges associated with understanding and generation must be addressed. Heuristics relevant to the particular domain are often utilized at this point. For example, in Eurotra preferences for PP attachment are expressed with heuristics such as: "a PP which is not a modifier is preferred over the same PP when it is a modifier" [23].

As the grammar is being written it must be continually tested and refined in order to arrive at a reasonably good result for most of the expected inputs. Herein lies two major challenges: determining what a reasonably good result is and predicting the most likely inputs. Since grammar writing requires a great deal of linguistic expertise, even a small adjustment to the grammar is a development issue and not a user maintenance issue. Even as a development problem, this is one of the more time consuming tasks and one for which not many tools have been created.

Another operational consideration is the type of text to be translated. Handling a wide range of styles and sources of published text present vastly different degrees of operational difficulties for MT systems. Literary texts, such as novels and poetry, make frequent use of metaphor, have complex and unusual sentence structure, and assume a wide world and social context; these are all outside the competence of current MT systems. This is also true of popular journalistic texts, which, in addition, use (or create) the most fashionable words and social context. The problem is exacerbated by the fact that authors of these texts assume their audiences are knowledgeable about the general world and in some cases about the technical field underlying their writings. Often, the text cannot be understood without this type of knowledge, referred to as *world knowledge*.

MT systems fail for texts that rely heavily on metaphor and world knowledge because they have great difficulty in representing and using complex and subtle metaphors or understanding social context and interactions, and it is nearly impossible for them to keep up with the rapid changes in vocabulary. MT systems work best for texts that are written using simple syntax, make little or predictable use of metaphor, and have a stable vocabulary and a limited domain. Scientific and technical documents fall into this category and thus far have represented the most successful applications of MT. Text fragments such as tables of contents and sentence fragments present a different situation in that the syntactic rules must be relaxed to deal with incomplete sentences and possibly ungrammatical phrases. Since there may be little context as a basis for translating the fragments, lexical selection becomes an important and difficult problem.

Another operational consideration is the necessity for the MT system to be designed in such a way that it can be effectively integrated with other user software such as optical character recognition (OCR) and document publishing tools [4], [81], [95], [97], [143]. An application such as OCR might utilize some of the linguistic information that is available to the MT system; thus, this information should be handled in such a way that it is easily retrievable and usable independently from the MT system. Research MT systems tend to be modular and this operational consideration provides further motivation and challenges in designing the system.

A final operational consideration is how to evaluate and test the MT system. This applies to both users and developers of systems. When a system is being extended or when a purchase is being considered, there must be a way to test the effectiveness of a system in meeting the user's requirements. Further, when building research systems, one needs to be able to evaluate the effectiveness of the approach. As mentioned earlier in the above discussion on grammar writing, predicting the inputs is a challenge. In the case of evaluation, the question is whether the testing adequately covers the possible inputs when it is not clear what all the inputs will be. A second difficulty is determining the correctness of a translation. The correctness depends on the intended usage of the translation. Along with this, correctness is not a single binary judgement but a set judgements which may or may not be binary. An important issue in evaluation is that of choosing the appropriate judgement for a particular use of translation. We elaborate on these evaluation issues and the research in this area later in section 6.

4 Architectures

Current architectures for MT may be roughly organized into the following three classes: (1) Direct; (2) Transfer; and (3) Interlingua. These levels have been characterized in terms of a 'pyramid' diagram (see figure 3) which first appeared (in a slightly different form) in [218] and has since become classic.

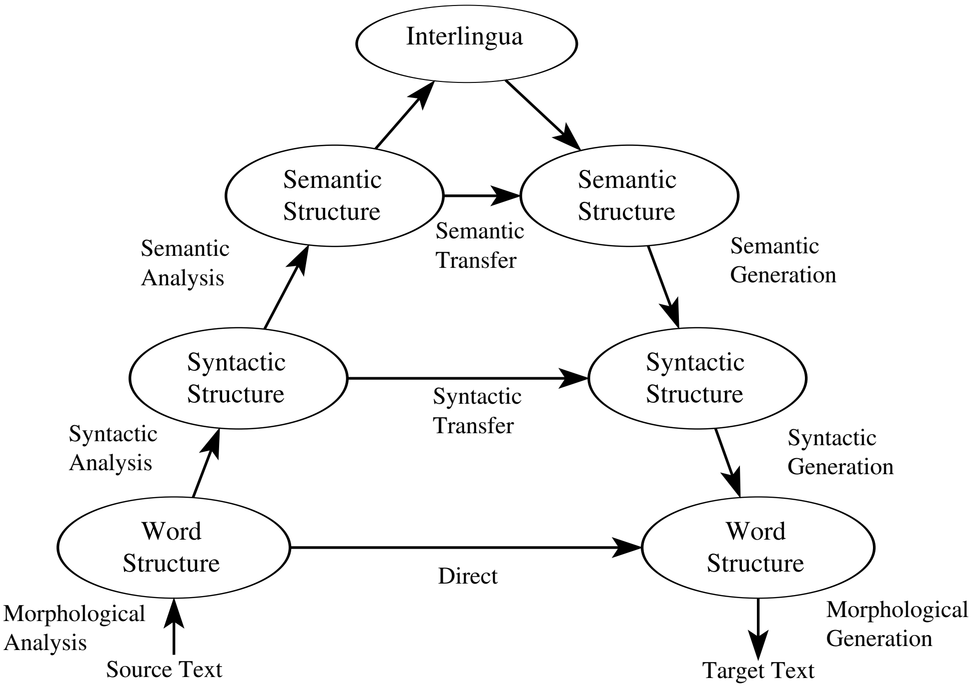

Figure 3: Types of MT Systems

The three levels correspond to different levels of transfer, depending on the depth of analysis provided by the system. At the bottom of the pyramid is the direct approach which consists of the most primitive form of transfer, i.e., word-for-word replacement. At the top of the pyramid is the interlingual approach which consists of the most degenerate form of transfer, i.e., the transfer mapping is essentially non-existent. Most translation systems fall somewhere between these two extremes ranging from a shallow (syntactic) analysis to a deeper (semantic) analysis. We will discuss, in turn, the MT architectures corresponding to these three levels.

4.1 Direct Architecture

The result of a direct translation architecture is a string of target-language words directly replaced from the words of the source language. Generally the word order of the target-language text is the same as that of the source-language, even in cases where the target-language does not permit the same word order. Unless the reader has a good knowledge of the source-language structure, this text can be very difficult to understand.

Some systems based on the direct architecture recognize special source-language syntactic forms and reorder the words to acceptable forms in the target language. These syntactic improvements increase the readability of the target text. For example, a direct approach might handle the thematic divergence of example (10) given earlier by means of a rule such as the following:

(15) **Direct Mapping:**
X LIKE Y → Y GUSTAR X

As we will see below, such rules are closer in nature to those used in the transfer approach. Unfortunately, without a detailed syntactic analysis, only simple forms can be recognized; consequently, complex structures, such as clauses and verb separations (as are frequently found in German), are left in the original syntax. Moreover, when more difficult cases arise, e.g., (11) above, it is impossible to construct direct mapping rules. The result is that this approach typically generates very literal translations, e.g., *I eat likingly* for the German sentence *Ich esse gern*.

A more serious problem with systems based on the direct architecture (as well as with some versions of transfer architecture systems) is selection of the correct target-language words for source-language words (lexical ambiguity). Recall from section 3.1 that many words, such as *ball*, have distinctly different meanings (homography) and translations; others, such as *kill*, have subtle related meanings (polysemy) that are frequently represented by distinct words in the target language. Direct architecture systems cannot cope with this lexical selection problem since they cannot relate a word to the way it is used in a sentence. The best that can be done is to restrict the textual domain and include in the lexicon only the translation most likely for that domain. Direct architecture systems produce, at best, poor translations. However, for limited domains and simple text (such as tables of contents or text fragments where correct syntax is less critical), they sometimes produce translations useful to domain experts.

4.2 Transfer Architectures

As shown in figure 3, transfer architectures lie on a spectrum ranging from direct to interlingual architectures: at the direct architecture end of the spectrum is the syntactic transfer architecture; at the interlingual end of the spectrum is the semantic transfer architecture. The initial intent of transfer architecture systems was to provide syntactically correct target-language text by transforming source-language representations into suitable target-language syntactic representations. Although the transfer rules that perform this conversion depend on both the source and target languages, some of the rules may need only slight modification when an MT system is developed for a new target language linguistically related to an existing one.

Both the transfer and the interlingual approaches require "linking rules" that map between the surface (source- and target-language) text and some form of internal representation. What distinguishes these two approaches is that the internal representations used in the transfer approach are assumed to vary widely from language to language. Thus, transfer rules must be constructed to map between these two representations. As we will see in the next section, no transfer rules are needed in the interlingual approach because the same internal representation is used for both the source and target languages.

Although the internal representations used for the source and target languages are not the same, the primitives (e.g., SUBJ, OBJ1) used in these representations are often similar, or even identical. The use of similar primitives allows for more general mapping rules than that of the direct approach. For example, the rule for mapping between the sentences in example (10) would be more general than the analogous rule in (15):

(16) **Transfer Mapping:**
gustar(SUBJ(ARG1:NP),OBJ1(ARG2:PREP)) →
like(SUBJ(ARG2:NP),OBJ1(ARG1:NP))

The effect of this rule is to swap the subject and object arguments and to change the category of the object from a preposition (in Spanish) to a noun phrase (in English).

Unlike the direct approach, the transfer architecture accommodates more complex mappings such as that of example (11). (We will discuss specific approaches to handling such cases in section 5.1.3.) However, a common criticism of this approach is that a large set of transfer rules must be constructed for each source-language/target-language pair; a translation system that accommodates n languages requires n^2 sets of transfer rules. This shortcoming has been noted by a number of researchers (see, e.g., [24], [64]).

Despite the drawbacks associated with the use of transfer rules, the syntactic transfer architecture has the advantage that ambiguities that carry over from one language to another are handled with minimal effort. Consider the example given earlier: *John hit the dog with a stick*. In this example, the syntactic ambiguity is not resolved by the syntactic analysis because there is no way to determine from syntax alone whether the dog had a stick or John used a stick to hit the dog. We have already seen that, if we are translating between similar languages, it may not be necessary to resolve the ambiguity; the source-language syntactic ambiguity may transfer to the target language and still be understandable to human readers. In an attempt to improve performance, some syntactic transfer architecture systems take advantage of this phenomenon and refrain from doing a complete syntactic analysis of these structures.

Transfer approaches are also able to resolve certain lexical ambiguities since the syntactic analysis can usually determine the lexical category (part of speech) of a source text word. For example, as mentioned earlier, it is possible to determine whether the English word *book* would be translated in Spanish to the noun *libro* or to the verb *reservar*, depending on the local context.

The overall translation quality of syntactic transfer architecture systems tends to be lower than those that employ a deeper analysis of the source-language text. Many lexical and syntactic ambiguities are not resolvable; consequently, long and complex sentences may not be understandable. In an attempt to improve translation quality by considering the meaning of the sentences, most transfer architecture systems have moved to the semantic transfer end of the spectrum by adding semantic analysis and semantic transfer rules as needed (i.e., ambiguities such as the *ball* and *with a stick* cases above would be resolved). The result of this combined syntactic and semantic analysis is a representation of the source text that combines translation-relevant syntactic and semantic information. Since this is usually done to solve specific language pair problems, the semantic analysis remains incomplete and, to some extent, language pair-dependent. That is, the addition of a new target language may well require modification of the source-language semantic analysis.

In principle, semantic transfer architecture systems have the capability to produce excellent translations, provided that a context (discourse and pragmatic) analysis is done in addition to a deep semantic analysis. In practice, little or no discourse or pragmatic analysis is done, and only enough semantic analysis is done to meet the translation goals of the system. Semantic transfer architecture systems can produce good translations when the analysis and rules are complete, and the bilingual lexicon covers the domain of interest.

A perceived difficulty with transfer architecture systems is that the transfer rules and, to some extent, the source-language analysis are dependent on both the source and target language. Thus a new system would have to be developed for each language pair of interest. This is not as problematic as might be expected. First, target-language generation can be expected to need little augmentation when a new source language is added. Second, much of the source-language analysis will not change as new target languages are added; only newly discovered semantic and structural differences need be resolved. Finally, it is true that new transfer rules will be required. However, the addition of a new source or target language will affect only the recognition or production parts of the rules, respectively; if the language is being replaced by one similar in structure, many of the transfer rules need not be changed. Of course, the addition of radically different languages (e.g. the first Asian language added to a system working between European languages) will require a major effort.

At the semantic-transfer end of the spectrum there is a final category of transfer architecture that could be viewed as a "special-case interlingual" design, i.e., one that defines a single syntactic and semantic representation for several related languages, such as the Romance languages. This approach is termed "multilingual". In figure 3, the multilingual representation takes the place of two semantic structure nodes; no transfer rules are necessary, yet the representation is not interlingual since, as in standard transfer systems, it relies on the characteristics of the source and target languages. In this approach the analysis and generation processes depend only on the respective source and target languages. In practice, this approach is being exploited by a number of systems.

To summarize, transfer architecture systems produce higher-quality results than direct architecture systems, but at the expense of having to develop extensive source-language analysis techniques and sets of transfer rules.

4.3 Interlingual Architectures

The basic idea of the interlingual (sometimes called pivot) architecture for MT is that the analysis of the source-language text should result in a representation of the text that is independent of the source language. The target-language text is then generated from this language-neutral, interlingual representation. This model has the significant advantage that analysis and generation development need be done only once for each language, and a translation system can be constructed by joining the analysis and generation through the interlingual representation.

This is currently a very active area of research, although a few commercial systems are based on this approach [16], [52], [79], [160], [161], [187]. The research issues center on the feasibility of specifying an interlingua that is adequate for all languages and on the depth of semantic analysis required to produce acceptable translations. The latter is also an issue for the more ambitious systems based on the semantic transfer architecture.

The interlingual approach to example (10) would be to assume that there exists a single underlying concept for the meaning of the main verb in both sentences, i.e., a representation such as the following:

(17) **Interlingual Representation:**
like/gustar: [CAUSE (X, [BE (Y, [PLEASED])])][4]

This representation conveys the idea that something or someone (X) causes someone (Y) to be pleased. An approach that adopts this representation would not require transfer rules since the representation would be the same for the source and target languages. Instead, all that would be needed is to define "linking rules" that map between the surface (source- and target-language) text and the interlingual form.

An issue raised with respect to this approach is that, because interlingual representations are generally independent of the syntax of the source text, the generation of the target language text from this representation often takes the form of a paraphrase rather than translation (see, e.g., [10], [101], [107].) That is, the style and emphasis of the original text are lost. However, this is not so much a failure of the interlingua as it is a lack of understanding of the discourse and pragmatics required to recognize style and emphasis. In some cases, it may be an advantage to ignore the author's style. Moreover, many have argued that, outside the field of artistic texts (poetry and fiction), preservation of the syntactic form of the source text in translation is completely superfluous. (See, e.g. [156], [222].) For example, the passive voice constructions in the two language may not convey identical

---

[4] This is a simplified, generic version of a Lexical Conceptual Structure (LCS) representation that could be attributed to a number of researchers including Schank ([184], [185], [186]) and Jackendoff ([104], [105]) among others. See [64] for a more detailed treatment of such cases.

meanings. The current state of the art seems to be that it is possible to produce interlinguas that are adequate between language groups (e.g., Japanese and western European) for specialized domains.

Another issue concerns a point raised earlier, i.e., that authors of source texts assume their audiences are knowledgeable about the general world and in some cases about the technical field underlying their writings. Many researchers (e.g., [154]) who adopt the interlingual approach aim to employ a deep semantic analysis that requires extensive world knowledge; the performance of deep semantic analysis (if required) depends on the (so far unproven) feasibility of representing, collecting, and efficiently storing large amounts of world and domain knowledge. This problem consumes extensive efforts in the broader field of artificial intelligence.

5 Paradigms of MT Research Systems

The architectural basis of the system is only one of many axes along which one might compare MT systems. Another important axis of comparison is that of research paradigm. It is important to understand the difference between the type of architecture and the type of paradigm: one does not presuppose the other. The former refers to the actual processing design (i.e., direct, transfer, interlingual), whereas the latter refers to informational components that aid the processing design (knowledge-based, example-based, statistics-based, etc.).

This section enumerates and discusses some of the more recent classes of MT paradigms that researchers are currently investigating. This list is, by no means, exhaustive. It is intended to cover most of the approaches that have been introduced in recent years, a vast majority of which were reported in the last five years at a number of conferences including the Annual Meeting of the Association for Computational Linguistics (ACL), the International Conference on Theoretical and Methodological Issue in Machine Translation (TMI), the Conference of the Association for Machine Translation in the Americas (AMTA), the International Conference on Computational Linguistics (COLING), and the Machine Translation Summit (MT-Summit).

There may be some disagreement about the boundaries of the classification. For example, the S&BMT paradigm has been viewed as Constraint-Based (CBMT) (see, e.g., [223], [224]) in that the translation process is taken to be a collection and resolution of sets of constraints. It has also been viewed as lexical-based (LBMT) (see e.g., [21], [22]) in that a bilingual lexicon is used to put into correspondence pairs of monolingual lexical entries. Frequently, researchers employ techniques from several categories. An example of such a case is an approach described in [90] which proposes to combine techniques used in Example-Based (EBMT), Statistics-Based (SBMT), and Rule-Based (RBMT).

We will discuss research paradigms in terms of three different categories: (1) those that propose to rely most heavily on linguistic techniques; (2) those that do not use any linguistic techniques; and (3) those that use a combination of the two. The separation of linguistics-based and non-linguistics-based approaches illustrates an emerging dichotomy among MT researchers that first became evident at the TMI in 1992. This is the confrontation dubbed the ‘rationalist-empiricist’ debate, which divides researchers into two groups, those who advocate well-established methods of rule-based/constraint-based MT (linguistic-based MT) and those involved in newer corpus-based MT (including EBMT, SBMT, and Neural Network Based (NBMT)). Many of these same issues have continued as hot topics of debate during the TMI in 1992. Several researchers have now acknowledged the need for a hybrid or integrated approach to MT that makes use of techniques from both types of paradigms, combining the best that each paradigm type has to offer.

5.1 Linguistic-Based Paradigms

Until very recently, most MT researchers studied Linguistic-based MT, i.e., translation on the basis of principles that are well-grounded in linguistic theory. Systems based on linguistic theory strive to use the constraints of syntax, lexicon, and semantics to produce an appropriate target-language realization of the source-language sentence. This section presents several of the most recent linguistic-based paradigms.

5.1.1 Constraint-Based MT

Constraint-Based (CBMT) techniques have shown up in several different MT approaches (see, e.g., [14], [75], [114], [115], [178], [223], [224]). For example, the Shake-and-Bake paradigm (discussed below) demonstrates the full utility of constraint application.

In this section, we will discuss one of the earliest MT approaches to use constraints on combination of lexical items, i.e., the LFG-MT system [114], [115]. This system translates English, French, and German bidirectionally based on *lexical functional grammar* (LFG) [113]. In the LFG formalism, f-structure (functional structure) is a fundamental component of the translation. For example, the f-structure for the sentence *I gave a doll to Mary* is:

$$(18)\quad \begin{bmatrix} \text{PRED} & \text{`GIVE}\langle(\uparrow \text{SUBJ})(\uparrow \text{OBJ})(\uparrow \text{TO OBJ})\rangle\text{'} \\ \text{SUBJ} & \begin{bmatrix} \text{NUM} & \text{SG} \\ \text{PRED} & \text{`I'} \end{bmatrix} \\ \text{TENSE} & \text{PAST} \\ \text{OBJ} & \begin{bmatrix} \text{SPEC} & \text{A} \\ \text{NUM} & \text{SG} \\ \text{PRED} & \text{`DOLL'} \end{bmatrix} \\ \text{TO} & \begin{bmatrix} \text{PCASE} & \text{TO} \\ \text{OBJ} & \begin{bmatrix} \text{PRED} & \text{`MARY'} \end{bmatrix} \end{bmatrix} \end{bmatrix}$$

The LFG-MT system is capable of handling difficult translation cases such as the head-switching case given earlier in (11). An example from [114] is the following:

(19) **Head-Switching Divergence:**
E: The baby just fell
F: Le bébé vient de tomber
'The baby just (verb-past) of fall'

Here, the English *just* is translated as the French main verb *venir* which takes the falling event as its complement *de tomber*. The f-structures that correspond, respectively, to the English and French sentences in this example are the following:

(20) (i)

$$\left[\begin{array}{ll} \text{PRED} & \text{`JUST}\langle(\uparrow \text{ARG})\rangle\text{'} \\ \text{ARG} & \left[\begin{array}{ll} \text{PRED} & \text{`FALL}\langle(\uparrow \text{SUBJ})\rangle\text{'} \\ \text{TENSE} & \text{PAST} \\ \text{SUBJ} & \left[\begin{array}{ll} \text{PRED} & \text{`BABY'} \\ \text{NUM} & \text{SG} \\ \text{SPEC} & \left[\begin{array}{ll} \text{DEF} & + \\ \text{PRED} & \text{`THE'} \end{array}\right] \end{array}\right] \end{array}\right] \end{array}\right]$$

(ii)

$$\left[\begin{array}{ll} \text{PRED} & \text{`VENIR}\langle(\uparrow \text{SUBJ})(\uparrow \text{XCOMP})\rangle\text{'} \\ \text{SUBJ} & \left[\begin{array}{ll} \text{PRED} & \text{`BÉBÉ'} \\ \text{GENDER} & \text{MASC} \\ \text{NUMB} & \text{SG} \\ \text{SPEC} & \left[\begin{array}{ll} \text{DEF} & + \\ \text{PRED} & \text{`LE'} \end{array}\right] \end{array}\right] \\ \text{XCOMP} & \left[\begin{array}{ll} \text{PRED} & \text{`TOMBER}\langle(\uparrow \text{SUBJ})\rangle\text{'} \\ \text{COMPL} & \text{DE} \\ \text{TENSE} & \text{INF} \\ \text{SUBJ} & \left[\text{BÉBÉ}\right] \end{array}\right] \end{array}\right]$$

Because the LFG-MT system is based on construction-specific representations, the mapping operations required in the transfer are performed by transfer equations that relate source- and target-language f-structures. The transfer equations that relate the f-structures (20)(i) and (ii) are the following:

(21) **LFG Transfer Rule:**
$(\tau \uparrow \text{PRED `JUST}\langle(\uparrow \text{ARG})\rangle\text{'}) = \text{VENIR}$
$(\tau \uparrow \text{XCOMP}) = \tau\ (\uparrow \text{ARG})$

This equation identifies *venir* as the corresponding French predicate, and it constrains the argument of *just* to be a complement that is headed by the prepositional complementizer *de*.

As illustrated here, the LFG-MT framework makes an association between the syntactic structure and the f-structure using a set of mediating selectional constraints that are encoded as lexical entries. The disadvantage to this approach is that the f-structure is tightly coupled with the syntactic structure of the language; thus, if a particular concept can be syntactically expressed in more than one way, there will be more than one f-structure in this framework.

A drawback to this approach, discussed in [179], is the handling of cases like (19) in the context of embedded clauses. In particular, if the English sentence in example (19) were realized as an embedded complement such as *I think that the baby just fell*, it would not be possible to generate the French output. The reason for this is that the LFG-MT system breaks this sentence down into predicate-argument relations that conform (roughly) to the following logical specification:

(22) **LFG Predicate-Argument Relation:**
think(I,fall(baby))
just(fall(baby))

The problem is that the logical constituent fall(baby) is predicated of two logical heads, *think* and *just.* The LFG-MT generator is unable to determine how to compose these concepts and produce an output string.

More recently, this difficulty has been addressed in [115] where a new description-language operator, *restriction*, is used to provide a more adequate account of head-switching. A refined version of this approach is described in [58] and currently used in the Verbmobil MT project [57].

5.1.2 Knowledge-Based MT

Knowledge-based machine translation (KBMT) is a paradigm adopted by developers of numerous U.S. systems including Pangloss at Carnegie Mellon's Center for Machine Translation [45], [147], [154], [160], [166]; ULTRA at New Mexico State's Computing Research Laboratory [80], [82]; Cyc-NL at CYCORP [130]; Japangloss and GAZELLE at University of Southern California and ISI [122], [123]; UNICON at the University of Pennsylvania [169]; and several European systems including an MT system developed at UMIST's Department of Language Engineering [8]. More recent extensions to the ULTRA and Pangloss systems have resulted in a Multi-Engine Machine Translation (MEMT) architecture [96]—based on techniques investigated in Mikrokosmos [135]—in which the strengths of many different approaches (KBMT, EBMT, SBMT, etc.) are exploited.

The KBMT paradigm concentrates on the development of knowledge intensive morphological, syntactic, and semantic information for the lexicon. Tools have been developed for creating ontologies and have provided a framework for automatic lexical acquisition [124]. In general, the focus of research in KBMT has been on the development of underlying KR concepts. Little attention has been devoted to providing a systematic "linking" of these concepts to the syntax; thus, KBMT is complementary to the (less knowledge-intensive) LBMT and PBMT paradigms described below.

An example of a case addressed in KBMT is the translation of a sentence such as the following (taken from [154]):

(23) Drop by your old favorite Dunkin' Donuts shop.

The underlying representation for this sentence is an interlingual text (ILT) representation involving a set of frames that look like the following:

```
(24) (i)    (make-frame text_1
               (clauses (value clause_1))
               (relations (value relation_1))
               (attitudes (value attitude_1))
               (producer-intentions (value producer_intention_1)))
     (ii)   (make-frame clause_1
               (head (value %visit_1))
               (aspect (duration prolonged) (phase begin) (iteration 1))
               (time (value time_2)))
     (iii)  (make-frame %visit_1
               (is-token-of (value *visit))
               (agent (value *consumer*))
               (destination (value %shop_1)))
     (iv)   (make-frame relation_1
               (type (value domain-conditional))
               (first (value %visit_1))
               (second (value %involuntary-perceptual-event_1)))
     (v)    (make-frame attitude_1
               (type (value &evaluative))
               (attitude-value (value .9))
               (scope (value %shop_1))
               (attributed-to (value *consumer*))
               (time (value (since time_5))))
     (vi)   (make-frame producer-intention_1
               (is-token-of (value *commissive-act))
               (scope (value relation_1)))
```

KBMT requires a syntactic frame (in the lexicon) for each conceptual event that occurs in the ILT. The frame that provides the meaning for the *drop by* action (which is really a form of *visit*) looks like the following:

```
(25) (drop
       (make-frame
         +drop-v1
         (cat (value v))
         (syn-struc
           ((root $var0)
            (subj ((root $var1) (cat n)))
            (obliques ((root $var2) (prep by)))))
```

```
(sem
  (lex-map
   (%visit
    (agent (value ^$var1))
    (theme (value ^$var2)
    (sem *location)
    (relaxable-to *object)))))))
```

Note that the syntactic realization information appears in the hand-coded syn-struc slot. The linking rule that relates this information to the ILT is implicit in the specification that appears in the sem slot. In particular, this information specifies that the agent (var1) maps into syntactic subject position and the theme (var2) maps into a syntactic oblique position.

This encoding of the linking rule identifies the precise syntactic realization for the underlying concept (i.e., the concept is realized as the verb *drop* with a subject and an oblique argument). The reason such specific syntactic information is required is that the conceptual representation diverges significantly from the surface realization. In other words, there is no single systematic mapping between a concept and its surface realization. (In the PBMT paradigm described below, we will see that the syntax-semantics mapping may be generalized.)

As described in [96], the KBMT paradigm provides high-quality, fully-automated translations; however, the domains used for this approach are narrow and well-defined since the required amount of information would otherwise be large and cost-prohibitive with respect to indexing and searching.

5.1.3 Lexical-Based MT

Lexical-Based MT (LBMT)[5] overlaps heavily with several other paradigms including RBMT ([9], [11], [13], [10]), PBMT ([60], [61], [63]), and S&BMT ([21], [22]); in general, a lexical-based system refers to any system that supplies rules for relating the lexical entries of one language to the lexical entries of another language. Several researchers have adopted the lexical-based paradigm, but at different degrees of generality. (See, e.g., [1], [30], [59], [62], [64], [69], [74], [77], [78], [79], [80], [181], [211], [213], [227].)

One such system is the LTAG system [1] for English-French and French-English. The system is a transfer approach that uses synchronous tree-adjoining grammars (as described in [194]) to map shallow tree-adjoining grammar (TAG) [112] derivations from one language onto another. The mapping is performed by means of a bilingual lexicon which directly associates source and target trees through links between lexical items and their arguments. Roughly, each bilingual entry contains a mapping between a source-language sentence and a target-language sentence.

---

[5] The acronym LBMT has also been used for linguistic-based MT, which is a more general term that refers to techniques that belong to any category outside of EBMT, NBMT, or SBMT. It is used in a more specific sense here, i.e., it refers to those linguistic-based systems that are driven primarily by the lexicon.

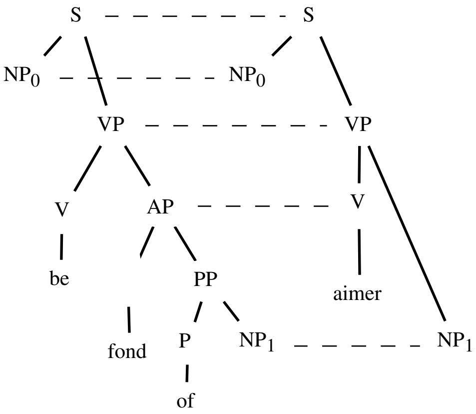

Figure 4: Mapping Source-Language Trees to Target-Language Trees in LTAG

This approach handles cases such as the following:

(26) **Categorial Divergence:**
E: John is fond of music
F: John aime la musique
'John loves the music'

Here, the source language concept is realized as the adjectival form *be fond of* in English, whereas the French translation realizes this concept as the verb *aimer*. The transfer rule that accounts for this mapping directly links the adjectival phrase *fond of* in the source-language tree with the verb *aimer* in the target-language tree as shown in figure 4. This translation mapping relates the AP node in the English tree to the V node of the French tree.

The advantage of this approach is that it accommodates modifying phrases in cases such as the following:

(27) **Categorial Divergence:**
E: John is very fond of music
F: John aime beaucoup la musique
'John loves very much the music'

Here, the English adverb *very* is associated with the predicate *fond of* (instead of with the main verb) whereas in French, the corresponding adverbial *beaucoup* is associated with the main verb *aimer*. The mechanism that permits this modification to be appropriately executed is the linking between the adjectival phrase *fond of* and the verb *aimer*: since the English main verb *be* has no link associated with it, the modifier must instead be associated with the adjectival phrase.

One disadvantage to this approach is that it requires *entire* trees to be stored in the transfer dictionary for each source-to-target pair, which is significantly burdensome as the number of source and target languages begin to add up. However, the approach is powerful in its treatment of cases such as (27), which have remained otherwise unaddressed in the literature.

5.1.4 Rule-Based MT

The Rule-Based MT (RBMT) paradigm is associated with systems that rely on different linguistic levels of rules for translation between the source and target language [9], [10], [11], [13], [54], [83], [90], [114], [115], [128], [139], [163], [176], [210], [215]. The prototypical example is Rosetta [176], an interlingual system which divides translation rules into two categories: (1) M-rules, which are "meaning-preserving rules," map between syntactic trees to underlying meaning structures; and (2) S-rules, which are "non-meaningful rules," map lexical items to syntactic trees. The former are used for compositional (or "regular") phenomena and the latter are used for non-compositional (or "exceptional") phenomena.[6]

Consider the following case:

(28) **Head-Switching Divergence:**
E: Mary happened to come
D: Mary kwam toevallig
'Mary came by chance'

The interlingual representation used for this case is a canonical form corresponding to '`by-chance(Mary,come)`.' Thus, the English syntactic structure parallels the canonical form: the verb construct corresponding to '`by-chance`' (*happen to*) takes as its argument the clause corresponding to '`(Mary,come)`' (*come*). The Dutch syntax, on the other hand, does not correspond to the structure of the canonical form since the verb corresponding to '`(come,Mary)`' (*kwam*) takes as its argument the adverbial construct corresponding to '`by-chance`' (*toevallig* or 'by chance'). In order to handle such cases compositionally, a "switch rule" is triggered by the adverb *toevallig* and normal processing is interrupted; control is then passed to a module that derives a new category (from an *argument* of the canonical head) that takes over the role of syntactic head.

This approach differs from the LBMT approach in that the idiosyncratic requirements of individual lexical items are encoded in the control mechanism rather than in the lexicon. In the case above, the control mechanism is governed by grammatical rules that interpret adverbials such as *toevallig* as a "deviant," i.e., a non-head in the syntax but a head in the canonical form. In the LBMT approach, such information is encoded in the lexicon, thus allowing only *certain* adverbials to act as interrupt triggers, namely those associated with a lexical marker.

---

[6] The separation of these two rule types is analogous to the notion of relaxed compositionality in systems such as Eurotra and its descendant, MiMo [11], [13]. Both groups address the same phenomena, e.g., head-switching divergences, but Rosetta uses an interlingual representation instead of transfer rules.

A difficulty with this approach is the accommodation of head-swapping cases where the "deviant" serves as a head in the syntax but a non-head in the canonical form. Consider the following example:

(29) **Head-Switching Divergence:**
E: Mary usually goes to school
S: Mary suele ir a la escuela
'Mary is accustomed to go to school'

In this example, the canonical form could, arguably, be '`go(Mary,school,usually)`,' i.e., the English syntax parallels the canonical form. By contrast, the Spanish syntax includes the "deviant" verb *suele*, which is a head in the syntax but a non-head in the canonical form; this is the inverse of the previous example. The head-swapping rules in Rosetta do not include such a case, perhaps because this would force an interrupt to occur too late—*after* the syntactic structure corresponding to the logical head has already been built. Even though it might not be linguistically justified, the canonical representation for the Spanish sentence would have to be '`be-accustomed(Mary,go(school))`.' By contrast, the English sentence would map into the canonical form given above which means that the two would never be translation equivalents. Instead the system would force the following, more literal, translation pairs (in both directions):

(30) **Literal Translation:**

(i) E: Mary usually goes to school
S: Mary usualmente va a la escuela

(ii) E: Mary is accustomed to going to school
S: Mary suele ir a la escuela

Taking a grammar-driven approach forces the Rosetta developers to regard such cases of mismatch as purely grammatical. The possibility of extending the notion of compositionality into the lexicon is a possible future direction for Rosetta, which underlyingly is a well-developed system that covers a wide range of linguistic phenomena.

5.1.5 Principle-Based MT

The Principle-Based MT (PBMT) paradigm has been investigated by a number of researchers [60], [61], [63], [71], [193] [196], [197], [221]. PBMT is an alternative to RBMT in which rules are replaced by a small set of *principles* that cover morphological, grammatical, and lexical phenomena in a general fashion. An example of a construction that is derived from general principles is the *passive* construction (e.g., *The ball was hit by John*), as described by Berwick in [29, pp. 1-2]:

> "... constructions like passive *follow* from the deductive interactions of a relative small set of primitives. On this view there is no 'passive rule'. Passive constructions result from the interaction of deeper morphological and syntactic operations."

A canonical example of the PBMT paradigm is the PRINCITRAN MT system [71], which is based on syntactic principles of Government-Binding (GB) Theory [50] and lexical-semantic principles of Lexical Conceptual Structure (LCS) [64], [105], [104]. This system was originally designed for Korean-English machine translation; lexicons for Spanish and Arabic were subsequently developed for related NLP applications (e.g., foreign language tutoring) [66]. In this system, structure building is deferred until underlying descriptions satisfy a number of linguistic principles that are parameterized for ease of porting to new languages [65]. The processing mechanism is language independent and accommodates structurally different languages (e.g., head-initial vs. head-final) with equally efficient run times.

PRINCITRAN encodes the grammar for each language as a network of nodes that represent grammatical categories (e.g., NP, Nbar, N) or subcategories, such as V:NP (i.e., a transitive verb that takes an NP as complement). Input sentences are parsed by passing messages in the grammar network. The network nodes are computing agents that communicate with each other by sending messages in the reverse direction of the links. GB principles are implemented as constraints attached to and links. All information at a node must satisfy the node's local constraint. A message can be sent across a link only if it satisfies the link's constraint. For example, there is a GB principle that requires every NP to be assigned abstract case. The Case Filter rules out sentences containing an NP with no case. A Case Filter violation is detected if a node containing `-cm` percolates its attribute to a sister node containing `-ca`, thus ruling out sentences such as *I believe John pleased* (but allowing, for example, *I believe John is pleased*).

This framework is well-suited to an interlingual design since the linking rules between the syntactic representations and the underlying lexical-semantic representation are well-defined. PRINCITRAN uses LCS as its interlingual representation and handles translational divergences (such as those described above in Section 3.1) by means of the parameter-setting approach. For example, there exists a single underlying concept for the meaning of the main verb in both the English and Spanish sentences in (10) given earlier, yet there are lexical-semantic divergences between the two languages:

(31) **Thematic Divergence:**
S: [S [NP X'] [VP [gustar [NP Y']]]]
E: [S [NP Y'] [VP [like [NP X']]]]

The relation between the lexicon entry for the verb and the syntactic representation is captured by means of a general "linking rule"—the main principle by logical arguments are mapped into syntactic positions. In Spanish, the default linking rules are used (e.g., the logical subject X is mapped into the syntactically external position X'). In English, the lexical entry is parameterized to override this linking (e.g., the logically external subject X is mapped into the syntactically internal position X'):

(32) **Lexical Representations for Thematic Divergence:**
like: [CAUSE (X :INT, [BE (Y :EXT, [PLEASED])])]

gustar: [CAUSE (X, [BE (Y, [PLEASED])])]

The PBMT paradigm is complementary to the KBMT and EBMT approaches in that it provides broad coverage of many different linguistic phenomena, but lacks deeper knowledge about the translation domain. Work is currently underway to combine the benefits of the PBMT and KBMT approaches in a collaborative large-scale Chinese-English MT effort between the University of Maryland and New Mexico State University [69], [164], [165].

5.1.6 Shake and Bake MT

One of the newest linguistic-based translation paradigms is Shake and Bake MT (S&BMT) [21], [22], [35], [223], [224], which is a perfect example of why we distinguish between research paradigm and MT architecture. Although the originators of this approach claim that S&BMT is an alternative to the transfer architecture (and also to the interlingual architecture), in fact, transfer rules are precisely the mechanism through which the translation is achieved. However, while the mapping between lexical items is achieved through standard transfer rules, the algorithm for combining these items to form a target-language sentence is nonconventional.

The transfer rules are defined on the basis of "bilingual lexical entries" which relate monolingual lexical entries. After the source-language sentence is parsed, the source-language words are mapped to target-language words by means of the bilingual entries. The algorithm used for combining the target language words attempts to order the words based on syntactic constraints of the target language.

S&BMT is motivated by the need to handle complex translations such as the head-switching case given above in section 3.1. Unlike the transfer approach, the S&BMT algorithm overcomes the difficulty of constructing non-compositional mapping rules for such cases by selecting target-language words from a bilingual lexicon and trying different orderings of these words (the 'Shake' of S&BMT) until a sentence is produced (the 'Bake' of S&BMT) that satisfies all syntactic constraints. Consider the following case for English/Dutch, which is analogous to example (11) given earlier:

(33) **Head-Switching Divergence:**
E: John enjoys swimming
D: Jan swemt graag
'John swims likingly'

The statements of equivalence in the bilingual entries for the words used in this example are spelled out as follows:

(34) **Bilingual Entries for Head-Switching Divergence:**

(i) $\mathrm{X_E}$ & $\mathrm{X'_E} \equiv \mathrm{X_D}$
<$\mathrm{X_E}$ cite> = enjoy
<$\mathrm{X'_E}$ cite> = prespart

<X_D cite> = graag
<X_E sem index> = <X_D sem index>
<X'_E sem index> = <X_D sem index>
<X_E sem exp index> = <X_D sem exp index>
<X_E sem obj index> = <X_D sem obj index>

(ii) $Y_E \equiv Y_D$
<Y_E cite> = swim
<Y_D cite> = zwemen
<Y_E sem index> = <Y_D sem index>
<Y_E sem agt index> = <Y_D sem agt index>

These bilingual rules form the basis of the transfer between the English lexical entries X_E (and X'_E) and Y_E to the Dutch lexical entries X_D and Y_D, respectively. The **cite** feature uniquely picks out an entry in the dictionary. The **sem** feature associates a semantic representation with different components of the entry. The first two usages of **sem** indicate that the semantics of *enjoy* and the present participial (i.e., the "-ing" form in English) are mapped to the semantics of *graag*. The **sem** feature is also used to associate thematic relations (i.e., **exp**eriencer and **obj**ect in (34i); and **agt**ent in (34ii)). The words *swim* and *zwemen* are also related through the use of the **sem** feature.

One point to note here is that the two lexical entries in English (X_E and X'_E) are related to one lexical entry in Dutch (X_D). The relation of X'_E to X_E need not be specified in the bilingual rules since this information can be determined from grammatical constraints during translation (i.e., that the present participial morpheme "-ing" must be associated with the verbal argument *swim*, not with the main verb *enjoy*). There are also other types of information inherent in the translation pairs that need not be specified in the bilingual rules; in particular, the fact that the equivalent tense morphemes (pres) occur on non-equivalent stems (*enjoy* and *zwemen*) follows immediately from the mechanics of generation.

The idea behind this approach is that, once the bilingual elements correctly identify the indices of the lexical entries, the S&BMT algorithm has the job of "combining" of these elements. Translation equivalence is stated between *bags* of lexical constituents. For example, the full bags for the source- and target-language sentences given above in (33) are the following:

(35) {jan,pres,zwemen,graag} ≡ {john,pres,enjoy,prespart,swim}

The relation between the constituents in those bags need not be explicitly stated in the lexicon since these can be determined from the grammatical restrictions of the two languages.

The benefit to this design is that the bilingual lexicographer need only specify contrastive knowledge between the two languages; the monolingual grammars used for parsing and generation take care of the rest. The creators of this design have proposed that the bilingual mappings are restricted enough to allow for the possibility of automated acquisition of bilingual correspondences from aligned corpora (see [22], [224]).

A disadvantage to this approach is that, as described in [35], S&BMT generation is an NP-complete problem. Thus, there is no tractable general algorithm for generating within the S&BMT framework. However, it is possible to impose restrictions on the target-language bag which forms the input to generation. For example, heuristic control might be provided from the structure of the source language. Brew has shown that a heuristic approach based on constraint propagation provides considerable improvements in practice. In addition, refinements have been proposed in [76, p. 365] for handling translation ambiguity.

5.2 Non-Linguistic-Based Paradigms

In the past few years, researchers have investigated MT paradigms that are not based on linguistic theories or even linguistic properties of language. This investigation has been made possible by the rapid advances in computational power and the availability of machine readable dictionaries and monolingual and bilingual text corpora. The approaches all depend on the existence of large text corpora which are used either for training data or databases of existing translations. This section describes some of the recent research in non-linguistic-based paradigms.

5.2.1 Statistical-Based MT (SBMT)

The production of translations based on statistical prediction techniques depends heavily on statistical analysis of bilingual parallel corpora. While there were some early investigations [116] of the SBMT approach, the modern efforts were initiated by IBM in 1988 in the Candide French-English Machine Translation Project [37], [38], [39], [41]. Additional investigations of SBMT have been reported in [48], [56], [90], [126], [136], [158], [159], [205].

SBMT was derived from speech processing techniques. In particular, a variant of Bayes' Rule is used to show that the probability that a string of words (T) is a translation of a given string of words (S) is proportional to the product of the probability that a string of target words is a legal utterance in the target language and the probability that a string of words in the source language is a translation of the string of words in the target language. That is:

$$(36)\quad P(T \mid S) \sim P(T) * P(S \mid T)$$

If the right hand probabilities are known, the translation is obtained by choosing T such that the left hand probability is maximized. Obviously, the probabilities for all strings in both languages cannot be known and consequently must be estimated for the approach to be tractable. The usual approach is to define approximate probabilistic models constructed from probabilities that can be directly estimated from existing data.

The Candide language model is a trigram model asserting that the probability that any word in a target language (English) string is part of a legal sentence depends only on the two previous words. Knowing these probabilities, an estimate of the probability that a string of words is a legal sentence is the product of all the trigrams in the string. The trigram probabilities can be estimated by counting the frequency of word triples in a large corpus of English text.[7]

The probabilistic model underlying the Candide system assumes that the probability that a source language (French) word is a translation of a given English word depends only on the English word. A single translation allows for 0 to 10 French words. It is also assumed that the English equivalents of the French words might be ordered differently in the target sentence. The estimation of these probabilities is considerably more difficult than for the trigrams. In this case a very large bilingual parallel corpus is required.[8] The problem is to align each English word in a target sentence with the French equivalent(s) in the source sentence.[9] The approach is to assume values for the alignment probabilities and compute the transfer probabilities of the sentence pairs in the corpus. Depending on the alignment and translation probabilities, a given sentence pair may have several transfer probabilities. Each occurrence of the alignments and translations are counted and weighted by the transfer probability. These weighted counts are used to make a new estimate of the individual probabilities and the process is repeated. This iterative approach converges to a local equilibrium of probabilities and is used to compute $P(S \mid T)$.[10]

This approach has been tested in the laboratory for French-English MT and produces translations approaching the quality of those from syntactic transfer systems. The grand claim made for this approach is that no lexicons or grammars are used. Everything comes from statistical analysis of corpora. While this is the strength of this approach, it is also the weakness since the corpora must exist. Also the translations are very dependent on the domain of the corpora. An even more serious problem is that the only way to improve the quality of the translation is to improve the accuracy of the probabilistic models of the target language and of the translation process. Unfortunately, this would add many more parameters to the millions required by the simple models described above. Recognizing this difficulty, IBM has applied a number of techniques from classical computational linguistics to form a hybrid system. Morphological analysis, part of speech tagging, syntactic regularization, limited grammatical analysis, and contextual marking are used. Most of these techniques are parameterized with the values derived from analysis of the corpora. Some improvement was achieved; however, they were unable

---

[7] Of course, no corpus will be large enough to contain all possible triples and some smoothing method is required to assign a (small) probability to unseen triples.

[8] Fortunately French and English versions of the Canadian parlimentary proceedings (called the Hansards) are available.

[9] The search for source text words is non-trivial (see [37]).

[10] The details of this process are outside the scope of this paper (see [49]). Obviously smoothing is required.

to match the best of the commercial MT systems [226].

IBM continued to work on this system until 1995 when both internal and external support were withdrawn. Two quotes from Yorick Wilks [226] best summarize the impact of this work:

> "Another way of looking at this is how much good IBM is doing us all: by showing us, among other things, that we have not spent enough time thinking about how to acquire, in as automatic a manner as possible, the lexicons and rule bases we use." (p. 144)
>
> "Brown et al.'s retreat to incorporating symbolic structures shows the pure statistics hypothesis has failed." (p. 145)

Since this earlier work on SBMT, there have been improvements to IBM's original model of translation. Speed, a long-standing issue for statistical methods, has been addressed by techniques involving dynamic programming [150] and structural constraints [228], [230].

Another interesting statistics-based MT approach to using statistical techniques to generate MT systems is LINGSTAT developed by Dragon Systems, Inc. [19], [232], [233]. This work started in 1992 as a translation aid as a direct substitution system with a simple, hand-generated finite state grammar for Japanese. The English glosses were based on bilingual dictionaries and the grammar was used to assign Japanese phrase attachment. This was quickly seen to be unsatisfactory and a number of statistical steps were taken to provide improved complete translation. The finite state grammar was expanded to a probabilistic context free grammar (PCFG) and trained on Japanese text. The PCFG was used to provide a gross parse and a lexicalized grammar (also trained on Japanese text) was used to assist with attachments. Hand-generated reordering rules were provided to assist the transfer to English. Further, a trigram probabilistic language model was developed for English to assist in gloss selection. Some improvement was obtained with these changes. These techniques were ported to Spanish/English translations with somewhat better results than for the Japanese.

The LINGSTAT project ended in 1995 when support was withdrawn. Interestingly, linguistic-based extensions were planned as a continuation of this work. One extension involved the assignment of case frame categories to the source-language verbs in order to improve the parse. The probabilities of these sub-categories were to be learned by iterative parsing of source text. Additional extensions involved experimentation with extraction of phrase translations from parallel bilingual corpora. While the LINGSTAT group never committed to "pure" statistical MT as did the Candide group, they were strongly committed to statistical training and extraction of more symbolic approaches. This intersection of statistical and symbolic paradigms is relevant to the hybrid techniques discussed in Section 5.3 below.

5.2.2 Example- (or Case-/Memory-) Based MT (EBMT)

Example-Based MT (EBMT), first suggested by Nagao [145], emulates human translation practice in recognizing the similarity of a new source language sentence or phrase to a previously translated item and using this previous translation to perform "Translation by Analogy". Sato and Nagao [183] implemented an experimental EBMT system to demonstrate the translation of simple Japanese sentences into English. Additional investigations of EBMT have been reported in [84], [90], [109], [137], [140], [153], [158], [159], [163], [175], [201], [206], [207], [234].

The basic idea of EBMT assumes a database of parallel translations which is searched for the source language sentences and phrases closest matching a new source language sentence. The translations of the matched phrases are then modified and combined to form a transfer translation of the new sentence. This technique is quite similar to Case Based Reasoning used in Artificial Intelligence (see [125]). A simple match would be an identical phrase (especially in the function words) except for a similar content word. The closeness of the match would be determined by the semantic "distance" between the two content words as measured by some metric based on a thesaurus or ontology. The translation would be the substitution of a translation of the different word in the translation of the best match.

The accuracy and quality of the translation depends heavily on the size and coverage of the parallel data base. While the database need not be as large as required for SBMT (since the full vocabulary need not be covered), the required coverage of syntactic and semantic divergences results in a size difficult to store and search. Phrasal matching requires at least a rough syntactic analysis of the parallel translations as well as some semantic analysis to determine the closeness of the match.[11] In order to avoid matching improper divergences, Collins and Cunningham [51] weight phrasal translations by their frequency in the database.

Sentence translation in EBMT requires that in addition to phrasal matching the syntactic structure of the source sentence must be matched with sentences in the database. While full sentence matching has shown some success [85], most uses of EBMT are restricted to subproblems such as function words [207], noun phrases [182], and prepositional phrase attachment [207].

5.2.3 Dialogue-Based MT

Dialogue-Based MT (DBMT) is a paradigm aimed at a user who is the original author of text to be translated. Systems based on this approach provide a mechanism for entering into a dialogue with the user about the translation [33], thus allowing the user to disambiguate the input text and to incorporate culture-specific stylistic detail for high quality MT [180]. DBMT is similar to EBMT in that a representation of the user's input text is built and, as it is iteratively revised through dialogue with the user, attempts are made to match the updated representation against those pre-stored in a translation database. Interaction with the user may take place during the

---

[11] Nirenburg, Domashnev, and Grannes [151] argue that such analysis defeats the purpose of EBMT and propose substring pattern matching using synonyms and hyponyms.

translation process or before text is input to the system. In the former, the interaction provides a user-guided online disambiguation mechanism; in the latter, the interaction provides a user-guided revision process that results in text that the system is capable of handling.

The first implementation of DBMT was the ENtran project [108] which served as the basis of work at UMIST for British Telecom [110], [201], [214]. The system was designed to provide a means for constructing an input text that it is capable of translating, leaving the remaining phenomena to be processed by the user. In an extension to that work [202], a more flexible system was developed for Japanese ATR research. This augmented system acted as an intermediary for the user in a dialogue with a conference office, thus providing a dialogue MT system in two senses (as described in [203]): (1) it enters into a dialogue with the user about translation; and (2) the object of the translation is the user's contribution to a dialogue.

The type of phenomena addressed by DBMT systems is more knowledge-intensive than the divergences described in Section 3.1. In [110], certain contextual information is encoded in a constrained domain (business correspondence) so that the system can determine the likely intention of the user. Using this information, the user is guided through a series of choice points, thus providing a means for constructing a representation that is offered to the system as a translation candidate.

Consider a case where the user has produced a text that is analyzed as two representations:

(37) (i) $\texttt{complain(X1)}_{agent}\texttt{(X2)}_{goal}\texttt{(X3)}_{recipient}$

(ii) $\texttt{not\_receive(X1)}_{agent}\texttt{(X2)}_{goal}\texttt{(X3)}_{source}$

The system has pre-stored knowledge of a relation between `complaint` and `not_received`; the user's intention is thus deduced to be part of a "wrong order" or "faulty goods" context, in which the writer and addressee are assumed to be related. This information is presented to the user in template form as a translation candidate. Such templates may be further embedded in additional templates, which are then used as the basis of novel textual expressions.

Somers [203] summarizes this work as follows:

> "This is an interesting development away from the current situation where the MT system makes the best of what it is given (and cannot really be sure whether or not its translation is good) towards a situation where quality can be assured by the fact that the system knows what it can do and will steer the user to the safe ground within those limitations." (p. 241)

DBMT has been used primarily for small domains such as the case given above. For a larger domain, the required amount of information would be too large to store and search. Similar observations have been made for the KBMT and EBMT paradigms described above.

5.2.4 Neural Network Based MT (NBMT)

Experiments have been done with neural network technology for such MT functions as parsing [106], lexical disambiguation [94], and learning of grammar rules. The incorporation of neural networks and connectionist approaches into MT systems is a relatively new area of investigation [102]. Most recently, Castaño et al. [157] have run some tests with very small vocabularies (about 30 words) and simple syntax. Handling large vocabularies and grammars dramatically inflates the size of the neural networks and the training set, as well as the training time. In addition, dealing with word sequences requires an explicit representations of time, further complicating the neural network representation. McLean [140] uses neural nets to find similar sentences in an EBMT system. But again, a small vocabulary (30 words) and short sentences are used. It is not clear that this approach can be extended to existing EBMT systems. In contrast with the other approaches described in this paper, no realistic MT Systems have been built based solely on neural network technology. This technology is thus more of a technique than a system approach.

5.3 Hybrid Paradigms

In the previous section on non-linguistic-based paradigms it was mentioned that many of those paradigms had difficulty with some aspects of the MT process. For example: SBMT does not handle long range contextual dependencies and EBMT has difficulties with complex sentence structure. It was quickly recognized that these non-linguistic paradigms could be combined with linguistic paradigms to exploit the strengths of each [6], [44],[84], [90], [120], [129], [153], [220], [229]. The hybrid paradigm involves a mixing of MT paradigms (as well as mixing of the MT architectures). The usual approach is to use linguistic methods to obtain parses of the source text and to use statistical or example techniques to resolve dependencies and phrasal translations [84]. Statistical trigram target language models have been used for lexical selection [42]. Statistically generated decision trees have been used to insert English articles into article free translations of Japanese text [121]. The Pangloss system [155] is a hybrid of both MT paradigms and MT architectures.

6 Evaluation of MT Systems

The evaluation of MT systems is also an active area of research. Once an MT system or portion of an MT system is built, how does one evaluate whether it is working correctly and whether it is a promising approach with which to continue? As noted in Hutchins [100], it is clear that fully automatic high quality translation is no longer the current goal of most MT experts. In fact, it is expected that revision is required for all translations, whether done by humans or computers. Thus, in order to decide what the evaluation criteria for a machine translation system should be, we must first determine what the intended use of the MT output will be, and then decide whether the output is satisfactory for this purpose. Hutchins argues that "There can be valid uses of poor quality output in unedited form if it is produced quickly, cheaply, and is not intended for publication. If better quality is required then collaboration of man and machine is essential" [100, p. 303].

Given that "perfect" translation is not within our grasp now, if ever, we still need to decide how we can judge whether the output is high quality or low quality. Hutchins claims that the concept of good quality MT output is an elusive concept. As observed by Van Slype [216], it is difficult to find an objective measure of any type of translation, machine-aided or otherwise. (In fact, there is no quality control metric for human translators.)

In this section, we will first briefly discuss why the evaluation of translations is so elusive and then describe current solutions to evaluation of translations and MT systems. We do this by first outlining the various approaches that can be taken for defining evaluation criteria and then the techniques that can be applied within these approaches.

6.1 Evaluation Challenges

NL applications, such as MT, have some unique problems that must be accounted for when doing evaluations. The biggest problem with evaluating NL applications is minimizing the subjectivity that, to date, has proven unavoidable due to the nature of natural language itself. Standard software evaluation techniques must be enhanced to allow for the multiple "correct" answers that frequently occur with natural language. It is not clear what constitutes a correct answer especially when dealing with translations. It is because of this that judging the correctness of the output for MT still retains a degree of subjectivity.

As pointed out in [15] there are no neighboring disciplines to which we can look for criteria and techniques. There is no general, well-developed methodology for evaluating software systems but as we will see in the next section there are some evaluation criteria that generally apply to software systems. Besides the lack of a general evaluation methodology, there are no clear measures for human translations to guide us and for that matter it is questionable whether MT systems should even be attempting to simulate the behavior of human translators. According to Krauwer [127], the human translator metaphor is questionable because, while the output quality may improve for a short time, it most likely will hit a point of little or no improvement given our current technology. He further suggests that it is better for designers and users to negotiate the specifications for specialized systems. The evaluation can then be based on the specifications. Admittedly, it is still not an easy task to come up with the specifications but it would enable better evaluations.

In keeping with the idea of writing specifications for MT systems, we must keep in mind that we need to produce an output that suffices for the intended use (most desirably this would be according to some specification), and we must do this cost-effectively.

6.2 Evaluation Approaches

The approach one takes when evaluating software systems (in general) is two-fold: (1) evaluation of the accuracy of the input/output pairs; and (2) evaluation of the architecture of the system and the data flow between the system components. The former (external) view of software evaluation is called "black-box" evaluation, and the latter (internal) view is referred to as "glass-box" evaluation [168]. Black-box evaluation covers engineering issues such as reliability, productivity, user learnability and user friendliness. Glass-box evaluation also considers reliability (at the component-level) as well as maintainability, improvability, extendibility, compatibility and portability.

Black-box evaluation, in the case of MT, tends to focus on evaluating the translation-quality of the output. Essentially it is an attempt to measure the acceptability of the translation to users. To produce the most objective measure possible, a standard test-suite of input/output pairs should be established for judging whether the system is performing "correctly" or not and whether it will be cost effective. In light of the above discussion, this is a very costly undertaking and has yet to be satisfactorily accomplished in any evaluation of an MT system.

Another difficulty in applying a black-box evaluation approach is the number of dimensions along which MT developers must limit their systems. These systems can be thought of as shells that are customized to apply to a particular domain, language pair, and type of text. The evaluation criteria (i.e. how well does it translate these texts) must also be limited along the same dimensions, but there is no common range among the systems. Because of this lack of commonality, some systems will need to be customized for the chosen ranges in order to do comparative evaluations. Comparative evaluations would be the goal for users looking to purchase an MT system. Researchers are also interested in comparative evaluations to determine the effectiveness of their MT paradigm or micro-theory. However, the most useful information in this case tends to result from glass-box approaches to evaluation.

The glass-box approach attempts to evaluate the system's internal processing strategies to measure how well the system does something. According to the ideas for evaluating NLP systems [168], this type of evaluation should include a determination of the system's linguistic coverage, and an examination of the linguistic theories used to handle the linguistic phenomena. Determining the linguistic coverage means testing what linguistic phenomena are handled and to what degree. The examination of the linguistic theories used includes how closely these theories were followed in the implementation and noting what modifications had to be made to the theories. In addition, the performance of the system's various modules must be examined and the evaluation of each of these modules should be treated as individual black-box evaluations. Under the glass-box evaluation approach, techniques for measuring improvability have received the most attention.

Considering these basic evaluation approaches, what then are reasonable and useful evaluation criteria for MT systems? There are a number of dimensions along which one can make a judgement of the quality of MT output. Carbonell et al. [46] enumerate the following external evaluation criteria:

1. Semantic Invariance: Is the "meaning" of the source text preserved in the target text?
2. Pragmatic Invariance: Is the implicit intent or illocutionary force (e.g., politeness, urgency, etc.) of the source text preserved in the target text?
3. Structural invariance: Is the syntactic structure of the source text preserved in the target text?
4. Lexical invariance: is there a one-to-one mapping of words or phrases from source to target texts?
5. Spatial invariance: are the external characteristics of the source text, such as length, location on page, etc. preserved in the target text?

Semantic invariance is today a more dominant criterion (in contrast to the early days when MT systems primarily sought to preserve lexical invariance). In general, MT systems currently seek to preserve meaning and style.[12]

Other researchers argue that, in order to determine which criteria are important in evaluating an MT system, we must first know what type of text we are translating. In [146], the criteria for evaluation are determined on the basis of a classification of the different types of text that are to undergo transformation into a foreign language. For example, if we are translating poetry, we would want to preserve pragmatic invariance, whereas if we are translating technical and scientific material, we would want to preserve semantic, lexical, and possibly spatial invariance.

If translation is to be confined to technical and scientific matter, then the text is generally from very narrowly defined fields that restrict the lexicon and grammar and constitute a sublanguage. In this case full "understanding" is less likely to be a necessity since the set of constructs is bounded and the vocabulary is limited; thus, a small set of simple mappings may be used.

On the other hand, translating free-text is a much harder problem than that of translating texts that are restricted to a particular sublanguage. In order to make an evaluation of a system that is intended to translate free-text, we need to look at the degree to which a machine translator might make mistakes if we are lenient with our "understanding" requirement. We can then decide if it is possible to get around these mistakes without adding a high degree of "understanding."

Van Slype [216] (quoted from Hutchins) offers additional evaluation criteria for black-box approaches to evaluation and has identified a number of metrics for evaluating the degree of success of an MT system:

---

[12] Preserving style is a questionable practice considering the findings of work in multilingual generation [170].

1. Intelligibility of output text, e.g., via readability scales.
2. Fidelity to the SL original, e.g., via measures of information transfer.
3. Acceptability to recipient of translation.
4. Time spent in revision (post-editing).
5. Number of errors corrected, and type.

A paper by Slocum and Justus [199] addresses some of the engineering measures described under the black-box evaluation approach as well as some of the measures described under the glass-box evaluation approach in addition to the usual focus on improvability. The criteria derived from this paper are:

1. Cross-linguistic applicability: the MT system must support several human languages. This means the system must be easily extensible. In particular, adding coverage for a new language should be facilitated.
2. Performance: the MT system must support implementation on a parallel architecture, or perform decently on non-parallel machines.
3. Eased acquisition: the MT system must be built on top of syntactic, semantic, and lexical information sources that are easily updated, perhaps automatically.
4. Uniform analysis and synthesis: the MT system should have rules that are used during both types of processing.
5. Fault-tolerant, fail-soft: the MT system should have adequate error recovery and it should be able to provide an understandable explanation for failures (e.g., misspelled word, unanalyzable syntax, etc.).
6. Suitability for Speech Input/Output: the MT system should provide support for speech processing (e.g., it should provide for the possibility that word boundaries are often ignored in speech).

Additional evaluation criteria provided in a paper by King [117] that also fall under the black-box and glass-box approaches to evaluation are:

1. Practicality: the MT system must have fall-back mechanisms. The interface structure must include information on the valency boundedness of constituents, on their surface syntactic function, etc. so that when no semantic interpretation is available, the system can provide some translation rather than none at all. (In the worst case, the translation would be word-for-word.)
2. Collaboration: the MT system should be built by means of joint teams that define and construct the sharable components (e.g. the interlingua or the transfer rule language). There must be an agreement to use a common basic software, that manipulates an agreed upon data structure.

3. Extensibility: the MT system must provide the ability to add new language pairs at any time without having to re-write the pre-existing system.

Summarizing all the criteria given above into a final list is difficult since the criteria need to be further adapted to the particular type of text that is being translated. This comment notwithstanding, we consider the following criteria to be crucial in the evaluation of MT systems:

1. Intelligibility – must be readable and reasonably "natural."
2. Fidelity – must preserve certain characteristics of the source text (e.g., must support structural invariance).
3. Acceptability – must be satisfactory for intended purpose (*e.g.*, must conform to properties of relevant sublanguage).
4. Speed – must have reasonable run-time.
5. Cost – must be cost-efficient.
6. Time spent for revision – must require as little post-editing as possible.
7. Number of errors – must not have an unreasonably large number of errors (e.g., every other sentence on the average).
8. Cross-linguistic applicability – must support several languages in a uniform fashion.
9. Extensibility – must provide ability to easily add new languages.
10. Uniform analysis and synthesis – must use same data structures for both parsing and generation.
11. Fault-Tolerance – must handle errors gracefully, and must provide some translation rather than none at all.
12. Collaboration – all languages must operate on basis of common software and data structures.

Some of these evaluation criteria require further definition depending on the intended purpose of the MT system. For example, what is "natural" in one domain may not be "natural" in another domain. In addition, various measures must be specified: "reasonable run-time" might be different for on-line processing vs. off-line processing; "cost-efficient" might mean one thing to one end user and something else to another; and "unreasonably large number of errors" might mean every other word in one domain and every other sentence in another domain. Also, the "graceful" handing of errors depends on what purpose the system serves (e.g., whether the system is intended to operate interactively as in a tutorial situation, or whether the system is intended to operate as a batch job).

In conjunction with the intended use of an MT system is the notion of the different purposes behind doing an evaluation. An end-user of an MT system will approach evaluation differently from a developer or a researcher. Not only will the most appropriate techniques for these different types of evaluators differ but so will the goals they have for doing an evaluation. Researchers typically work with test-suites since usually they are focusing on one aspect of the whole problem of translation at one particular time (e.g. a theory about translation, an architecture, or a technique for handling a difficult phenomena within a particular theoretical framework). A developer will use test-suites to ensure modifications have not effected sentences that were previously correctly translated (regression testing) as well as whether the targeted sentences that motivated the modification are now correctly handled by the change. An end-user will use test-suites to comparatively evaluate MT systems when considering a purchase and will also use test-suites after acquiring a system and arranging for system extension in lexical or grammatical coverage.

A final, frequently overlooked point is that the MT paradigm has a significant impact on the choice of evaluation criteria [15]. Today's statistical-based and example-based paradigms should not be expected to rate as well on fidelity, intelligibility and acceptability, for example, as the linguistics-based paradigms. On the other hand, we would expect the linguistics-based paradigms to be less fault-tolerant. This idea meshes well with the intended use of an MT system. A statistical-based approach would be expected to provide rough translations more cost-effectively.

6.3 Evaluation Techniques

Test-suites are often proposed as a way to determine a system's linguistic coverage and can be useful for both black-box and glass-box approaches to evaluation. When one is more interested in the types of errors produced by a system than the total number of errors, test-suites are most often the technique used.

To construct a test-suite one must attempt to predict the linguistic constructions and legal combinations of these constructions that will be encountered in the input. In addition, it is important to include illegal constructions as well since an inability to recognize the construction as illegal can result in poor quality output as well. So a test suite could contain sentences with different verb forms and auxiliaries or various complex sentence structures such as sentences with restrictive or non-restrictive relative clauses, or conjoined clauses.

However, determining the appropriate constructions to include in a test-suite is difficult and the size of the test-suite grows quickly. To bound the problem, the test-suite developers must know what linguistic phenomena are of greatest importance to the users and be well-versed in linguistics and the languages of interest [119].

Test suites have also been proposed by King and Falkedal [119] as a way to test the improvability of an MT system. Improvability tests assume that either the evaluator is working closely with the developer or that the evaluator is able to modify the system. The caveats mentioned earlier on bounding the problem, apply here as well.

The simplest use of a test suite is to run the system on it and record the successes and failures. This then gives developers and perhaps potential end-users an idea of what constructions are problematic as well an idea of the overall progress being made in the development of the MT system. However, unless a clear record is made of what constructions and interactions the input is intended to test, one can only get an indication of the overall progress in development. The developer will then have to spend time examining each failure and determining exactly what went wrong in the system.

Some problems with test suite construction as noted in Arnold et. al. [15] are:

1. The projection assumption: the assumption is that it is possible to determine the behavior of the system on the real input from the behavior on the test suite. The test suite may not include all of the phenomena encountered in a real input.

2. Weighting of phenomena: a test suite does not indicate the weighting of the phenomena according to what one would expect to encounter in the real input. So the inability to handle a large number of low frequency phenomena will lead one to expect a worse performance than if there is a problem with one high frequency phenomena.

3. It is necessary to take source and target languages into account. For example, "John went into the house" would test past tense and location prepositional phrases whereas "John entered the house" would test past tense as well as structural divergences in the case of Spanish.

A final technique used in evaluation is to collect the output of the system and evaluate it. While it may be difficult to get representative test material, this tends to be the most reasonable means end-users have today for evaluating MT systems. One approach for evaluating the output is to mark and categorize the errors. Another approach is to rate the output according to intelligibility and fidelity scales. In both cases this evaluation is done by hand and tends to be expensive, tedious, error-prone and subjective. Marking and categorizing errors requires that a category of errors be defined beforehand and a score associated with each. The weighting of the errors tends to be subjective unless something such as the frequency of the construction in the real input is the basis of the scoring. Likewise, the rating of intelligibility and fidelity requires a rating scheme and will be subjective. In addition, this type of rating requires many test inputs and evaluators to get a statistically significant result.[13]

---

[13] It might be useful to consider a test such as the Kappa coefficient of agreement [47] as used in discourse annotation research for assessing inter-coder reliability. In this case, one would be testing inter-evaluator reliability. By establishing such a measure, the evaluators would learn what improvements are needed in the evaluation scheme to get consistency

Test-suite evaluations can also be time-consuming, tedious and error prone. A number of tools are being researched to help with test-suite evaluation. Shiwen [195] has a tool for scoring test-suite results. This tool provides a language to associate input strings with patterns representing acceptable outputs and scores. Arnold et al. [12] describe a tool for test-suite construction which uses a simple grammar to generate a test-suite. The tool described by Nerbonne et al. [149] records in a relational database the phenomena tested by a particular construction so that a test-suite can be built from this database by indicating the grammatical constructions that need to be tested.

6.4 Examples of Machine Translation Evaluations

We have examined the task of MT system evaluation from three different perspectives: (1) End-user evaluation; (2) Developer's evaluation; and (3) Researcher's evaluation. Each evaluation perspective is described in turn.

6.4.1 End-User Evaluation

We will look at two examples of end-user evaluation to illustrate the different depths of evaluation a potential customer of an MT product might undergo. The first example is the evaluation methodology of the Union Bank of Switzerland (UBS) [2] [3] which assessed MT products on behalf of end-users with well-defined requirements. The second is an evaluation methodology of our Mitre study, conducted on behalf of a wide-range of end users in the US government with disparate requirements [111]. In the case of the UBS evaluation, the well-defined requirements pre-selected a smaller candidate pool of MT products, making an in-depth evaluation more feasible. In the case of the Mitre evaluation, the end-users' requirements were open-ended in that they were interested in all language pairs and all domains of interest. The result was that no MT system could be excluded from the evaluation; thus, the Mitre evaluators undertook a higher-level evaluation to narrow down the possible candidates.

Both evaluations considered operational and economic issues. For example, both examined the costs associated with adapting the system to the current work environment and the types of text that needed to be translated. In addition, both took into account the quality of support provided by the MT system. While the UBS evaluation answered these questions by bringing the systems in-house and adapting them enough to do testing, the Mitre evaluation attempted to determine this by means of a survey questionnaire. The difference in the approaches was a matter of time and cost.

Another area considered by both studies was the evaluation of translation output. Both used representative input texts, instead of collections of unrelated sentences, that were customized to test particular features of

---

among the evaluators. Once reliability is achieved, one evaluator is all that should be necessary. However, establishing an interpretation of the Kappa measure is not without its problems [55].

the system. Both evaluations acknowledged the need to test the system on what it had been customized to do. In the UBS evaluation, the system was customized to handle the input texts. In the Mitre evaluation, test suites were selected for each system; these were tailored for the domain, coverage, and end-user needs for each system.

The studies were distinguished by the metrics used for assessment of the quality of MT output. The UBS study categorized the complexity of the input test sentences on a scale of 1 to 4 and scored the mistakes in the raw translations based on: (1) whether the error was easily corrected; (2) the degree to which the error interfered with understanding; and (3) whether the error violated basic grammatical structure. In addition, dictionary errors and failures to translate were counted separately. All the criteria (linguistic capabilities, technical environment, and organizational changes) were weighted according to user requirements; a final rating was designed for each system evaluated.

The Mitre study evaluated output quality by comparing it to human-produced translations; multiple evaluators were used for each test case. To lessen the influence of possible exposure to the "good" translation in rating the meaning invariance, the evaluators first paraphrased the meanings before comparing the MT output to the human produced translation.[14] After paraphrasing, the evaluators completed the evaluation by comparing it to the human produced translation. With the widely ranging end-user requirements (from rough translations to publication-quality translations), categorization of the types of errors was impractical. Thus, the criteria results were listed separately so that each end-user group could weight the criteria according to their specific needs.

End-users also have a role to play in the evaluation of an MT system once it is purchased and customized by the developer. Although the UBS evaluation touches upon this situation to a small extent—since they customized the system before their selection evaluations—the example in the next section discusses this from the perspective of a longer-term customization process.

6.4.2 Developer's Evaluation

Typically the end-user and the developer of an MT system must develop an on-going relationship in order to customize the system. During this long-term customization process, techniques for evaluating the progress of the customization must be established. Moreover, the integrity of the system must not being compromised by incremental modifications, e.g., during regression testing. One example that addresses such considerations is the evaluation procedure used by B'VITAL and SITE for their CAT system [177]. Although the procedure was not fully deployed to the extent planned at the time of this writing (i.e., only 100 pages of the 5000 intended pages were

---

[14] The rationale was that if the "answer" were present it would be easier to see the MT system's output as being closer in meaning to the human output, i.e., there might be a false sense of understanding even when something was badly translated.

tested), it serves as a useful example since it is representative of what many developers reportedly do according to the initial Mitre survey.

B'VITAL's procedure involves both the end-user who, as post-editor, reports output problems and the developer who, as linguist and lexicographer, modifies the grammar and the dictionary. First a balanced corpus of pages from technical manuals is selected and run through CAT. The post-editor revises the system output and records the time required to do so. This gives a measure of efficiency of the system. Presumably as the customization progresses, the efficiency improves. For a random sampling of the test corpus, the post-editor also fills out a detailed problem report, categorizing the types of problems encountered by counting surface form errors. In addition, the post-editor assesses the clarity of the incorrectly translated sentences and gives the developer the context needed to reproduce the problems (i.e., input, MT output, and the expected output).

Next, the developer reviews each problem and either corrects the problem or categorizes the reasons it could not be corrected (e.g., a phenomenon that is rare, low priority, or too complex). Categorizing the reasons for which problems are left uncorrected provides feedback to the end-user on the limits of the system, frequently resulting in a refinement of user requirements. For the problems that are corrected, the time required for the repair is recorded along with the type of modification needed (e.g., dictionary or grammar modification). Keeping track of the repair time gives an indication of how difficult the system is to customize.

At the end of the procedure, once the entire test corpus has been translated and corrections implemented, a subset of the corpus is again tested as a higher level form of regression testing.

Developers also undertake a type of evaluation that does not directly involve the end-user when the system is still being built or when some aspect of it is changed. This evaluation typically relies on test suites and overlaps to a large extent with the researchers' evaluation, the subject of the next section.

6.4.3 Researcher's Evaluation

MT evaluation in the research community focuses on the handling of different linguistic phenomena by various MT systems. Mitre's later evaluation work [173], [235] is a current example of such an evaluation. This work is one of several contributions in the 1998 AMTA Workshop on Interlinguas [7] which examined the translation of a text from the UNESCO Courier [167] into multiple languages. The objective of the workshop was to determine how different aspects of the text would be represented using an Interlingual Representation.[15] Specific challenges addressed by the group were the handling of basic predicate-argument structure, noun phrases/referents, proper nouns, prepositional meaning, non-literal language, temporal relations, tex-

---

[15] This text was available in 13 languages. Our examples focus on the English and Spanish versions of the text.

tual organization, lexical divergences, syntactic divergences.

The more general issue examined in the evaluation work of [235] was whether an interlingua would improve the system output over that of a transfer design. Consider the following sample bilingual text examined in the workshop:

(38) (i) E: ACCION International is a U.S.-based private non-profit organization that currently provides technical assistance to a network of institutions in thirteen countries in Latin America and six cities in the United States. Its network of eighteen independent organizations in Latin America has lent over \$1 billion to microenterprises in the last five years, in loans averaging less than \$500.

(ii) S: ACCIÓN Internacional es un organismo privado sin fines de lucro con sede en Estados Unidos, que brinda actualmente asistencia técnica a una red de establecimientos de microcrédito en trece países de América Latina y seis grandes ciudades de Estados Unidos. En América Latina la red abarca dieciocho organizaciones independientes que desembolsaron mil millones de dólares en los últimos cinco años en forma de préstamos de una cuantía media inferior a quinientos dólares.

Two transfer-based commercial MT systems were used to translate the English excerpts above to Spanish: (1) Logos [190], [191], [192]; and (2) Systran [86], [188], [189]. The following results were produced:[16]

(39) (i) Logos: ACCION—organización privada basada no beneficio que proporciona en la actualidad la asistencia técnica a una red de instituciones en trece países en la América Latina y seis ciudades en los Estados Unidos. Su red de dieciocho organizaciones independientes en la América Latina ha prestado más de 1 mil millones de dólares a microenterprises en los cinco últimos años, en préstamos que promedian menos de 500 dólares.

(ii) Systran: ACCION internacional es una organización no lucrativa privada de U.S.-based que proporciona actualmente a asistencia técnica a una red de instituciones en trece países en América latina y seis ciudades en los Estados Unidos. Su red de dieciocho organizaciones independientes en América latina ha prestado concluío \$1 mil millones a los microenterprises en los cinco años pasados, en los préstamos que hacían un promedio menos de de \$500.

Overall, the translations into Spanish were readable by a native speaker, but not entirely grammatical and not at all similar to the version produced

---

[16] In [235], Globalink [98], [132], [174] was used in place of Logos, and the translation direction was Spanish to English. Despite this difference, the findings in that work were similar to those described here.

by the human speaker. For example, the phrase *provides technical assistance to a network of institutions* is translated as *proporciona en la actualidad la asistencia técnica a una red de instituciones* ('supplies at the present time the technical assistance to a web of institutions') by Logos and *proporciona actualmente a asistencia técnica a una red de instituciones* ('supplies presently to the technical assistance to a web of institutions') by Systran. Neither of these corresponds exactly to the human-generated Spanish version *brinda actualmente asistencia técnica a una red de establecimientos* ('offers currently technical assistence to a web of establishments'). In particular, the mapping between *provides* and *brindar* appears to be missing in these two systems.

This hypothesis was further tested by translating the Spanish version of the clause, *que brinda actualmente asistencia técnica a una red*, and verifying that *provides* is not the main verb in the target-language sentence. The test was done with Systran and Globalink only since Logos does not translate in the Spanish-to-English direction:

(40) (i) Systran: ... that offers currently technical assistance to a network
(ii) Globalink: ... that at the moment offers to technical attendance to a network

From the translations above, the internal mappings are deduced to be of the following form:

(41) **Transfer Mapping for Systran:**
brindar(Y) → offer(Y)

(42) **Transfer Mapping for Globalink:**
brindar(Y) → offer(to(Y))

From these we see three things:

- The verb *provide* is not an English equivalent of *brindar* in either system.[17]
- The phrase *to a network* is viewed as a modifier that is taken to be outside of the predicate-argument structure of either verb.
- In the Globalink translation, the direct object of the verb is taken (erroneously) as the beneficiary of the main verb as in 'toast to the bride', which forces two prepositional phrases to be generated.[18]

---

[17] Interestingly, Spanish-speaking informants at the AMTA commented that the source text is, in fact, awkward—that the verb *brindar* is used in a less conventional context. Perhaps this evaluation acts as somewhat of a diagnostic for acceptability of the human-generated input. We set aside this possibility for the purpose of this discussion, but see related work on DBMT (above) in which the user's input is iteratively revised through interaction with the translation system.

[18] Spanish-speaking informants indicated that *brindar* translates most typically to the verb *toast*.

Given that an interlingual representation is designed to accommodate complex predicate-argument structure mappings, these examples provide evidence that an interlingual approach would improve the system output over that of a transfer design. Even a simplified form of interlingua, such as one based on thematic roles [69], provides the basis of a more systematic mapping into the target-language.[19] For example, the following thematic-role representations are used for the English and Spanish verbs used above:

(43) (i) provide: [`ag th goal(to)`]
(ii) offer: [`ag th ben(to)`]
(iii) brindar: [`ag ben`, `ag th goal(a)`]

A thematic-role analysis of the English sentence containing *provide* would not result in a beneficiary reading of *technical assistance* because, even if *brindar* were chosen in the target-language, the thematic specification forces the direct object into *theme* (`th`) position; thus reserving the remaining argument for *goal* (`goal`) position. Unlike the transfer case, this alignment of thematic information allows *to a network* to be analyzed as an argument rather than as a modifier and the verb *brindar* is selected as a possible translation candidate. Similarly, a thematic-role analysis of the Spanish sentence containing *brindar* would allow either *offer* or *provide* to be selected, with a preference for the latter because of the alignment of the three roles. In either case, this analysis differs from that of the transfer approach: *technical assistance* is not associated with a spurious preposition because the only argument that allows this preposition has been "saturated" by the phrase *a network*.

This evaluation has demonstrated the potential benefit of adopting interlingual over transfer. However, one must take care to recognize the potential risks when making such an assessment. As described in [96], a deeper analysis runs the risk of prohibitive storage and access costs as systems are scaled up for broader coverage. On the other hand, lexical-transfer systems provide lower-quality translations and require a great deal of human assistance. A more recent trend is to prioritize the development of systems that exploit the strengths of different approaches or that vary in functionality according to the task at hand. This trend is associated with increasingly higher recognition of the importance of establishing the requirements of the consumer of the translation output [204], [217]. Evaluations metrics are, correspondingly, tuned to the objectives of the final application [72], [162].

7 Summary and Conclusions

In this survey we have presented the major challenges for MT in the areas of linguistics, operations, system and translation evaluation, acquisition of linguistic and domain knowledge, and the utilization of the empirically

---

[19] Thematic roles in this system are taken from a standardized set, e.g., Agent (ag), Theme (th), Source (src), Goal (goal), Location (loc), Beneficiary (ben) [103].

available linguistic and domain knowledge in bilingual, and source and target language corpora.

To summarize the linguistic challenges: MT research must deal with the unanswered questions of how much understanding is enough to deal with the various ambiguities that arise in natural language, whether an assumption of pragmatic invariance will be enough to generate a useful text, and how one can best deal with the various linguistic divergences between languages. Building general core MT systems and tools that can be easily extended to other languages is another of the major challenges. In conjunction with this is the issue of how to acquire and maintain all of the information that will adapt a general core MT system to particular language pairs and translation needs. This in turn influences the basic approach of the general core MT systems. As for the system and translation evaluation challenges, MT research must find ways to evaluate the efficiency, maintainability, portability, and quality of translation output relative to user needs. We expect to see more research on comparative evaluation approaches as well as on the evaluation of *working* prototypes in actual translation settings.

Since large amounts of knowledge are necessary (but not sufficient) for better quality results and broader coverage, this has led to research on methods for acquiring and utilizing information more easily. The traditional rationalist approach is to try for high quality translations first and then to find ways to broaden the coverage while the empiricist approach is to try for broad coverage first and then upgrade the quality of the result. One predicts then that hybrid approaches might provide the next step forward for advances in machine translation research. Indeed, some research is moving in this direction, and current trends indicate a heightened awareness of the utility of multi-engine approaches.

Over the next several years, advances in efficiency and portability are expected as researchers discover more about cross-linguistic generalizations in language processing and develop efficient techniques for utilization of existing knowledge resources. Moreover, synergism between systems and humans is expected as technologies are combined to enhance MT throughput and/or quality. Finally, as the need for international access to multi-lingual text continues to grow, advances in MT technology will become increasingly more important and, correspondingly, the utility of MT paradigms will be measured in terms of their impact with respect to a wide range of cross-language applications.

Acknowledgements

The first author has been supported, in part, by DOD Contract MDA904-96-C-1250, DARPA/ITO Contract N66001-97-C-8540, Army Research Laboratory contract DAAL01-97-C-0042, NSF PFF IRI-9629108 and Logos Corporation, NSF CNRS INT-9314583, and Alfred P. Sloan Research Fellowship Award BR3336. The second author has been supported, in part, by the Mitre Corp, the Center for Machine Translation at Carnegie Mellon

University, Mitsubishi Electric Research Laboratory, the National Science Foundation under Grant No. IRI-9314961, an Andrew Mellon Predoctoral Fellowship and the Intelligent Systems Program at University of Pittsburgh. Useful guidance and commentary have been provided by several members of the Laboratories for Computational Linguistics and Information Processing (CLIP) and Language and Media Processing (LAMP) at the University of Maryland. Many thanks especially to Doug Oard, Mari Olsen, Clare Voss, Amy Weinberg, and David Zajic. We are forever indebted to Edna Walker at the University of Maryland, who spent countless hours assisting us in the compilation of the bibliography.

References

[1] A. Abeillé, Y. Schabes, and A.K. Joshi. Using Lexicalized Tags for Machine Translation. In *Proceedings of Thirteenth International Conference on Computational Linguistics*, pages 1–6, Helsinki, Finland, 1990.

[2] Doris Albisser. Evaluation of MT Systems at Union Bank of Switzerland. In Kirsten Falkedal, editor, *Proceedings of the Evaluators' Forum*, pages 51–53, Les Rasses, Vaud, Switzerland, April 21-24 1991. ISSCO, Universite' de Gene've, Geneva, Switzerland.

[3] Doris Albisser. Evaluation of MT Systems at Union Bank of Switzerland. *Machine Translation*, 8(1–2):25–27, 1993.

[4] Jeff Allen. Lexical Variation in Haitian Creole and Orthographic Issues for MT and OCR applications. In *Proceedings of Embedded MT Workshop, AMTA-98*, Langhorne, PA, October 1998.

[5] Automatic Language Processing Advisory Committee (ALPAC); Language and Machines: Computers in Translation and Linguistics, 1966.

[6] Hiyan Alshawi, Adam L. Buchsbaum, and Fei Xia. A Comparison of Head Transducers and Transfer for a Limited Domain Translation Application. In *Proceedings of the 35th Annual Meeting of the Association for Computational Linguistics (ACL-97)*, pages 360–365, Madrid, Spain, July 7-12 1997.

[7] AMTA Special Interest Group for Interlinguas. *Proceedings of the SIG-IL Second Workshop on Interlinguas, AMTA-98, October 28-31, Langhorne, PA, New Mexico State University Technical Report MCCS-98-316.* Computing Research Laboratory, New Mexico State University, Las Cruces, New Mexico, 1998.

[8] Margherita Antona and Jun-ichi Tsujii. Treatment of Tense and Aspect in Translation from Italian to Greek – An Example of Treatment

of Implicit Information in Knowledge-based Transfer MT. In *Proceedings of the Fifth International Conference on Theoretical and Methodological Issues in Machine Translation of Natural Languages*, pages 131–153, 1993.

[9] L. Appelo. A Compositional Approach to the Translation of Temporal Expressions in the Rosetta System. In *Proceedings of Eleventh International Conference on Computational Linguistics*, pages 313–318, Bonn, Germany, 1986.

[10] D. Arnold and L. des Tombe. Basic Theory and Methodology in Eurotra. In S. Nirenburg, editor, *Machine Translation: Theoretical and Methodological Issues*, pages 114–135. Cambridge University Press, Cambridge, England, 1987.

[11] D. Arnold, S. Krauwer, L. des Tombe, and Louisa Sadler. Relaxed Compositionality in Machine Translation. In *Proceedings of the Second International Conference on Theoretical and Methodological Issues in Machine Translation of Natural Languages*, Carnegie Mellon University, Pittsburgh, PA, 1988.

[12] D. Arnold, D. Moffat, L. Sadler, and A. Way. Automatic Test Suite Generation. *Machine Translation*, 8(1–2):29–38, 1993.

[13] D. Arnold and L. Sadler. Theoretical Basis of MiMo. *Machine Translation*, 5(3):195–222, 1990.

[14] D. Arnold and L. Sadler. Rationalism and the Treatment of Referential Dependencies. In *Proceedings of the Second International Conference on Theoretical and Methodological Issues in Machine Translation of Natural Languages*, pages 195–204, Montreal, Canada, 1992.

[15] D. Arnold, L. Sadler, and R.L. Humphreys. Evaluation: An Assessment. *Machine Translation*, 8(1–2):1–24, 1993.

[16] ATLAS II: A Machine Translation System for Multilingual Translation/ Japanese to German. Sales document from FUJITSU, 1986. **50**.

[17] B. Ballard and D.E. Stumberger. Semantic Acquisition in TELI. In *Proceedings of the 24th Annual Meeting of the Association for Computational Linguistics*, pages 20–29, 1986.

[18] Y. Bar-Hillel. The Present Status of Automatic Translation of Languages. In *Advances in Computers, 1*, pages 91–163. Academic Press, New York, 1960.

[19] J. Barnett, J. Cant, A. Demedts, T. Dietzel, B Gates, E. Hays, Y. Ito, and J. Yamron. LINGSTAT: State of the System. In *ARPA Workshop on Machine Translation*, Vienna, Virginia, November 1994.

[20] M. Bates and R. Bobrow. A transportable natural language interface. In *Proceedings of the Sixth Annual ACM SIG Conference on Research and Development in Information Retrieval*, 1983.

[21] J. Beaven. *Lexicalist Unification-Based Machine Translation.* PhD thesis, University of Edinburgh, Edinburgh, 1992.

[22] J. Beaven. Shake and Bake Machine Translation. In *Proceedings of Fourteenth International Conference on Computational Linguistics*, pages 603–609, Nantes, France, 1992.

[23] P. Bennett and P. Paggio. Preference in Eurotra. In *Studies in Machine Translation and Natural Language Processing, Vol. 3.* Commission of the European Communities, 1993.

[24] P.A. Bennett, R.L. Johnson, J. McNaught, J.M. Pugh, J.C. Sager, and H.L. Somers. *Multilingual Aspects of Information Technology.* Gower, Brookfield, VT, 1986.

[25] W.S. Bennett. How Much Semantics is Necessary for MT Systems? In *Proceedings of the Third International Conference on Theoretical and Methodological Issues in Machine Translation of Natural Languages*, pages 261–269, Linguistics Research Center, The University of Texas, Austin, TX, 1990.

[26] John W. Benoit, Pamela W. Jordan, and Bonnie J. Dorr. Machine Translation Action Plan. Technical Report MP-91W00023, The Mitre Corporation, 1991.

[27] John W. Benoit, Pamela W. Jordan, and Bonnie J. Dorr. Machine Translation Technology Survey and Evaluation Report. Technical Report MP-91W00021, The Mitre Corporation, 1991.

[28] Sabine Bergler. Generative Lexicon Principles for Machine Translation: A Case for Meta-Lexical Structure. *Machine Translation*, 9(3–4):155–182, 1995.

[29] Robert C. Berwick. Principles of Principle-Based Parsing. In R.C. Berwick, S.P. Abney, and C. Tenny, editors, *Principle-Based Parsing: Computation and Psycholinguistics*, pages 1–37. Kluwer Academic Publishers, 1991.

[30] B. Bläaser, U. Schwall, and A. Storrer. A Reusable Lexical Database Tool for Machine Translation. In *Proceedings of the Fourteenth International Conference on Computational Linguistics*, pages 510–516, Nantes, France, 1992.

[31] B. Boguraev and E.J. Briscoe, editors. *Computational lexicography for natural language processing.* Longman, London, 1989.

[32] Branimir Boguraev and James Pustejovsky. Lexical Ambiguity and the Role of Knowledge Representation in Lexical Design. In *Proceedings of Thirteenth International Conference on Computational Linguistics*, pages 36–41, Helsinki, Finland, 1990.

[33] C. Boitet. Speech Synthesis and Dialogue Based Machine Translation. In *ATR Symposium on Basic Research for Telephone Interpretation*, pages 6–5–1–9, Kyoto, Japan, 1989.

[34] Michael Brent. From Grammar to Lexicon: Unsupervised Learning of Lexical Syntax. *Computational Linguistics*, 19(2):243–262, 1993.

[35] C. Brew. Letting the Cat out of the Bag: Generation for Shake-and-Bake MT. In *Proceedings of Fourteenth International Conference on Computational Linguistics*, pages 610–616, Nantes, France, 1992.

[36] E.J. Briscoe and A.A. Copestake. Enjoy the Paper: Lexical Semantics Via Lexicology. In *Proceedings of Thirteenth International Conference on Computational Linguistics*, pages 42–47, Helsinki, Finland, 1990.

[37] P.F. Brown. A Statistical Approach to Machine Translation. *Computational Linguistics*, 16(2):79–85, 1990.

[38] P.F. Brown, J. Cocke, S.A. Della Pietra, V.J. Della Pietra, F. Jelinek, R.L. Mercer, and P.S. Roossin. A Statistical Approach to French/English Translation. In *Proceedings of the Second International Conference on Theoretical and Methodological Issues in Machine Translation of Natural Languages*, Carnegie Mellon University, Pittsburgh, PA, 1988.

[39] P.F. Brown, J. Della-Pietra, F. Jelinek, R.L. Mercer, and P.S. Roossin. A Statistical Approach to Language Translation. In *Proceedings of Twelfth International Conference on Computational Linguistics*, pages 71–76, Budapest, Hungary, 1988.

[40] P.F. Brown, J.C. Lai, and R.L. Mercer. Aligning Sentences in Parallel Corpora. In *Proceedings of the 29th Annual Meeting of the Association for Computational Linguistics*, pages 169–176, University of California, Berkeley, CA, 1991.

[41] P.F. Brown, S.A. Della Pietra, V.J. Della Pietra, J.D. Lafferty, and R.L. Mercer. Analysis, Statistical Transfer, and Synthesis in Machine Translation. In *Proceedings of the Fourth International Conference on Theoretical and Methodological Issues in Machine Translation of Natural Languages*, pages 83–100, Montreal, Canada, 1992.

[42] R. Brown and R. Frederking. Applying Statistical English Language Modeling to Symbolic Machine Translation. In *Proceedings of the Sixth International Conference on Theoretical and Methodological Issues in Machine Translation*, pages 221–239, Leuven, Belgium, 1995.

[43] R.J. Byrd, N. Calzolari, M.S. Chodorow, J.L. Klavans, M.S. Neff, and O.A. Rizk. Tools and Methods for Computational Linguistics. *Computational Linguistics*, 13:219–240, 1987.

[44] J. Carbonell, T. Mitamura, and E. Nyberg. The KANT Perspective: A Critique of Pure Transfer (and Pure Interlingua, Pure Statistics, ...). In *Proceedings of the Fourth International Conference on Theoretical and Methodological Issues in Machine Translation*, pages 225–235, Montreal, Canada, 1992.

[45] J.G. Carbonell. Knowledge-based Machine Translation. In *Tutorial at 25th Annual Meeting of the Association for Computational Linguistics*, Stanford University, Stanford, California, 1987.

[46] J.G. Carbonell, R.E. Cullingford, and A.V. Gershman. Steps Toward Knowledge-Based Machine Translation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI-3(4), 1981.

[47] Jean Carletta. Assessing Agreement on Classification Tasks: The Kappa Statistic. *Computational Lingustics*, 22(2):249–254, June 1996.

[48] J.S Chang and K.Y. Su. A Corpus-Based Statistics-Oriented Transfer and Generation Model for Machine Translation. In *Proceedings of the Fifth International Conference on Theoretical and Methodological Issues in Machine Translation*, pages 3–14, Kyoto, Japan, 1993.

[49] E. Charniak. *Statistical Language Learning.* MIT Press, Cambridge, Massachusetts, 1993.

[50] Noam Chomsky. *Knowledge of Language: Its Nature, Origin and Use.* The MIT Press, Cambridge, MA, 1986.

[51] B. Collins and P. Cunningham. Adaptation Guided Retrieval: Approaching EBMT with Caution. In *Proceedings of the 7th International Conference on Theoretical and Methodological Issues in Machine Translation*, pages 119–1126, Santa Fe, New Mexico, 1997.

[52] Computing Research Laboratory, New Mexico State Univ. *The Interactive Text Translation System User's Manual (ULTRA).* **225**.

[53] Ann Copestake, Ted Briscoe, Piek Vossen, Alicia Ageno, Irene Castellon, Francesc Ribas, German Rigau, Horacio Rodríguez, and Anna Samiotou. Acquisition of Lexical Translation Relations from MRDS. *Machine Translation*, 9(3–4):183–219, 1995.

[54] L. Danlos and P. Samvelian. Translation of the Predicative Element of a Sentence: Category Switching, Aspect, and Diathesis. In *Proceedings of the Fourth International Conference on Theoretical and Methodological Issues in Machine Translation of Natural Languages*, pages 21–34, Montreal, Canada, 1992.

[55] Barbara Di Eugenio, Pamela W. Jordan, Johanna D. Moore, and Richmond H. Thomason. An empirical investigation of proposals in collaborative dialogues. In *ACL/COLING 98, Proceedings of the 36th Annual Meeting of the Association for Computational Linguistics* (joint with the 17th International Conference on Computational Linguistics), Montreal, Canada, 1998.

[56] S. Doi and K. Muraki. Translation Ambiguity Resolution Based on Text Corpora of Source and Target Languages. In *Proceedings of Fourteenth International Conference on Computational Linguistics*, pages 525–531, Nantes, France, 1992.

[57] Michael Dorna and Martin C. Emele. Semantic-Based Transfer". In *Proceedings of the 16th International Conference on Computational Linguistics*, pages 316–321, Copenhagen, Denmark, 1996.

[58] Michael Dorna, Anette Frank, Josef van Genabith, and Martin C. Emele. Syntactic and Semantic Transfer with F-Structures. In *Proceedings of the 36th Annual Meeting of the Association for Computational Linguistics (ACL'98)*, pages 341–347, Montreal, Quebec, Canada, 1998.

[59] Bonnie J. Dorr. Solving Thematic Divergences in Machine Translation. In *Proceedings of ACL-90*, pages 127–134, University of Pittsburgh, Pittsburgh, PA, 1990.

[60] Bonnie J. Dorr. Principle-Based Parsing for Machine Translation. In R.C. Berwick, S.P. Abney, and C. Tenny, editors, *Principle-Based Parsing: Computation and Psycholinguistics*, pages 153–184. Kluwer Academic Publishers, Norwell, MA, 1991.

[61] Bonnie J. Dorr. Parameterization of the Interlingua in Machine Translation. In *Proceedings of Fourteenth International Conference on Computational Linguistics*, pages 624–630, Nantes, France, 1992.

[62] Bonnie J. Dorr. The Use of Lexical Semantics in Interlingual Machine Translation. *Machine Translation*, 7(3):135–193, 1992.

[63] Bonnie J. Dorr. Interlingual Machine Translation: a Parameterized Approach. *Artificial Intelligence*, 63(1&2):429–492, 1993.

[64] Bonnie J. Dorr. *Machine Translation: A View from the Lexicon.* The MIT Press, Cambridge, MA, 1993.

[65] Bonnie J. Dorr. Machine Translation Divergences: A Formal Description and Proposed Solution. *Computational Linguistics*, 20(4):597–633, 1994.

[66] Bonnie J. Dorr. Large-Scale Acquisition of LCS-Based Lexicons for Foreign Language Tutoring. In *Proceedings of the ACL Fifth Conference on Applied Natural Language Processing (ANLP)*, pages 139–146, Washington, DC, 1997.

[67] Bonnie J. Dorr. Large-Scale Dictionary Construction for Foreign Language Tutoring and Interlingual Machine Translation. *Machine Translation*, 12(4):271–322, 1997.

[68] Bonnie J. Dorr, Joseph Garman, and Amy Weinberg. From Syntactic Encodings to Thematic Roles: Building Lexical Entries for Interlingual MT. *Machine Translation*, 9:221–250, 1995.

[69] Bonnie J. Dorr, Nizar Habash, and David Traum. A Thematic Hierarchy for Efficient Generation from Lexical-Conceptal Structure. In *Proceedings of the Third Conference of the Association for Machine Translation in the Americas, AMTA-98, in Lecture Notes in Artificial Intelligence, 1529*, pages 333–343, Langhorne, PA, October 28-31 1998.

[70] Bonnie J. Dorr, Jim Hendler, Scott Blanksteen, and Barrie Migdalof. Use of LCS and Discourse for Intelligent Tutoring: On Beyond Syntax. In Melissa Holland, Jonathan Kaplan, and Michelle Sams, editors, *Intelligent Language Tutors: Balancing Theory and Technology*, pages 289–309. Lawrence Erlbaum Associates, Hillsdale, NJ, 1995.

[71] Bonnie J. Dorr, Dekang Lin, Jye-hoon Lee, and Sungki Suh. Efficient Parsing for Korean and English: A Parameterized Message Passing Approach. *Computational Linguistics*, 21(2):255–263, 1995.

[72] Bonnie J. Dorr and Douglas W. Oard. Evaluating resources for query translation in cross-language information retrieval. In *Proceedings of the First International Conference on Language Resources and Evaluation*, pages 759–763, Granada, Spain, 1998.

[73] Bonnie J. Dorr and Mari Broman Olsen. Deriving Verbal and Compositional Lexical Aspect for NLP Applications. In *Proceedings of the 35th Annual Meeting of the Association for Computational Linguistics (ACL-97)*, pages 151–158, Madrid, Spain, July 7-12 1997.

[74] Bonnie J. Dorr and Clare Voss. Machine Translation of Spatial Expressions: Defining the Relation between an Interlingua and a Knowledge Representation System. In *Proceedings of Twelfth Conference of the American Association for Artificial Intelligence*, pages 374–379, Washington, DC, 1993.

[75] K. Eberle, W. Kasper, and C. Rohrer. Contextual Constraints for Machine Translation. In *Proceedings of the Fourth International Conference on Theoretical and Methodological Issues in Machine Translation of Natural Languages*, pages 213–224, Montreal, Canada, 1992.

[76] Martin C. Emele and Michael Dorna. Ambiguity Preserving Machine Translation Using Packed Representations. In *Proceedings of the 36th Annual Meeting of the Association for Computational Linguistics (ACL'98)*, pages 365–371, 1998.

[77] David Farwell, Louise Guthrie, and Yorick Wilks. The Automatic Creation of Lexical Entries for a Multilingual MT System. In *Proceedings of the Fourteenth International Conference on Computational Linguistics*, pages 532–538, Nantes, France, 1992.

[78] David Farwell, Louise Guthrie, and Yorick Wilks. Using Machine Readable Dictionaries for the Creation of Lexicons. In *Working Notes, Building Lexicons for Machine Translation, AAAI-93 Spring Symposium, Technical Report SS-93-02*, page 106, Stanford, CA, 1993.

[79] David Farwell and Yorick Wilks. Ultra: a Multi-lingual Machine Translator. Technical Report MCCS-90-202, Computing Research Laboratory, New Mexico State Univ, 1990. **227**.

[80] David Farwell and Yorick Wilks. ULTRA: A Multi-Lingual Machine Translator. In *Proceedings of Machine Translation Summit III*, pages 19–24, Washington, DC, July 1991.

[81] F. Fisher and C. R. Voss. FALCON, An MT System Support Tool for Non-linguists. In *Proceedings of the Advanced Information Processing and Analysis Conference*, McLean, VA, 1997.

[82] Robert Frederking and Sergei Nirenburg. Three Heads are Better than One. In *Proceedings of the Fourth Conference on Applied Natural Language Processing (ANLP-94)*, Stuttgart, Germany, 1994.

[83] Y. Fujii, K. Suzuki, F. Maruyama, and T. Dasai. Analysis of Long Sentence In Japanese-English Machine Translation System. In *Proceedings of Information Processing Society of Japan*, 1990.

[84] O. Furuse and H. Iida. An Example-Based Method for Transfer-Driven Machine Translation. In *Proceedings of the Fourth International Conference on Theoretical and Methodological Issues in Machine Translation of Natural Languages*, pages 139–150, Montreal, Canada, 1992.

[85] O. Furuse and H. Iida. Constituent Boundary Parsing for Example-Based Machine Translation. In *COLING-94*, 1994.

[86] Denis A. Gachot. The SYSTRAN Renaissance. In *MT SUMMIT II*, pages 66–71. Deutsche Gesellschaft fuer Dokumentation e.V. (DGD), August 1989. **111**.

[87] W.A. Gale and K.W. Church. A Program for Aligning Sentences in Bilingual Corpora. *Computational Linguistics*, 19:75–102, 1993.

[88] D. Gates and P. Shell. Rule-Based Acquisition and Maintenance of Lexical and Semantic Knowledge. In *Proceedings of Sixth Conference of the European Chapter of the Association for Computational Linguistics*, pages 149–157, Utrecht, Holland, 1993.

[89] J. Ginsparg. A Robust Portable Natural Language Data Base Interface. In *Proceedings of the Conference on Applied Natural Language Processing*, 1983.

[90] R. Grishman and M. Kosaka. Combining Rationalist and Empiricist Approaches to Machine Translation. In *Proceedings of the Fourth International Conference on Theoretical and Methodological Issues in Machine Translation of Natural Languages*, pages 263–274, Montreal, Canada, 1992.

[91] Ralph Grishman. *Computational Linguistics: An Introduction.* ACL Studies in Natural Language Processing. Cambridge University Press, New York, 1986.

[92] Barbara J. Grosz, Douglas E. Appelt, Paul. A. Martin, and Fernando C.N. Pereira. TEAM: An Experiment in the Design of Transportable Natural-Language Interfaces. *Artificial Intelligence*, 32(2):173–243, 1987.

[93] G. Guida and C. Tasso. IR-NL1: An Expert Natural Language Interface to Online Data Bases. In *Proceedings of the Conference on Applied Natural Language Processing*, 1983.

[94] D. Higinbotham. Semantic Cooccurrence Networks. In *Third International Conference on Theoretical and Methodological Issues in Machine Translation of Natural Language*, Austin Texas, 1990. LRC University of Texas at Austin.

[95] Christopher Hogan. Embedded Spelling Correction for OCR with an Application to Minority Languages. In *Proceedings of Embedded MT Workshop, AMTA-98*, Langhorne, PA, October 1998.

[96] Christopher Hogan and Robert E. Frederking. An Evaluation of the Multi-engine MT Architecture. In *Proceedings of the Third Conference of the Association for Machine Translation in the Americas, AMTA-98, in Lecture Notes in Artificial Intelligence, 1529*, pages 113–123, Langhorne, PA, October 28-31 1998.

[97] V. M. Holland and C. Schlesinger. FALCON, Lessons Learned in the Field. In *Proceedings of the Advanced Information Processing and Analysis Conference*, McLean, VA, 1998.

[98] Klaus L. Hopp. Globalink Translation System Equal Goes It Loose...The Problem of Computer-Assisted Translation. *Computerland*, 3(8), 4th Quarter 1990. **160**.

[99] J.W. Hutchins and H. Somers. *An Introduction to Machine Translation.* Academic Press, London, England, 1992.

[100] W.J. Hutchins. *Machine Translation: Past, Present, Future.* Ellis Horwood, Ltd., Chichester, England, 1986.

[101] W.J. Hutchins. Prospects in Machine Translation: Proceedings of MT Summit I. In *Proceedings of Machine Translation Summit*, Japan, 1987.

[102] M. Ishikawa and R. Sugimura. Natural Language Analysis Using a Network Model: Modification Deciding Network. In *Proceedings of the Fourth International Conference on Theoretical and Methodological Issues in Machine Translation of Natural Languages*, pages 55–66, Montreal, Canada, 1992.

[103] Ray Jackendoff. Grammatical Relations and Functional Structure. In *Semantic Interpretation in Generative Grammar.* The MIT Press, Cambridge, MA, 1972.

[104] Ray Jackendoff. *Semantics and Cognition.* The MIT Press, Cambridge, MA, 1983.

[105] Ray Jackendoff. *Semantic Structures.* The MIT Press, Cambridge, MA, 1990.

[106] A.N. Jain. Parsing Complex Sentences with Structured Connectionist Networks. *Neural Computation*, 3:110–120, 1991.

[107] R. Johnson, M. King, and L. des Tombe. EUROTRA: A Multilingual System under Development. *Computational Linguistics*, 11:155–169, 1985.

[108] Roderick L. Johnson and Pete Whitelock. Machine Translation as an Expert Task. In S. Nirenburg, editor, *Machine Translation: Theoretical and Methodological Issues*, pages 136–144. Cambridge University Press, Cambridge, England, 1987.

[109] D. Jones. Non-Hybrid Example-Based Machine Translation Architectures. In *Proceedings of the Fourth International Conference on Theoretical and Methodological Issues in Machine Translation of Natural Languages*, pages 163–172, Montreal, Canada, 1992.

[110] D. Jones and J. Tsujii. High Quality Machine-Driven Text Translation. In *Proceedings of the Third International Conference on Theoretical and Methodological Issues in Machine Translation of Natural Languages*, pages 43–46, Linguistics Research Center, The University of Texas, Austin, TX, 1990.

[111] Pamela W. Jordan, Bonnie J. Dorr, and John W. Benoit. A First-Pass Approach for Evaluating Machine Translation Systems. *Machine Translation*, 8(1–2):49–58, 1993. Mitre Best Paper Incentive Awar.

[112] A.K. Joshi. How Much Context-Sensitivity is Necessary for Characterizing Structural Descriptions — Tree Adjoining Grammars. In D.

Dowty and L. Karttunen and A. Zwicky, editor, *Natural Language Processing: Theoretical, Computational, and Psychological Perspectives.* Cambridge University Press, New York, 1985.

[113] R. Kaplan and J. Bresnan. Lexical-Functional Grammar: A Formal System for Grammatical Representation. In J. Bresnan, editor, *The Mental Representation of Grammatical Relations*, pages 173–281. The MIT Press, Cambridge, MA, 1982.

[114] R. Kaplan, K. Netter, and A. Zaenen J. Wedekind. Translation By Structural Correspondences. In *Proceedings of Fourth Conference of the European Chapter of the Association for Computational Linguistics*, pages 272–281, Manchester, 1989.

[115] R.M. Kaplan and J. Wedekind. Restriction and Correspondence-based Translation. In *Proceedings of Sixth Conference of the European Chapter of the Association for Computational Linguistics*, pages 193–202, Manchester, 1993.

[116] G. King. Stochastic Methods of Mechanical Translation. *Mechanical Translation*, 1956.

[117] M. King. Design Characteristics of a Machine Translation System. In *Proceedings of the Seventh International Joint Conference on Artificial Intelligence*, pages 43–46, University of British Columbia, Vancouver, B.C., Canada, 1981. Volume 1.

[118] M. King. *Machine Translation: The State of the Art.* Edinburgh University Press, Edinburgh, 1987.

[119] M. King and K. Falkedal. Using Test Suites in Evaluation of Machine Translation Systems. In *COLING-90*, Helsinki, 1990. **328**.

[120] K. Knight and Y. Al-Onaizan. Translation with Finite-State Devices. In *Proceedings of the Third Conference of the Association for Machine Translation in the Americas, AMTA-98, in Lecture Notes in Artificial Intelligence, 1529*, pages 421–437, Langhorne, PA, October 28-31 1998.

[121] K. Knight and I. Chander. Automated Postediting of Documents. In *Proceedings of the Twelfth National Conference on Artificial Intelligence*, pages 779–784, Seattle, Washington, 1994.

[122] K. Knight, I. Chander, M. Haines, V. Hatzivassiloglou, E.H. Hovy, M. Iida, S.K. Luk, R.A. Whitney, and K. Yamada. Filling Knowledge Gaps in a Broad-Coverage MT System. In *Proceedings of the 14th International Joint Conference on Artificial Intelligence*, Montreal, Quebec, Canada, August 1995.

[123] K. Knight and S. Luk. Building a large knowledge base for machine translation. In *Proceedings of AAAI-94*, 1994.

[124] Kevin Knight. *Integrating Knowledge Acquisition and Language Acquisition.* PhD thesis, School of Computer Science, Carnegie Mellon University, Pittsburgh, PA, 1991. CMU-CS-91-209.

[125] J. Kolodner. *Case-Based Reasoning.* Morgan Kauffman, San Mateo, California, 1993.

[126] E. Komatsu, C. Jin, and H. Yasuhara. A Mono-lingual Corpus-based Machine Translation of the Interlingua Method. In *Proceedings of the Fifth International Conference on Theoretical and Methodological Issues in Machine Translation*, pages 24–46, Kyoto, Japan, 1993.

[127] S. Krauwer. Evaluation of MT Systems: A Programmatic View. *Machine Translation*, 8(1–2):59–66, 1993.

[128] J. Landsbergen, J. Odijk, and A. Schenk. The Power of Compositional Translation. *Literary and Linguistic Computing*, 4(3), 1989.

[129] H. Lehmann and N. Ott. Translation Relations and the Combination of Analytical and Statistical Methods in Machine Translation. In *Proceedings of the Fourth International Conference on Theoretical and Methodological Issues in Machine Translation of Natural Languages*, pages 237–248, Montreal, Canada, 1992.

[130] Douglas B. Lenat and R.V. Guha. *Building Large Knowledge-Based Systems.* Reading, MA, Addison-Wesley, 1990.

[131] Deryle Lonsdale, Teruko Mitamura, and Eric Nyberg. Acquisition of Large Lexicons for Practical Knowledge-Based MT. *Machine Translation*, 9:251–283, 1995.

[132] M. Carmen Lozano. Globalink Translation System. In *American Translators Association Conference*, pages 489 – 490, 1989. **205**.

[133] Steven L. Lytinen. Integrating Syntax and Semantics. In S. Nirenburg, editor, *Machine Translation: Theoretical and Methodological Issues.* Cambridge University Press, Cambridge, England, 1987.

[134] Current Research in Machine Translation. Special Issue, 1993. Volume 7, No 4.

[135] Kavi Mahesh and Sergei Nirenburg. A situated ontology for practical NLP. In *Proceedings of IJCAI-95 Workshop on Basic Ontological Issues in Knowledge Sharing*, August 1995.

[136] H. Maruyama, S. Ogino, and M. Hidano. The Mega-Word Tagged-Corpus Project. In *Proceedings of the Fifth International Conference on Theoretical and Methodological Issues in Machine Translation*, pages 15–23, Kyoto, Japan, 1993.

[137] H. Maruyama and H. Watanabe. Tree Cover Search Algorithm for Example-Based Machine Translation. In *Proceedings of the Fourth International Conference on Theoretical and Methodological Issues in Machine Translation of Natural Languages*, pages 173–184, Montreal, Canada, 1992.

[138] D. Maxwell, K. Schubert, and T. Witkam. *Recent Developments in Machine Translation.* Foris, Dordrecht, Holland, 1988.

[139] M.C. McCord. Design of LMT: A Prolog-Based Machine Translation System. *Computational Linguistics*, 15:33–52, 1989.

[140] I.J. McLean. Example-Based Machine Translation Using Connectionist Matching. In *Proceedings of the Fourth International Conference on Theoretical and Methodological Issues in Machine Translation of Natural Languages*, pages 35–43, Montreal, Canada, 1992.

[141] R.L. Mercer. Inflectional Morphology Needs To Be Authenticated By Hand. In *Working Notes, Building Lexicons for Machine Translation, AAAI-93 spring symposium, Technical Report SS-93-02*, page 105, Stanford, CA, 1993.

[142] K.J. Miller and D.M. Zajic. Lexicons as Gold: Mining, Embellishment and Reuse. In *Proceedings of the Third Conference of the Association for Machine Translation in the Americas, AMTA-98, in Lecture Notes in Artificial Intelligence, 1529*, pages 484–493, Langhorne, PA, October 28-31 1998.

[143] E. Mittendorf, P. Schäuble, and P. Sheridan. Applying Probabilistic Term Weighting to OCR Text in the Case of a Large Alphabetic Library Catalogue. In *Proceedings of the 18th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 328–335, Seattle, Washington, July 9-13 1995. Available at http://www-ir.inf.ethz.ch/Public-Web-Pages/sheridan/sheridan.html; http://trec.nist.gov (TREC-3 and TREC-4 confusion track papers).

[144] Simonetta Montemagni and Lucy Vanderwende. Structural Patterns vs. String Patterns for Extracting Semantic Information from Dictionaries. In *Proceedings of the Fourteenth International Conference on Computational Linguistics*, pages 546–552, Nantes, France, 1992.

[145] M. Nagao. A Framework of a Mechanical Translation Between Japanese and English By Analogy Principle. In A. Elithorn and R. Banerji, editor, *Artificial and Human Intelligence.* North-Holland, 1984.

[146] M. Nagao. *Machine Translation: How Far Can It Go?* Oxford University Press, 1989. (Translated from Japanese by Norman D. Cook.).

[147] M. Nagao. Dependency Analyzer: A Knowledge-Based Approach to Structural Disambiguation. In *Proceedings of Thirteenth International Conference on Computational Linguistics*, pages 282–287, Helsinki, Finland, 1990.

[148] M.S. Neff, B. Blaser, J.M. Lange, Hubert Lehmann, and Isabel Zapata Dominguez. Get It Where You Can: Acquiring and Maintaining Bilingual Lexicons for Machine Translation. In *Working Notes, Building Lexicons for Machine Translation, AAAI-93 spring symposium, Technical Report SS-93-02*, page 104, Stanford, CA, 1993.

[149] J. Nerbonne, K. Netter, A. K. Diagne, J. Klein, and L Dickmann. Diagnostic Tool for German Syntax. *Machine Translation*, 8(1–2):85–108, 1993.

[150] S. Niessßen, S. Vogel, H. Ney, and C. Tillmann. A DP based Search Algorithm for Statistical Machine Translation. In *ACL/COLING 98, Proceedings of the 36th Annual Meeting of the Association for Computational Linguistics* (joint with the 17th International Conference on Computational Linguistics), pages 960–967, Montreal, Canada, August 10-14 1998.

[151] S. Nirenburg, C. Domashnev, and D.J. Grannes. Two Approaches to Matching in Example-Based Machine Translation. In *Proceedings of the Fifth International Conference on Theoretical and Methodological Issues in Machine Translation*, pages 47–57, Kyoto, Japan, 1993.

[152] Sergei Nirenburg. *Machine Translation: Theoretical & Methodological Issues.* Cambridge University Press, Cambridge, England, 1987.

[153] Sergei Nirenburg. A Direction of MT Development. In *Proceedings of Machine Translation Summit IV*, pages 189–194, Kobe, Japan, 1993.

[154] Sergei Nirenburg, Jaime Carbonell, Masaru Tomita, and Kenneth Goodman, editors. *Machine Translation: A Knowledge-Based Approach.* Morgan Kaufmann Publishers, San Mateo, CA, 1992.

[155] Sergei Nirenburg, David Farwell, and Yorick Wilks. Multi-Engine, Adaptive MT. In *Proceedings of Fifteenth International Conference on Computational Linguistics*, 1994.

[156] Sergei Nirenburg and Kenneth Goodman. Treatment of Meaning in MT Systems. In *Proceedings of the Third International Conference on Theoretical and Methodological Issues in Machine Translation of Natural Language*, pages 171–187, Linguistic Research Center, University of Texas at Austin, 1990.

[157] M. A. Casta no, F. Casacuberta, and E. Vidal. Machine Translation using Neural Networks and Finite-State Models. In *Proceedings of the 7th International Conference on Theoretical and Methodological Issues in Machine Translation*, pages 160–167, Santa Fe, New Mexico, 1997.

[158] H. Nomiyama. Lexical Selection Mechanism Using Target Language Knowledge and Its Learning Ability. Technical Report IPSJ-WG, ?, 1991. NL86-8.

[159] H. Nomiyama. Machine Translation by Case Generalization. In *Proceedings of Fourteenth International Conference on Computational Linguistics*, pages 714–720, Nantes, France, 1992.

[160] E. Nyberg and T. Mitamura. An Efficient Interlingua Translation System for Multilingual Document Production. In *Proceedings of Machine Translation Summit III*, pages 55–61, Washington, DC, July 1991.

[161] Eric Nyberg and Teruko Mitamura. The KANT system: Fast, accurate, high-quality translation in practical domains. In *Proceedings of Fourteenth International Conference on Computational Linguistics*, Nantes, France, 1992.

[162] Douglas W. Oard. A Comparative Study of Query and Document Translation for Cross-Language Information Retrieval. In *Proceedings of the Third Conference of the Association for Machine Translation in the Americas, AMTA-98, in Lecture Notes in Artificial Intelligence, 1529*, pages 472–483, Langhorne, PA, October 28-31 1998.

[163] A. Okumura, K. Muraki, and K. Yamabana. A Pattern-Larning Based, Hybrid Model for the Syntactic Analysis of Structural Relationships among Japanese Clauses. In *Proceedings of the Fourth International Conference on Theoretical and Methodological Issues in Machine Translation of Natural Languages*, pages 45–54, Montreal, Canada, 1992.

[164] Mari Broman Olsen. Translating English and Mandarin Verbs with Argument Structure (Mis)matches Using LCS Representation. In *Proceedings of the SIG-IL Second Workshop on Interlinguas, AMTA-98, New Mexico State University Technical Report MCCS-98-316*, pages 45–58, Langhorne, PA, 1998.

[165] Mari Broman Olsen, Bonnie J. Dorr, and Scott C. Thomas. Enhancing Automatic Acquisition of Thematic Structure in a Large-Scale Lexicon for Mandarin Chinese. In *Proceedings of the Third Conference of the Association for Machine Translation in the Americas, AMTA-98, in Lecture Notes in Artificial Intelligence, 1529*, pages 41–50, Langhorne, PA, October 28-31 1998.

[166] Boyan Onyshkevich and Sergei Nirenburg. A Lexicon for Knowledge-Based MT. *Machine Translation*, 10(1–2):5–57, 1995.

[167] Mariá Otero. América latina: Radiografiá de una proeza. *UNESCO Courier*, January 1997.

[168] M. Palmer and T. Finin. Workshop on the Evaluation of Natural Language Processing Systems. *Computational Linguistics*, 16(3), 1990.

[169] Martha Palmer and Zhibao Wu. Verb Semantics for English-Chinese Translation. *Machine Translation*, 10(1–2):59–92, 1995.

[170] Cécile Paris and Donia Scott. Stylistic Variation in Multilingual Instructions. In *Seventh International Workshop on Natural Language Generation*, pages 45–52, Kennebunkport, Maine, 1994.

[171] J. Pustejovsky, S. Bergler, and P. Anick. Lexical Semantic Techniques for Corpus Analysis. *Computational Linguistics*, 19(2):331–358, 1993.

[172] James Pustejovsky and Sabine Bergler. The Acquisition of Conceptual Structure for the Lexicon. In *Proceedings of the Sixth Conference of the American Association of Artificial Intelligence*, pages 566–570, Seattle, Washington, 1987.

[173] F. Reeder and D. Loehr. Finding the Right Words: An Analysis of Not-Translated Words in Machine Translation. In *Proceedings of the Third Conference of the Association for Machine Translation in the Americas, AMTA-98, in Lecture Notes in Artificial Intelligence, 1529*, pages 356–363, Langhorne, PA, October 28-31 1998.

[174] Christian-Thomas Retinger. Globalink - Foreign Languages Made Easy Do You Speak a Foreign Language? Excerpt from COM (austria) **159**.

[175] S.D. Richardson, L. Vanderwende, and W. Dolan. Combining Dictionary-Based and Example-Based Methods for Natural Language Analysis. In *Proceedings of the Fifth International Conference on Theoretical and Methodological Issues in Machine Translation*, pages 69–79, Kyoto, Japan, 1993.

[176] M.T. Rosetta. *Compositional Translation.* Kluwer Academic Publishers, Dordrecht, The Netherlands, 1994.

[177] Brigitte Roudaud, Maria Claudia Puerta, and Otemea Gamrat. A Procedure for the Evaluation and Improvement of an MT System by the End-User. *Machine Translation*, 8(1–2):108–116, 1993.

[178] L. Sadler and D. Arnold. A Constraint-Based Approach to Translating Anaphoric Dependencies. In *Proceedings of Fourteenth International Conference on Computational Linguistics*, pages 728–234, Nantes, France, 1992.

[179] L. Sadler and H.S. Thompson. Structural Non-Correspondence in Translation. In *Proceedings of Fifth Conference of the European Chapter of the Association for Computational Linguistics*, pages 293–298, Berlin, Germany, 1991.

[180] H. Saito and M. Tomita. On Automatic composition of Sterotypic Documents in Foreign Languages. Technical Report CMU-CS-86-107, Department of Computer Science, Carnegie Mellon University, 1986.

[181] A. Sanfilippo, T. Briscoe, A. Copestake, M.A. Marti, M. Taule, and A. Alonge. Translation Equivalence and Lexicalization in the ACQUILEX LKB. In *Proceedings of the Fourth International Conference on Theoretical and Methodological Issues in Machine Translation of Natural Languages*, pages 1–11, Montreal, Canada, 1992.

[182] S. Sato. Example-Based Translation of Technical Terms. In *Proceedings of the Fifth International Conference on Theoretical and Methodological Issues in Machine Translation*, pages 58–68, Kyoto, Japan, 1993.

[183] S. Sato and M. Nagao. Toward Memory-Based Translation. In *Proceedings of Thirteenth International Conference on Computational Linguistics*, pages 247–252, Helsinki, Finland, 1990.

[184] R.C. Schank. Conceptual Dependency: A Theory of Natural Language Understanding. *Cognitive Psychology*, 3:552–631, 1972.

[185] Roger Schank. Identification of Conceptualizations Underlying Natural Language. In Roger Schank and Kenneth Colby, editors, *Computer Models of Thought and Language*, pages 114–151. Freeman, San Francisco, CA, 1973.

[186] Roger C. Schank, editor. *Conceptual Information Processing.* Elsevier Science Publishers, Amsterdam, The Netherlands, 1975.

[187] K. Schubert. The architecture of DLT — Interlingual or Double Direct? In D. Maxwell, K. Schubert, and T. Witkam, editors, *Recent Developments in Machine Translation*, pages 131–144. Foris, Dordrecht, Holland, 1988.

[188] Science Applications International Corporation (SAIC). *SYSTRAN Administrator's Manual*, September 1988. **22**.

[189] Science Applications International Corporation (SAIC). *SYSTRAN User's Manual*, September 1989. **23**.

[190] Bernard E. Scott. The LOGOS System. In *MT SUMMIT II*, pages 174–179. Deutsche Gesellschaft fuer Dokumentation e.V. (DGD), August 1989. **132**.

[191] Bernard E. Scott. Biological Neural Net for Parsing Long, Complex Sentences. Review draft, Logos Corporation Technical Center, 111 Howard Blvd., Mt. Arlington NJ 07856, November 1990. **361**.

[192] Bernard E. Scott. Tender Offer in Response to European Commision Tender Call #6. Response to call, Logos Corporation, January 1990. **362**.

[193] Randall M. Sharp. A Model of Grammar Based on Principles of Government and Binding. Master's thesis, Department of Computer Science, University of British Columbia, 1985.

[194] S.M. Shieber and Y.Schabes. Synchronous Tree Adjoining Grammars. In *Proceedings of Thirteenth International Conference on Computational Linguistics*, pages 253–258, Helsinki, Finland, 1990.

[195] Y. Shiwen. Automatic Evaluation of Output Quality for Machine Translation Systems. *Machine Translation*, 8(1–2):117–126, 1993.

[196] B. Sigurd. Translating to and from Swedish by SWETRA — a Multilanguage Translation System. In Dan Maxwell and Klaus Schubert and Toon Witkam, editor, *Recent Developments in Machine Translation*, pages 205–218. Foris, Dordrecht, 1988.

[197] B. Sigurd and M. Eeg-Olofsson. Prolog Implementations of English and Swedish GB Grammars. Technical Report Working Papers 38, Department of Linguistics, Lund University, 1991.

[198] J. Slocum. *Machine Translation Systems.* Cambridge University Press, Cambridge, 1988.

[199] J. Slocum and C.F. Justus. Transportability to Other Languages: The Natural Language Processing Project in the AI Program at MCC. *ACM Transactions on Office Information Systems*, 3(2):204–230, 1985.

[200] Jonathan Slocum. A Survey of Machine Translation: Its History, Current Status and Future Prospects. *Computational Linguistics*, 11, January-March 1985. **1**.

[201] H. Somers. Interactive Multilingual Text Generation for a Monolingual User. In *Proceedings of the Fourth International Conference on Theoretical and Methodological Issues in Machine Translation of Natural Languages*, pages 151–161, Montreal, Canada, 1992.

[202] H. Somers, J. Tsujii, and D. Jones. Machine Translation Without a Source Text. In *Proceedings of Thirteenth International Conference on Computational Linguistics*, pages 271–276, Helsinki, Finland, 1990.

[203] Harold Somers. Current Research in Machine Translation. *Machine Translation*, 7(4):231–246, 1993.

[204] E. Steiner. Producers—Users—Customers: Towards a Differentiated Evaluation of Research in Machine Translation. *Machine Translation*, 7(4), 1993. Special Issue: Current Research in Machine Translation.

[205] K.Y. Su and J.S. Chang. Why Corpus-Based Statistics-Oriented Machine Translation. In *Proceedings of the Fourth International Conference on Theoretical and Methodological Issues in Machine Translation of Natural Languages*, pages 249–262, Montreal, Canada, 1992.

[206] E. Sumita, H. Iida, and H. Kohyama. Example-Based Approach in Machine Translation. In *Proceedings of InfoJapan*, 1990.

[207] H. Sumita, O. Furuse, and H. Iida. An Example-Based Disambiguation of Prepositional Phrase Attachment. In *Proceedings of the Fifth International Conference on Theoretical and Methodological Issues in Machine Translation*, pages 80–91, Kyoto, Japan, 1993.

[208] M. Templeton and J. Burger. Problems in Natural Language Interface to DBMS with Examples for EUFID. In *Proceedings of the Conference on Applied Natural Language Processing*, pages 3–16, 1983.

[209] B. Thompson and F. Thompson. Introducing ASK, a Simple Knowledgeable System. In *Proceedings of the Conference on Applied Natural Language Processing*, pages 17–24, 1983.

[210] G. Thurmair. Complex Lexical Transfer in METAL. In *Proceedings of the Third International Conference on Theoretical and Methodological Issues in Machine Translation of Natural Languages*, pages 91–107, Linguistics Research Center, The University of Texas, Austin, 1990.

[211] A. Trujillo. Locations in the Machine Translation of Prepositional Phrases. In *Proceedings of the Fourth International Conference on Theoretical and Methodological Issues in Machine Translation of Natural Languages*, pages 13–20, Montreal, Canada, 1992.

[212] Arturo Trujillo. Towards a Cross-Linguistically Valid Classification of Spatial Prepositions. *Machine Translation*, 10(1–2):93–141, 1995.

[213] Jun-ichi Tsujii and Kimikazu Fujita. Lexical Transfer Based on Bilingual Signs: Towards Interaction During Transfer. In *Proceedings of the European Chapter of the Association for Computational Linguistics*, pages 275–280, Berlin, Germany, 1991.

[214] Jun-ichi Tsujii and Makoto Nagao. Dialogue Translation *vs.* Text Translation: Interpretation Based Approach. In *Proceedings of Twelveth International Conference on Computational Linguistics*, pages 688–693, Budapest, Hungary, 1988.

[215] G. van Noord, J. Dorrepaal, P. van der Eijk, M. Florenza, and L. des Tombe. The MiMo2 Research System. In *Proceedings of the Third International Conference on Theoretical and Methodological Issues in Machine Translation of Natural Languages*, pages 213–233, Linguistics Research Center, The University of Texas, Austin, TX, 1990.

[216] G. van Slype. Conception d'une Méthodologie Générale d'évaluation de la Traduction Automatique. *Multilingua*, 1:221–237, 1982.

[217] M. Vasconcellos. What Do We Want from MT? Special Issue: Current Research in Machine Translation, 1993.

[218] B. Vauquois. A Survey of Formal Grammars and Algorithms for Recognition and Transformation in Machine Translation. In *IFIP Congress-68*, pages 254–260, 1968.

[219] Clare Voss and Bonnie J. Dorr. Toward a Lexicalized Grammar for Interlinguas. *Machine Translation*, 10(1–2):143–184, 1995.

[220] Ye-Yi Wang and Alex Waibel. Modeling with Structures in Statistical Machine Translation. In *ACL/COLING 98, Proceedings of the 36th Annual Meeting of the Association for Computational Linguistics* (joint with the 17th International Conference on Computational Linguistics), pages 1357–1363, Montreal, Canada, August 10-14 1998.

[221] E. Wehrli. The IPS System. In *Proceedings of Fourteenth International Conference on Computational Linguistics*, pages 870–874, Nantes, France, 1992.

[222] Pete Whitelock. Why Transfer and Interlingua Approaches to MT are Both Wrong: A Position Paper. In *Presented at MT Workshop, Into The 90's*, Manchester, England, 1989.

[223] Pete Whitelock. *A Lexicalist Unification Grammar of Japanese*. PhD thesis, Language and Linguistics, UMIST, Manchester, 1991.

[224] Pete Whitelock. Shake-and-Bake Translation. In *Proceedings of Fourteenth International Conference on Computational Linguistics*, pages 784–791, Nantes, France, 1992.

[225] Y. Wilks. Comparisons with US/Europe. In *JTEC Machine Translation Workshop*, Washington, D.C., March 1991. National Science Foundation.

[226] Y. Wilks. Corpora and Machine Translation. In *Proceedings of Machine Translation Summit IV*, pages 137–146, Kobe, Japan, 1993.

[227] Yorick Wilks, Dan Fass, Cheng-Ming Guo, James E. McDonald, and Tony Plate. Providing Machine Tractable Dictionary Tools. *Machine Translation*, 5(2):99–154, 1990.

[228] Dekai Wu. A Polynomial-Time Algorithm for Statistical Machine Translation. In *Proceedings of the 34th Annual Conference of the Association for Computational Linguistics*, pages 152–158, Santa Cruz, CA, June 1996.

[229] Dekai Wu. Stochastic Inversion Transduction Grammars and Bilingual Parsing of Parallel Corpora. *Computational Linguistics*, 23(3):377–400, 1997.

[230] Dekai Wu and Hongsing Wong. Machine Translation with a Stochastic Grammatical Channel. In *ACL/COLING 98, Proceedings of the 36th Annual Meeting of the Association for Computational Linguistics* (joint with the 17th International Conference on Computational Linguistics), pages 1408–1415, Montreal, Canada, August 10-14 1998.

[231] Dekai Wu and Xuanyin Xia. Large-Scale Automatic Extraction of an English-Chinese Translation Lexicon. *Machine Translation*, 9(3–4):285–313, 1995.

[232] J. Yamron, J. Baker, P. Bamberg, H. Chevalier, T. Dietzel, J. Elder, F. Kampmann, M. Mandel, L. Manganaro, T. Margolis, and E. Steele. LINGSTAT: An Interactive, Machine-Aided Translation System. In *ARPA Workshop on Human Language Technology*, Plainsboro, New Jersey, 1993.

[233] J. Yamron, J. Cant, A. Demedts, T. Dietzel, and Y. Ito. The Automatic Component of the LINGSTAT Machine-Aided Translation System. In *ARPA Workshop on Machine Translation*, Vienna, Virginia, March 1994.

[234] H. Yasuhara. An Example-Based Multilingual MT System in a Conceptual Language. In *Proceedings of Machine Translation Summit IV*, pages 189–194, Kobe, Japan, 1993.

[235] David M. Zajic and Keith J. Miller. Where Interlingua Can Make a Difference. In *Proceedings of the SIG-IL Second Workshop on Interlinguas, AMTA-98, New Mexico State University Technical Report MCCS-98-316*, pages 69–78, Langhorne, PA, 1998.