LANGUAGEFORCE INC

UNIVERSAL TRANSLATOR ®

Presented are the Executive Summary and White Paper which outline the Universal Translator Core Technology invented by LanguageForce, Inc.

We are 100% dedicated towards the future development of machine translation, natural language processing and understanding systems. We welcome all inquires concerning co-development and research.

All inquires should be made to:
Yuri Mordovskoi, Ph.D. Chief Technical Officer (yuri@languageforce.com)
Blaine Leckett, Ph.D. President (blaine@languageforce.com)

LANGUAGEFORCE, Inc
1601 E. Lincoln Ave., Orange, CA 92865
USA

Table of Contents

EXECUTIVE SUMMARY PAPER 1

TECHNOLOGY EVALUATION PAPER 1

1. INTRODUCTION 1
1.1 Direct Architecture 2
1.2 Transfer Architectures 3
1.3 Interlingual Architectures 4

2. OVERVIEW OF MEETING: UT FRAMEWORK 4
2.1 Interlingual Design of UT 5
2.2 Components of the UT System 6

3. TECHNOLOGY EVALUATION: BENEFITS AND LIMITATIONS OF THE UT DESIGN 12
3.1 Potential for Cross-linguistic Coverage 13
3.2 Extensibility 14
3.3 Potential for Testing Environment 14
3.4 Broad Scale Potential 14
3.5 Potential for Development of New Applications 15
3.6 Handling Linguistic Challenges 17
3.7 Summary 30

4. APPLICATION-ORIENTED EVALUATION 30
4.1 Evaluation Metrics for UT 31
4.2 Comparison of UT Output to Output of Other Systems 34

5. SUMMARY AND CONCLUSIONS 37

6. BIBLIOGRAPHIC REFERENCES 39

A. APPENDIX: TEST SUITE FOR SOME LEVEL II LANGUAGES

German Testing Sentences 40
Swedish Testing Sentences 41
Spanish Testing Sentences 42
Russian Testing Sentences 43
Japanese Testing Sentences 44
Chinese Testing Sentences 45

LANGUAGEFORCE INC
UNIVERSAL TRANSLATOR ®

EXECUTIVE SUMMARY

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

September 28, 1999

This summary reports the overall results of a meeting involving Dr. Bonnie Dorr, an associate professor and MT Expert at the University of Maryland, Mr. Yuri Mordovskoi, CTO of LanguageForce (LF), and LF Programmers and Linguists regarding the evaluation of the Universal Translator (UT) technology. Dr. Dorr (the author of this report) is an independent researcher in machine translation and has no involvement with the development of the UT system.

To the best of my knowledge, UT is the first commercial *interlingual* system (i.e., one that uses a single language-independent representation—the IR—as the basis of translation into multiple languages) that is broad scale (i.e., 40 languages with dictionaries ranging in size from 30K to 1200K entries). Its design is distinguished from the traditional interlingual approach—where analysis and generation components differ from language to language—in that the same analysis and generation process is used for all languages of the system. The IR is a dependency structure that conveys the essential relations associated with each event. During the analysis process, the syntax of the source language is mapped into the IR; during the generation process, the syntax of the target language is produced from this very same representation. This mapping is achieved by means of linking rules that are defined in a formalism called Universal Translator Language (UTL).

The motivation behind using a single underlying representation is that it allows UTL rules to be stated on a uniform representation of the sentence. To give an idea of what is gained is here, take n to be the number of languages. Typically, commercial systems adopt the transfer approach, in which a large set of transfer rules must be constructed for each source-language/target-language pair. Thus, there are n^2 sets of transfer rules for n languages. This amounts to 1600 for 40 languages. Early interlingual approaches attempted to decrease this number, yielding $2n$ mappings. This amounts to 80 mappings for 40 languages. UT is distinguished from these alternative approaches in that it uses a single analyzer and generator for all 40 languages; thus, there are 2 mappings regardless of the number of languages handled by the system.

While there is still quite a bit of effort necessary for improving the *quality* of UT output, the principled handling of 40 languages bidirectionally is clearly a breakthrough in the commercial MT world. From an engineering point of view, the interlingual design of UT is advantageous because the analyzer and

generator need not be reprogrammed when new languages are added to the system. The development of lexical and grammatical knowledge is, of course, necessary for each language added to the system—as in any machine translation system. However, the standards for the development of this knowledge have been set in advance and the modules that *use* this knowledge during the translation process need not be reprogrammed. In addition, there are no transfer lexicons or bilingual mappings required for the system. Each language-dependent lexicon may be used in either direction—for analysis and generation.

An important point here is that much of the development that would generally be required for *most* MT systems—i.e., the information that needs to be added as each new language is adopted into the system—has been done in advance, during the early planning years (starting in 1995), when all morphological, syntactic, and semantic features and attributes were standardized to one large, uniform set designed to accommodate a very broad range of language types.

In the short term, the UT design does not guarantee a higher translation quality than that of other commercial designs. The main design benefit is not this—at least not initially—but that of *extensibility*, i.e., ease of ramping up to handle a wide range of languages quickly. An important point here is that translation quality is not the current focus of the UT design, at least not in the short term. It is only in the long term that the potential for higher quality is evident: as each language is added, the time that is typically spent building and/or reprogramming system modules and for each language pair (i.e., n^2 sets of bilingual rules) can be spent fine-tuning the lexicons and grammars of the new language.

UT includes a sophisticated development environment that lends itself readily to easy maintenance and upgrades/extensions. The development methodology includes tools that accommodate incremental improvements—adding new words to dictionaries and fine-tuning grammar rules—by computational linguists whose training time is approximately one month. Rules are created in UTL, a standardized format that is designed to be easily used by linguists; a working set for a particular language can be developed in 6 person-months, on average.

The internal modules of UT can be used as part of an interlingual testing environment that allows for error detection in the UTL rules. In particular, the *UT Quality Control Module* points to possible candidate errors in the rules when a sentence fails. This module includes specialized software that runs a comparison between the specification of each language and structures required by the IR and displays all differences that exist in the language dictionary. The types of mismatches it detects are: inappropriate usage of an attribute for a particular part of speech, missing translations for some senses, and inappropriate word sense information.

UT has what LanguageForce developers call an *Open Architecture*, i.e., it is not only easy to use UT in different applications, but individual system modules can be used for principled development of new types of systems that are not yet implemented. This architectural design provides a means for porting modules into a number of different applications including document retrieval, summarization, database question-answering, natural language understanding, internet-based application (web page, chat, e-mail translation programs) and translation services.

Developers of UT have taken the approach of attempting broad coverage first; the next step is to upgrade the quality of the result. This directly contrasts with approaches taken in other commercial systems, where the focus is typically on generating one particular language pair—or a small set of pairs—without regard to what it would take to ramp up quickly to other (potentially diverse) languages. In summary, UT is worthy of consideration as the "MT framework of the future" in the endeavor to handle multiple languages broadly and uniformly.

LANGUAGEFORCE ® INC
UNIVERSAL TRANSLATOR ®

TECHNOLOGY EVALUATION

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

September 24, 1999

ABSTRACT: This report summarizes the results of a meeting involving Dr. Bonnie Dorr, an associate professor and MT Expert at the University of Maryland, Mr. Yuri Mordovskoi, CTO of LanguageForce (LF), and LF Programmers and Linguists regarding the evaluation of the Universal Translator (UT) technology. Dr. Dorr (the author of this report) is an independent researcher in machine translation and has no involvement with the development of the UT system. A short history of machine translation is presented as background for the framework and technological decisions underlying the approach adopted in the UT system. An overview of the meeting is presented next, including a description of the theoretical and technological framework underlying UT. Following this, a linguistic-based **technology evaluation** of the UT system is presented in the context of representative examples. The examples are described in detail along with a discussion of the features of the UT system that allow it to handle several types of complex phenomena, as well as a discussion of the practicalities of scaling up the UT system for cases it does not currently handle. Summarizing remarks and recommendations for future research and development are included. Finally, the results of an application-oriented assessment using 12 well-established criteria and a comparison with other MT systems are presented.

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.

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 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.

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 have now emerged from this effort. 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. This expansion has been stimulated by the desire for more foreign markets; international and domestic corporations are developing systems that are competing in the world market.

Current architectures for MT may be roughly organized into the following three classes: (1) Direct; (2) Transfer; and (3) Interlingua. The three levels correspond to different levels of transfer, depending on the depth of analysis provided by the system. The direct approach is the most primitive form of transfer, i.e., word-for-word replacement. The interlingual approach is the most degenerate form of transfer, i.e., the transfer mapping is essentially non-existent. Most translation systems are called transfer-based, falling somewhere between these two extremes ranging from a shallow (syntactic) analysis to a deeper (semantic) analysis.

1.1 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. 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 it is impossible to construct direct mapping rules. The result is that this approach typically generates very literal translations. 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). 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.

1.2 Transfer Architectures

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.

Unlike the direct approach, the transfer architecture accommodates more complex mappings. 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.

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. 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 major 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. In a system where only a few languages are under development, this may not be problematic as might be expected, especially if the language being added is similar in structure to ones that are already handled by the system. However, in a system where there are numerous languages of different types, 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.

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.

1.3 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. In general, commercial systems are not based on this approach. 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 assumes 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 for *like* and *gustar*:

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

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. 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. 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 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.

2. OVERVIEW OF MEETING: UT FRAMEWORK

Dr. Bonnie Dorr met with Mr. Yuri Mordovskoi, CTO of LanguageForce (LF), and LF Programmers and Linguists regarding the evaluation of the Universal Translator (UT) technology, on Friday, 9/10/99. Dr. Dorr (the author of this report) is an independent researcher in machine translation and has no involvement with the development of the UT system. The discussions included an overview of the UT Framework as well as a presentation by Dr. Dorr that provided an overview of different MT approaches and techniques.

(Throughout this report, 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.)

2.1 Interlingual Design of UT

As mentioned above, the interlingual approach is currently a very active area of research and is generally not used in commercial systems. Typically, commercial systems adopt the traditional transfer approach; see Figure 1. This approach requires the construction of a large set of transfer rules for each source-language/target-language pair.

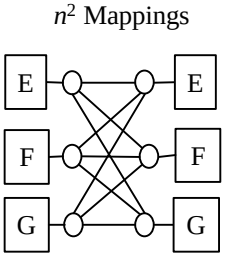


Figure 1: Most Commercial Systems Adopt the Traditional Transfer Approach

UT is the first commercial broad-scale interlingual system, to my knowledge (i.e., 40 languages with dictionaries ranging in size from 30K to 200K entries). Its design is distinguished from the traditional interlingual approach—where analysis and generation components differ from language to language (see Figure 2a)—in that the same analysis and generation process is used for all languages of the system (see Figure 2b).

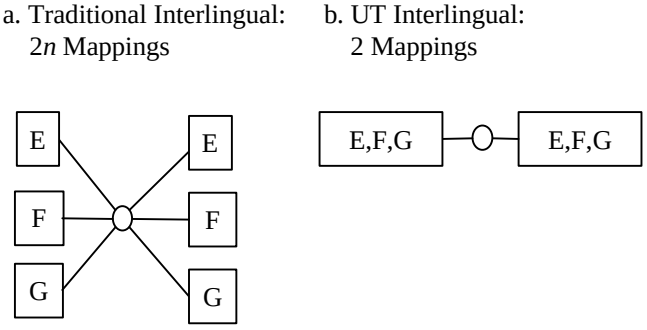


Figure 2: UT Differs from Early Interlingual Approaches

The Interlingual Representation (IR) used in UT is a dependency structure that conveys the essential relations associated with each event. During the analysis process, the syntax of the source language (SL) is mapped into the IR; during the generation process, the syntax of the target language (TL) is produced from this very same representation. This mapping is achieved by means of linking rules that are defined in a formalism called Universal Translator Language (UTL).

The motivation behind using a single underlying representation (the IR) is that it allows UTL rules to be stated on a uniform representation of the sentence. To give an idea of what is gained is here, take *n* to be the number of languages. Recall that there are n^2 sets of transfer rules for *n* languages. In Figure 1, this amounts to 9 mappings for 3 languages. The number would be 1600 for 40 languages. Early interlingual approaches attempted to decrease this number, yielding 2*n* mappings. In Figure 2a, this amounts to 6 mappings for 3 languages. The number would be 80 for 40 languages. Because the UT system uses a single analyzer and generator for all 40 languages (or for any number of languages), there are 2 mappings in Figure 2b.

2.2 Components of the UT System

The components that are accessed in the construction of the IR are the following: (1) the Universal Translator Dictionary (UTD)—one for each of the 40 languages translated by the system—in which words are associated with their interlingual concepts (i.e., word senses such as `like=2-1`); (2) the Universal Translator Rules (UTR)—one set for each of the 40 languages translated by the system—in which UTL is used to associate a word with its interlingual relations (such as `predicator` or `verb_direct`) and attributes (such as `+human`); and (3) the Interlingual Representation Dictionary (IRD), a language independent module that serves as the interface between interlingual concepts in the UTD and interlingual relations in the UTR.

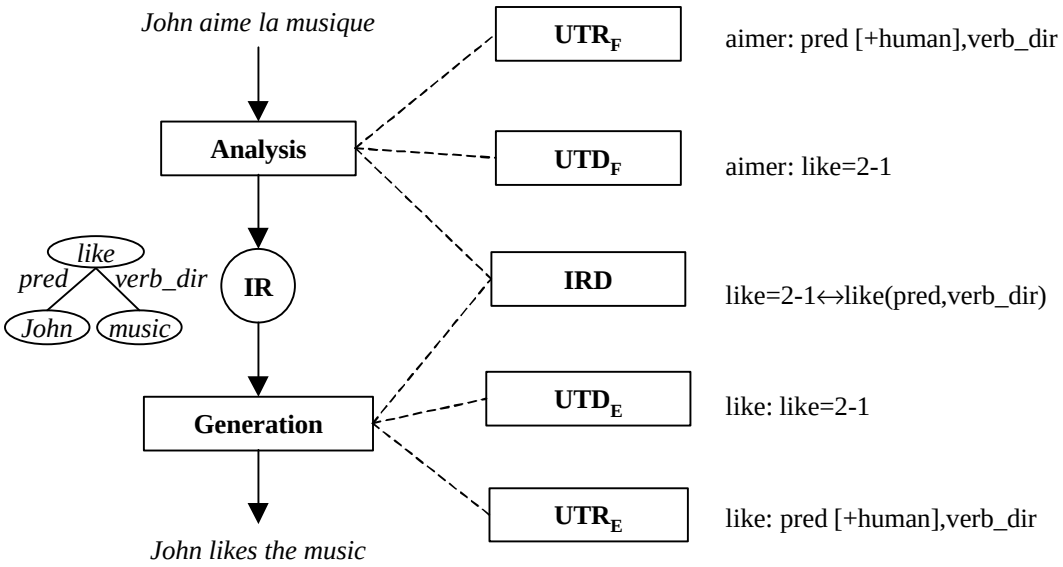


Figure 3: Universal Translator Design

Figure 3 illustrates the full mapping process that incorporates each of these three components. A particular translation example is given in the figure, mapping the French sentence *John aime la musique* into the English sentence *John likes music*. The IR (to the left of the IR circle) is a tree structure that represents the `predicator` and `verb_direct` roles associated with `John` and `music`. The word sense `like=2-1` is an interlingual concept stored in the IRD. The French and English UTDs link this concept to their corresponding verbs, *aimer* and *like*, respectively. The label 2-1 in this interlingual concept corresponds the verbal form 2 and its associated word sense 1.

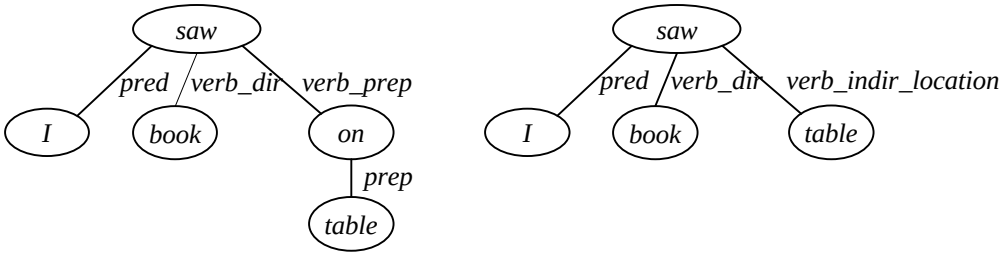
2.2.1 Interlingual Representation (IR) and Interlingual Representation Dictionary (IRD)

While not particularly sophisticated, the word senses underlying the IR provide a uniform "pivot point" for a wide range of different languages. The total number of word senses stored in the IRD is approximately 1,200,000; the UTDs for each of the 40 languages link these word senses into words and phrases for general and 9 specific areas (aerospace, medical, law, oil, etc.). The process of building the IRD took 4 years—an effort that involved 30 on-site people plus partners from Eastern Europe and Asia. All languages were taken into consideration in the development of the full set of word senses.

The IR is, more or less, a hybrid grammatical/semantic structure in the sense that, when a grammatical relation is common to all languages, that term is used in place of its thematic counterpart. For example, the grammatical structure for *I saw the book on the table* includes language-specific information about prepositional phrases as shown below in (2)(i) below. The final IR abstracts incorporates the thematic notion of Location in place of this prepositional phrase, but retains the grammatical labels `pred` and `verb_dir` (for Agent and Theme, respectively), as shown below in (2)(ii).

(2) E: I saw the book on the table

(i) Grammatical Structure **(ii) IR**



Additional examples and related discussion are given below in Section 2.2.3.

2.2.2 Universal Translator Dictionary (UTD)

The size of each language-particular UTD varies significantly, ranging from the low end of 30,000 entries (for 20 languages) to the mid-level end of 70,000 (for 14 languages) to the high end of 200,000 (for 6 languages). The following illustrates the format of the English UTD:

```
father:  father=1-1
daddy:   father=1-1
book:    book=1-1
book:    book=2-1
reserve: book=2-1
table:   table=1-1
table:   table=1-2
```

The corresponding Spanish UTD excerpt is the following:

```
padre:    father=1-1
papa:     father=1-1
libro:    book=1-1
reservar: book=2-1
mesa:     table=1-1
tabla:    table=1-2
```

These links into IRD word senses were created semi-automatically in a two-stage process. First, 40 bilingual dictionaries between English and each of the other languages were used to produce initial hypotheses about the word-sense links for each language. This initial process takes approximately 1 person-month per language.

The 40 bilingual dictionaries were taken from different sources. The first 28 bilingual dictionaries were acquired in 1996; the remaining dictionaries were acquired between 1997 and 1998. Certain of the dictionaries were purchased from small independent companies who are developing bidirectional MT systems, learning products, and electronic dictionaries. Others were donated by private individuals; most were from countries in which the native language is spoken. Those from Europe were found to be much cheaper than the dictionaries sold in the United States.

The next step of the dictionary linking process is a manual cleaning for the most common words (generally on the order of 24,000). This takes 3 person-months per language. Approximately 30 person-months were required for an additional cleaning of 70,000 words. This cleaning process involved outside consulting with companies and individuals for manual editing/adding of new words/phrases; most were done by internet. After 4-5 revisions, the resulting dictionaries were inverted for development of an IRD to fit all languages for analysis and generation.

This second step of the dictionary linking process was very labor-intensive. Most of the dictionaries had widely varying structure and quality. Several had imperfect data and no morphology; in-house morphological development was required in several cases. Some Eastern European languages are still missing. A future direction proposed by LanguageForce is the development of acquisition procedures based on Web data mining using the UT word senses as a knowledge-based "seed."

2.2.3 Universal Translator Rules (UTR)

The size of each language-particular UTR also varies, although not as significantly as that of the UTDs, ranging from 1121 rules in Chinese to 1546 rules in English. These rules are based on UTL, a standardized format that is designed to be easily used by linguists for access to words, relations between words, attributes, and the complete IR structure. The methodology underlying UTL is based on 14 different stages of standardization that constrain the rule format for linguists.

The average development time for a first-draft rule set is 6 person-months by a native speaker (with good computational skills) who has had 1 month training. An additional 4 person-months is necessary for the implementation of advanced rules. Certain of these are developed for more general areas—e.g., numerals and punctuation—and are stored in a Rule Library that is accessible to the entire language set. Rules in the Rule Library may be copied to individual UTRs in cases where modifications are necessary for accommodating language-specific characteristics. The entire set of UTRs were developed over a period of 2 years, starting in November of 1997. The final 10 months of this development period were devoted to data clean-up.

The UTR for each language specifies the linking between the language-specific syntactic structure and the IR using two types of knowledge: (1) relations and (2) attributes. There are 94 relations falling into two general categories, *grammatical* (48) and *thematic* (46). Grammatical relations are syntactic links between events and their participants in the syntactic structure. Each language uses approximately 25 of the 48 possible grammar links. Some of the grammatical relations are: nominative, accusative, Sentence_Coordinator, Sentence_Subordinator, Linking_Verb, Modifier, Clause, Prep, Subject_Description, Verb_Direct, Verb_Indirect, and Name.

Thematic relations are semantic links between events and their participants in the IR. The thematic relations are: Agent, Theme, Predicate, Instrument, Location. For example, both *with fork* and *by hammer* are analyzed as the same thematic relation in the IR—an instrument—in the phrases *eat with fork* and *break*

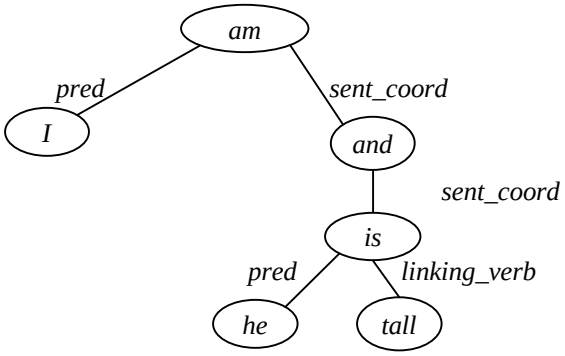
by hammer, even though these are distinguished by a `prep` link in the syntactic representation (i.e., *by* vs. *with*).

Examples of the use of the grammatical relations (in the syntactic tree) and their thematic counterparts (in the IR) are given below in (3) for different language phenomena. In each case, the transformation between the syntactic tree and its corresponding IR is achieved by iterative application of rules in the UTR.

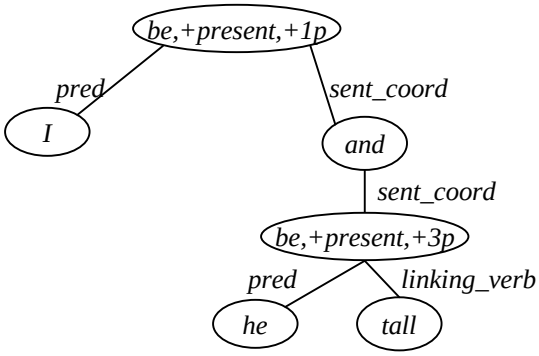
(3) **Phenomena handled by UT:**

(i) **Conjunction:** I am short and he is tall

Grammatical Structure

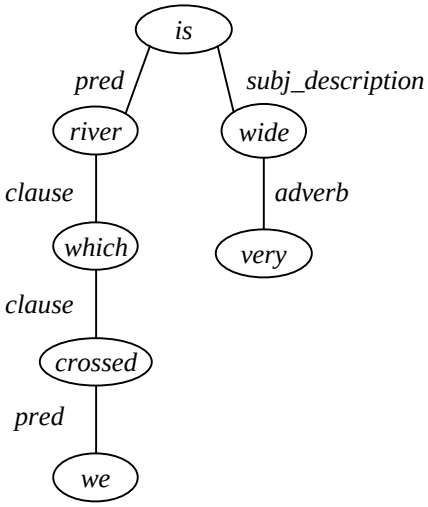


IR

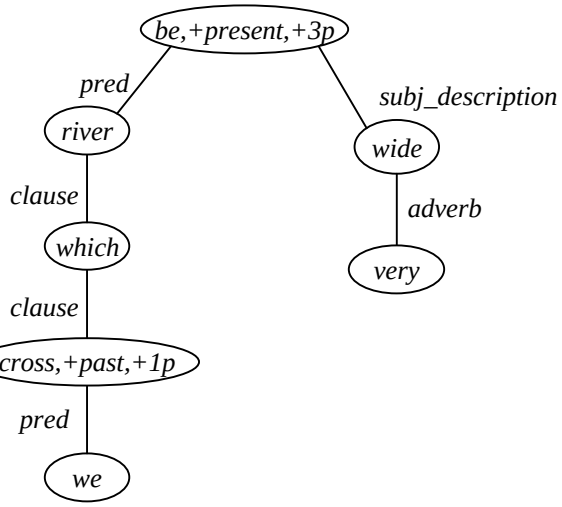


(ii) **Relative Clause:** The river which we crossed is very wide.

Grammatical Structure

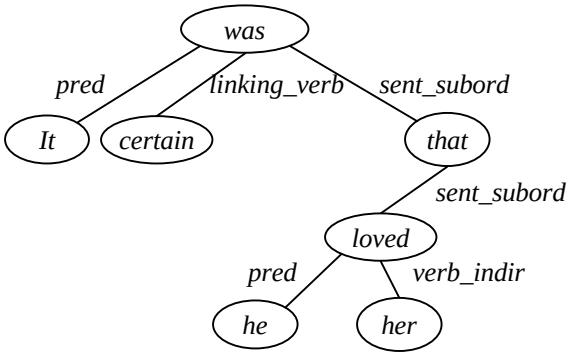


IR

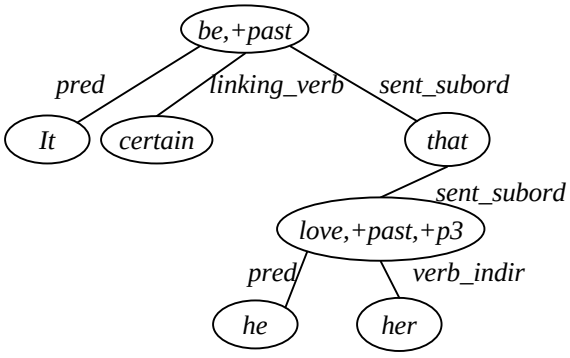


(iii) **Adjectival Complementizers:** It was certain that he loved her

Grammatical Structure



IR



In addition to grammatical/thematic relations, the UTRs access 900 different types of attributes, falling into three categories: Morphological, Grammatical, and Semantic. Morphological attributes are related to the actual form of the word. The analysis process detects morphological attributes, e.g., `+past` for *wrote*. The generation process uses these for appropriate realization of word forms, e.g., that the root form *write* with `+past` features is *wrote*. Other morphological attributes are: `+present`, `+nominative`, `+dative`, `+singular`, `+plural`, `+feminine`, and `+masculine`. There are 100 morphological attributes grouped into 15 independent groups, e.g., Tense, Case, Plurality, and Gender.

Grammatical attributes are related to words that can influence other words in the sentence, e.g., with/by markers associated with the instrument relation above. Other grammatical attributes are: `+verb`, `+adj`, `+adverb`, `+noun`, `+prep`, `+prep_to_event`, `+prep_from_event`, `+prep_from_space`, and `+prep_to_space`. (These last four are used for associating prepositions with events; in this case, `+event` is a grammatical attribute for a preposition and a semantic attribute for a noun.) In addition, there are certain types of language-specific attributes, e.g., `+silentH` (for associating appropriate determiners with nouns that start with `h'). There are 150 grammatical attributes grouped into 60 categories, e.g., PartofSpeech, Case, Tense, and Plurality. (Certain of these are also treated as morphological attributes since they contribute toward both grammatical and morphological processing.)

Semantic attributes are used primarily for word-sense disambiguation during analysis and generation, e.g., `+abstract` or `+physical_object` for the noun *table*. Other semantic attributes are `+place`, `+abstract`, `+human`, `+title`, `+nation`, `+language`, and `+event`. There are 100 semantic attributes grouped into 30 categories, e.g., Animacy (animal,human,notanimal), Entity (abstract/physical_object/energy/place), Humanity (title,family_member,proper_name), and Modality (about 15 values).

Japanese and Korean are special cases, with 300 attributes for morphology due to autosegmentation. In addition, there are some special attribute sets for Chinese markers (50), German separable prefixes (200), and English particles such as on, off, up, etc. (40).

Adding a new language entails manual creation of a set of morphological attributes for that language. Existing attributes are primarily used in creating this set; sometimes new attributes are needed, but these can be added easily to the universal morphological set without modification to the overall format of the UTL. The grammatical attributes differ from morphological attributes in that they have remained generally stable. However, completion of Arabic and Farsi will undoubtedly require new attributes. In general, grammatical attributes are expected to stabilize as one fixed set except for open-classed items: verbs, adjectives, and nouns.

2.2.4 Technology Sequence

In mapping from the source text to the target text, rules in the relevant UTRs are used for each step of the "technology sequence." Each step resolves one type of problem in the overall translation process.

There are two overall processes that produce a target-language sentence from a source-language input: analysis and generation. During analysis, a major focus is on disambiguation at the syntactic and semantic levels. At the syntactic level, disambiguation consists of appropriate grouping of words (phrase analysis); at the semantic level, ambiguous lexical items must be resolved (homonym analysis). The output of these two processes are used to generate a tree (Link Creation) which is then filtered and normalized to produce the IR. The full sequence consists of the following stages:

1. String Analysis: Convert between equivalent strings, e.g., *I've* becomes *I have*.

2. Phrase Analysis: Group words into phrases according to information in the UTD and UTR, e.g., *control system* is analyzed as a noun phrase.

3. Homonym Analysis: Disambiguate word senses according to context, word order, and attributes, e.g., *book* is resolved to either a verb or a noun, depending on its context.

4. Grammar Tree Creation: Produce a syntactic dependency tree according to the highest ranking links in the UTR. The output is a language-dependent grammatical tree.

5. Filtering: Reduce total number of syntactic trees down to one according to weights on the links in the grammatical tree.

6. Normalization: Replace grammar links with semantic links, e.g., eliminate Determiner links. Make final corrections to structure (erasing words, changing attributes, etc.), e.g., replace *good* with *well* in *It goes good.* The result of normalization is a language-independent structure—the IR. (All morphological and syntactic markers are eliminated.)

The two focii in generation are the selection of target-language words (phrase generation and semantic generation) and their corresponding structural configuration (syntactic generation and order generation) based on information stored in the IR. Following this, the surface form is produced by morphological and string operations. The full sequence consists of the following stages:

1. Phrase generation: Replace each IR concept with one or more words.

2. Semantic generation: Select appropriate lexical forms from UTD, e.g., *have read* for present perfect form of *read*.

3. Syntactic Generation: Associate attributes with words that are related to other words, e.g., change the verb *eat* to *eats* if the subject is *he*. The output of this module is a language-specific structure—a grammatical tree, including attributes.

4. Order generation: Sub-trees inside of the overall grammatical tree are cut and replaced with appropriate ordering, e.g., the noun phrase sub-tree corresponding to *good book white* reordered to produce *good white book*.

5. Morphological generation: Produce appropriate inflection, e.g., the Spanish verb *puedo* is produced from the root form *poder*.

6. String generation: Transliterate, where translation is not appropriate; produce contracted forms, e.g., the Spanish form *al* is produced from *a el*.

3. TECHNOLOGY EVALUATION: BENEFITS AND LIMITATIONS OF THE UT DESIGN

This section describes the benefits and limitations of the UT design. It is important to point out that this is an evaluation of the **technology** underlying the UT system, not the **quality** of the translation output. In particular, the focus is on the innovations behind the system and the potential for future development and use.

3.1 Potential for Cross-linguistic Coverage

UT is an implemented interlingual system for 40 languages that operates entirely bidirectionally.[1] While there is still quite a bit of effort necessary for improving the *quality*, the principled handling of 40 languages bidirectionally is clearly a breakthrough in the commercial MT world.

From an engineering point of view, the interlingual design of UT is advantageous because the analyzer and generator need not be reprogrammed when new languages are added to the system. The development of lexical and grammatical knowledge (i.e., the UTDs and UTRs) is, of course, necessary for each language added to the system—as in any machine translation system. However, the standards for the development of this knowledge have been set in advance—during the 4-year effort mentioned in Section 2.2.1—and the modules that *use* this knowledge during the translation process need not be reprogrammed. In addition, there are no transfer lexicons or bilingual mappings required for the system. Each UTD (i.e., each language-dependent lexicon) may be used in either direction—for analysis and generation.

An **important point** here is that much of the development that would generally be required for *most* MT systems—i.e., the information that needs to be added as each new language is adopted into the system—has been done in advance, during the early planning years (starting in 1995), when all morphological, syntactic, and semantic features and attributes were standardized to one large, uniform set (the IRD) designed to accommodate a very broad range of language types. This contrasts with the approach adopted by many other commercial systems, where the focus is typically on generating one particular language pair—or a small set of pairs—without regard to what it would take to ramp up quickly to other (potentially diverse) languages.

UT operates bidirectionally on all languages below i.e., 1600 language pairs. These are grouped together by level of translation quality:

1. Level I Quality (Word Phrase translation, limited sentence and grammar support, contextual translation) is available for: Arabic, Czech, Dutch, Esperanto, Farsi, Finnish, Greek, Hebrew, Hungarian, Indonesian, Latin, Polish, Romanian, Slovak, Swahili, Tagalog, Thai, Turkish, Ukrainian, Vietnamese.

2. Level II quality (full sentence and grammar support, content translation) is available for: Chinese, Danish, English, French, German, Italian, Japanese, Korean, Norwegian, Portuguese, Russian, Spanish, Swedish. The current quality of Level II languages is comparable to that of other commercial systems.

3. Seven additional languages and/or dialects are being added to the current release. The Level II dialects are UK English and additional dialects for French, Portuguese, Spanish, and Chinese. (These are primarily differences in spelling and some grammatical distinctions.) The new Level I languages are Norwegian/Bokmal and Zulu.

4. Levels higher than this (e.g., Level III, which would have full sentence translation, grammar support, and extensive technical dictionaries) are not currently available. With more testing and optimization of lexicons, three languages will be released at Level III in the near future: Chinese, English, and Russian.

[1] The only languages examined for the purposes of this evaluation were Dutch (Level I), Chinese (Level II), English (Level II), French (Level II), German (Level II), and Spanish (Level II). Examples from other languages are given in Appendix A.

3.2 Extensibility

In the short term, the UT design does not guarantee a higher translation quality than that of other commercial designs. The main design benefit is not this—at least not initially—but that of *extensibility*, i.e., ease of ramping up to handle a wide range of languages quickly. An **important point** here is that translation quality is not the current focus of the UT design, at least not in the short term. It is only in the long term that the potential for higher quality is evident: as each language is added, the time that is typically spent building and/or reprogramming system modules and for each language pair (i.e., n^2 sets of bilingual rules) can be spent fine-tuning the lexicons and grammars of the new language.

UT includes a sophisticated development environment that lends itself readily to easy maintenance and upgrades/extensions. The development methodology includes tools that accommodate incremental improvements—adding new words to dictionaries and fine-tuning grammar rules—by computational linguists whose training time is approximately one month. Rules are created in UTL, a standardized format that is designed to be easily used by linguists; a working set for a particular language can be developed in 6 person-months, on average.

3.3 Potential for Testing Environment

The internal modules of UT can be used as part of an interlingual testing environment that allows for error detection in the UTL rules. In particular, the *UT Quality Control Module* points to possible candidate errors in the rules when a sentence fails. This module includes specialized software that runs a comparison between the specification of each language and structures required by the IR and displays all differences that exist in language dictionary. The types of mismatches it detects are: inappropriate usage of an attribute for a particular part of speech, missing translations for some senses, and inappropriate word sense information. The module is also able to detect mismatches between homonyms and their allowable parts of speech—e.g., *which* is not allowed to be any part of speech other than an adjective or pronoun.

The error-detecting capability of the Quality Control module provides an environment for incremental development of more sophisticated rule sets and dictionaries. In addition, this testing environment has the potential of providing eventual assistance in the evaluation of output, given its current capability of detecting errors in translation rules. (More discussion on this point is given below in Section 3.5.)

3.4 Broad Scale Potential

The question addressed in this section is whether UT goes beyond toy examples, i.e, whether there is significant potential for quality improvement *over time*.

Characterizing the potential for quality improvement is not an easy task. One way to do this is to examine the difference between Level I coverage and Level II coverage. An example of earlier Level I French-English coverage was recently posted (and questioned) in an internet mailing:

(4) **Level I French-English Translation:**

F: La monnaie nippone se trouve a son plus haut niveau depuis trois ans face au dollar. Inquite pour la sante des enterprises exportatrices et la solidite de la reprise economique, la Bourse de Tokyo a fortement baise jeudi 16 September.

E: Currency the nippone himself is has her higher above level since three years in the face of the dollar. Inquite for the sante of the exportatrices and the the economique, the Tokyo Grant repeat solidite enterprises has to a considerable degree kiss September Thursday 16.

This translation is clearly a word-for-word replacement, i.e., a Level I result, as advertised in an earlier version of UT. The more recent Level II UT translation of this paragraph, taken verbatim, is given below:

(5) **Level II French-English Translation:**

> E: The currency nipponne is at his higher above level since three years in the face of the dollar. Worried for the health of the exporters and the solidity of the economic repeat, companies Tokyo's Grant has to a considerable degree dropped Thursday 16 September.

Without touching the input or modifying the lexicon, the translation above already exhibits significantly better output quality. There are still errors, however. The two most glaring are the missing lexical entry for *nipponne* and the awkward structure of the sentence due to missing accents in the input text. Once *nippone* is added to the dictionary and missing accents are added to the original French text, the following translation is produced:

(6) **Level II French-English Translation (refined):**

> F: La monnaie nippone se trouve a son plus haut niveau depuis trois ans face au dollar. Inquiète pour la santé des entreprises exportatrices et la solidité de la reprise économique, la Bourse de Tokyo a fortement baissé jeudi 16 Septembre
>
> E: The Japanese currency is at his higher above level since three years in the face of the dollar. Worried for the health of the exporters and the solidity of the economic repeat, companies Tokyo's Grant has to a considerable degree dropped Thursday 16 September.

There are still some minor errors in the text above, e.g., the pronoun *his* instead of *its* and the use of *highest* instead of *higher above level*, but the translation certainly is much more acceptable in (6) than that of (4) or (5) above.

What is of interest here is the time it took to get from (4) to (5) and then from (5) to (6). The first transition, from Level I to Level II, was a full two-year effort—24 person-months. Section 2.2.3 outlines the UTR construction and cleaning that are necessary for bringing about Level II quality. This time period is not unreasonable, as long as quality enhancements are achieved with ease beyond this point—i.e., a significant reduction in time for incremental improvements. As it turns out, this is the case: the second transition above—from (5) to (6)—was achieved in a matter of minutes. Adding a new lexical entry was quick, simple, and significantly reduced the output awkwardness.

On the other hand, a great deal of amount of work would be needed to bring the level of translation up to Level III, including the addition of semantic features to correct the erroneous use of *his*. Section 3.2 above describes the status of Level III languages, which are not yet released.

An additional "Test Suite" is given in Appendix A for determining the potential for quality of Level II languages. The focus here is on technology assessment, i.e., determining the potential for quality improvement—not an evaluation of the quality itself. An evaluation of quality on the cases in Appendix A, and other examples, will appear in a forthcoming report on quality assessment.

3.5 Potential for Development of New Applications

The question addressed in this section is whether the system modules are designed in a way that they can be used in other types of systems. UT has what LanguageForce developers call an *Open Architecture*, i.e., it is not only easy to use UT in different applications but individual system modules can be used for principled development of new types of systems that are not yet implemented.

The porting of UT modules into certain types of applications would be straightforward. Some examples of these are:

- Document retrieval: In this application, the analysis component and associated dictionaries (UTD, IRD) would be used.

- Summarization: For topic clustering and summaries, the generation component and associated UTRs would be used.

- Database Question-Answering: For transforming natural language into structured SQL commands, the analysis component and associated dictionaries (UTD, IRD) would be used.

- Translation services: For interfacing with other translation services, the testing and translation modules in the UT Quality Control module could be used by computational linguists for drafting rules in any language.

Other applications are not as straightforward with respect to the porting of UT modules. Future system development for these applications would require a significant amount of research and development for incorporation of the UT modules. Some examples of these are:

- Conversational/Dialogue applications: The goal would be to build a model of discourse in which information from the complete text is used for the resolution of anaphora. An application in which this would be useful would be Dialogue-Based MT (DBMT), in which the system provides a mechanism for entering into a dialogue with the user about the translation, thus allowing the user to disambiguate the input text and to incorporate culture-specific stylistic detail for high quality MT. Interaction with the user may take place during the 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.

- Understanding Systems for graphical, animation, virtual reality, and speech systems: The idea would be to produce an IR that would then be used for deciding what action the agent would do in a particular scene, e.g., in a computer-aided tutoring system. Ultimately, the goal would be to allow the user to speak in their native language into the UT dictation software, providing commands that would ultimately be executed according to the structure of the IR. (For example: "Go and retrieve the document" would cause the agent to walk over to a table and pick up a document.)

- Automatic Knowledge Acquisition: This area deserves significant attention. The process could be better streamlined, potentially incorporating statistical or corpus-based techniques for a more automatic approach to dictionary building.

- Interlingual Development Environment: A future direction proposed by LanguageForce is the development of a collective linguistic dictionary that allows users to acquire linguistic data based on their needs, while simultaneously providing data for improving the existing UT system.

The conversational and understanding applications will require significant effort toward the development of a deeper semantic structure, particularly in the area of spatial relations (for navigation in a scene), which are somewhat superficially defined—by attributes such as marker-with, marker-by—in the existing IR. The current UT system is designed to allow a wide range of languages to be added quickly—even if it means compromising quality in several cases. However, for true understanding, a deeper analysis is required.

Automatic knowledge acquisition will require the development of techniques for linking UT word senses to language-specific terms. An initial database of "seed terms" (which are also useful for development of an interlingual development environment) would provide a basis for automatic routines, including statistical web-based techniques. The development of automatic acquisition routines enables the addition of new languages and significant enhancements of existing languages.

The interlingual development environment is the most promising direction for use of the UT modules; the idea is to develop an online web-based interface that allows access to the UT word senses stored in the UT Knowledge Base (UTKB). Access will be restricted to qualified linguists as ascertained in advance by a questionnaire. Certain tasks that are currently used for UT development will be enhanced by this online KB system including: adding/editing translations to include new words and correct morphology; corrections to attachment decisions in syntactic processing; modifications to semantic and grammatical attributes; and testing of word/sentence level translations. Statistical smoothing will be required for ensuring consistency among users. For this system to be fully operational, there would still be human labor required—particularly for initial construction of the "seed terms" in the KB to be accessed by linguists. Beyond this, the idea is to provide the resulting KB to translation agencies for machine-aided human translations.

The KB-based framework could be further incorporated into tools that provide additional translation capabilities. For example, portions of the UT Quality Control Module could be reconfigured to work in tandem with the KB information for assessment of translation quality. Although such a system would require several years of effort, the UT framework provides a high degree of modularity and enabling technology for bringing this about.

3.6 Handling Linguistic Challenges

There are three main categories into which translation challenges fall; language understanding, language generation, and the mapping between language pairs. In all three categories, 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. 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. This is the basic idea behind the interlingual representation used in the UT system. We will examine some of the key issues in each of these three categories and the degree of success achieved by the approach adopted in UT.

3.6.1 Analysis: Disambiguation Potential

The most formidable problem during the analysis process is ambiguity: Lexical, Semantic, and Contextual. Each of these ambiguity types will be discussed, in turn.

3.6.1.1 Lexical Ambiguity

Lexical ambiguity arises when there is a choice between two possible meanings of a source-language lexical item. This is often easily resolved if enough syntactic context is available. Consider the following example:

(7) **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*. The UT translation system is able to discern the difference between these two cases, producing the translation in (8).

(8) **UT Resolution of Lexical Ambiguity:**

E: I booked a flight to California
S: Reservé a California un vuelo
`(I) reserved to California a flight'

3.6.1.2 Semantic Ambiguity

Semantic ambiguity is a more formidable problem than Lexical ambiguity; the resolution of this type of ambiguity falls outside of the realm of syntactic and lexical knowledge as in the following examples:

(9) **Semantic Ambiguity (Homonymy):**

(i) E: table
S: mesa, tabla

(ii) E: be
S: ser, estar

(iii) S: esperar
E: expect/expect, wait

Many words, such as *table*, have distinctly different meanings (homonymy); MT systems are forced to choose the correct meaning of the source-language constituent in these cases (e.g., whether *table* corresponds to the physical object (mesa) or the graphical object (tabla)). Other problems arise for words like *be* or *esperar* which have subtly related meanings in different contexts (e.g., *Estoy feliz* vs. *Soy un obrero* or *I wait for John* vs. *I hope/expect that John has a car*) and are frequently represented by distinct words in the target language. Some UT translations that include such cases are given below:

(10) **UT Resolution of Semantic Ambiguity:**

(i) mesa vs. tabla

E: I put books on the table
S: Yo puse los libros en la mesa
`I put the books on the table'

E: I put data in the table
S: Yo puse los datos en la tabla
`I put the data in(to) the table'

(ii) estar vs. ser

E: I am happy
S: Estoy feliz
`(I) am happy'

E: I am a factory worker
S: Soy un obrero de fábrica
`(I) am a worker of a factory'

E: I am tall
S: Soy alto
`(I) am tall'

(iii) hope/expect vs. wait

S: Espero que Juan tenga un carro
E: I hope that Juan has a car

S: Espero a Juan
E: I hope Juan

In two out of three cases, the ambiguity is resolved appropriately by the homonymy analysis process described earlier. For example, in (i), *data* has a single word sense associated with a `+abstract` attribute; *table* has two word senses, one with a `+physical` attribute and one with a `+abstract` attribute. The preposition *in* is associated with the attribute `+to_space`. UT recognizes *data* as a direct object and then chooses the word sense `table=1-2` (i.e., as `+abstract`) because *data* and *table* have a matching `+abstract` attribute. (In addition, a Mexican Spanish version is generated in the second case in (i): *Yo puse la informacion en la tabla.* In general, Spain Spanish is the default, but global attributes may be set for translation into one dialect over the other.)

The second case above is handled analogously, where *be* is resolved to a word sense associated with an attribute indicating a conditional emotion, as in *estar feliz*, and to an alternative word sense otherwise.

The one case that fails above is (iii), where the main verb should be translated as *wait* since its source verb *esperar* is associated with the preposition *a*. Such cases are easily corrected with a simple change to the UTR rule that links *esperar* with its direct object. This rule should be modified to include a restriction on *esperar* that resolves it to the word sense associated with *wait* when it is followed directly by a noun phrase or by the preposition *a*.

More complex semantic ambiguities have 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. The following well-known examples illustrate the difficulty of more complex semantic ambiguity:

(11) **Complex Semantic Ambiguity:**

(i) E: He wrote with a pen
S: Él escribió con un bolígrafo

(ii) E: The box was in the pen
S: La caja estaba en el corral / *el bolígrafo
`The box was in the pen (enclosure) / *pen (writing)'

In the example above, 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). UT fails to resolve this ambiguity:

(12) **UT Handling of Complex Semantic Ambiguity:**

(i) E: The box was in the pen
S: La caja fue en el bolígrafo

(ii) E: He wrote with a pen
S: Él escribió con un bolígrafo

Assigning attributes in such cases could be a possible solution, once again, e.g., the attribute `+enclosure`, which would then be a preferred selectional restriction on the preposition *in*. Note, however, that such an approach would fail in the following case:

(13) **Complex Semantic Ambiguity:**
The ink in the pen was running dry

In order to handle such cases, one could imagine more and more attributes being added to the system. UT could benefit from serious effort in the development of *preferred* selectional restrictions so that the system does not become unwieldy.

3.6.1.3 Contextual Ambiguity

The importance of this effort becomes more evident when one considers cases of inter-sentential contextual ambiguity, which occurs routinely in translated texts (e.g., computer manuals). Consider the following examples:

(14) **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'

Currently UT does not resolve such cases:

(15) **UT Handling of Contextual Ambiguity:**

(i) E: The computer outputs the data; it is fast
S: La computadora outputs los datos; eso es rapidamente

(ii) E: The computer outputs the data; it is stored in ascii
S: La computadora outputs los datos; eso es almacenado el in ascii

Determining the appropriate antecedent for the word *it* could, perhaps, be solved by distinguishing between storable objects and non-storable objects (±storable) and between objects with a speed attribute and those without (±speed). However, this proliferation of distinguishing features contributes further to the unwieldiness of the growing attribute set. Fine-tuning attributes for a specific domain is one way of tackling this problem. For example, 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.

The development of attribute sets could be further constrained by the use of systems like those proposed in Section 3.5—particularly the Conversational/Dialogue and interlingual development environments. These may provide a more automatic means of deriving the information necessary to handle contextual ambiguities of the type described above.

Even 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. Here again is where additional research would be needed. 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:

(16) **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'

For this particular case, the UT solution is adequate since the Spanish output is equally ambiguous:

(17) **UT Handling of Complex Contextual Ambiguity:**

E: John hit the dog with a stick
S: John pega el perro con una batuta

However, in cases where the target language depends on the contextual information, this approach would not work. A more sophisticated approach that incorporates context might resolve this ambiguity 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, but these are topics for additional research and development in the MT community.

3.6.2 Generation: Adequacy of Lexical Selection and Syntactic Realization

UT developers have adopted a strategy that coincides with the opinion of most MT researchers 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. Even so, generation

is a non-trivial exercise since it is often difficult to select target-language words—*lexical selection*—and their corresponding structural configuration—*syntactic realization*—for the conceptual knowledge underlying the IR. We address each of these problems in turn.

3.6.2.1 Lexical Selection

Lexical selection is the flipside of semantic disambiguation, e.g., the cases given above in (10):

(18) **Lexical Selection:**

(i) mesa vs. tabla

E: I put books on the table
S: Yo puse los libros en la mesa
`I put the books on the table'

E: I put data in the table
S: Yo puse los datos en la tabla
`I put the data in(to) the table'

(ii) estar vs. ser

E: I am happy
S: Estoy feliz
`(I) am happy'

E: I am a factory worker
S: Soy un obrero de fábrica
`(I) am a worker of a factory'

E: I am tall
S: Soy alto
`(I) am tall'

(iii) hope/expect vs. wait

S: Espero que Juan tenga un carro
E: I hope that Juan has a car

S: Espero a Juan
E: I hope Juan

The focus earlier was on disambiguation of the source-language input, i.e., assignment of appropriate attributes to the ambiguous term. Recall that, for the sentence *I put data in the table*, UT recognizes *data* as a direct object and then chooses the word sense `table=1-2` (i.e., as `+abstract`) because *data* and *table* have a matching `+abstract` attribute. Once this has been done, the target-language term can be selected directly from the word sense. Since *mesa* is stored as `table=1-1` and *tabla* is stored as `table=1-2`, the latter is chosen as the target-language term.

A similar strategy is used for the *ser* vs. *estar* case, i.e., *estar* is selected directly from the word sense associated with an attribute indicating a conditional emotion and *ser* is selected in the case of the alternative word sense.

These two cases distinguish UT from non-interlingual approaches in that the word senses—not the attributes associated with those word senses—are used to select the appropriate translation. A non-interlingual forces **both** the source- and target-language to include all distinguishing features, e.g., `±abstract`, `±emotion`, even if those features are irrelevant to the particular language. With 1600 source-target pairs, the proliferation of features in a non-interlingual approach would cause superfluous slowdowns due to extensive checking of irrelevant attributes.

In case (iii) above, the analysis phase is slightly different in that the syntactic structure—not attributes—is what provides the disambiguating information. Recall that this case fails due to the lack of distinguishing information in the UTR. If the appropriate word sense were selected during the analysis phase, lexical selection would operate similarly to the cases above.

Even though this third case is not handled in UT, the potential for correcting it quickly for all 40 languages is much higher than it would be in a non-interlingual approach. A non-interlingual approach would require that the MT designer build a construction-specific mapping rule relating *esperar a* to *wait* and *esperar que* to *hope*—which would not carry over to the other 1599 language pairs. Resolving the ambiguous term to a particular word sense during the analysis phase eliminates construction-specific rules of this type.

3.6.2.2 Syntactic Realization

We now turn to the handling of syntactic realization. The particular realization problem that will be examined for the remainder of this section is Tense Generation. This is a difficult area for most MT systems because, in many cases, the linguistic information required in the target language is not explicit in the source language sentence.

Consider the following example:

(19) **Tense Generation (past/present):**

C: Wo bèi California de fengjing xiyinzhù le
E: I was captivated by the scenery of California
E: I am captivated by the scenery of California

(The transliteration above is a romanized phonetic format used for expositional purposes only. The UT system provides full support of Chinese fonts and, for web browsing and other applications, Unicode is used.)

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.

Currently, UT handles this sort of decision by making an arbitrary choice, with past tense being the default:

(20) **UT Handling of Tense Generation (past/present):**

C: Wo bèi California de fengjing xiyinzhù le
E: I was attracted by Californian scenery

While this choice is perfectly reasonable, there are more problematic cases of tense generation in other languages. For example, 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:

(21) **Tense Generation (preterit/imperfect):**

(i) E: When I was young, I liked to dance
S: Cuando yo era joven, me gustaba bailar

(ii) E: When I went home, I saw the book
S: Cuando yo fui a casa, vi el libro

In the first example, *was* and *liked* are translated in the imperfect past since these are ongoing states. In the second example, *went* and *saw* are translated in the preterit past since these are momentary events.

UT falls short in its handling of such cases:

(22) **UT Handling of Tense Generation (preterit/imperfect):**

(i) E: When I was young, I liked to dance
S: Cuando yo fui joven, me gustó bailar

(ii) E: When I went home, I saw the book
S: Cuando yo fui a, casa, yo vi el libro

A full theory of tense and aspect would be needed to handle such cases. This is an area that is worthy of further investigation by UT developers. There is currently a great deal of research in this area (see recent ACL papers), and progress is being made toward handling of such cases using the notion of "compositionality." Because non-interlingual approaches are generally non-compositional in nature, these are not as amenable to the incorporation of current theoretical frameworks of tense and aspect as the interlingual design.

3.6.3 Mapping Between Language Pairs: Divergences

A third type of linguistic challenge in 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*. (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.)

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. Because the UT approach is interlingual, construction-specific rules are minimized or eliminated, but fundamental research will be necessary in several areas for complete handling of divergences. This section addresses each divergence type, in turn; suggestions for areas of future research and development in UT are given, where appropriate.

3.6.3.1 Thematic Divergences

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

(23) **Thematic Divergence:**

E: I like the books
S: Los libros me gusta
`The books (to) me pleases'

Here, *the books* appears in object position in English and in subject position in Spanish; analogously, the subject *I* appears as the object *me*. UT handles such cases adequately:

(24) **UT Handling of Thematic Divergence:**

S: Me gustan los libros
E: I like the books

E: I like books
S: Los libros me gustan

The IR is designed to support such cases since it is based on a predicate-argument structure, i.e., it includes canonical positions that are related to source- and target-language lexical items by rules in each of the respective UTRs. By contrast, a non-interlingual approach would require construction-specific mapping rules for each language pair, even in cases where surface positions would otherwise line up with canonical IR positions.

3.6.3.2 Head-Switching Divergences

Head-switching divergences are a much more difficult phenomenon. These occur commonly across language pairs, as illustrated in the following examples, where a source-language main verb is subordinated as an adverbial in the target language:

(25) **Head-Switching Divergence (French, Spanish, Dutch):**

(i) E: The baby just fell
F: Le bébé vient de tomber
`The baby just (verb-past) of fall'

(ii) S: Mary suele ir a la escuela
`Mary is accustomed to go to school'
E: Mary usually goes to school

(iii) E: John enjoys swimming
D: Jan swemt graag
`John swims likingly'

(iv) E: Mary happened to come
D: Mary kwam toevallig
`Mary came by chance'

In all four of these cases, the UT system selects a more literal translation. For completeness, the bidirectional translations are given below:

(26) **UT Handling of Head-Switching Divergence (French, Spanish, Dutch):**

(i) E: The baby just fell
F: Le bébé tomba justement

F: Le bébé vient de tomber
E: The baby comes falling

(ii) S: Mary suele ir a la escuela
E: Mary is used to go to the school

E: Mary usually goes to school
S: Mary va usualmente a escuela

(iii) E: John enjoys swimming
D: John genieten zwemmen

D: John swemt graag
E: John swim gladly

(iv) E: Mary happened to come
D: Mary gebeurd jegens komen
`Mary happened on/against come'

D: Mary kwam toevallig
E: Mary come accidental

In (i) and (ii), the literal translation is acceptable, although not preferred. In (iii) and (iv), the translations are not acceptable. As it turns out, Dutch is the only Level I language represented in the examples above. In principle, it would not be be difficult to extend UT to handle such cases for Level II coverage. In fact, the German analog to (iii) is already handled since German is currently a Level II language. Consider the following case:

(27) **Head-Switching Divergence (German):**

E: I like to eat
G: Ich esse gern
`I eat likingly'

The English sentence is translated into the perfectly acceptable German literal equivalent using the main verb *mögen* (*to like*) and the reverse direction produces a correct, non-literal translation:

(28) **UT Handling of Head-Switching Divergence (German):**

(i) E: I like to eat
G: Ich mag essen

(ii) G: Ich esse gern
`I eat likingly'
E: I like to eat

As in the thematic divergence cases, the UT IR is designed to support such cases through the use of canonical positions that are related to source- and target-language lexical items by rules in each of the

respective UTRs. A non-interlingual approach would require a construction-specific mapping rule to relate source- and target-language structures, e.g., one like the following:

(29) **Canonical Transfer Rule for `gern/like':**
G: PRED GERN<ARG> = E: LIKE<SUBJ,XCOMP(TO)> PRED

This equation identifies the adverbial *gern* as the corresponding English main verb *like*. The disadvantage to this approach is that it forces the source- and target-language structures to be 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 internal structure in this framework. For example, if the *gern esse* clause is further embedded as *John glaubt, daß Ich gern esse*, the rule-mapping approach becomes inordinately complicated because the PRED (= *essen*) constituent on the German side of equation (27) is associated not just with *gern*, but with *glauben*. In effect, another canonical transfer rule would be needed to bring about this second association:

(30) **Canonical Transfer Rule for `glauben/believe':**
G: GLAUBEN<SUBJ,XCOMP(DAß)> PRED = E: BELIEVE<SUBJ,XCOMP(TO)> PRED

The problem is that the PRED on the German side of each equation above corresponds to one lexical item, *essen*, but the PRED on the English side corresponds to two lexical items, *eat* (in (29)) and *like* (in (30)).

The interlingual design of UT accommodates such cases by first mapping the arguments of *gern* into their canonical IR positions (effectively eliminating the need for rule (29)) and then selecting the appropriate word senses before generating the target-language form (effectively eliminating the need for rule (30)):

(31) **UT Handling of Embedded Head-Switching Divergence (German):**

G: John glaubt, daß ich gern esse.
E: John believes that I like to eat.

UT developers need to focus on handling head-swapping cases like this for languages other than German. This will involve modifying other UTRs so that they accommodate the mapping from non-canonical syntactic positions into canonical IR positions for such cases.

3.6.3.3 Structural Divergences

Structural divergence is another obstacle to straightforward mappings between language pairs. In structural divergence, a verbal argument has a different syntactic realization in the target language:

(32) **Structural Divergence:**

E: John entered the house
S: John 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. UT fails to handle this case:

(33) **UT Handling of Structural Divergence (Spanish):**

(i) E: John entered the house
S: John entró la casa

(ii) S: John entró en la casa
E: John entered in the house

In principle, such cases are easily corrected in UT with a simple change to the Spanish UTD so that *entrar* is associated with the preposition *a*.

3.6.3.4 Categorial Divergences

Categorial divergence is a more complicated case. This 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:

(34) **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 category change forces the generator to select a different main verb.

Although UT is designed to accommodate certain types of category changes (e.g., replacing *good* with *well* in *It goes good.*), the case above is a bit more complicated and will require a more compositional representation of the main verb than a simple word sense. For now, UT handles this case with a less preferred translation of the English, and an awkward translation of the German:

(35) **UT Handling of Categorial Divergence (German):**

(i) E: I am hungry
G: Ich bin hungrig

G: Ich habe Hunger
E: I have hunger

Because such cases arise frequently, e.g., *Yo tengo calor* (`I am hot'), UT developers should put some effort toward the handling of categorial divergence. The payoff for doing so will come later, when handling even more complicated cases of categorial divergence. Consider, for example, the following case:

(36) **Complex Categorial Divergence (French):**

E: John is fond of music
F: John aime la musique
`John likes 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*. Currently, such cases are handled with a French-English translation that is much less natural than its English-French counterpart:

(37) **UT Handling of Complex Categorial Divergences (French):**

(i) E: John is fond of music
F: Jean est affectueux de musique

(ii) F: John aime la musique
E: John likes the music

Even more complex cases arise with additional modifying phrases:

(38) **More Complex Categorial Divergence (French):**

E: John is very fond of music
F: John aime beaucoup la musique
`John likes 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*.

For such cases, the UT translations are not as natural as the sentences above:

(39) **UT Handling of More Complex Categorial Divergence (French):**

(i) E: John is very fond of music
F: Jean est très affectueux de musique

(ii) F: John aime beaucoup la musique
E: John likes a lot the music

Transfer systems are forced to handle such cases by storing entire trees in the transfer dictionary for each source-to-target pair. This is significantly burdensome as the number of source and target languages begin to add up. Although UT does not currently resolve divergences this complex, using an interlingual representation is the first step toward successful handling of such cases without the overhead of building an unwieldy rule set. UT could benefit significantly from devoting resources toward accommodating such cases.

3.6.3.5 Conflational Divergences

Conflational divergence is another obstacle to straightforward development of language-pair mappings. 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:

(25) **Conflational Divergence (Spanish):**

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.

UT handles such cases without taking into account argument incorporation:

(26) **UT Handling of Conflational Divergence (Spanish):**

(i) E: I stabbed John
S: Le apuñalé a John

(ii) S: Yo le di puñaladas a Juan
E: I gave stabs to Juan

As in the case of Categorial divergences, UT could benefit significantly from devoting resources toward accommodating such cases.

3.7 Summary

A summary of the technology evaluation described above is given the following table:

Capability or Characteristic	Assessment and/or Recommendation
Bidirectional	Yes
Number of languages	40
Quality	Level III (0), Level II (13), Level I (27)
Ease of Upgrades/Extensibility	Yes
Iterative Development/Testing	Yes (UT Quality Control Module)
Time/Effort for Scaling up	I→II (average), II→III (average),
Time/Effort for Incremental Improvement	Low
Framework for reusing modules	Yes (Open Architecture)
Future system development	Significant effort required
Lexical Ambiguity	Covered
Semantic Ambiguity	Most cases covered
Complex Semantic Ambiguity	Needs significant research effort
Contextual Ambiguity	Needs significant research effort
Complex Contextual Ambiguity	Needs significant research effort
Lexical Selection	Most cases covered; overcomes transfer unwieldiness
Syntactic Realization	Some cases covered; Needs significant research effort (e.g., tense)
Thematic Divergence	Covered
Head-switching Divergence	Some cases covered; extension to others straightforward
Structural Divergence	Some cases covered; extension to others straightforward
Categorial Divergence	Not covered; Needs significant research effort
Complex Categorial Divergence	Not covered; Needs significant research effort
Conflational Divergence	Not covered; Needs significant research effort

Table 1: Assessment and Recommendations for UT System

4. Application-Oriented Evaluation

Evaluation of MT systems is an active area of research. Given that "perfect" translation is not within our grasp now, if ever, we still need to decide how we can judge the efficacy of a given MT system. 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. 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.

However, as stated by Jordan in Dorr et al. (1999), there are some important criteria that should be imposed on any MT system. This section focuses on these and presents a brief comparison of a UT translation to that of other commercial systems.

4.1 Evaluation Metrics for UT

Section 3 assessed different components, algorithms, and I/O behavior of the UT system, combining *black box* techniques (i.e., examining the accuracy of the input/output pairs) with *glass box* techniques (i.e., examining the architecture of the system and the data flow between the system components). The black-box component of the assessment involved an examination of translation quality on certain types of examples in order to evaluate UT's linguistic coverage with respect to ambiguity, word choice and realization, and divergence classes. The black-box component of the assessment involved an examination of the system's internal representations and processing strategies in order to measure how well the system handled these examples. Throughout this assessment, we have implicitly and explicitly compared the potential for the UT interlingual approach to handle complex cases to that of the non-interlingual approaches.

Considering these basic evaluation approaches, what then are reasonable and useful evaluation criteria—beyond quality assessment—for MT systems? 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."
Taking into account these comments, we adopt the metrics shown in Table 2—as given in Dorr et al. (1999)—for an overall evaluation of the UT system. Each of these will be addressed, briefly, in turn.

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

Table 2: Metrics for Evaluation of UT

4.1.1 Intelligibility

In general, the intelligibility of UT is somewhat low. In many cases, the UT translations are not as natural as the preferred translation (or as the translation in the opposite direction). Often, a literal translation is used, even in Level II languages; this translation appears to be acceptable about half the time (see, e.g., the four cases in (26) above).

4.1.2 Fidelity

UT scores high on fidelity. This metric might be seen to conflict with the last one, in that the the preservation of source-language characteristics—particularly structural ones—often results in a very "unnatural" translation. On the other hand, if high-quality output is not guaranteed, as it currently is not in the UT system, adhering to fidelity is the next best option.

4.1.3 Acceptability

The acceptability of UT with respect to particular domains and applications is not yet known since it is currently set up to be a general translator by the population at large. In the cross-language information retrieval (CLIR) world, UT is as useful as many of the other commercial systems, e.g., for selecting documents in a language other than the query language. Because UT supports an Open Architecture, it is expected that individual system modules can be used for principled development of other types of applications as well.

4.1.4 Speed

UT runs at an average speed, as commercial systems go. A 15-word sentence takes under 5 seconds. The speed benefits should become more apparent as UT system development progresses. UT is likely to come out ahead of transfer systems which, if forced to handle 40 languages at once, would require inordinately large rule sets that would inevitably result in significant slowdowns. Another component to the speed-efficiency of UT is the early resolution of word senses in analysis. Performing disambiguation prior to construction of the syntactic dependency tree pares down the number of possible syntactic combinations and quickly reduces candidate trees to a small enough number that one can be selected during the filtering stage.

4.1.5 Cost

Because the interlingual approach does not use transfer rules, much of the cost associated with laborious development of pairwise mappings is entirely eliminated. In addition, the UTL accommodates a universal set of attributes and relations, thus enabling cost-efficient management of a large number of languages. In particular, the addition of new language-specific phenomena does not rely on regeneration of underlying constructions from scratch.

4.1.6 Time spent for revision

There is a great discrepancy between Level I and Level II languages as we saw earlier in (4), (5), and (6). Significant time would be required for Level I languages—most likely one minute per one sentence, assuming 1–2 errors per sentence. For Level II languages, the time spent for revision is average in comparison to other commercial systems, assuming one error in every other sentence (see next metric). Note that there is a separate time factor for developers that should be considered: the time spent in revision of the system for quality enhancements. As noted earlier, a Level II language is enhanced relatively quickly—in a matter of minutes—for incremental improvements because the addition of a language-specific construction to one language does not require augmentations to any of the other languages.

4.1.7 Number of errors

Again, there is a great discrepancy between Level I and Level II languages: Level I languages are likely to have 1–2 errors per sentence; Level II languages are likely to have one error every other sentence, i.e., an average number of errors in comparison to other commercial systems.

4.1.8. Cross-linguistic applicability

Cross-linguistic applicability of the system is high. The interlingual design accommodates a wide range of languages using standards that have been set in advance. There are no transfer lexicons or bilingual mappings required for the system. Each UTD (i.e., each language-dependent lexicon) may be used in either direction—for analysis and generation. Much of the development that would generally be required for *most* MT systems has been done in advance; this contrasts with approaches adopted by most other commercial systems, where the focus is typically on generating one particular language pair. (Additional details were covered above in Section 3.1.)

4.1.9 Extensibility

Extensibility of the system is high. UT includes a sophisticated development environment that lends itself readily to easy maintenance and upgrades/extensions. The development methodology includes tools that accommodate incremental improvements—adding new words to dictionaries and fine-tuning grammar rules. Rules are created in UTL, a standardized format that is designed to be easily used by linguists. (Additional details were covered above in Section 3.2.)

4.1.10 Uniform analysis and synthesis

All UTDs (i.e., all language-dependent lexicons) may be used in either direction—for analysis and generation. Rules used in analysis are different from those used in synthesis, but both rule types are based on the same standardized format (UTL).

4.1.11 Fault-Tolerance

In general, errors are handled by the fall-back position that a literal translation is provided; in all cases, a translation is always provided rather than none at all.

4.1.12. Collaboration

All languages in the system operate on the basis of the same data format for dictionaries, rules (and their associated attributes and relations), and interlingual representation.

4.1.13 Summary

A summary of the results of applying the metrics above to the UT system is given the following table:

Metric	Assessment
Intelligibility	Low
Fidelity	High
Acceptability	To be determined (Open Architecture evaluation required)
Speed	Average (should come out ahead of transfer as development proceeds)
Cost	Low
Time spent for revision	Level I (high); Level II (average)
Number of errors	Level I (high); Level II (average)
Cross-linguistic applicability	High
Extensibility	High
Uniform analysis and synthesis	Yes
Fault Tolerance	Yes
Collaboration	Yes

Table 3: UT Assessment

4.2 Comparison of UT Output to Output of Other Systems

MT evaluation in the research community focuses on the handling of different linguistic phenomena by various MT systems. This was the focus of the 1998 AMTA Workshop on Interlinguas which examined the translation of a text from the UNESCO Courier into multiple languages. The objective of the workshop was to determine how different aspects of the text would be represented using an Interlingual Representation. Specific challenges addressed in this forum were the handling of basic predicate-argument structure, noun phrases/referents, proper nouns, prepositional meaning, non-literal language, temporal relations, textual organization, lexical divergences, syntactic divergences.

We will adopt an approach taken by Mitre in which an excerpt from the UNESCO Courier was examined for determining 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) **English-Spanish Translation of UNESCO Courier:**

(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 and (2) Systran. The following results were produced:

(39) **Logos/Systran Translation into Spanish:**

(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.

We ran this same text through the English-Spanish component of UT and arrived at the following translation:

(40) **UT Translation into Spanish:**

ACCION Internacional es un U.S.-based que organización de non-profit privado corrientemente provee a una cadena de las instituciones en trece países en América Latina y seis ciudades en los EU la ayuda técnica. Su cadena de dieciocho organizaciones independientes en América Latina tiene cuaresma sobre $1 a microenterprises en el last cinco 1000000000 en préstamos años, prorrateando menos de $500.

All three translations into Spanish were readable by a native speaker, but not entirely grammatical and not at all similar to the version produced 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, *proporciona actualmente a asistencia técnica a una red de instituciones* (`supplies presently to the technical assistance to a web of institutions') by Systran, and *corrientemente provee a una cadena de las instituciones en trece países en América Latina y seis ciudades en los EU la ayuda técnica* (`currently provides to a chain of institutions ... technical help') by UT. None of these three translations corresponds exactly to the human-generated Spanish version *brinda actualmente asistencia técnica a una red de establecimientos* (`offers currently technical assistance to a web of establishments'). In particular, the mapping between *provides* and *brindar* appears to be missing in all three 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. Since Logos does not translate in the Spanish-to-English direction, this test was performed by Systran, Globalink (which has since been bought by Lernout & Hauspie), and UT:

(40) **Systran/Globalink/UT Translation into English:**

(i) Systran: ... that offers currently technical assistance to a network

(ii) Globalink: ... that at the moment offers to technical attendance to a network

(iii) UT: ... that offers technical assistance currently purposes a net

From the translations above, the transfer mappings (for the first two) and the interlingual entries for the last one are deduced to be of the following form:

(42) **Transfer Mapping for Systran:**

brindar(Subj,Dir_Obj,(a Ind_Obj)) → offer(Subj,Dir_Obj,(to Ind_Obj))

Transfer Mapping for Globalink:

brindar(Subj,Dir_Obj) → offer(Subj,(to Dir_Obj))

Interlingual entries for UT:

brindar(Subj,Dir_Obj), offer(Subj,Dir_Obj)

From these we see three things:

1. The verb *provide* is not an English equivalent of *brindar* in any of the three systems.

2. The phrase *a una red* (`to a network') is viewed as a modifier in Globalink and UT and is taken to be outside of the predicate-argument structure of the verb in both cases.

3. 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.

4. In the UT translation, the preposition *a* is taken (erroneously) to be a purpose clause, which is the reason for the *purposes a net* clause.

The Globalink and UT translations fare equally poorly in the translation of this particular phrase. Moving in the direction of a more complete interlingual approach would significantly improve the output of UT. Even a simplified form of interlingua, such as one based on thematic roles (`ag`, `th`, etc.), provides the basis of a more systematic mapping into the target-language. For example, the following thematic-role representations might be used for the English and Spanish verbs used above:

(43) **Thematic-Role Representations:**

(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*. Moreover, the preposition *a* is not erroneously taken to be a purpose because the existence of a `goal(a)` clause forces this to be selected as an internal argument of the verb.

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. A deeper analysis runs the risk of prohibitive storage and access costs as systems are scaled up for broader coverage. On the other hand, transfer systems provide lower-quality translations and require a great deal of human assistance. UT is representative of a more recent trend toward 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.

5. SUMMARY AND CONCLUSIONS

This report has assessed the UT system in the face of major challenges including architectural adequacy, depth of linguistic coverage, and cross-linguistic applicability. Developers of UT have taken the approach of attempting broad coverage first; the next step is to upgrade the quality of the result. Since large amounts of knowledge are necessary (but not sufficient) for better quality results and broader coverage, it is highly recommended that UT developers conduct further research on methods for acquiring and utilizing information more easily.

Over the next several years, if UT developers are willing to put a significant amount of effort toward understanding and characterizing cross-linguistic generalizations, it is expected that the system will achieve broader coverage. Toward that end, some linguistic challenges that have been recommended for future investigation are certain types of ambiguity (semantic and contextual), syntactic realization (e.g., a theory of tense), and certain types of divergences (e.g., head-switching, structural, categorial, and conflational). The enhancement of UT to include a richer approach to these challenges would address the issue of intelligibility (overall quality), which received the lowest rating in the final UT Assessment.

Another area that deserves attention is the development of new applications using modules that already exist in the UT system. The potential for developing certain types of applications is high, e.g., document retrieval and summarization, database question-answering, and translation services. Other types of applications would require a significant amount of research and development, e.g., Conversational/Dialogue applications, Understanding Systems, Automatic Knowledge Acquisition, and an Interlingual Development Environment.

The interlingual design of UT is advantageous from the point of view of cross-linguistic coverage. The UT approach contrasts with that of other commercial systems, where the focus is typically on generating one particular language pair—or a small set of pairs—without regard to what it would take to ramp up quickly to other (potentially diverse) languages. In UT, the analyzer and generator need not be reprogrammed when new languages are added to the system; moreover, there are no transfer lexicons or bilingual mappings required for the system. Each language-dependent lexicon may be used in either direction—for analysis and generation.

In the short term, the UT design does not guarantee a higher translation quality than that of other commercial designs. The main design benefit is not this—at least not initially—but that of *extensibility*, i.e., ease of ramping up to handle a wide range of languages quickly. It is only in the long term that the potential for higher quality is evident: as each language is added, the time that is typically spent building and/or reprogramming system modules and for each language pair (i.e., n^2 sets of bilingual rules) can be spent fine-tuning the lexicons and grammars of the new language.

Although the transition from a Level I language to a Level II is somewhat in duration (approximately 2 years), this time period is not unreasonable given that quality enhancements are achieved with ease beyond this point (i.e., a significant reduction in time for incremental improvements). For example, adding a new lexical entry is quick, simple, and significantly reduced the output awkwardness. On the other hand, a great

deal of amount of work would be needed to bring the level of translation up to Level III; the transition between Level II and Level III will require a great deal of research and development.

UT includes a sophisticated development environment Quality Control Module that lends itself readily to easy maintenance (e.g., error detection) and upgrades/extensions. The development methodology includes tools that accommodate incremental improvements—adding new words to dictionaries and fine-tuning grammar rules in 6 person-months—by computational linguists whose training time is approximately one month. The reason that this development time is significantly reduced is that much of the development that would generally be required for most MT systems—i.e., the information that needs to be added as each new language is adopted into the system—has been done in advance, during the early planning years (starting in 1995), when all morphological, syntactic, and semantic features and attributes were standardized to one large, uniform set (the IRD) designed to accommodate a very broad range of language types.

6. BIBLIOGRAPHIC REFERENCES

Space limitations preclude the complete list of citations underlying the discussion of research approaches, architecture, and evaluation techniques described above. For a complete listing (233 bibliographic entries), see the following chapter:

Dorr, Bonnie J., Pamela W. Jordan, and John W. Benoit, "A Survey of Current Research in Machine Translation," Advances in Computers, Vol 49, M. Zelkowitz (Ed), Academic Press, London, 1999.

The 12 criteria used in the MT evaluation above are the ones that are considered by Dorr, Jordan, and Benoit to be the most critical in the evaluation of broadscale MT systems. These are a distillation of several different criteria developed by numerous researchers in the field; see their chapter above for additional discussion.

A. APPENDIX: TEST SUITE FOR SOME LEVEL II LANGUAGES

The format for all examples below is:

L_1 [~ L_1]* # L_2 [~ L_2]*

Sentences to the left and right of the sharp signs (#) are both accepted and generated by UT. The (optional) sentences separated by tilde (~) correspond to alternates that are accepted and/or generated by the system.

German Testing Sentences

If she wrote it, I would go. ~ If she wrote it I would go. # Wenn sie es schriebe, würde ich gehen.
He is the boy who isn't writing. ~ He is the boy who does not write. # Er ist der Junge, der nicht schreibt.
I'm writing a good book. ~ I write a good book. # Ich schreibe ein gutes Buch.
He gives me the book. # Er gibt mir das Buch.
I write in the book. # Ich schreibe in dem Buch. ~ Ich schreibe im Buch.
It is always written. ~ It always is written. # Es wird immer geschrieben.
He is the boy who isn't writing the book quickly. ~ He is the boy who does not write the book quickly. # Er ist der Junge, der das Buch nicht schnell schreibt.
Mr. Jones writes the book. # H. Jones schreibt das Buch.
The big, white, and beautiful book is good. # Das große, weiße, und schöne Buch ist gut.
It is and was good. # Es ist und war gut.
Our big, white, and beautiful book is good. # Unser großes, weißes, und schönes Buch ist gut.
The book is mine. ~ The book belongs to me. # Das Buch gehört mir.
Which white letters are good? ~ What white letters are good? # Welche weißen Briefe sind gut?
Whose white books are good? ~ Who has white books that are good? # Wer hat weiße Bücher, die gut sind?
The boy's white book is good. # Das weiße Buch des Jungen ist gut.
He does not write very quickly. # Er schreibt nicht sehr schnell.
Don't write! ~ Do not write! # Schreiben Sie nicht!
He is good, isn't he? # Er ist gut, nicht wahr?
He doesn't write, does he? # Er schreibt nicht, oder?
What's that? ~ What's that? # Was ist das?
Who is this? ~ Who is that? # Wer ist das?
Why does the boy write? # Warum schreibt der Junge?
Where does the boy write? # Wo schreibt der Junge?
Where is the boy? # Wo ist der Junge?
When does the boy write? # Wann schreibt der Junge?
How does the boy write? # Wie schreibt der Junge?
Which boy writes? ~ What boy writes? # Welcher Junge schreibt?
I see him and her. # Ich sehe ihn und sie.

Swedish Testing Sentences

We had not been good. # Vi hade inte varit goda.
If she wrote it I would go. # Om hon skrev det skulle jag gå .
If she were good I would go. ~ If she was good I would go. # Om hon vore god skulle jag gå .
I write and he sleeps. # Jag skriver och han sover.
I know he is good. # Jag vet han är god.
He is the boy who writes. # Han är pojken som skriver.
They are the dogs which sleep. ~ They are the dogs that sleep. # De är hundarna som sover.
They are the dogs that sleep. # De är hundarna som sover.
He is a boy who writes. # Han är en pojke som skriver.
He is a dog which sleeps. ~ He is a dog that sleeps. # Han är en hund som sover.
He is a dog that sleeps. # Han är en hund som sover.
They are boys who write. # De är pojkar som skriver.
They are dogs which sleep. ~ They are dogs that sleep. # De är hundar som sover.
They are dogs that sleep. # De är hundar som sover.
He is the boy who isn't writing. ~ He is the boy who does not write. # Han är pojken som inte skriver.
She is the girl who writes. # Hon är flickan som skriver.
I'm writing a good book. ~ I write a good book. # Jag skriver en god bok.
I write books. # Jag skriver böcker.
I write good books. # Jag skriver goda böcker.
You see me. # Du ser mig.
I see her. # Jag ser henne.
I see you. # Jag ser dig.
I write this. # Jag skriver det här.
I write that. # Jag skriver det där.
I write these. # Jag skriver de här.
I write those. # Jag skriver de där.
He gives me the book. # Han ger mig boken.
I give you the book. # Jag ger dig boken.
I give him the book. # Jag ger honom boken.
I give her the book. # Jag ger henne boken.
I give it the book. # Jag ger det boken.
I give us the book. # Jag ger oss boken.
I give them the book. # Jag ger dem boken.
I write in the book. # Jag skriver i boken.
He is always good. # Han är alltid god.
It is always written. ~ It always is written. # Det blir alltid skrivet.
I write the book quickly. # Jag skriver boken snabbt.
I don't write the book quickly. ~ I do not write the book quickly. # Jag skriver inte boken snabbt.
He is the boy who writes the book quickly. # Han är pojken som skriver boken snabbt.
He is the boy who isn't writing the book quickly. ~ He is the boy who does not write the book quickly. # Han är pojken som inte skriver boken snabbt.
Mr. Jones writes the book. # Herr Jones skriver boken.
The big white book is good. # Den stora vita boken är god.
The big, white, and beautiful book is good. # Den stora, vita, och vackra boken är god.
It is good and white. # Det är gott och vitt.
It is and was good. # Det är och var gott.
A big white book is good. # En stor vit bok är god.
A big, white, and beautiful book is good. # En stor, vit, och vacker bok är god.
Our big, white, and beautiful book is good. # Vå r stora, vita, och vackra bok är god.
The book and the boy are good. ~ The book and boy are good. # Boken och pojken är goda.
The book is mine. # Boken är min.
The book is yours. # Boken är din.
The book is hers. # Boken är hennes.
The book is his. # Boken är hans.
The book is its. # Boken är dess.
The book is ours. # Boken är vå r.

Spanish Testing Sentences

I see him and her. # Veo a él y ella.
I see a woman, a man, and a child. # Les veo a una mujer, un hombre, y un niño.
A woman, a man, and a child are there. # Una mujer, un hombre, y un niño están ahí.
If the woman who is here writes it, I will go. # Si la mujer que está aquí lo escribe, yo iré.
If the woman who is always here writes it, I will go. # Si la mujer que está siempre aquí lo escribe, yo iré.
If the woman who is not always here writes it, I will go. # Si la mujer que no está siempre aquí lo escribe, yo iré.
I see the house he bought. ~ I see the house that he bought. # Veo la casa que él compró.
It is the house he lives in. ~ It is the house in which he lives. # Es la casa en que él vive.
It is the house in which he lives. # Es la casa en que él vive.
It is the house in which he has his furniture that is old. ~ It is the house in which he has his furniture which is old. # Es la casa en que él tiene su menaje que es viejo.
The house he bought is big. ~ The house that he bought is big. # La casa que él compró es grande.
She is old, not him. ~ She is old, not he. # Ella es vieja, no él.
She is older than he is. ~ She is older than he. # Ella es más vieja que él.
She is older than he. # Ella es más vieja que él.
She is older than him. ~ She is older than he. # Ella es más vieja que él.
She is the oldest of the women. # Es la más vieja de las mujeres.
She is the oldest of them all. ~ She is the oldest of them. # Es la más vieja de ellas.
I know that man. # Le conozco a ese hombre.
I know that fact. # Sé ese hecho.
I know that it's true that she is sleeping. ~ I know that it is true that she is sleeping. # Sé que es verdadero que ella está durmiendo.
I know that it's true she is sleeping. ~ I know that it is true that she is sleeping. # Sé que es verdadero que ella está durmiendo.
I know it's true that she is sleeping. ~ I know that it is true that she is sleeping. # Sé que es verdadero que ella está durmiendo.
I know it's true she is sleeping. ~ I know that it is true that she is sleeping. # Sé que es verdadero que ella está durmiendo.
I want to go if you want to go. # Quiero ir si usted quiere ir.
I want to go if you want to. ~ I want to go if you want. # Quiero ir si usted quiere.
He came early to prepare. ~ He came to prepare early. # Vino temprano para preparar.
I go to the house from the school. ~ I go from the school to the house. # Voy de la escuela a la casa.
I go from the school to the house. # Voy de la escuela a la casa.
I will be there from five to six. # Estaré ahí de las cinco a las seis.
The man he saw was tall. The man that he saw was tall. # El hombre que él vio fue alto.
The Russian speaks Russian. # El ruso habla ruso.
Let's see the house! # ¡Veamos la casa!
I had him go. ~ I had it go. # Lo hice ir.
I got it fixed. # Lo hice arreglar.
How many cats do you have? # ¿Cuántos gatos tiene usted?
How much does it cost? ~ How much does he cost? # ¿Cuánto cuesta?
I know why she is sleeping. # Sé por qué ella está durmiendo.
I know where she is sleeping. # Sé donde ella está durmiendo.
I know how she is sleeping. # Sé como ella está durmiendo.
I know when she is sleeping. # Sé cuando ella está durmiendo.
I know who is sleeping. # Sé quién está durmiendo.
I know how much she is sleeping. # Sé cuánto ella está durmiendo.
We go whenever we have the chance. ~ We go whenever we have the opportunity. # Vamos siempre que nosotros tengamos la oportunidad.
We go wherever we choose. # Vamos adondequiera que nosotros escojamos.
I like him, whoever he is. # Él me gusta, quienquiera que él sea.
I don't know what it is. ~ I do not know what it is. # No sé lo que eso es.
I don't know what he was doing. ~ I do not know what he was doing. # No sé lo que él estaba haciendo.
I want that one. ~ I want that. # Quiero eso.
That one is good. ~ It is good. # Eso es bueno.

Russian Testing Sentences

I see him and her. # Я вижу его и ее.
I see a woman, a man, and a boy. ~ I see woman, man and boy. # Я вижу женщину, мужчину и мальчика.
A woman, a man, and a boy are there. ~ Woman, man and boy are there. # Женщина, мужчина и мальчик - там.
If the woman who is here writes it, I will go. ~ If woman who is here writes this, I will go. # Если женщина, которая здесь, это напишет, я пойду. ~ Если женщина, которая здесь это напишет, я поеду.
If the woman who is always here writes it, I will go. ~ If woman who is always here writes this, I will go. # Если женщина, которая всегда здесь, это напишет, я пойду. ~ Если женщина, которая всегда здесь это напишет, я поеду.
If the woman who is not always here writes it, I will go. ~ If woman who is not always here writes this, I will go. # Если женщина, которая не всегда здесь, это напишет, я поеду. ~ Если женщина, которая не всегда здесь это напишет, я поеду.
I see the house he bought. ~ I see house that he bought. # Я вижу дом, который он купил.
It is the house he lives in. ~ This is that house in which he lives. # Это - тот дом, в котором он живет.
It is the house in which he lives. ~ This is that house in which he lives. # Это - тот дом, в котором он живет.
This is the house in which he has his furniture that is old. ~ This is that house in which he has his furniture that is old.# Это - тот дом, в котором у него есть своя мебель, которая старая.
The house he bought is big. ~ House that he bought is big. # Дом, который он купил - большой.
She is old, not him. ~ She is old, not he. # Она - старая, не он.
She is old, he isn't. ~ She is old, he is not. # Она - старая, он - нет.
She is older than he is. ~ She is older than he. # Она - старше, чем он.
She is older than he. # Она - старше, чем он.
She is older than him. ~ She is older than he. # Она - старше, чем он.
She is the oldest of the women. ~ She is the oldest of women. # Она - самая старая из женщин.
She is the oldest of them all. # Она - самая старая из них всех.
I know that. ~ I know this. # Я это знаю.
I know that man. # Я знаю того мужчину.
I know that fact. # Я знаю тот факт.
I know that it's true that she is sleeping. ~ I know that this is true that she sleeps. # Я знаю, что это - правдиво, что она спит.
I know that it's true she is sleeping.~ I know that this is true that she sleeps. # Я знаю, что это - правдиво, что она спит.
I know it's true that she is sleeping. ~ I know that this is true that she sleeps. # Я знаю, что это - правдиво, что она спит.
I know it's true she is sleeping. ~ I know that this is true that she sleeps. # Я знаю, что это - правдиво, что она спит.
I want to go if you want to go. # Я хочу поехать, если вы хотите поехать.
I want to go if you want to. ~ I want to go if you want. # Я хочу поехать, если вы хотите.
He came early to prepare. ~ He came to prepare early. # Он пришел рано подготовить.
I go to the house from the school. ~ I go from school to house. # Я еду в дом из школы.
I go from the school to the house. ~ I go from school to house. # Я еду из школы в дом.

Japanese Testing Sentences

I have written. ~ I wrote. # 私は書いた。 ~ 私は書いたことがある。
I wrote. # 私は書いた。
You wrote. # あなたは書いた。
I was writing. # 私は書いていた。
I had written. ~ I wrote. # 私は書いた。 ~ 私は書いたことがあった。
I will write. # 私は書くだろう。
I will be writing. # 私は書いているだろう。
I will have written. ~ I will be writing. # 私は書いているだろう。 ~ 私は書いただろう。
I shall write. ~ I will write. # 私は書くだろう。
I shall be writing. ~ I will be writing. # 私は書いているだろう。
I shall have written. ~ I would have written. # 私は書いただろう。
I would write. ~ I will write. # 私は書くだろう。
I would be writing. ~ I will be writing. # 私は書いているだろう。
I would have written. # 私は書いただろう。
I should write. # 私は書くべきだ。
I should be writing. # 私は書いているはずだ。 ~ 私は書いているべきだ。
I should have written. # 私は書くべきだった。
I ought to write. ~ I should write. # 私は書くべきだ。
I ought to be writing. ~ I should be writing. # 私は書いているはずだ。 ~ 私は書いているべきだ。
I ought to have written. ~ I should have written. # 私は書くべきだった。
I must write. # 私は書くにちがいない。 ~ 私は書かなければならない。
I must be writing. # 私は書いていなければならない。 ~ 私は書いているにちがいない。
I must have written. # 私は書いたにちがいない。
I may write. # 私は書くかもしれない。
I may be writing. # 私は書いているかもしれない。
I may have written. # 私は書いたかもしれない。
You may have written. # あなたは書いたかもしれない。
I might write. ~ I may write. # 私は書くかもしれない。
They might be writing. # 彼らは書いているかもしれなかった。
I might have written. ~ I may have written. # 私は書いたかもしれない。
I can write. # 私は書くことが出来る。
I can be writing. ~ I may be writing. # 私は書いているかもしれない。

Chinese Testing Sentences

Now I cut the potato into pieces. ~ I cut potato into pieces now. # 我現在切馬鈴著成片。
A very happy girl came in. ~ Very happy girl comes in. # 非常快樂的女孩進來。
I got a postcard from Helen last week. ~ Last week I got postcard from Helen. # 我上星期從海倫得到明信片。
It seems quite natural that he should have failed in the examination. ~ That he should fail in the exam seems quite natural. # 他應該沒有通過考試似乎相當自然。
She is seventeen years old. # ・是十七　。
He is two years older than she. ~ He is two years older than her. # 他是比・老二　。
All the paintings that he collected are now on sale. ~ All the paintings which he collected are on sale now. # 全部他收集過的繪畫現在是在打折。
How hot it is! ~ It is so hot! # 它是如此地熱！
What a man he is! ~ He is such a man! # 他是如此的男人！
Please stop your complaining. ~ Stop your complaint please! # 請停止・的抱怨！
He was such a good runner that I could not catch him. ~ He is such a good runner that I can't catch him. # 他是如此好的賽・者以致我無法・上他。
It was so hot that I could not sleep. ~ It is so hot that I can't sleep. # 它是如此熱以致我無法睡。
The factory employs both male and female workers. ~ Factory employs male and female workers. # 工廠雇用 男與女工人。
You had better chain up the dog in case he bites. ~ You should chain up dog in case he bites. # ・應該・好狗以防他咬。
As the lion is king of beasts, so is the eagle king of birds. ~ As lion is king of beasts, so is eagle king of birds. # 就像獅子是萬獸之王，同樣地鷹是萬禽之王。
This is the habit I want to get rid of. ~ This is the habit which I want to get rid of. # 這是我要去除的習慣。
She was born at 10 a.m. on January 7th in the year 1979. ~ She was born on January 7th at 10 am in the year 1979. # ・在1979年一月7日早上10點出生。
She was born on the 26th of October. ~ She was born on October 26th. # ・在十月26日出生。
The tower is 333 meters high. ~ Tower is 333 meters tall. # 高樓是333公尺高。
The lake is fifteen meters deep. ~ Lake is fifteen meters deep. # 湖是十五公尺深。
If the woman who is here writes it, I will go. # 如果是在這裡的女人寫它，我會去。
If the woman who is always here writes it, I will go. # 如果是始終在這裡的女人寫它，我會去。
If the woman who is not always here writes it, I will go. # 如果不是始終在這裡的女人寫它，我會去。
I see the house he bought. ~ I see the house which he buys.　# 我看見他買的房子。
It is the house in which he has his furniture that is old. ~ It is the house in which he has his old furniture.　# 它是他有他的老的家具的房子。
The house he bought is big. ~ The house which he buys is big.　# 他買的房子是大的。
She is old; he isn't. ~ She is old; he is not.　# ・是老的；他不是。
She is older than he. # ・是比他老的。
She is the oldest of the women. # ・是女人中最老的。