[*MT Summit III*, Washington, DC, USA, July 1991]

Applications of MT Technology Panel Discussion

Chair: Sergei Nirenburg, Carnegie Mellon University, USA
Rod Johnson, IDSIA, Lugano, Switzerland
Richard Kittredge, University of Montreal, Canada
Lori Levin, Carnegie Mellon University, USA
Hiroshi Uchida, Fujitsu, Japan
Steve Weinstein, Reuters Ltd., UK
Yorick Wilks, New Mexico State University, USA

1 Introduction, Sergei Nirenburg

Machine translation is not the only application at the intersection of linguistics and computer science (or, if you wish, language technology and software engineering). Other fields of applied research and development in this general area include computer-aided language instruction systems, text skimming and abstracting systems, information retrieval packages, grammar and style checking systems, lexicographic database management systems and natural language text generators. Many of such systems include components that are germane to MT systems—both processing modules, such as parsers, semantic interpreters or various user interfaces and "static knowledge sources" such as the various grammars and dictionaries. To reuse a system component in a different application is certainly in the interests of applied researchers and commercial product developers. One purpose of this panel is to discuss which MT system components can be considered multipurpose and what kinds of systems can benefit from this.

Another purpose of this panel is to discuss applications in which entire MT systems are included as single components. Environments of this sort can include multilingual message processing systems or multilingual document production systems. One of the typical comprehensive text processing tasks is production of technical documentation in a variety of languages. This task involves technical writing, translation, art production and typesetting. Progressive automation of all of these processes and the need to make the multiple interactions among them more efficient provide a good example of the potential for "situated MT." When an MT system is developed for such an application, it is best to design it so that it fits the general production process.

Indeed, adapting independently developed MT systems to a particular environment is not a trivial task, and not only from the standpoint of the necessary modifications to the dictionary and grammar content. This seems to be true not only of fully automatic MT systems but also of machine-aided translation environments of varying sophistication. The greatest benefits are to be expected when such environments are integrated with the computer support for the other tasks in the application. Thus, in an environment, where authoring and translation are integrated, a significant number of translation problems can be eliminated at authoring time. This can be achieved in a number of ways. One can equip the authoring system with a preediting filter which would check whether the author uses words and syntactic constructions from a "canonical" set. This set need not be too constrained (it has been shown in practice that it is impossible to force writers to use a very restricted sublanguage) but any level of protection against unexpected input will be more than welcome in the MT component! An authoring system can also be equipped with a filter which detects ambiguities in the text and requests their clarification before the text is submitted to the translation component.

Yet another example of "situated MT" is a mixed automatic/human translation environment, a workstation that can support many different levels of automation in translation. Such an environment will support purely human translation, with a word processor and various reference access capabilities. But it will also include an option of machine-translating a passage and then providing support for postediting. It will also provide an interface for interactive editing when the translation system is put into the mode of trying to resolve problems through interaction with the user *during* translation. Finally, it will also include a developer's interface which will facilitate acquisition and testing of dictionaries and grammars supporting the resident MT system. The above seems to form the set of functionalities to be supported by a "top-of-the-line" translator's workstation.

A number of reusability ideas and MT technology applications will be discussed by the panelists. It is clear, though, that there are many more such ideas and applications, from those which could be implemented today to quite futuristic ones. It is my hope that the members of the audience will contribute to the discussion by recounting their own experiences and putting forward new reusability and integration ideas,

2 Rod Johnson's Statement

To summarize very crudely the state of the art in MT technology, one could say that so far machines have turned out to be pretty good at things that people are often less good at (getting the agreements right, spelling, remembering to use the preferred terminology, ...), while people excel where computers typically fall down (understanding the real message, figuring metaphors and anaphors, using encyclopaedic knowledge, ...) Translation is an area where, on balance, failings of the first kind are probably less critical in achieving an acceptable result than those of the second kind.

There are, however, circumstances where an inexorably high quality capability in syntax, morphology and terminological accuracy is of major importance. One such arises in the case where that capability is being exploited alongside human expertise to teach other humans to make fewer mistakes in their syntax, morphology, spelling and so on.

My thesis here is that the MT community has built up over the years a vast resource and enormous competence in the formal parts of language description, evident even in MT systems of mediocre quality and scant success as translation engines, and far exceeding the limited linguistic capability which seems to be available to typical current CALL systems. I would like to examine the extent to which our MT know-how has been and could be put to use in improving the quality of existing CALL.

Many of the remarks applied above to CALL hold also for text critiquing, especially to grammar/morphology/spelling checkers. Though I have little experience which authorizes me to speak on this topic, it seems to me that a detailed knowledge of the grammar of a language, while insufficient for first rate machine translation, must surely be a very useful basis for a prescriptive text analyzer. It is probable that in MT we have built grammars of several languages with a wider coverage than is available anywhere else, so a priori I guess some effort in technology transfer from one field to the other must be desirable.

It is also worth observing that the criterion of satisfying the constraints imposed by a text critiquing system is a good one for determining the eventual translatability of a text.

Traditionally preferred domains for MT have always been characterized as those where a small number of utterance types are used to describe some finitely describable conceptual universe – cf. the oft-cited weather reports, avalanche reports, technical manuals. It could be, indeed has been argued that many of these domains are in fact more suitable for multilingual generators, systems where the input is not expressed in any natural language at all but where the output appears as a set of texts in several languages. Typical inputs are abstract structures derived from interaction with expert users via menus and on-screen forms, or directly from external devices like thermometers, hygrometers, or other environmental sensors.

The general idea of integrated document processing including MT is not new - notable examples are the various systems from Brigham Young University which flourished around a decade ago. Perhaps they were oversold as MT systems, undersold and overpriced as document processing workstations. Also, they were not able to exploit the same degree of connectivity as is commonplace today.

I still think the view of MT as a component of some larger document-oriented system is the right one, particularly in the light of the (relatively poor) quality of current (and foreseeable) MT and the increasing tendency to produce and distribute text electronically.

Perhaps this is a European view: in an environment where working in several languages is the norm the idea of having access to good multilingual document processing can be at least as attractive as off-line high quality translation. This is especially the case when you can compensate for relative weakness in one or more of the languages you work with by easy and immediate access to good dictionaries and critiquing software at various levels, not to mention quick draft-quality translation.

This is not to say that we should not continue to strive for better and better MT. But even if all we had to show for four decades of MT research were the spin-offs it would still have been worth while. Such, after all, is the premise on which this panel is based.

3 Richard Kittredge's Statement

One of the most important application areas of MT technology has turned out to be natural language generation, which deals with the planning and grammatical realization of texts (in one or more languages) out of some meaning representation of those texts. Researchers in text generation often distinguish two major classes of applications: (1) report generation, where relational data tables can provide the basic input from which meaning representations are built to produce texts with stereotyped structure and style, and (2) explanation generation, where texts must be planned and generated "on the fly" to satisfy the changing communication goals of application programs such as expert systems.

From the viewpoint of MT, we should be particularly interested in cases, occurring especially within reporting domains, where bilingual or multilingual generation can actually do a better job than existing MT. Such cases are relatively infrequent, but may help clarify the problem of relating language to more abstract forms of information that are needed for language processing in general. These cases also show report generation as a "spinoff technology", which would probably not have been possible without the insights gained from work on report translation.

Since 1985 there has been a project within the Canadian Environment Ministry to apply language generation technology to the automatic composition of weather forecasts [1,2]. This has led to the FoG system [3], operational since October 1990 for the production of daily marine forecasts in three weather centers (Halifax, Gander and Toronto). This project is quite distinct (except for my participation) from the TAUM-METEO project at the University of Montreal [4] in the mid-1970s, which resulted in an operational system for translating English forecasts into French (a system which still translates well over 10 million words of manually composed forecasts each year). Automatic forecast translation requires that someone write a source language text. Moreover, such machine translation is typically possible for fewer than 90% of forecast sentences. The new FoG system, which is currently limited to marine forecasts, generates bilingual forecasts from a single set of forecast data, without human intervention. The forecast data coming from an atmospheric modeling program are reviewed by a small expert system which builds from the data a set of abstract "facts" corresponding to predicted significant changes in the weather situation over the forecast period. FoG's automatic composition directly from these abstract facts also assures that all selected facts are verbalized simultaneously in both English and French, and according to the best professional style in each of the forecasting sublanguages. Although it will undoubtedly take several years for bilingual generation of forecasts to be implemented for all varieties of forecast, and hence displace MT, this is clearly a preferable approach to bilingualism for such restricted, data-driven sublanguages [5].

A second application area of bilingual generation is currently being explored by CoGenTex in Montreal in co-operation with Communications Canada and Statistics Canada. The aim of the project is to generate English and French reports on employment statistics, and other reporting domains, from relational data available on-line. Here the first effect will be to replace manual composition and human translation of existing texts, but the long-term goal will be to provide bilingual texts in a wide variety of statistical reporting areas where data is so volatile, or so specialized, that no attempt can be made (economically) to compose source language reports on the data.

Let me add several words about operational environments where MT can be used. If bilingual or multilingual text generation (MLTG) is included as a variant (mutant?) form of MT, the range of operational environments for MT is considerably enlarged. Reports might be generated simultaneously in a variety of languages in any recurrent kind of situation where on-line data is available and sufficiently rich to determine the content of a useful report. This is of course easier where texts already exist, so that stylistic and grammatical norms can be simply canonized into the system's rules. But it is also possible where no norms exist, by borrowing from norms of existing texts in other domains which share features of purpose, etc. with the new domain.

In the near future, one can also expect that other kinds of texts besides reports will be produced multilingually from more articulated knowledge structures. There is no reason to doubt, for example, that wherever expert system explanation generation becomes practical in one language, it will soon become a reality in others. Other kinds of structures, such as plan structures, which have proved to be appropriate structures for input to text planners and generators, can also become multi-lingual.

References

[1] R. Kittredge, A. Polguère and E. Goldberg, "Synthesis of Weather Forecasts from Formatted Data" *Proceedings of COLING-86,* Bonn, pp. 1986.

[2] E. Goldberg, R. Kittredge and A. Polguère, "Computer Generation of Marine Weather Forecast Text", *Journal of Atmospheric and Oceanic Technology,* vol.5, no.4, pp.473-483, 1988.

[3] L. Bourbeau, D. Carcagno, E. Goldberg, R. Kittredge and A. Polguère, "Bilingual Generation of Weather Forecasts in an Operations Environment", Proceedings of the 13th International Conference on Computational Linguistics (COLING-90), pp. 318-320, Helsinki, 1990.

[4] M. Chevalier et al. *TAUM-METEO,* Groupe de recherches pour la traduction automatique, Université de Montréal, 1978,

[5] R. Kittredge, "Multilingual Text Generation as an Alternative to Machine Translation" *Proceedings of the 30th Annual Conference of the American Translators Association* (D.L.Hammond, ed.), pp.465-469, Learned Information, 1989.

4 Lori Levin's Statement

I will address applications of machine translation technology to Computer Assisted Instruction. I will concentrate on the notion of a workstation—a set of tools including a user interface that can be configured for various applications. I will compare a Translator Workstation (TWS) to a Language Instruction Workstation (LIWS) and suggest a design for a modular language workstation (LWS) that can be configured to meet the needs of translators, foreign language students, and anyone needing to read or write in a language in which he or she is not fluent.

The LWS will include a number of natural language processing (NLP) tools including a machine translation engine, reference material, corpora of texts and videodiscs, and tools for teachers and system designers to use in configuring the LWS for specific applications. Figure 1 shows the components of the LWS. Section 2 describes the content of these components and what functionality they should provide for a TWS, Sections 3 does the same for a LIWS. The components and functionality described here are a combination of those from the TWS

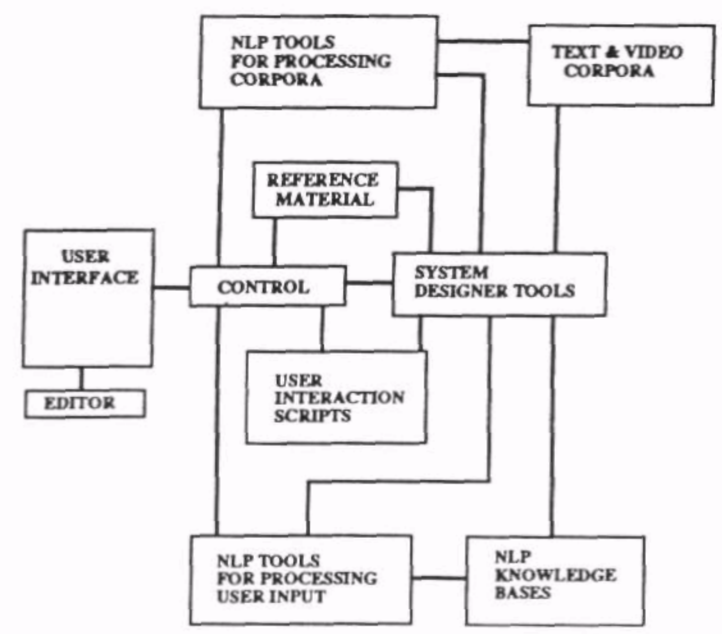

Figure 1: Components of a Language Workstation

of Nirenburg et al. (in press) and the LIWS of Levin et al. (1991).

4.1 The LWS as a TWS

Obviously, the primary function of a TWS is to provide tools that help a translator translate. This section describes how the components of the LWS portrayed in Figure 1 can be filled to meet the needs of translators.

The editor of a TWS allows the translator to view and edit aligned source and target language documents side by side. In addition to providing access to the editor the **user interface** lets the translator interact with other workstation toots, for example, reference material, concordances and corpora of previously translated documents, and human-assisted MT, The **control** box interprets and executes the user's commands (such as looking up words in a dictionary or selecting the correct reading of a sentence during human-assisted MT) and also controls system communications to the translator (such as queries about the correct reading of an ambiguous sentence during human-assisted translation or errors detected in the target language document during machine-assisted human translation).

The corpora include source language documents, target language documents, and pairs of corresponding source and target language documents that are translations of each other. Some of these will be from a library of documents provided by the system designer and some will be the translator's own files. **NLP tools for processing corpora** will include monolingual and bilingual concordance-making programs and automated indexing and retrieval tools. These tools will allow the translator to examine how a word was translated in previously translated documents, or simply to see how it has been used in monolingual source or target language documents. The translator will also be able to survey the format and style of various types of documents such as business letters, contracts, and newspaper articles, and he or she will be able to copy chunks of previously translated documents that are similar to the one currently being translated. The system designer will use the corpora in the process of building user-readable dictionaries and NLP grammars and lexicons.

Reference Material in the TWS consists of monolingual and bilingual dictionaries, term banks, gazetteers, and reference grammars. Instead of presenting illustrative examples out of context, the reference material can contain pointers into the corpora. This should insure that illustrative examples will be natural and typical of actual usage. Also, by examining the portion of the text that the example came from, the translator can pick up subtle contextual cues surrounding the example. Both the system designer and the translator should be able to update the reference material.

The TWS will contain a library of NLP tools for processing user input. For a human-assisted MT application, the NLP tools will include all of the components of an MT system. (See, for example, Goodman and Nirenburg, 1991.) For a machine-assisted human translation application, the NLP tools might consist of grammar checking for the target language document. A selection of MT and grammar-checking components might be available to suit the speed, quality, coverage and level of interaction required by the user. The **NLP knowledge bases** include analysis and generation grammars, domain models, and machine-readable dictionaries. Again, a selection of these could be available to meet different user needs. A system designer will frequently update and adapt the NLP knowledge bases for new domains and new applications.

The **system designer** tools support a number of tasks involved in building a specific system for a customer. These tasks might include building NLP knowledge bases, entering and updating reference material, formulating queries for an augmentor (Brown, 1991) in human-assisted MT, entering corpora, and configuring NLP components.

4.2 The LWS as LIWS

Our goal in building a LIWS is to have a general set of tools and components that can be adapted to various instructional activities (reading, writing, games, grammar drills, etc.) and to various levels of proficiency. This section describes how the LWS of Figure 1 can serve as a platform for the implementation of various types of instructional programs. I will concentrate on the use of the LIWS a reading and writing assistant, but I will also mention how it can be used for drills and simulations in which the computer plays the role of a conversation partner for the student.

In the **editor window**, a student can enter a composition that he or she is writing, enter responses to a conversation simulator, or enter answers to drill questions. The editor might also display a foreign language text that the student is reading. It will usually not be desirable to show aligned native and foreign language texts (using the capability for aligning source and target texts in the TWS) because translation is not considered to be conducive to foreign language acquisition. That is, the student should learn the foreign language in its own terms, not as a translation of the native language. However, if the system is serving as a reading assistant program for someone whose goal is to read technical material in a certain field, diving into foreign language texts with aligned native language translations might be an efficient way to get started. In addition to these capabilities for viewing and editing text, the user interface will allow the student to access reference material, retrieve excerpts from the corpora, and run instructional programs provided by the system designer.

The **corpora** for a LIWS should consist of videodiscs as well as textual corpora. If the LIWS is serving as a reading or writing assistant, the corpora will serve as reservoirs of examples of lexical usage, usage of moods and tenses, appropriate performance of speech acts in various situations, and cultural background on the foreign language. In order to use the corpora as reservoirs of examples, there must be **NLP tools for processing the corpora** that can index and retrieve appropriate examples at the student's request or at appropriate points in a lesson.

The **reference material** for a LIWS will be learners' dictionaries and reference grammars. As in the TWS, the reference material can have pointers to the corpora so that examples of usage will be realistic and will be presented in context.

The functions of **NLP tools** in a LIWS will include components for error detection for a grammar drill or writing assistant application, and natural language understanding within limited semantic domains for applications that involve simulations of conversations. As in the TWS, the **NLP knowledge bases** for the LIWS will include analysis and generation grammars, domain models, and machine readable dictionaries. A variety of these will be available for different types of instructional programs and different levels of student proficiency.

In computer-assisted language instruction systems, **system designer tools** are usually known as an **authoring system**, a language in which teachers can write language instruction programs. In the case of intelligent language tutoring systems, however, the system design task will be shared by a teacher and a computational linguist because it will involve design of instructional material as well as NLP tools and knowledge bases. The teacher will design reference material, choose examples from the corpora to illustrate instructional points, and formulate exercises and feedback for errors. The output of the teacher's authoring will be a set of **user interaction scripts** which, when executed by the **controller**, will display exercises, feedback, and other system responses on the student's screen and then take student input and send it to the appropriate natural language processor. The computational linguist's job in system design is to configure the NLP tools and to build and maintain NLP knowledge bases.

References

Brown, R. 1991. "Automatic and Interactive Augmentation." In K. Goodman and S. Nirenburg (eds.), 189-229.

Goodman, K. and Nirenburg, S. (eds.) 1991. *The KBMT Project: A Case-Study in Knowledge-Based Machine Translation.* San Mateo, Calif.: Morgan Kaufmann.

Levin, L., Evans, D.A., Gates, D., & Delon, L. 1991. "Capturing Context in a Computer-Based Foreign Language Assistant." In D. Ross and D. Brink (eds.) *ACH/ALLC '91: "Making Connections",* Conference Handbook. Humanities Computing Facility, Arizona State University, 279-284.

Nirenburg, S., J. Carbonell, M. Tomita and K. Goodman. In press. Machine Translation: The Knowledge-Based Approach. San Mateo, Calif.: Morgan Kaufmann.

5 Steven Weinstein's Statement

The value of information lies not in the information itself but in what use one can make of it. That said, we at Reuters have concentrated on building systems in this area that help us manage the flow of information (news, in the cases listed below) to subscribers and which have the equal appeal of holding the line on costs in our production centers.

Topic Identification System, for example, was developed with Carnegie Group Inc to index news stories on the fly, so that they could be made available quickly and with highly consistent and accurate coding at a reasonable cost. The system uses proprietary pattern-matching technology that, in essence, skims the text for words and phrases, then makes some conclusions about content.

Similar skimming functions are available in a range of systems from standard Boolean databases (which typically index every word in a text so that it may be retrieved) through to more "intelligent" systems such as the Verity Inc "Topic" product running in various sites such as the National Security Agency, the Joint Chiefs of Staff, and Chase Manhattan Bank, This product allows a user to specify areas of interest so that as information is received it can be routed instantly to the appropriate recipient.

These "skimming" systems look merely for patterns in text and do not contain a significant degree of text understanding. Their usefulness as standalone applications in the pursuit of Machine Translation must be seen, therefore, as limited. As a component, however, the ability to isolate a domain for an MT system may, however, prove crucial,

Another system we have under development now will read free text and extract key facts from it. It is from this area that some multi-lingual facilities may emerge. Our view of the technology today is that it might be used in translation activities, in updating databases, in providing structured feeds, or in providing unstructured information such as abstracts and headlines to accompany news stories.

Once the key facts can be identified, they may be written into a shell and output in many languages, albeit in fairly structured form with little opportunity for flair in language. Similarly, we may require structured input for information, such as equity fundamental data, which can then be massaged into many languages and formats.

We are viewing this current development as a next level of text understanding. Boolean systems can index on a word by word basis; TIS can index on a conceptual basis with a high degree of accuracy, provided the ideas can be specified in advance; our new system will be able to understand the text to the degree that A's relationship to B can be understood (A bought B, for example, is an entirely different effect than A has been bought by B) and database entries can be created for buyer and seller, for example. We still will not be able to preserve the flavor of the language, but these applications do appear to be a step in the right direction.

Reuters and other multi-language information providers clearly have a critical need to make their translators more productive, and fulfilling this need effectively and economically is one factor that can extend as far as the bottom line. More traditional business, however, has a difficult time quantifying the value of translated information. A letter produced in a "foreign" language may bring in a large contract, or it might not. And the speculative nature of the translation makes it all the more important to bring its cost as low as possible.

Text skimming, therefore, becomes important as a cost-effective tool to handling standard translation. The typical business document, be it letter or proposal or internal memorandum, will contain mostly non-specialist words and phrases, and these may be identified through skimming techniques. Having been identified and "understood" (i.e. whether A is buying B or being bought by B) a document may be translated. This translation will be at least an order of magnitude better than a simple word swapping routine, and probably more.

The problem of translating the specialist language, then, becomes the remaining only custom work (because the standard sentences should be provided as part of the standard package). This knowledge may be added to the skimming algorithm over time and translations generated from it. The obvious drawback of using these skimming techniques is that the translations may lose some of the flavor of the original document and perhaps a bit of nuance. The commercial advantages of using these technologies, however, will move business toward them.

Grammar and style critiquing systems are viewed at Reuters as tools that can help us convey news consistent and correctly, and they will help us in the same way spelling checkers can. Additionally, by aiding the input in a proactive manner, we can enhance our ability to translate. It is clear that clear writing is a must, both for our present mono-lingual activities and for any automation to be added to translation in future,

6 Yorick Wilks's Statement

A natural place to begin this discussion of how to advance the field of MT technology, would be to consider the linguistic roles that could be combined if one could create a "dream workstation". Sergei Nirenburg is one researcher who is actively concerned with the specification and construction of such a project: in this case, a Translator's workstation. And at CRL/NMSU, we are working on ongoing projects to create workstations for Lexicographers and Linguists. If they were installed in a single laboratory environment, all of the above could share functions. And, as Sergei has noted, it would also be an excellent opportunity for translators to learn the language as they go along: There are more low-level professional, almost rudimentary, translators in the field than was once believed. But I will leave the exploration of these connections to my colleague.

Speaking for CRL/NMSU, our laboratory is engaged in ongoing projects in MT, Text Skimming (the DARPA TIPSTER project), Computer Aided Language Instruction, and Lexical Extraction of semantic entries from dictionaries and related techniques. I will not describe any of these in detail here (nor do you expect me to) but instead will make some very general remarks about the way in which these tasks are related.

First of all, there is MT and CALI, Any approach to MT should have CALI implications; and it may be that, while it is customary to pursue a single standard approach to MT at a given time (e.g. interlingual, transfer, or statistics-based), it is possible that all these approaches could be integrated within a single CALI system. Conversely, on what one might call a Whorfian approach to the question, where cultures and their associated languages form closed wholes, there can be no translation into and out of any language; a fortiori no MT, and (I suspect) no CALI that uses any element of MT technology. But let us assume for the moment that no one here actually believes that.

In that case, if you are someone who takes an interlingua] approach to MT (as I happen to do), one might also assume that aspects of that technology are reusable within a CALI system. And we do, indeed, make that assumption. In one of our CALI projects, we have tried to implement Ogden's "stick picture" approach to language learning; one in which the picture sets and their pair-wise contrasts (the heart of the method) are virtually the same for all languages, and a language is taught directly, without employing the learner's first language. These picture sets and their contrasts form an interlingua whose primitives are not the pictures themselves, but rather contrasted cross-cultural items within them. And, as is the case with all interlinguas, the method frays a little at the edges, since some items and contrasts are simply not expressed in every language. But if you take a tough-minded approach to interlinguality, that proves nothing: it simply goes to show that languages contain lexical gaps. The assumption that such an interlingual method makes is that all languages share certain pragmatic purposes that their users wish to avail themselves of. To me, this seems to be the most common-sense link between CALI and a classical approach to MT.

For instance, the Transfer Approach to MT may be incorrect, but would certainly lead one to focus on issues such as systematic syntactic errors by the learner (based on inappropriate transfer from the first language) if linked with CALI. Any CALI system should have a way of correcting these. From a non-transfer view, these simply constitute more errors to be corrected, but from a transfer view, the system is made richer and more effective if it knows the source-based error patterns to look for. Since I do not know the relevant facts, I have an open mind about the empirical issues here; but 1 can easily believe that there is a place for such an approach within a battery of CALI techniques.

In a similar way, the recently fashionable resurrection at IBM of the statistics-only-based approach to MT (based on trigram analysis of vast quantities of English-French bilingual text) could also prove useful to CALL Whatever its adequacy for MT, the role of a statistical-norm approach, in which the learner is prompted word-by-word to produce the most plausible, natural-looking/sounding sequence in the target language at every point, is easy to imagine within the context of CALI. It is precisely this statistical naturalness that a non-native speaker often lacks, even when his speech is syntactically and semantically correct. The well-known example of the statistical solecism "powerful coffee" instead of "strong coffee", would mark one as a foreign speaker immediately, even though it contains no grammatical or semantic error.

At CRL, we have championed a very strong connection between MT and text skimming as consumers of large-scale lexical extraction with maximal automation. Use of what I call the "interlingual assumption" also leads us to seek a theoretically neutral type of lexicon, one that is uncommitted as to any given theory or language. We have developed such a working, multi-lingual MT system (ULTRA) that functions with that assumption; but we are also aware of how unacceptable many people find both forms of assumed neutrality to be.

In my view, with text skimming there is more contrast and less connection with MT. It is obvious that although the skimming and MT technologies could be linked (with an MT system translating only the skimmed gists of a text, as many have proposed), one probably does not want to use the same skimming parser on a text that one intended to translate; unless a very gist-like paraphrase translation was requested. And as always in MT, one can describe or lower one's translation needs in such a way that just about any method of MT will do something for one.