MARKET AND INDUSTRY STUDY FOR COMPUTER-ASSISTED TRANSLATION SYSTEMS

September 28, 1987

Submitted to: Ms. Bev Hillman
Directorate of Technology Assessment
and Development
Department of Communications

Submitted by: Dr. George A. Neufeld
Dr. Alan W. Underdown
Coopers & Lybrand

Mr. Benoit Thouin
Computational Linguistics Consultants Ltd.

The Coopers & Lybrand Consulting Group

MARKET AND INDUSTRY STUDY FOR COMPUTER-ASSISTED TRANSLATION SYSTEMS

September 28, 1987

Submitted to: Ms. Bev Hillman
Directorate of Technology Assessment and Development
Department of Communications

Submitted by: Dr. George A. Neufeld
Dr. Alan W. Underdown
Coopers & Lybrand

Mr. Benoit Thouin
Computational Linguistics Consultants Ltd.

The Coopers & Lybrand Consulting Group

Management Consultants

EXECUTIVE SUMMARY

INTRODUCTION

The overall objectives of this study were to:

* identify the user needs and market opportunities for Computer Assisted Translation (CAT) systems; and

* assess the Canadian industrial capabilities to capitalize on these opportunities.

The study took into account what is happening in the industrialized world concerning CAT systems, but concentrated on the Canadian market.

The results of this study are based on:

* a review of existing reports;

* interviews with selected persons in government, industry, and universities;

* two focus groups that were held during the study; and

* a survey of firms, in-house groups, and persons providing translation services.

The study examined:

* the size and nature of the market for translation services; the needs of translation suppliers and translators; and

* the potential for CAT systems, both generally and for Canada.

Finally, the strategic considerations from an industrial development viewpoint were identified.

THE MARKET FOR TRANSLATION SERVICES

The demand for translating material is driven by the need to overcome the language barrier experienced in a broad range of human endeavour having economic, social, political, or military significance. The demand is also dependent upon cost, turnaround time, and quality of translation services.

The size of the Canadian market for translation services is estimated to be in the 550-750 million words-per-year range. English-French translation work accounts for 94% of the Canadian translation market. The Canadian market for English-French translation represents a significant share (between 5 and 19%) of the world market for English-French translation; this becomes important for language pair dependent CAT system products. Overall, the Canadian market represents between 1 and 4% of the world market for translation service.

The demand for translation services would probably be larger if material could be translated at significantly reduced cost, reduced turnaround time, or improved quality. This latent demand is estimated to be about 30% of the total current demand. While our survey was not aimed at end-users of translated material, this seems reasonable, based on the information we obtained from translation suppliers.

The demand for translation services is growing about 10% per year, which is greater than the growth rate for the world economy. This appears to be reasonable, based on a qualitative assessment of economic, political, and social trends, both in Canada and elsewhere.

THE SIZE AND NEEDS OF THE TRANSLATION INDUSTRY

The four main types of translation suppliers and the estimated population size in Canada for each type are:

* private sector translation firms (380);
* translations units within corporations (260);

* freelance translators (1,000 to 1,100); and
* translation units within governments (110).

The Translation Bureau at Secretary of State is by far the largest translation organization in Canada. The average firm/unit (for all types surveyed except freelance translators) has 15 persons.

Translation suppliers are already making extensive use of technology. Of the translation firms/units surveyed, 71% have at least one computer; 14% use local area networks; 16% use a mini or mainframe computer; 46% use terminology databases; 39% have modems; and 34% have facsimile machines.

The translation suppliers' concerns centre around providing a high level of service at the lowest possible cost. In particular, they are concerned about the time and cost associated with terminological research and about the difficulties associated with using their existing automation equipment/systems as well as interfacing with that of their clients. CAT systems that address these concerns would be welcomed by suppliers.

Based on a combination of direct experience and communications within their industry, translation suppliers are less optimistic about CAT systems that automatically translate material. However, about 30% of all survey respondents translate material which has a simple format or structure compared to normal prose and which may some day be translated by machine.

The surveyed translation firms/units, on average, spend only $17,500 per year on automation equipment, and 85% spend less than that. Furthermore, CAT systems that are attractive to translators in the federal government may be different from those which are attractive to other translation suppliers because government translators make greater use of dictation equipment.

THE POTENTIAL FOR COMPUTER-ASSISTED TRANSLATION (CAT) SYSTEMS

Computer-based products aimed at the translation industry fall into two categories:

* general purpose software products which are not specifically designed for translation but which assist translation and supporting operations such as information management, document production, and production scheduling; and
* translation specific products whose features are aimed at the translation market, including terminological databases, research tools, and systems that automatically produce a rough translation of either general text or text with a limited structure/vocabulary.

None of the currently available general purpose products has all the features required by translators. Also, the terminological research features of currently available general purpose and translation-specific products are weak. Finally, the feasibility and cost effectiveness of developing systems that can automatically translate material may well be possible but has not yet been accomplished; the claims of current suppliers of such systems are generally regarded as vastly exaggerated. The identification of applications with sufficiently simple and large enough volume of text to make automated translation systems feasible represents a challenging problem. Also, there may be applications where the material is currently not written in simple or limited language form, but where the users may consider doing so if they could get the material translated at significantly less than current costs, for example, military equipment manuals.

The overall conclusion of the study is that there are opportunities in the area of Computer-Assisted Translation (CAT) systems for the translation supply industry. There are immediate commercial opportunities for selling software packages that integrate features now available in different systems and that have better translation support capabilities than do currently available systems. There are potential commercial opportunities for automated translation systems and terminological banks required by such systems, although the technical and economic viability of such systems has not yet been established. Persons who write professionally in a second

language may comprise a second-tier market for suppliers of CAT systems that have been developed for the translation supply industry.

The total potential size of the market will depend upon the extent to which the product is language pair and/or subject matter dependent. The adoption of any product will depend on current methods used for translation (e.g., dictation versus keyboard input or handwritten), cost, and the target customers' existing equipment/systems. Products that are low cost and compatible with existing equipment will have greater potential, particularly given that the typical translation supplier's expenditures on automation equipment is relatively small.
While the feasibility of developing products aimed at automated translation has yet to be determined and depends upon technological advances in artificial intelligence and natural language processing, the potential market for these products is likely to be relatively small because of their high cost and the high volume of work that any one system will be able to translate.

The analysis of the commercial attractiveness of CAT systems (under various assumptions concerning product price, potential market size, supplier market share, and development costs) shows that Canadian suppliers of CAT systems would need to penetrate international markets in order to achieve an acceptable return on investment. Suppliers focusing only on the Canadian market would be working on the basis of unrealistic assumptions concerning development costs, price, product adoption rate, and market share -- making it a high-risk venture from an investor perspective.

STRATEGIC CONSIDERATIONS

The fundamental issue is whether the federal government should assist in the development of a Canadian industrial supplier base for CAT system.

On the positive side, there are several factors that make CAT systems an attractive opportunity for Canada, namely:

* there is a need for CAT products that can increase the cost effectiveness of translation services;

* the currently available CAT products do not perform well and their quality is questionable;

* the current suppliers of CAT products have not penetrated the translation market to any great extent;

* there is a significant domestic and international market for translation services involving English and French languages;

* Canada has world-class expertise in universities and government research labs, although on a fairly small scale, in research pertaining to the technologies underlying CAT;

* Canada possesses the expertise required to pursue ventures in CAT systems;

* there is a small but growing Canadian software industry which could exploit CAT opportunities; and

* it increases an opportunity for the government to use its own needs to support Canadian industry.

But there are risks as well; the negative factors are:

* foreign companies in the U.S. and Europe have established a reasonably strong position in the CAT market and have at least a two-year lead on a new entrant;

* existing companies have recently introduced new products that are directed at the opportunities identified in this study;

* Canada has no real CAT system supplier base;

* new Canadian supplier(s) would face stiff competition, even in the Canadian market;

* the market for CAT systems is much more limited than for more general purpose products such as personal computers or spreadsheet packages;

* the cost effectiveness of automated translation systems has yet to be proven; and

* the scale of Canada's R&D capabilities and existing industrial base, relevant to CAT systems, is significantly smaller than that of other countries.

While there are risks associated with each of these downside factors, CAT systems provide a moderately attractive opportunity for Canada, particularly given the world-wide demand for English-French translation. However, it will require that Canadian suppliers evolve and move quickly as well as catch up and outsmart the competition.

Given that the federal government is the single largest buyer of translation services in Canada, it has and will continue to influence the development of a Canadian supplier base of CAT systems. This influence is exercised through the systems it chooses to test or acquire, and through how it deals with its translation service suppliers. For example, General Motors uses its influence to motivate its suppliers to reduce their costs and improve the quality of their products -- including through the adoption of technology. However, if the government wishes to use its buying power to help develop a Canadian CAT system supplier base, it will need to move quickly since foreign companies are already past the starting line. It also implies action-oriented coordination among government departments that have an interest in CAT, particularly the Department of Communications, the Secretary of State, and possible the Department of Industry, Science and Technology.

Another important consideration for the government is whether it should encourage Canadian suppliers to focus on more advanced systems such as automated translation or less ambitious products based on system integration and/or improvements to terminological tools. Product development, including specifications, should be developed by industry and be market-driven. Otherwise, it is government defining a commercial opportunity for industry, which could result in developing a supplier base that is overly dependent on government for contracts. If it is concluded that the

product(s) required to meet the government's needs for CAT systems are different than those of other translation suppliers, the reasons for these differences need to be examined. From an industrial and market development perspective, focusing only on advanced CAT systems could be an R&D success but a commercial failure. Thus R&D for the development of advanced CAT systems should only be supported if the government decides to assist in the development of a Canadian industrial supplier base for CAT systems that focus on the immediate needs in order for suppliers to establish themselves in the marketplace.

SOMMAIRE À LA DIRECTION

INTRODUCTION

Cette étude avait pour objectifs principaux d'identifier les besoins des utilisateurs et de cerner le marché potentiel des systèmes de traduction assistée par ordinateur (TAO), et d'évaluer la capacité de l'industrie canadienne à tirer parti de ce marché. Nous avons tenu compte de l'évolution de la TAO dans le monde industrialisé, en accordant une attention particulière au marché canadien.

Les résultats de cette étude reposent sur l'examen de rapports existants, sur des entrevues avec des personnes choisies du gouvernement, du secteur privé et du milieu universitaire, sur deux réunions avec des représentants du milieu et sur un sondage auprès de particuliers, de services internes et d'entreprises indépendantes fournissant des services de traduction.

L'étude a porté sur la taille et les caractéristiques du marché des services de traduction, sur les besoins des traducteurs et des services ou entreprises de traduction et sur les possibilités des systèmes de TAO, plus particulièrement au Canada. Enfin, nous avons mis en lumière certaines considérations stratégiques relatives au développement d'une industrie dans ce domaine.

LE MARCHÉ DES SERVICES DE TRADUCTION

La demande de traduction est motivée par le besoin de franchir la barrière linguistique dans un grand nombre d'activités de nature économique, sociale, politique ou militaire. La demande dépend également des coûts, des délais et de la qualité des services de traduction.

On estime à entre 550 et 750 millions de mots par année la taille du marché canadien de services de traduction. La traduction de l'anglais au français et du français à l'anglais représente environ 94% de ce volume. Pour ce qui est de la traduction de l'anglais au français, le marché canadien représente une part significative (entre 5 et 19%) du marché mondial. Il s'agit là d'une donnée importante pour les systèmes de

TAO qui dépendent du couple de langues. Le marché canadien représente globalement entre 1 et 4% du marché mondial des services de traduction.

La demande serait probablement encore plus considérable si la traduction pouvait se faire à des coûts ou dans des délais sensiblement moindres ou avec une meilleure qualité. On estime que cette demande latente représente environ 30% de la demande totale actuelle. Même si nous n'avons pas interviewé les utilisateurs des traductions, nous croyons que ce chiffre est raisonnable d'après les données obtenues des fournisseurs de services de traduction.

La demande de services de traduction connaît une croissance d'environ 10% par année, soit une croissance supérieure ? celle de l'économie mondiale. Ce chiffre semble raisonnable d'après une évaluation qualitative des tendances économiques, politiques et sociales, tant au Canada qu'ailleurs dans le monde.

LA TAILLE ET LES BESOINS DE L'INDUSTRIE DE LA TRADUCTION

Voici les quatre principaux types de fournisseurs de services de traduction et leur nombre approximatif au Canada: entreprises de traduction du secteur privé (380); services de traduction au sein d'entreprises (260); traducteurs à la pige (1000-1100); services de traduction du secteur public. Le Bureau des traductions du Secrétariat d'État est de loin l'organisme de traduction le plus important au Canada. Si l'on excepte les traducteurs à la pige, l'entreprise ou le service de traduction moyen compte 15 personnes.

Les fournisseurs de services de traduction font déjà un usage considérable de la technologie. Parmi les entreprises et les services de traduction qui ont répondu au sondage, 71% possèdent au moins un ordinateur; 14% utilisent un réseau local; 16% se servent d'un mini ou d'un maxi-ordinateur; 46% utilisent une banque de données terminologiques; 39% possèdent un modem; 34% possèdent un télécopieur.

Les fournisseurs désirent offrir un service de qualité au moindre coût possible. En particulier, il se préoccupent du temps et des coûts consacrés à la recherche terminologique, des difficultés d'utilisation de leurs systèmes actuels et d'interface

avec ceux de leurs clients. Des systèmes de TAO répondant à ces préoccupations seraient bien accueillis par les fournisseurs de services de traduction.

D'après leur propre expérience et l'information dont ils disposent, les fournisseurs de services de traduction sont moins optimistes à l'égard de systèmes faisant une traduction automatique de documents. Cependant, environ 30% des répondants traduisent des documents d'une structure simplifiée et qui pourraient donc un jour être traduits automatiquement.

Les entreprises et services de traduction ayant répondu au sondage ne dépensent en moyenne que 17 500 $ par année en matériel informatique, et 85% d'entre eux dépensent moins que ce montant. De plus, les systèmes de TAO qui conviennent aux traducteurs du gouvernement fédéral ne sont pas nécessairement les mêmes que ceux qui conviennent aux autres fournisseurs, étant donné que les traducteurs du gouvernement font davantage usage de la machine à dicter.

LES POSSIBILITÉS DES SYSTÈMES DE TRADUCTION ASSISTÉE PAR ORDINATEUR (TAO)

Les produits informatiques servant à la traduction se répartissent en deux catégories:

* des logiciels généraux qui ne sont pas spécifiquement conçus pour la traduction mais qui peuvent venir en aide à la traduction et aux opérations connexes comme la gestion de l'information, la production de documents et l'ordonnancement des tâches; et

* des produits spécifiquement destinés au marché de la traduction, notamment des bases de données terminologiques, des outils de recherche et des systèmes produisant une première traduction d'un texte général ou d'un texte ayant une structure ou un vocabulaire simplifiés.

Aucun des logiciels généraux actuellement sur le marché ne possède toutes les caractéristiques souhaitées par les traducteurs. De plus, les possibilités des logiciels

spécialisés sont encore limitées. Enfin, la faisabilité et la rentabilité de systèmes de traduction automatique sont rarement établies hors de tout doute: les prétentions des fournisseurs actuels de tels systèmes sont en général considérées comme grandement exagérées. D'autre part, il n'est pas facile de trouver des situations où le volume et la structure des textes à traduire rendent possible la traduction automatique. Par contre, il peut exister des situations où les textes ne sont pas écrits à l'heure actuelle dans un langage simplifié, mais où les utilisateurs pourraient fort bien envisager de le faire si cela pouvait entraîner une réduction sensible des coûts de traduction. On pourrait citer à cet égard les manuels d'équipement militaire.

La conclusion générale de l'étude est qu'il existe des possibilités en matière de systèmes de traduction assistée par ordinateur (TAO) destinés aux fournisseurs de services de traduction. Il existe des possibilités immédiates de commercialisation de progiciels qui intégreraient les caractéristiques de différents systèmes disponibles à l'heure actuelle et qui constitueraient un meilleur soutien aux activités de traduction. Il y a également un certain potentiel commercial pour des systèmes de traduction automatique et pour les banques de terminologie qu'ils exigeraient, mais la viabilité technique et économique de tels systèmes reste encore à établir. Les personnes appelées à écrire dans une langue seconde pourraient constituer un autre segment de marché pour les fournisseurs de systèmes de TAO destinés à l'origine aux traducteurs.

La taille potentielle du marché est différente selon que le produit dépend ou non du couple de langues ou du domaine. Le choix de tout produit dépendra des méthodes de traduction actuellement utilisées (p. ex. dictée, utilisation d'un clavier ou manuscrit), du coût et des systèmes actuellement utilisés par les clients potentiels. Les produits dont le coût est faible et qui sont compatibles avec le matériel actuel auront de meilleures possibilités de pénétration, étant donné que les fournisseurs de services de traduction dépensent relativement peu d'argent pour le matériel informatique.

Tant que la faisabilité de systèmes de traduction automatique restera à démontrer et dépendra des progrès de l'intelligence artificielle et du traitement des langues

naturelles, le marché potentiel de tels produits restera relativement limité en raison du coût élevé de ces systèmes et du volume considérable que chacun sera capable de traiter.

Une analyse de l'intérêt commercial des systèmes de TAO (sous diverses hypothèses à propos du prix, du marché potentiel, de la part de marché et des coûts de développement) montre que les fournisseurs canadiens de sytèmes de TAO devront pénétrer les marchés internationaux pour obtenir un rendement acceptable. Des fournisseurs qui se concentreraient uniquement sur le marché canadien le feraient sur la base d'hypothèses non réalistes sur les coûts de développement, le prix, la rapidité de pénétration de leur produit et leur part du marché - ce qui, du point de vue d'un investisseur, serait fort risqué.

CONSIDÉRATIONS STRATÉGIQUES

La question fondamentale est de savoir si le gouvernement fédéral devrait contribuer au développement d'une industrie canadienne des systèmes de TAO.

De nombreux facteurs positifs militent en faveur du développement d'une industrie des systèmes de TAO au Canada:

* il existe un besoin pour des produits de TAO capables d'augmenter la productivité des services de traduction;

* les produits actuels de TAO ont des performances médiocres et leur qualité peut être mise en doute;

* les fournisseurs actuels de produits de TAO n'ont pas encore fortement pénétré le marché de la traduction;

* il existe un marché assez considérable au Canada et à l'étranger pour des services de traduction de l'anglais au français ou du français à l'anglais;

* même s'ils sont relativement peu nombreux, les laboratoires de recherche gouvernementaux et les universités travaillant sur les technologies sous-jacentes à la TAO confèrent au Canada une compétence internationalement reconnue dans ce domaine;

* le Canada possède les compétences nécessaires à la poursuite de projets de TAO;

* le Canada possède une industrie du logiciel, petite mais en pleine croissance, qui pourrait exploiter les perspectives nouvelles de la TAO;

* une industrie canadienne des produits de TAO pourrait profiter des besoins du gouvernement en matière de traduction.

D'un autre côté, il y a aussi des facteurs de risque qu'il faut prendre en considération:

* des entreprises américaines et européennes sont déjà établies sur le marché de la TAO et possèdent au moins deux ans d'avance sur un éventuel nouveau concurrent;

* des entreprises déjà établies ont récemment mis sur le marché de nouveaux produits qui vont dans le sens des orientations présentées dans cette étude;

* le Canada ne possède aucun fournisseur réel de systèmes de TAO;

* les nouveaux fournisseurs canadiens feraient face à une forte concurrence, même sur le marché canadien;

* le marché des systèmes de TAO est beaucoup plus restreint que celui de produits plus généraux comme les ordinateurs individuels ou les tableurs;

* la rentabilité des systèmes de traduction automatique reste à établir;

* les installations canadiennes de recherche et de développement et les entreprises proches du domaine de la TAO sont d'une taille beaucoup plus petite que celles d'autres pays.

En dépit des risques énumérés ci-dessus, les systèmes de TAO représentent des perspectives raisonnables pour le Canada, en particulier à cause de la demande mondiale pour la traduction de l'anglais au français et du français à l'anglais. Il faudra cependant que les fournisseurs canadiens progressent rapidement pour rattraper et dépasser leurs concurrents.

Étant donné que le gouvernement fédéral est l'acheteur le plus important de services de traduction au Canada, il a et continuera d'avoir une influence importante sur le développement d'une industrie canadienne de systèmes de TAO. Cette influence se manifeste par les systèmes qu'il choisit d'essayer ou d'acquérir et par la façon dont il traite avec les fournisseurs de services de traduction. Pour faire une comparaison, General Motors utilise son influence pour inciter ses fournisseurs à réduire leurs coûts et à améliorer la qualité de leurs produits - y compris par des changements technologiques. Cependant, si le gouvernement désire utiliser son pouvoir d'achat pour contribuer au développement d'une industrie canadienne de systèmes de TAO, il devra agir rapidement puisque des compagnies étrangères occupent déjà le terrain. Il devra également y avoir une coordination entre les ministres ayant un intérêt pour la TAO, en particulier le ministère des Communications, le Secrétariat d'État et, le cas échéant, le ministère de l'Industrie, des Sciences et de la Technologie.

Le gouvernement devra également décider s'il souhaite encourager les fournisseurs canadiens à mettre au point des systèmes plus évolués comme des systèmes de traduction automatique, ou bien des produits moins ambitieux intégrant ou améliorant diverses fonctions dont la recherche terminologique. La définition et le développement des produits devraient émaner de l'industrie et correspondre à la demande du marché. Si, au contraire, c'est le gouvernement qui définit les possibilités de l'industrie, celle-ci risque de se mettre en situation de dépendance. Si l'on en arrive à la conclusion que les systèmes de TAO répondant aux besoins du gouvernement sont différents de ceux qui répondent aux besoins des autres fournisseurs de services de traduction, il faudra examiner en détail les raisons de ces

différences. D'un point de vue industriel et commercial, le fait de se concentrer uniquement sur des systèmes perfectionnés de TAO amènerait des succès au plan de la recherche et du développement, mais un échec au plan commercial. Par conséquent, la recherche et le développement en matière de systèmes perfectionnés de TAO ne devraient être encouragés que si le gouvernement décide de soutenir une industrie canadienne produisant des systèmes capables de se tailler une place sur le marché en répondant aux besoins immédiats.

I. INTRODUCTION

A. BACKGROUND

The application of computers to language translation is currently the subject of considerable commercial and research activity in several countries including the United States, Great Britain, West Germany, Japan, and France. This application of computer technology is often referred to as machine translation but is more accurately described by the term: Computer-Assisted Translation (CAT).

The most ambitious form of CAT involves inputting a written text in the language to be translated from (source text), and having the computer perform the translation and produce the translated text (target text). The CAT user can, in principle, benefit from the computer's extremely high output rate and reproducible performance. The computer algorithms which perform the translation function can potentially be based on theories from computational linguistics, natural language processing, and artificial intelligence. In recent years, several such computer-assisted translation systems have been introduced to the world market, almost all by U.S. firms. These systems are generally considered to be based on relatively unsophisticated algorithms including simple substitution of a word in the source language by the corresponding target language word. The cost effectiveness of this type of CAT is currently a controversial area. Research into more sophisticated CAT is being carried out by several nations (1). In some cases, this research forms a small part of their national research programs in artificial intelligence and Fifth Generation computer technology. The manufacturers of commercial CAT systems are also actively engaged in applied research and product development to improve their products.

Another type of CAT product is the terminological database or "term bank", of which there are several on the market. These products, which

allow a translator to access terminological information required for translation, employ computer database technology.

In addition to these two classes of product, there are many types of office automation and document production related computer products which can or do play a role in assisting translation operations. Although not unique to the translation field, the role of these products in assisting translation will be considered part of CAT technology in the present study.

There is a small but growing market for translation services in Canada both in the public and private sectors. These services are provided by in-house translation services in government and industry, commercial translation agencies and freelance translators. The Canadian research community possesses capabilities in several of the underlying technologies which contribute to CAT. There have also been some technically successful research projects which involved computer generated translation (2). There are, however, almost no CAT products developed by Canadian firms at the present time.

The potential for computer assisted translation has been the subject of several previous studies both in Canada and abroad. Recent surveys of translation practices including the use of CAT have been performed on behalf of the Commission of the European Communities (3) and the Digital Equipment Corporation (4). The coverage of both these surveys was centered on western Europe, the United States, and Japan. In Canada, the Department of Communications and the Secretary of State have supported investigations into CAT. These include a study by Cognos Incorporated to examine the potential of Canadian industry to become active in the CAT market as well as possible Federal Government strategies to facilitate this (5, 6). A study concerning the specific requirements of translators with respect to computers and automation was carried out by Socioscope Limited for the Department of Communications and Secretary of State (7). This area is also the subject

of continuing investigation at the Department of Communication's Canadian Workplace Automation Research Centre (CWARC). The Department of the Secretary of State is also participating in this work. Recent activities include surveying translators to determine their needs with respect to a translator's workstation. CWARC has plans to contract the private sector to develop a prototype. Work with this prototype is expected to provide more information on the utility of various CAT features.

The present study builds on previous ones to acquire and interpret the key information necessary to assess the opportunities for computer assisted translation products from a Canadian supplier perspective.

B. OBJECTIVES OF THIS STUDY

The overall objectives of the study were to:

* identify the user needs and market opportunities for CAT systems; and

* assess the Canadian industrial capabilities to capitalize on these opportunities.

Sub-objectives of the study were to: obtain detailed information on user's current and future requirements for computer assisted translation systems; estimate the size, growth rate and time frame for the markets; determine whether the requirements of the various market segments (i.e. federal government, provincial and municipal governments, private sector, and international market) are of a similar nature; document the factors and constraints affecting the evolution of these markets; assess the economic feasibility of computer assisted translation from the user's point of view; analyze the capabilities of existing (or planned) operational systems to determine how they meet the identified user's requirements; analyze the capabilities of Canadian industry in meeting this potential

market demand; and provide a base for information to support and guide policy in terms of R&D requirements and Canadian industry support.

Emphasis was placed on the Canadian market. The international market was examined from the perspective of the total market size.

C. STUDY METHODOLOGY

The fact finding and analysis conducted as part of the study focused on the following:

1. Size and nature of the market for translation services. This was necessary since the commercial potential of CAT systems is ultimately driven by the demand for translation services.

2. The nature of the translation process. It was first necessary to examine the processes involved in providing translation services in order to identify any needs of translation service suppliers which can be addressed by CAT systems. The processes were studied in terms of what resources they consume, how they are carried out, and the important characteristics of their inputs and outputs.

3. The needs, size, and nature of the translation service supply industry. Rather than focus on current demand for and attitudes towards specific CAT products, emphasis was placed on learning the translation industry's inherent needs and relevant characteristics with respect to CAT technology. The needs were assessed to identify potential opportunities for CAT systems. The size and nature of the translation service supply industry were examined to assess its readiness and ability to adopt CAT systems and to assess the size of the potential market. While the fact finding focused on the Canadian supply industry, comparisons with that of foreign suppliers were made, particularly as it became clear that the Canadian translation supply industry was relatively small.

In assessing the market for CAT products, we have focused on the translation supply industry, that is, those firms or departments in the public and private sector whose stated function is to produce translation work. We recognize that there is another potential market for CAT, namely anyone who reads or writes in a second language as part of their job. We have concentrated on the translation industry because it is more easily defined and has a significant need for CAT technology. It has therefore been selected as the key target market for new commercial CAT ventures.

4. Currently available CAT systems. The current situation was evaluated with respect to the capabilities of current and future commercial systems, the underlying technologies, on-going research, and the capabilities of Canadian industry.

Based on a comparison of the supply and market factors, several opportunities for CAT systems were identified and evaluated. The opportunities were assessed taking into account: technical feasibility, economic attractiveness from both a buyer and supplier perspective, and attractiveness to Canada, taking into account Canadian competitive strengths and weaknesses.

The sources of information utilized for the study included the published literature, discussions with people in the translation and computer-assisted translation industries, and a survey of translation organizations. Key literature sources included reports on two previous international surveys concerning the translation industry (3, 4) and previous Federal Government/Reports (5, 6, 7). The results of a previous study of the Canadian translation industry done on behalf of the Secretary of State were also utilized (8). Individuals interviewed included representatives of CAT suppliers, other software developers, and managers of translation organizations involved in evaluating CAT technology. Discussions with translators and translation service managers were held in the form of two focus group sessions, each one involving about eight participants.

Appendix II provides the moderator's guide during the focus group discussions as well as the questionnaire completed by the focus group participants and a summary of their responses.

The focus group discussions as well as discussions amongst team members and between the team members and representatives from DOC and SOS provided the basis for planning a market survey of approximately 150 Canadian translation suppliers. The initial approach was to include testing the demand for 3-4 products defined at a conceptual level. However, the results of the focus group discussions showed that the market survey would be better used to assess the needs and potential opportunities and that it would be premature to test particular product concepts.

The overall sample size used for the market survey was 150, selected on the basis of providing an allowable error of $\pm$ 8%, 95 times out of 100 (assuming maximum variances) and assuming a randomly selected sample from the overall target population. The overall sample was allocated among four types or groups of respondents: independent translators, translation firms, translation departments within private sector firms and translation departments with federal and provincial governments. The allocation of the sample among these four types of translation suppliers was based on the estimated size of each group and volume estimates of material translated by each group (see Tables A and B in Appendix IV). The allocation resulted in a smaller number of survey interviews amongst independent translators than would a totally random selection from the overall population of translation suppliers. However, the resulting allocation is more representative in terms of assessing the need for and potential of CAT systems. While the reliability of the survey results for each type of translation supplier is somewhat lower than $\pm$ 8%, 95 times out of 100, it was also necessary to make a trade-off between the number of interviews and the length of the questionnaire. The budget in our terms of reference allowed for 10-15 minute interviews, whereas the actual interviews required 20-25 minutes to obtain the critically required

information. Appendix III provides additional information on the survey and questionnaire design, including the research issues and a copy of the questionnaire.

We believe that the results of the market survey, together with information obtained from the focus group discussions, other interviews and previous studies, provide a sound basis for assessing the opportunities associated with CAT systems.

The following sections of this report progress from a discussion of the overall demand for translation services, to a description of the translation supply industry, to a discussion on new opportunities for CAT systems and strategic considerations for pursuing these opportunities.

II. SIZE AND NATURE OF THE MARKET FOR TRANSLATION SERVICES

A. FACTORS DRIVING THE DEMAND FOR TRANSLATION SERVICES

The demand for translation services is driven by the need to overcome the language barrier experienced in a broad range of human endeavour having economic, social, political, or military significance. This language barrier interferes with the transfer of information and the resulting accrual of benefits to either the originator or recipient of the information. The cost of translation services is ultimately justified by these benefits.

The need for translation is driven by the use of native language, international trade and business, international intergovernmental relations and information technology. Each of these factors are addressed in the following sections.

1. Use of Native Language

There are, on both the Canadian and international scenes, opposing forces at work in this area. The extent of foreign language instruction, especially in English, throughout the world has been increasing significantly in recent years (3). Increased competency in foreign languages will naturally decrease the need to conduct business exclusively in one's own language. Opposing this are certain political and sociological developments which have resulted in the placing of a higher value on being able to conduct business in one's native language. In Canada, both English and French have status as Official Languages as defined by the Official Languages Act (9). The Federal Government and various governments at other levels are committed to provide their services to Canadians in the Official Language of their choice and to ensure that government business is conducted to a representative extent in both languages. The Office of the Commissioner of Official Languages is responsible for ensuring that these policies are implemented.

While it is difficult to quantify the outcome of the interaction between the above mentioned forces, it is clear that both private and public sector organizations are placing emphasis on providing services in their clients' own languages, both in Canada and elsewhere.

2. International Trade and Business

Improvements in transportation and communication are generating opportunities in this area by making markets and suppliers around the world more accessible to each other. Additionally, the high level of investment required to produce the increasingly sophisticated products and services demanded necessitates access to these world markets, as well as access to cheaper labour in less industrialized countries to produce goods at competitive prices.

These factors, which are economic in origin, give rise to the need for translation of a broad range of subject matter. In addition to product related material such as technical documentation and advertising, there is a wide range of business related material falling in categories such as legal, financial, and administrative.

3. International Inter-governmental Relations

Nations' social, economic, political, and military interests extend beyond their borders and create a need for translation services to overcome the language barriers. Translation needs stem from international associations such as the United Nations, from military alliances such as NATO, and from international organizations that develop standards (e.g. telecommunications), that deal with economic development (e.g., OECD and the World Bank) and that deal with trade. Needs for translation stemming from international relations also arise in countries' governments in order to, among other things, benefit from intelligence gathered, prepare for meetings with foreign governments, and carry out joint military maneuvers with other

countries. The level of international inter-governmental relations saw a significant increase as a result of two World Wars and, more recently, it has increased due to the development of a global economy. This has resulted in a sharp increase in demand for translation services.

4. Information Technology

The availability of ever improved and new information technology creates new expectations for foreign language material to be translated as well as creating some expectation from end-users that the computer must be able to translate selected materials they already have on computer, e.g., police reports. The advent of desk-top publishing will make foreign-language reports, as well as technical and business journals available in computer-readable format and hence more easily accessible and cheaper to retrieve. This in turn could result in increased demand for translation services.

Although all four factors influence the demand for translation services in Canada, the increased interest in and use of native language has been the single largest driver behind the increased demand seen in this country during the past twenty years. This has been accompanied by an increasing desire in both the public and private sectors to communicate with Canadians in the official language of their choice.

B. CHARACTERISTICS OF THE END-USERS' EXPECTATIONS

The benefits sought after by end users influence the demand for translation services as a function of characteristics such as cost, turnaround time, and quality.

The nature and extent of these benefits are based on the utility of the information contained in the material be translated. The utility of this information depends on three primary factors: relevance, quality, and

timeliness. The interactions between these factors and translation services are introduced next.

1. Relevance

The relevance or inherent value of translated material is derived from the material to be translated and hence the associated economic, social or political activity. This can affect the range of translation costs which are acceptable. Although the inherent value of the source text cannot always be quantified rigorously, the product of sale price and expected sales volume may provide a useful estimate of relative economic importance.

2. Quality

Quality will contribute to the total value of a translation to an extent determined by the accuracy requirements of the user. The greater the extent to which a decrease in quality degrades the inherent value of information, the more sensitivity to translation quality will be shown by the demand for translation. Quality of translation work is usually considered to consist of three components: accuracy, intelligibility, and style (3). Accuracy refers to the extent to which the information in the source text has been transferred to the target text without omission or distortion. While accuracy is always desirable, the consequences of poor accuracy in, for example, an aircraft maintenance manual will be considerably more severe than in a travel brochure. Intelligibility refers to the ease with which the translated text can be understood. Style is a subjective term dealing with the use of language to convey ideas.

Both intelligibility and style, although not unique to translation, affect the value of translated material to varying degrees, depending on the subject matters and situation. Style is clearly a more

significant factor in translation of literary works and advertising than in legal or financial documents.

3. Timeliness

The turnaround time for translation work affects its value under the following conditions:

a) when timeliness affects the value of the associated economic, social or political activity; and

b) when the turnaround time for translation is significant compared with the time scale of the associated activity.

For example, the production of technical manuals associated with engineering equipment in several languages cannot begin until product design is finalized and must be completed prior to equipment delivery. On the other hand, the conversion of older documents is often less sensitive to time constraints.

C. CURRENT COSTS OF TRANSLATION SERVICES

Translation costs in Canada vary from 12¢ to 35¢ per word, depending on the nature of the translated material, end users, and the source of translation supply. The Secretary of State estimates its average in-house translation costs at 28¢/word. SOS also has significant translation work done externally, paying an average of 18¢/word to private translation firms and freelancers. Private sector organizations generally pay from 15¢ to 30¢/word for translation work depending on the circumstances. Specific factors resulting in higher prices include specialized subject matter, high quality, and quick turnaround time. The lower costs charged by small translation firms and freelance translators may be partly due to their lower overhead but the fact that their relatively modest resources limit the effective size of their market may also be a contributing factor.

D. CURRENT DEMAND FOR TRANSLATION

1. Overall Size of the Market for Translation Services

Based on information in the literature, current world translation volume is believed to be between 20 and 60 billion words per year and increasing at a rate of 10% per year (3). However, there is considerable uncertainty regarding these estimates since:

* the response rates of international surveys have varied considerably from one country to another; and

* the size of the translator population is not known with uniform accuracy in all countries.

An estimate of 750 million words per year has been quoted for the Canadian translation market (6). While it is not clear upon which assumptions this estimate is based, it implies that the federal government translation service is responsible for about 50% of the total which is consistent with the beliefs of several persons in the translation industry.

The survey conducted during this study provides an estimate of 550 million words per year based on the following calculation:

	Total volume of translation work reported by survey respondents:	174 million words/year
+	Number of translators/revisors in respondent organizations:	1,585
X	Estimated population of translators:	5,000
=	Estimated Canadian Market	548 million words/year

This estimate of the Canadian market for translation services is very rough since:

a) It assumes the demographics of the respondent organizations is representative of the entire population in terms of such factors as: full versus part-time translators, productivity, and persons doing translation only versus translation and revision.

b) The total number of translators in Canada (5,000) used to derive the estimate is itself an estimate that was published by Statistics Canada, is somewhat dated (1984), and includes interpreters (believed to be a relatively small number - approximately 200).

c) The survey results may be providing a higher than actual estimate of the number of translators/revisors in the respondent organizations due to inclusion of freelancers.

The actual number of translators/revisors in the respondent organizations may be as low as 1,122 which, if used instead of 1,585 in the above-mentioned calculation, would result in an estimate of 775 million words/year for the Canadian translation market.

Based on the foregoing discussion, the current Canadian translation market is assumed to be in the 550-750 million words per year range. This represents an annual market of from $100 to $230 million on the basis of a cost per word range of 18¢ to 30¢ per word. As will be explained later in this report, this uncertainty regarding the current size of the Canadian translation market does not affect the conclusions reached regarding the potential, for Canada, concerning Computer Assisted Translation (CAT) systems.

2. An Analysis of the Current Market for Translation Services in Canada

The Canadian government is a major customer for translation services in Canada. Government sources indicate that the Translation Bureau at the Secretary of State (SOS) supplies just over 300 million words per year or 40-54% of the total Canadian market of which 60 million words per year or 8-11% of the Canadian market are contracted out by SOS to translation service firms and freelancers.

Outside the Federal government, no other sectors appear to dominate the demand for translation services. Table 1, opposite, shows the frequency of industry sectors cited by respondents in our survey as being major sources of business. No one sector is responsible for more than a quarter of the citations. Several of the most frequently cited sectors such as advertising/research/consulting, other manufacturing, and high tech require a broad range of subject matter to be translated.

There are several broad categories of subject matter that contribute significantly to the translation market. The relative amounts of translation work falling into various categories determined by our survey and the two previous surveys are summarized in Tables C and D in Appendix IV. While there is considerable variation between the results from the surveys, interpretation of these differences is difficult since different categorization schemes were used. All three surveys indicate that the market for translation work can be divided into several broad subject matter categories. From a supplier perspective, the following comparisons between the categories are useful:

* subject matter knowledge and writing ability corresponding to persons with very different backgrounds (scientific vs. advertising vs. legal vs. administrative) is required by the translation process in different categories;

* all of the categories, except for legal and financial, are broad and heterogeneous in terms of the terminological knowledge base required;

* within all the individual categories, there will be a broad variation in terms of requirements for quality and turnaround time;

* the end users of translated material belonging to the various categories will belong to distinct types of departments or organizations. They will have uniquely different and identifiable concerns/uses for the translated material; and

* there is some variation in the importance of various categories to different types of translation supplier. For example, from Table D, legal material is more important to independent translators than to the other types of respondents. Also, engineering/technical material is less important to both independent and government translators.

The aspects of the categories which are mentioned above will be relevant to translation system suppliers both in deciding whether to pursue a specific venture and in deciding on a market strategy.

The demand for translating material between English and French is significant both in Canada and internationally. As shown in Table E (Appendix IV), translations involving the English-French language pair account for approximately 20% of the international translation market covered by previous surveys. Other than the German-English language pair, which accounts for a similar percentage, no other language pair represents a comparable portion of the world market.

Results from our survey indicate the Canadian market is dominated by the English-French language pair to a much greater extent. The

relative amounts of the Canadian translation market accounted for by English-to-French and French-to-English translation work are 80% and 14% respectively. This is consistent with the relative translation volumes handled at SOS and the general perception among translators in other organizations that the English/French language pair dominates the Canadian market.

A comparison of the relative size of the Canadian and world markets for all language pairs and the English-French pair is shown in Table 2, opposite. Because the Canadian translation market is dominated by English and French, it represents a significant share (between 5% and 19%) of the world market for English-French translation.

These results presented in this section indicate that the demand for translation work is based on a broad range of endeavours with no single type, with the exception of the federal government, dominating the demand. One would not expect the growth in demand for translation in the private sector to be linked to that of any specific activity or industrial industry.

E. FUTURE DEMAND FOR TRANSLATION

The results of previous international surveys indicate that total demand for translation is increasing by about 10% per year, which is considerably greater than the growth rate of the world economy. This is consistent with the qualitative picture based on technological, economic, and social trends. In Canada, these trends have been evident over the last few years as the capability to carry out business of all kinds equally well in both official languages has become a sought after goal. There are no figures available for the overall growth rate of translation services in Canada and it was not possible to determine this in the current survey. Some indication concerning the rate of growth in demand for translation services can be obtained from SOS data. Prior to the early 1980's the volume of translation work at SOS had been increasing at an annual rate

of 10% per year. In order to keep costs down, client departments have been encouraged to work with SOS and Treasury Board to eliminate unnecessary and low priority job requests. As a result of this program, annual translation volume at SOS has remained constant over the last three years. The initial departmental forecasts, those prior to the cost reduction process, have been about 10% greater than those of the previous year.

F. LATENT DEMAND FOR TRANSLATION

The CEC survey indicated the existence of a so-called latent demand for translation equal to about 30% of the total current demand. This refers to the increase in demand for translation services which might result from corresponding performance improvements in the following areas:

* Decreased translation fees;
* Increased translation quality;
* Decreased turnaround time;
* Increased efficiency of receipt/delivery of work; and
* Capability to handle an increased range of languages and subject matter.

The method by which the 30% latent market estimate was arrived at is not given explicitly in the major reference on the survey but it is indicated that the responses to over one hundred questions were taken into account (3). There is bound to be considerable uncertainty in this estimate since the only available method of measurement involves asking survey respondents hypothetical questions requiring fairly prompt answers. In reality, the decision making process will be carried out under different circumstances.

Expressing the potential demand increases resulting from productivity improvements as a single number is somewhat of an oversimplification. These changes are more accurately described, at least in principle, by

demand elasticities with respect to each factor affecting performance. Our survey did attempt to collect qualitative information concerning the magnitude of latent demand and the relative importance of the various factors mentioned above. Respondents were not asked to differentiate between increased business due to a greater market for translation services and that due to a greater market share for their firm. Table 3, opposite, summarizes the respondents' estimated effects of cost reduction, turnaround time reduction, quality improvements, and other increased capabilities on business volume. The percentage of respondents indicating that significant potential exists ranges from 35 to 60% depending on the type of improvement. Factors such as improved methods for exchanging text with clients, reduced turnaround time, and the ability to handle a broad range of subject matter were cited as being likely to increase business by somewhat greater numbers of respondents than were factors such as cost and the ability to handle increased numbers of language pairs.

The potential demand for services based on automated translation was investigated by asking survey respondents if they believed that they could increase business by offering an "initial draft" translation at one half the cost and time required for a normal translation. The responses, shown in Table 3, indicated that about a third of respondents in all categories felt that a business increase was at least somewhat likely. The requirements of end users who do not deal with translation services because of cost or time constraints are not reflected in the results given here. The responses to this survey question, however, were sufficiently positive to justify further study in this area, preferably involving a survey of translation user needs directly.

Comparison of our survey results with the previous CEC survey concerning latent demand is difficult because of insufficient information and a lack of question-by-question correspondence. However, the current results do appear qualitatively consistent with the reported estimate of a 30% latent demand for translation services. It should also be noted that

respondents to the CEC survey, included a large number of firms without translation departments whose needs may not be fully understood by translation firms.

In spite of the uncertainties of measurement, the available evidence indicates that productivity improvements will result in some increased use of translation services. While these factors will have some positive effect on the economic viability of Computer Assisted Translation (CAT) systems, they will not be critical to the potential for a Canadian CAT product supply industry. The factors of competition and market penetration will contribute considerably more to the variability in commercial potential for Canadian firms. These factors will be discussed in Sections IV and V.

III. DESCRIPTION OF THE TRANSLATION PROCESS AND INDUSTRY

A. BASIC ALTERNATIVE METHODS USED TO TRANSLATE MATERIAL

For the purposes of examining the potential for Computer-Assisted Translation (CAT) products, the methods used to translate material can be described in terms of the following activities.

1. Pre-editing

The source text is read by the translator for general understanding and identification of terms or passages which require clarification. These items are marked in pencil and may be annotated.

2. Research

The goals of research are to:

* develop a complete understanding of the source text; and
* assemble the necessary information to produce a target text possessing appropriate quality and which is consistent with common terminological usage.

These include dictionaries, terminological databases, previously translated material and documents on the relevant subject matter. Also, other translators or subject matter experts may be consulted. Table 4, opposite, shows our survey results concerning the percentage of translators' time spent on various research activities. These results indicate a total percentage of approximately 60% of productive time is spent on research which is considerably higher than the 34% cited by respondents to the DEC survey for similar activities. The high percentage's in our survey may be due to a significant number of revisors being included in the sample. Nevertheless, these results provide strong qualitative evidence that

the various research activities constitute a significant fraction of the professional activity within a translation operation and represent costs comparable to those of translation itself.

Respondents were asked about the frequency with which they consult other people about a particular translation job. About 20% of the respondents cited consulting others more than five times per day and another 40% consult others from one to five times per day.

Consulting terminological and subject matter references is done using whatever method is available and considered cost effective by the individual translator or firm. Traditionally, this has involved paper files and index cards. There is also a significant and increasing use of computer technology for storage and access of this information.

3. Translation

This is the key process whereby the translator reads and understands the source text and produces the target text either by writing, typing or dictation. If the translator uses a dictaphone, the translation is converted to written text by a dictaphone typist.

Traditionally, the translation process was considered to involve substituting the source text words or phrases with equivalent constructions in the target language. A more modern view is that translation consists of a two-step process consisting of:

* determining the meaning of the source text; and
* constructing the target text to convey this meaning.

The difference between these viewpoints is significant with respect to the translation process itself and its associated prospects for automation. The method used to record a translation as it is being performed will impact on the feasibility of using CAT products. The

5. Target Text Document Production

The revised version of the target text must be produced on the appropriate medium in a form which is acceptable to the customer. This involves all aspects associated with visual presentation including type font, spatial formatting, and incorporation of drawings with translated annotations. While this activity is not strictly part of the translation process, it involves the utilization of some or all of the same people and equipment. It is useful to consider it part of the overall translation process in a practical sense.

6. Receipt and Delivery of Translation Work

As with document production, these activities are related to the translation process more in practice than in theory. Source and target text documents can be exchanged with clients as either hardcopy, magnetic media (tape or diskette), or by data telecommunications (modem or facsimile machine). These activities interact with the true translation activities through sharing resources (people, equipment) and data (source and target text). There are, consequently, certain limitations and opportunities for computer assistance resulting from this interaction. These will be discussed in Section IV.

Table 6, opposite, summarizes the survey results for all respondents concerning the receipt and delivery of translation work. While hardcopy is still the dominant medium, significant amounts of translation work are being exchanged via diskettes, data telecommunications, and facsimile machine. This is true for all respondent categories although freelance translators rely somewhat more but not exclusively on transfer by hardcopy.

three methods for recording translations are typing, dictation, and handwriting. The relative use, in terms of translation volume, of these methods by our survey respondents is given in Figure 1, opposite. The most commonly used method for recording translations in Canada, accounting for 46% of translation volume, involves the translator entering the target text at the keyboard of a word processing machine, a computer, or a typewriter. A smaller amount, 15%, is hand written by the translator during the translation process. Approximately 38% of translation work was, on average, reported to be performed using the dictaphone method. The corresponding results for the CEC and DEC surveys also indicated greater use of keyboards than the other two methods.

Based on our survey, Translators outside of government make significantly more use of keyboard devices whereas in government they make more use of dictaphone (see Figure 1, opposite). The potential significance of this difference will be deferred to Section V.

4. Revision

The initial target text, in written form, is edited to remove errors in grammar, terminology and meaning. The style may also be improved during this stage. The revisor may be a revision specialist, another translator or, in some cases, the initial translator. From our survey results in Table 5, opposite the next page, it is clear that the translation industry employs people in all of the following categories: people doing translation only; people doing both translation and revision; and people doing revision only. Although revisors constitute a significant fraction of all translation professionals there are almost twice as many translators who also do revisions.

7. Maintenance of Dictionaries and Terminological Databases

While this activity can be classified as an indirect activity, it is closely connected with the translation process and is based in part on the translator's professional body of knowledge. It has been established by previous surveys and verified by the current one that a significant fraction of translation firms' resources are consumed by this activity. Our survey respondents reported that their staff spent on average of 20% of their time maintaining terminological databases, although the variance of time spent on this activity among the respondents is quite broad. While the expression "terminological database" implies computer technology, most stores of terminological information are maintained on paper files or index cards. They require regular updating as a translation organization encounters new terminology, as conventions change, and as new terminology is developed in the relevant subject areas. The responsibility for database maintenance is divided between those people who function primarily as translators and those who are terminology specialists. The effect of computer technology on the cost and effectiveness of database maintenance will be discussed in Section IV.

B. CHARACTERISTICS OF THE TRANSLATION SUPPLY INDUSTRY

1. Main groups within the industry

In Canada and other industrialized countries, the translation industry is active in overcoming the language barrier for both private and public sector activities. The following shows the organizational structures in Canada that provide translation services and rough estimates of the number of such firms/units and translators involved.

Organization Type	Estimated No. of firms/units	Estimated No. of translators
Private sector translation firms	380	1400-2000
Translation units within corporations	260	1200-1700
Translation units within governments	110	1100-1200
Freelance translators	-	1000-1100
		4700-6000

These estimates for private sector translation firms and translation units within corporations and governments are based on lists compiled for conduct of the survey, government sources, and the results of our survey. The estimate for the number of freelance translators was made by subtracting our value for the number of translation firms from a previous estimate (8) of the combined number of translation firms and independent translators. The total number of translators previously estimated by Statistics Canada in 1984, 5000, falls within our estimated range. It should be emphasized the uncertainties associated with the definitions of part-time and independent translators limit the accuracy possible in this estimate.

Translation firms provide services to both public and private sector clients on a fee per word translated or fee per hour worked basis. The cost effectiveness of translation operations in these firms is based on the relationship between fees collected and the costs associated with translation and supporting operations.

Freelance translators, many of whom work only part-time as translators, provide their services to private and public sector organizations including the translation firms described above. They work at home or at their clients and generally spend very little on

advertising, receiving most of their business through word-of-mouth. Considerations governing the cost-effectiveness of translation operations for freelance translators are similar to those for translation firms.

In-house translation units of private sector organizations provide service to other departments within their firm or organization that need to disseminate information in more than one language to customers or other interested parties. For private sector firms, this includes intelligence and technology reports as well as advertising literature necessary for the promotion and use of a product or service. Besides commercial enterprises, this category includes associations which represent the interests of various groups and have requirements to provide information to interested parties. The central interest of the translation departments in this category is to deliver the organization's information in more than one language with consistent quality while meeting budget and timeliness constraints.

The cost effectiveness of translation operations in these departments is based on a comparison of the benefits derived by the overall organization with the translation unit's budget.

Unlike the translation firms and freelance translators, the translation departments are often able to share computer and automation resources with a larger organization located on the same premises.

Government translation units such as those within the Translation Bureau of the Department of the Secretary of State (SOS) provide translation services to support government departments and programs. Within the federal government, these units are located at Secretary of State (SOS) and in other federal government departments. The bulk of SOS translators are located in the National Capital region but there are some units working in other regions across the country. In addition to its in-house operations, SOS

contracts out a considerable amount of work to private sector translation firms and freelancers. Most of the provincial governments and large municipalities have similar, albeit much smaller, translation units.

For public sector translation units, the considerations associated with cost effectiveness will be similar to in-house translation units in the private sector, that is, meeting the parent organization translation needs using as small a budget as possible.

2. Characteristics of the translation supply industry

The size of a translation organization, either independent or departmental, affects the economic and operational feasibility of CAT products.

Our survey shows that the overall average translation unit has 15 persons, although it is somewhat higher for translation groups within the federal government, as shown in Table 5, opposite p. 24, which shows the average number of persons in various categories associated with each type of organization. The survey questions concerning numbers of people of different types within an organization were not asked to independent translators. Table 7, opposite, shows that the distribution of size based on several different employment categories is broad and that a significant fraction of translation groups employ several people. Comparisons with other surveys concerning translation group size are difficult because of the unstructured nature of the translation industry coupled with differences in survey questions. For example, the number of persons in each employment category reported in the C & L survey is higher than those reported in previous surveys. This may be due to the fact that our survey asked respondents to include associated freelancers, irregardless of the extent of this association. Previous survey results indicate that Canadian translation firms have an average of eight translators on

staff (three full-time and five part-time). Our survey results for the non-government organizations range from five to eighteen persons per organization and hence are in rough agreement with the previous results. It should also be noted that the numbers of word processors/typists most likely denotes a category which overlaps with the three translation related categories. Also, the numbers of persons working on and off site are roughly equivalent.

Table 8, opposite, summarizes our survey results on the amount of computer and office automation equipment used by respondent organizations. The methods used to translate material and the use made of computer and office automation equipment vary significantly with the size and type of translation unit. However, several observations based on our survey are noteworthy:

* A significant percentage of organizations have at least one computer (71%), use local area networks (14%), use a mini or mainframe computer (16%), use terminology databases (46%), modems (39%), and facsimile machines (34%) (See Table 8);

* Translation groups in the federal government make more use of dictation equipment than do other groups in which the translated material is entered at a computer, word processor or typewriter by the translator as it is being translated; and

* Respondents generally consider themselves to be reasonably knowledgeable and experienced in computers.

These observations are relevant for later discussion on the opportunities and market for Computer-Assisted Translation (CAT) systems.

C. LEVEL OF INVESTMENT IN AUTOMATION

The current level of automation reported by the survey respondents is fairly high which demonstrates that these organizations will invest in technology in order to satisfy their needs concerning productivity. To provide more information on the current level of investment, respondents were asked how much they spent on automation over the last year. The responses, shown in Figure 2, opposite, indicate that investment level is broadly distributed among the respondents and averaged $17,700 per organization last year. (Because SOS automation budgets are centralized, the survey responses for government spending were not meaningful and are not included in the survey.) Approximately one-half of respondents reported expenditures between one and twenty thousand dollars. Only about twelve percent spend above twenty thousand dollars. The highest reported expenditures were approximately a quarter of a million dollars. Table F (Appendix IV) shows the variation in responses among the respondent categories. The average amount spent by freelance translators, private translation firms, and translation departments within private firms spans a range from $10,000 to $21,000. Because SOS automation budgets are centralized, the survey responses for government spending were not meaningful. Departmental spending estimates for 1986-87 indicate that a total of $1,857,000 was spent on computers and automation to support the bureau's 868 translators and terminologists. Using this figure for government respondents, who were predominantly from SOS, and the survey data for the other respondent categories, the amounts spent on automation last year per translation professional working on-site were calculated and are shown below:

Translation Department within Company:	$ 4,068
Translation Group in Government Dept:	$ 2,164
Private Translation Firm:	$ 5,334
Independent Translator:	$ 10,616

To obtain a rough estimate of the extent to which the investment level in automation is changing, respondents were asked about their specific plans over the coming year. Table G (Appendix IV) lists the products which

respondents cited most frequently as having plans to purchase. Approximately forty percent of respondents cited no plans for automation over the coming year.

D. NEEDS FACED BY THE TRANSLATION SUPPLIERS

The discussion in this section focuses on the needs of translation suppliers, based primarily on the results of focus groups and the survey conducted during this study. The assessment of needs focused on determining the extent of translation suppliers' concerns regarding the efficiency of their operation and quality of their service.

The needs of translation suppliers which were identified fell into the following four classes:

a) Reducing translation costs.
b) Reducing the turnaround time.
c) Improving the quality of the end-product, i.e., the translated material.
d) Facilitating the receipt and delivery of translation work.

Although the relative emphasis on these four classes of needs may vary among respondent categories, the underlying concerns of providing a high level of service at the lowest possible cost are present in all cases. They are, in turn, based on the fundamental need to increase profitability and/or business for firms providing translation services.
To assess the extent of each class of need, the survey questionnaire was designed to include the following topics:

1. Identification of activities that are costly in terms of time required to perform them;

2. Major causes of quality problems;

3. Major problems encountered when exchanging work with clients;

4. - Problems encountered with the use of automated equipment;

5. Factors concerning translation services which clients consider important; and

6. Potential of various service improvements and cost reduction to increase business.

The responses on which activities are costly in terms of time are summarized in Table 9, opposite. Roughly 80% of the time, activities other than translation were cited as costly in terms of time. Of these non-translation activities, those involving research were cited most often among most respondents. The other translation related activities including revision and document production were also cited fairly often. Clearly, CAT products which can improve the productivity of any of these operations, especially research, would be useful.

Table 10, opposite the next page, presents a summary of the major causes of quality problems cited by the survey respondents. The most commonly cited identifiable cause was a lack of knowledge of the subject matter, with special reference being made to technical translation. This lack of subject matter knowledge means that more terminological research must be performed to attain a given level of quality. Unfamiliarity with the subject matter and lack of reference materials compound the problem further by limiting the efficiency with which research is carried out. A significant fraction of problems, approximately one quarter, are classified as "other problems" in the table due to their being too numerous or poorly described to mention explicitly.

Respondents were asked to comment on problems encountered exchanging work with clients. The results, summarized in Table H (Appendix IV), indicate that about 50% of respondents mentioned lack of system

compatibility as a widespread problem. This represents a need which is not inherent to CAT but which should be taken into account if CAT products are to deliver maximum benefit to customers.

Less than ten percent of respondents cited difficulties with automated equipment they are using. This does not mean that there are no limitations present with the equipment currently in use since the survey question did not probe this level of detail. Our analysis of currently available systems, discussed in the next section, indicate that they do have deficiencies.

Respondents were asked to indicate the importance to clients of various factors associated with translation services. The C&L survey responses which are given in Table I (Appendix IV), verify the importance of cost, quality, and turnaround time to clients of respondents in all categories except for the case of cost with government translation groups. Based on discussions with translation service management in the Federal Government, however, cost as seen in departmental budgets is a major concern both to the Translation Bureau and departments using its services. It therefore seems likely that this question was not interpreted by government respondents in the manner expected.

The factors of cost, quality, and turnaround time are to some extent interchangeable by shifting of resources within a translation operation. It is therefore likely that improvements derived by CAT products may affect all three of these areas.

The potential of various improvements in translation services to increase business was first introduced in Section II as part of the discussion on latent demand. From the viewpoint of the translation service supplier, these potential improvements also represent needs to the extent that they can increase profitability or business. The responses summarized in Table 3, opposite p. 19, indicate that from one third to one half of

respondents felt they would derive useful benefits from the improvements which are listed. Improvements such as decreased turnaround time, increased subject matter range, diskette compatibility with customers, and full computer communications with customers received positive responses from roughly half of all respondents. Of those improvements related to the translation process itself, decreased turnaround time received the greatest response. The largest response for any improvement was for full computer communications with clients (roughly 60% positive). While this may seem surprising since the issue does not involve translation itself, it indicates the practical significance of general office automation problems to specialist groups such as translation organizations.

Because of the potential usefulness of CAT for translating material having a simple format or structure compared to normal prose, respondents were asked if they currently handle this type of work. Table 11, opposite, shows, that about 30% of all respondents handle at least some material of this type, with the majority of such respondents being in the private sector. The types of subject matter which were most often cited as involving some simple format/structure are listed in Table J, (Appendix IV). The only type of material cited corresponding to a specific subject matter was legal. Other categories referred to document formats such as reports, notes, memos, etc. The category of "other material" was the most frequently cited. As will be discussed in Section V, the greatest potential for CAT in this area may lie in applications which are not currently impacting on translation suppliers' needs.

For an application involving a limited language domain to be promising for automated translation, it should have as simple a sentence structure and represent as large a volume of text as possible. The previous TAUM Aviation and METEO projects have provided experience in possible applications using these criteria. A recent Canadian study indicates that only a few promising applications are known within the types of text being

translated manually at the Translation Bureau (10). The identification of applications with sufficiently simple structure and large enough volume of text to make automated translation feasible represents a challenging problem.

The feasibility of CAT technology to address the needs discussed above will be discussed in Section IV from both technical and economic perspectives.

IV. POTENTIAL FOR COMPUTER-ASSISTED TRANSLATION (CAT) SYSTEM

A. CURRENT AVAILABILITY AND USE OF CAT SYSTEMS

The current market for CAT technology is difficult to assess because many of the products are general purpose products and are not marketed specifically to the translation industry. Those products which are translation specific represent a $10 million market in the U.S. alone which is growing at approximately 30% per year (11).

Both translation-specific and general purpose products can be analyzed by comparing the features which they offer to a common set of known, useful and potentially useful features. These features can be classified under the following headings:

A. Terminological research;
B. Automated Translation;
C. Input/Output;
D. Text Processing; and
E. Project Management.

By comparing features offered by existing products with all the useful features potentially available, it should be possible to determine the opportunities for new and improved products.

Currently Available General Purposes Products

Software packages which are not specifically designed for translation but which assist translation and supporting operations such as information management, document production, and production scheduling are widely

used throughout the translation industry. Available products fall into the following categories:

a) Foreign Language Kits
   (for character representation and i/o)
b) Word Processing Programs
c) Author Systems
   (grammar & style checking, text organization)
d) Diagramming Systems
   (produce drawings and integrate them with text)
e) Desk-top Publishing
   (word processing with advanced document production capabilities)
f) Project Management
g) Telecommunications
h) General Office Automation
i) Database Management Systems
j) Information Retrieval Systems
k) Full-text Indexing and Retrieval Systems

Explanations and examples for each of these categories are given in Appendix V.

Table 12, opposite, indicates which features are present in each class of product currently on the market. A "+" sign indicates that the feature has been more or less successfully implemented in the product class and a "-" sign indicates that the feature is present but considerable improvement is needed. The absence of a sign indicates that a feature has not been addressed.

Excluding Category B, which is associated with automated translation, there are products which provide most of the features in each of the other categories. The category where existing products are weakest is A, terminological research. Therefore, an opportunity exists to develop new CAT products in this area. In order to have the full range of available

features, it would be necessary to use at least seven different products on the same computer system. Accomplishing this without degrading system performance or creating a cumbersome user interface represents an opportunity to provide new CAT products based on improved system integration.

Currently Available Translation Specific Products

Several representative products of this class are listed below and are briefly described in Appendix V:

Public Term Banks:

a) TERMIUM (Secretary of State)
b) Banque de terminologie du Québec

Integrated tools for the individual language professional:

c) Termex-Mercury
d) ABC WORD
e) INK Text Tools

Automated Dictionary Lookup Tools:

f) ALPS AutoTerm

Commercial Machine Translation Systems:

g) ALPS TransActive
h) Logos
i) Smart
j) Systran
k) Weidner

Table 13, opposite, indicates the features which are present in each of the above systems. The "+" and "-" signs have the same meaning as in Table 12.

Categories of features for which these systems are particularly weak include terminological research tools, A; input/output, C; and project management, E. Also, none of the systems address all of the features associated with automated translation. Moreover, the field of automated translation may require considerable progress before truly cost effective products can be produced. Therefore, opportunities may therefore exist to improve automated translation systems by integrating a greater number of features.

From the focus group discussions it was learned that there is widespread feeling among translators and translation group managers that currently available automated translation systems are not cost effective due to their high acquisition and maintenance costs and the inferior quality of the translations which they produce. Major problems cited by the focus group participants are:

a. The currently available CAT systems produce rough translations which are vastly inferior to the unedited translations produced by human translators. In many cases, information is lost by automated translation and ideas can be badly distorted. Besides increasing the workload of revisors, this can cause reliability problems.

b. Current CAT systems require substantial direct and indirect labour cost. Foremost among these are the costs associated with maintaining the dictionaries used by the automated translation program. These dictionaries require more maintenance per word than conventional terminology dictionaries maintained by translators.

c. The claims of the suppliers are regarded as vastly exaggerated.

In spite of these difficulties, the focus group participants indicated an interest in future CAT systems if they can improve the cost effectiveness of translation services.

B. RESEARCH & DEVELOPMENT PROGRAMS RELEVANT TO COMPUTER ASSISTED TRANSLATION

National research and development programs in the technologies on which CAT products are based are in progress in the United States, Japan, the United Kingdom, and several European countries. The goals of these programs, levels of funding involved, and potential for CAT have been discussed in previous reports prepared by Cognos Limited for the Department of Communications (5, 6). The key characteristics of these programs which are relevant to CAT are:

a) The programs are directed at several diverse computer technologies which are the underlying technologies on which CAT products are based. These include Artificial Intelligence, Natural Language Processing, Knowledge Representation, office automation, and the human/computer interface.

b) The funding levels for these projects are in the hundred's of millions over the next three to five years.

c) The programs are long-term with intermediate sub-optional goals.

d) They call for rapid action and results.

e) They have been designed to relate to the nation's business climate.

R&D work on translation-specific CAT systems, primarily involving automated translation has been reported at a recent international symposium. (1) The research programs and experimental systems described are based on more sophisticated linguistic models than used in the currently available commercial systems. Because these experimental systems are designed to conduct research into different theories and techniques from computational linguistics, there is no guarantee that they can lead to a commercial product. On the other hand, it is likely

that at least some of these experimental systems could form the basis for future commercial systems. Also, the knowledge derived from system development, including proprietary knowledge, will provide a competitive advantage if applied to future commercial development projects.

Based on National Science and Engineering Research Council (NSERC) listings of strategic and operating grants, the funding level for Canadian university computer research and development in areas underlying computer-assisted translation was approximately $2.3 million in 1985-86. This R&D was distributed among most Canadian universities and involved all of the technologies underlying CAT including Natural Language Processing, Artificial Intelligence, and Knowledge Representation.

There are at least four Canadian Universities with Machine Translation programs and at least ten with programs in Natural Language Processing. All of the Machine Translation programs and most of the Natural Language Processing programs are fairly small having less than five researchers including students involved. Federal Government activities and interests in CAT related technologies are summarized in an earlier report prepared for DOC (6).

The computer industry in Canada is significant ($9.6 billion in 1986) and growing at about 10% per year. Although most of the industrial activity is by hardware manufacturers who are primarily foreign-owned, there is a growing proportion of Canadian-owned software producers. Some of these companies have been involved in relevant technologies such as Artificial Intelligence. Also, there are several small Canadian consulting firms with expertise in computer-assisted translation and related areas. These firms provide consulting services to industry and government on CAT systems and technology and have developed prototype systems under contract. In particular, the present implementation of METEO, operated for Environment Canada by SoS, has evolved from the University of Montreal TAUM METEO project due to improvements made by private sector consultants under contract. The overall Canadian effort is, however, not

comparable in size to those of countries such as the U.S., Japan, and Germany, even though the Federal Government also supports related software development by contracting out through the Department of Supply and Services. For example, companies such as ALPS and Logos of the U.S., which are dedicated solely to CAT, are larger than most software companies in Canada.

In summary, there are several characteristics of Canadian R&D capabilities relevant to CAT which should be recognized:

1) Canadian expertise and on-going research in the technologies germane to CAT resides mostly in the public sector, i.e., University and government research groups.

2) While the overall level of effort is less than that in the larger western nations, it has been sufficient to establish a base of world class expertise.

3) Several government departments are interested in CAT itself and the underlying technologies.

4) The Canadian software industry, although small, is growing rapidly; based on the successful commercial development of software products.

Therefore, while the scale of Canada's R&D capabilities and existing industrial base is significantly smaller than that of other countries such as the U.S., Canada possesses the necessary expertise required to pursue ventures related to CAT systems. While there is always a need to ensure that enough R&D is being done, the major challenge for Canada is to successfully commercialize technology.

C. NEW OPPORTUNITIES FOR CAT SYSTEMS

1. - Identification of opportunities

The characteristics of both the supply and demand for CAT technology have been examined in the previous sections. Four major areas of potential opportunity have been identified and are listed below:

a) System Integration

There is a significant opportunity to integrate the CAT features which have been found useful to translators into a single system. The ability to move from one application to another within the system without disrupting one's work is a key element of integration. Another key element is to provide compatibility with customers' equipment with respect to document exchange. System integration may be achieved through the extension or combination of existing products as well as the development of totally new, products, including software specifically for system integration. Features that are unique to the translation environment as well those that are of a general purpose nature associated with office automation should be considered for system integration. Also, the introduction of newly developed CAT products must take place within the current automation environment of translation organizations. Product developers will have to account for this to minimize incompatibility and redundancy resulting from the introduction of their products.

b) Terminological Research Tools

These are database applications which allow a translator to obtain information on correct terminology, preferably without disrupting the word processing application being used, while

translating or revising text. These databases may contain any information which helps establish correct usage of terminology based on common practice or accepted standards. The methods of database searching may vary in sophistication based on their ability to obtain and process contextual information from the source text.

c) Automated Translation of Simple Text

The development of a fully automated system which can produce translator-quality translations of any text for which it has a terminology database is not likely to be technically or economically feasible within the near future. However, an automated translation program which operates on source text written in a limited language domain is potentially useful for some applications. The translation algorithms would be less complex than those needed to produce similar quality translations of general text. These systems could, in principle, be designed to consist of two modules, one of which contains the translation algorithm itself and the other containing the subject matter dependent terminology database used by the algorithm. These modules could be sold separately and even supplied by different firms.

d) Automated Translation Producing an Inexpensive Draft

In some situations, it may be useful to obtain a rough translation produced by a fully automated system. These rough translations, produced with lower costs and turnaround times, will be useful to screen documents for general content. As in the type of automated translation system described in c), there may be the opportunity to develop the algorithm and terminology modules separately.

2. Design Considerations for Exploitation of CAT Opportunities

Each of these opportunities may potentially be exploited through the production and marketing of various types of product. In designing CAT products with a given set of specifications, several basic considerations relating to the type of technology to be incorporated into the product must be addressed:

a. Software package versus hardware/software system

Depending on the product specifications, it may be possible to implement the required functions in a software package which runs on available computers. Software packages will generally have lower development costs and sale prices than systems resulting in less risk to the developer and a potentially larger market.

b. Type of Computer System

Either software packages or complete systems which are designed to utilize available personal computers and peripherals will appeal to a larger market than those involving more specialized computers. The processing, input/output, and mass storage of the current and near-future generation of high-performance PC's will be more than sufficient to meet the CAT needs of individual and small groups of translators. The most powerful of these PC's have specifications which meet or exceed those of minicomputers introduced a few years earlier. (12) The factors limiting overall performance over the next few years will be the rates with which new operating systems and applications software are developed.

For CAT systems which have rigorous requirements concerning simultaneous information processing by many users, the more

well established capabilities or mini and mainframe computers may be required.

c. Compatibility with Existing Products

Compatibility with commercially available software and peripherals increases the attractiveness of a product from the user's standpoint. This can be achieved, in part, by incorporating as much off-the-shelf technology as possible into the new product.

Although there may be circumstances for each of the above cases where performance requirements dictate the use of a more specialized or costly approach, the subsequent loss of benefits must be considered.

D. FACTORS AFFECTING THE COMMERCIAL ATTRACTIVENESS OF THE OPPORTUNITIES

1. Market Size

The size of the market for CAT products is dependent upon product type and price. Products such as terminological research tools and software utilities to enhance integration can be offered at lower costs than the automated translation systems and should therefore have a larger potential market. Products whose purchase can be justified based on their increasing the efficiency of a single translator or small group of translators will have the largest market size. Products such as automated translation systems which must be utilized by larger groups of translators to be cost effective will have more limited markets because:

* Less units will have to be sold to service all translators requiring them; and

* Many translators who require the product will not be able to use it since they work in organizations whose size is too small to justify purchasing the product.

The market for CAT systems may be fragmented due to particular products being language pair and/or subject matter dependent. This may apply to some terminological research tools and the terminology database components of automated translation systems. Automated translation systems if the actual database is a significant part of the package sold, and in particular the algorithms required for these systems, will be language pair dependent in the short-term. In the longer-term, translation algorithms may consist of two modules, one for source text analysis and one for target text composition; these modules would have larger potential markets since they would encompass translation of all language pairs whose source/target language corresponds to that of any particular module (3).

The market size for automated translation systems for translating simple text will be highly application dependent. For example, the size of market for a system that translates weather reports is dependent upon the number of customers that have a requirement for translating weather reports, at least until such time that the system can be generalized or improved to handle other subject matter or limited language applications. This market is not well defined (10). It cannot be fully assessed by surveying translation service suppliers since the application may involve large amounts of material that are currently not translated for cost reasons. Also, there may be applications where the material is currently not written in simple or limited language form, but where the users may consider doing so if they could get the material translated at significantly less than current costs.

As stated in Section I, the scope of this study is restricted to the translation supply industry as this is more easily defined than other

markets and is the most promising on the short-term. Nevertheless, it will be advantageous to potential CAT suppliers to consider the applicability of proposed products to the much larger market consisting of people writing professionally in a second language.

Table 14, opposite, provides a framework for assessing the commercial attractiveness of CAT systems from a market perspective. The total potential market size shown in Table 14 corresponds to the market for French-English translation since CAT systems will tend to be language pair dependent in the foreseeable future. The total potential market size was estimated, assuming that:

* 85% of the 5,000 translators in Canada do English-French translation;
* the Canadian market for English-French translation represents 20% of the world market for English-French translation; and
* one unit (of the product) is adopted amongst translation suppliers for every single, third and tenth translator.

The percent of total potential customers, used in Table 14, for whom the product is useful is assumed to be either 20% (e.g., due to the product being subject matter dependent) or 100% (i.e., useful to all potential customers. Finally, the product prices used in Table 14 cover a range consistent with many product scenarios and are tied in with the economic benefits derived by CAT system buyers, which are shown in Table 15 and discussed later.

From Figure 2, between 10% and 40% of the survey respondents have annual expenditures on automation corresponding to the upper and lower limits respectively of the range of product prices used in Table 14. These product prices are therefore not inconsistent with the current financial structure of the translation industry.

In order for a venture to be attractive, the projected revenues must cover the costs of developing, producing, marketing and distributing the product and provide a rate of return on investment which is reasonable in view of the associated risks.

Successful North American software developers spend an average of about 8% of their total revenue on Research and Development with 60% of them spending 10% or less. It is therefore reasonable to assume that product development costs should be no more than 10% of total revenues over the product life cycle. The approximate range of product development costs implied by Table 14 is $50,000 to $2 million. Assuming at least $200,000 is required for product development, nine of the twelve ventures would be viable.

If, however, only the Canadian market was addressed, revenues would be reduced by a factor of five. Acceptable product development costs could then be only 1/5 of the revenue figures for the international market shown in Table 14. In this case, only three of the ventures would be viable. Clearly any CAT related ventures undertaken by Canadian suppliers should be directed at the international market to achieve a reasonable chance of success.

2. Development Costs

Ideally, development costs should be as low as possible to limit the consequences of unsuccessful ventures and to increase the chances of success by permitting the sale of the finished product at the lowest possible price. Products which are more expensive will also take longer to penetrate the available market and this would slow down the development of a successful CAT industry in Canada. The products based on automated translation either of limited language domain material or to produce a rough draft will have significantly higher development costs than those based on terminological research. System integration costs can vary widely depending on

the scope of the product and the extent to which it uses off-the-shelf components. The production of terminological databases for automated modules may be a way to participate in this market without incurring the development costs for a complete system. The development costs must ultimately be paid from the revenue derived from sale of the product shown in Table 14. They should not be greater than ten percent of total projected revenue for the venture to be attractive relative to most other software R&D ventures.

3. Product Feasibility and Usefulness

The usefulness of CAT products is actively being studied by large user organizations and research and developments groups (13, 14). Although there is strong evidence from several sources that these products can improve translator efficiency, there are no quantitative results which are universally agreed upon and easily used for comparison purposes. Increases in translator productivity from the use of terminology research and system integration products are more certain at this point than those from the use of automated translation systems. The effects of increases in translator efficiency from one to ten percent on revenue of a translation organization are shown in Table 15, opposite. These revenue increases over three years of operation represent the price which a translation organization could pay for a CAT product and break even. From an investment perspective, prices of fifty percent of these revenue increases would be reasonable. The product prices used to calculate revenue from the sale of a CAT product, shown in Table 14, are based on the increased revenue to translation organizations derived from its purchase. For each group size, the product price is taken to be fifty percent of the increased revenue to the translation organization based on a five percent efficiency increase.

4. Foreign Competition

For all of the areas of opportunity discussed above, a new entrant into the market will face competition from existing suppliers. These firms have established themselves in the market with some existing products, learned from their shortcomings, and are actively pursuing the development of improvements. This process, which is interactive, is very effective in adapting available technology to market needs. The current CAT suppliers should be considered to have at least a two year lead on new entrant based on the time required to develop and begin marketing a new product.

On the other hand, their penetration of the market and the level of sophistication of their products have been very modest to date. The current suppliers have established virtually all of their business with large organizations such as government departments and agencies and large corporations such as General Motors and Siemens. Because of the large costs of mainframe and minicomputers which, until recently, have been necessary to run automated translation software, small and medium size translation organizations have not been promising customers.

It may be possible for new entrants to catch up to the existing suppliers while minimizing the effects of competitive pressures. This can be done by directing products at the large segments of the market which have not yet adopted CAT technology. These are the comparatively larger numbers of small to medium translation organizations for which today's automated systems cannot be justified economically.

Manufacturers of automated general purpose translation systems must compete with all other manufacturers of similar systems translating the same language pair. For the case of subject matter-dependent systems and automated terminology modules for general

purpose systems, the market will be more fragmented which may allow a new entrant to find a niche based on product differentiation. This also applies to limited language domain automated translation systems.

E. SUMMARY

From the above discussion it is clear that opportunities exist in four areas:

i. System integration
ii. Terminological research tools
iii. Automated translation of simple text
iv. Automated translation producing an inexpensive draft.

There are several key characteristics of the CAT market and industry which contribute to the viability of all of the above listed opportunities from the viewpoint of Canadian suppliers:

i. There is a need for CAT products that can increase the cost effectiveness of translation services.

ii. The currently available CAT products do not perform well.

iii. The current suppliers of CAT products have not penetrated the translation market to any great extent.

iv. There is a significant domestic market for translation services in Canada, primarily involving English and French. The international market involving these languages is even larger.

v. Canada has world class expertise, although on a fairly small scale, in research pertaining to the technologies underlying CAT.

vi. There is a small but growing Canadian software industry which could exploit these opportunities.

In evaluating which opportunities are the most attractive to a new entrant in the market, the factors of market size, development costs, product feasibility, and foreign competition must be considered.

A summary of the affects of these factors is given in Table 16, opposite. The market size for all the classes of opportunity will be limited to varying degrees determined by the effects of language pair, subject matter, product price, and system features. The development costs of automated translation systems will be greater than those of either terminological research tools or products designed to enhance system integration. They will, therefore, be less economically attractive to the comparatively large numbers of small and medium sized translation organizations. The feasibility and usefulness of automated translation systems is also uncertain at this time and may require technical breakthroughs to reach acceptable levels. With the exception of automated translation systems to produce rough drafts from completely general source text, products designed to exploit any of the opportunities can be designed to minimize competitive pressure through product differentiation.

Decisions made at the initial product design stage should endeavour to maximize the overall potential of a specific opportunity. The factors described above cannot be controlled independently and, in some cases, trade-offs must be made. For example, reducing competition by product differentiation will generally decrease overall market size. Increasing product usefulness may involve increasing development costs thus reducing market size. Based on a combination of these factors, products based on terminology research and system integration should have greater potential given the current market and state of technology.

Although none of the opportunities described above can be guaranteed to succeed, there are opportunities which present a favourable combination of risk and reward from the Canadian supplier standpoint.

V. STRATEGIC CONSIDERATIONS FOR THE GOVERNMENT

The key results presented in the previous sections are that:

* There are immediate commercial opportunities for selling software packages that integrate features now available in different systems and that have better translation research support capabilities than do currently available systems; and

* There is a potential commercial opportunity for selling automated translation systems and terminology banks required by such systems. Although the technical and economic viability of such systems has not yet been established, private sector suppliers in other countries have a lead of at least two years for such products.

The fundamental issue is whether the Canadian government should pro-actively assist the development of a Canadian industrial supplier base for Computer-Assisted Translation (CAT) systems, knowing that Canadian companies would have to compete with relatively large foreign companies in the Canadian marketplace as well as elsewhere since the Canadian market for CAT systems is too small to make it commercially viable.

On the positive side, it can be reasoned that the Canadian government should pro-actively pursue the development of a Canadian supplier base for CAT systems since:

* It would build on Canada's expertise in software and systems engineering, natural language processing, information technology, artificial intelligence and expert systems;

* It could result in avoiding Canadian translation service suppliers from having to procure CAT systems from foreign suppliers, and thus making a negative contribution, albeit relatively small, to Canada's balance of trade;

* The Canadian government itself is a significant user of translation services and hence a significant customer for CAT systems both directly for its in-house translation units and indirectly through the translation work that is contracted out; and

* It provides a unique opportunity for the Canadian government to foster an industrial capability in a niche that is also important to the government's own needs since it is a major user of translation services in Canada, and to do so in collaboration with provincial governments that have similar vested interests.

On the negative side, it can be reasoned that the Canadian government should not pro-actively pursue the development of a Canadian supplier base for CAT systems, but rather that Secretary of State continue to assess whatever CAT systems become available from either Canadian or foreign suppliers and procure those that will improve the productivity of its own translators. Reasons to support this course of action are:

* Foreign companies in the CAT systems business are relatively large compared to most Canadian software companies and have gained a sufficient lead, hence making it difficult for a new entrant to be successful even though their products still have numerous deficiencies;

* The market for CAT systems is small compared to general purpose products such as personal computers, word processing packages and spreadsheet packages; and

* The cost effectiveness of more ambitious CAT technology has yet to be proven.

While there are risks associated with each of these downside reasons, CAT systems provide a moderately attractive commercial opportunity. However, it will require that Canadian suppliers evolve and move quickly as well as catch up to and outsmart the competition.

The federal government will influence the establishment and success of Canadian companies in this area because it is the single largest buyer of translation services and CAT systems in Canada. To date, this purchasing need/power has not resulted in the development of Canadian suppliers of CAT systems, although this should be given consideration. However, in so doing, it will be important to recognize that the SOS Translation Bureau's requirements for CAT systems may be somewhat different than other suppliers of translation services. For example, our survey indicated that SOS translators make more use of dictaphones. Therefore, action will be required to ensure that systems developed to meet SOS's requirements are also attractive to (or can easily be modified for) other types of translation service suppliers in terms of features, compatibility with existing equipment and price. Not only does this maximize the size of the suppliers' potential market in Canada and elsewhere, it avoids influencing Canadian suppliers to focus primarily on major customers like the United Nations, the World Bank, the U.S. government, and NATO where competition with foreign suppliers of CAT systems is likely to be direct and fierce.

Another important consideration for the government in deciding how it uses its purchasing power is whether it requests Canadian suppliers to focus on more advanced systems such as automated translation, or less ambitious products based on system integration or improvements to terminological research tools or both. From an industrial and market development perspective, focusing only on advanced CAT systems could be an R&D success but a commercial failure: the likelihood of success is indeterminable; the development costs are likely to be high; the market is limited since a smaller number of translation service suppliers will be able to meet the demand for translation once such systems become available. The larger commercial opportunity once such advanced systems become available will be to supply the terminological database modules for different language pairs and subject matter. Therefore, R&D for the development of advanced CAT systems should only be supported if the government decides to assist in the development of a Canadian industrial supplier base for CAT systems that focus on the immediate needs and are already technically feasible. Besides increasing the probability of

early successes, this course of action will generate expertise in system integration, terminological databases, and text analysis which would be useful for later development automated translation systems.

The government should also be careful as to how it supports industry in developing such a supplier base. The Department of Communications is currently undertaking a computer-assisted translation research program that appears to include plans for developing specifications for a "translator workstation" and then having a private sector company develop a prototype based on those specifications. Aside from these specifications being based on SOS's requirements and hence possibly being different from those of other translation suppliers, having industry develop the specifications would enhance the likelihood of commercial success, particularly if the development is permitted to be oriented at a product that would be an attractive commercial venture and if the development involves at least some private sector investment. Otherwise, it is government defining a commercial opportunity for industry which could result in developing a supplier base that simply depends on the government for contracts.

If the resulting product definition is not useful to the Translation Bureau, then SOS should examine why its requirements are different and how it could continue to meet the government's needs for translation services and at the same time use its buying power to support the development of a Canadian CAT System supplier base. One approach would be for SOS to contract out more of its translation work. This approach is used by major corporations such as General Motors to reduce their own costs and improve the performance of their suppliers.

In summary, the federal government will have a significant influence on the establishment of a Canadian CAT system supplier base, and must be careful in how it exercises its buying power and industrial support.

VI. CONCLUSION

The overall conclusion of the study is that there are opportunities in the area of Computer-Assisted Translation (CAT) systems. There are immediate commercial opportunities for selling software packages that integrate features now available in different systems and that have better translation research support capabilities than do currently available systems. There is a potential commercial opportunity for selling automated translation systems and terminology banks required by such systems, although the technical and economic viability of such systems has not yet been established.

While the market for CAT systems is much more limited than for more general purpose products such as personal computers or spreadsheet packages, computer-assisted translation provides an attractive market niche for a small number of world-scale suppliers. However, while CAT-related development projects were undertaken in Canada during the late 1970s, Canada has no real CAT system supplier base. Furthermore, several companies in the U.S. and Europe have established a reasonably strong position in the CAT system business. Even though their products still have many shortcomings and have not been widely adopted by users, these companies have very recently introduced new products that are directed at the opportunities identified during this study. However, there are several factors that make CAT systems an attractive opportunity for Canada, namely:

* there is a need for CAT products that can increase the cost effectiveness of translation services;

* the currently available CAT products do not perform well;

* the current suppliers of CAT products have not penetrated the translation market to any great extent;

* there is a significant domestic market for translation services in Canada, primarily involving English and French. The international market involving these language is even larger;

* Canada has world class expertise, although on a fairly small scale, in research pertaining to the technologies underlying CAT; and

* There is a small but growing Canadian software industry which could exploit these opportunities.

There are several important strategic considerations concerning the pursuit of opportunities for CAT systems. From a commercial perspective, the main considerations pertain to target market, competition, product design, language pair and whether to concentrate on the shorter versus longer term opportunities, or both. It will be necessary to look at the global market for any particular CAT product since the Canadian market is not large enough to make it commercially viable.

The Canadian government also has some important strategic decisions to make. It will have a significant influence on if Canadian suppliers of CAT systems evolve since it is the single largest customer for translation services and CAT systems in Canada. Given this situation, it would be appropriate for the government to use its buying power to help establish a Canadian CAT system supplier basis, provided it is done in such a way that it is driven by market needs and involves private sector investment. However, if the government wishes to use its buying power for this reason, it will need to act quickly since foreign companies are already past the starting-line. It also implies action-oriented coordination amongst government departments that have an interest in this endeavor, particularly the Department of Communications and Secretary of State, and perhaps the new Department of Industry, Science and Technology.

THE COOPERS & LYBRAND CONSULTING GROUP

George A. Neufeld

APPENDIX I

MAIN REFERENCES

MAIN REFERENCES

1) State of the Art in Machine Translation: Proceedings of the IAI-MT86 Conference, T. C. Gerhardt, Ed. August 1986.

2) The Meteo System. B. Thouin; Practical Experience of Machine Translation: V. Lawson, Ed., North-Holland Publishing Company, Amsterdam, 1982.

3) Better Translation for Better Communications: G. Van Slype. Commission of the European Communities, Pergamon Press, (1984).

4) Translation Practices: D. Smith and D. Tyldesley. Digital Equipment Corporation, 1986.

5) Natural Language Processing and Artificial Intelligence:
Volume 1 - State-of-the-Art
Volume 2 - Technology Thrust
Volume 3 - Possible Social and Economic Impacts
Volume 4 - Opportunities for Canada
Volume 5 - A Program Plan
Volume 6 - Priorities for Canada
By: Cognos Inc. and sub-contractors.
For: Secretary of State and Department of Communications.

6) Machine Translation and Natural Language Processing: Opportunities For Artificial Intelligence in Canada,
By: Cognos Incorporated, Canadian Artificial Intelligence Products Corporation.
For: Department of Communications.
Date: Fecruary, 1985.

7) Functional Specifications for a Translator's Work Station:
By: Socioscope, Inc.
For: The Canadian Workplace Automation Research Centre.
Date: March 30, 1986.

8) Survey of the Official Languages Translation Market in Canada:
By: Bureau of Management Consulting, DSS.
For: Translation Bureau, SOS.
Date: 1983.

9) Revised Statutes of Canada, Chapter O - 2.

10) Sublanguage - Specific Computer Aids to Translation: A Survey of the Most Promising Application Areas:
By: R. Kittredge, University of Montreal.
For: Secretary of State.
Date: March, 1983.

11) Machine Translation Poised for Growth: C. Connell. High Technology Magazine, June 1986.

12) Personal Workstations Redefine Desktop Computing: J. Bairstow. High Technology Magazine, March 1987.

13) Evaluation of the Performance and Costs of the Systran System: J. Chaumier and G. Van Slype. Commission of the European Communities, Pergammon, 1977.

14) Trial of the Weidner Computer-Assisted Translation System:
By: Bureau of Management Consulting.
For: Secretary of State.
Date: October, 1985.

APPENDIX II

INSTRUMENTS USED FOR AND RESULTS FROM THE FOCUS GROUP DISCUSSIONS

APPENDIX II

INSTRUMENTS USED FOR AND RESULTS FROM THE FOCUS GROUP DISCUSSIONS

CONTENTS

I. MODERATORS' GUIDE FOR FOCUS GROUP DISCUSSIONS

A. INTRODUCTION

1. Warm-up Period

2. Group Leader Introduction

3. Equipment:
   * Audio recorder
   * One-way mirror .

4. Nature of Focus Group Discussion:
   * Informal
   * No right or wrong answers
   * Open/honest opinions
   * Role of moderator
   * Use of questionnaire

5. General Purpose of discussion:
   * Talk about translation methods and issues
   * Talk about machine assisted translation

6. Participant Introduction

7. Questions Before Sessions Begin

B. CURRENT TRANSLATION METHODS

1. How is translation handled now?
   * Type of end users and material
   * Organization and methods
   * Workload and structure

2. What factors are critical to improved methods?
   * Potentiel improvements

C. PERCEPTIONS OF INCREASED AUTOMATION

1. Current levels of automation:
   * How is it used?
   * What trends exist?

2. Attitudes towards increasing automation:
   * Is it required?
   * What barriers exist?

D. DISCUSSION OF FEATURES

1. Turn to; Automated "look up" in On-Line Dictionaries and Databases:
   * Review ratings
   * Rationale

2. Turn to; Computer Assisted Translation

3. Turn to; Features Related to Input/Output

4. Turn to; Other Word Processing and Administrative Functions

E. COMPUTER ASSISTED TRANSLATION SERVICE OPTIONS

1. Turn to F; A Translation Software Package for Existing Computers:
   * Review ratings
   * Rationale

2. Turn to G; A Translator's Workstation

3. Turn to H; Networking Software and Hardware

4. Turn to I; Automated Translation

F. OTHER OPTIONS

* Discuss possible improvements

F. CONCLUSION

* Thank participants and turn to last section of questionnaire.

QUESTIONNAIRE THAT FOCUS GROUP PARTICIPANTS WERE ASKED TO COMPLETE

Please tell us how useful you would consider the following computer assisted translation features to be to you personally or to the translators in your shop. The more useful you think a feature would be, the higher the number you should circle; the less useful you think a feature would be, the lower the number you should circle. Please circle one number for each feature.

A. Automated "Look-Up" in On-Line Dictionaries and Terminological Databases

Feature	Not at all Useful									Extremely Useful
1. Basic Word Processing	1	2	3	4	5	6	7	8	9	10

The translator simply types the translated version on a computer or word processor and checks difficult words, manually, in a dictionary if required.

2. On-Line Dictionary

While working on a computer, a translator would access a computerized bilingual dictionary. The translator would just highlight the word and suggested translations would show up in the corner of the screen. New words and their translations could be added as required to the dictionary.

3. Word Pairs

The ability to insert the preferred translated version of a word by typing in the original word and simply hitting a special "translation" key.

1 2 3 4 5 6 7 8 9 10

4. Historical Text Access

The ability to quickly search previously translated text using key words. Portions of the original or translated text could be imported to the current translation job. Historical databases would be updated and maintained by the translator.

B. Computer Assisted Translation

Again, please rate how useful you think the following features would be, on a scale from 1 to 10. Please circle one number for each feature.

Feature	Not at all Useful									Extremely Useful
1. List all Word Pairs	1	2	3	4	5	6	7	8	9	10

Prior to starting the translation of a target sentence or paragraph the computer could list all the target words in its dictionary along with a suggested translation. It would also list words it can't translate.

2. Rough Translation - Full Text

The computer provides a rough draft translation of source text. The user would then revise the translated text, sentence by sentence.

3. Rough Translation - Portions of the Text

The translator would identify specific sentences or paragraphs for a rough draft translation by the computer.

4. Building Terminological Database

The translator would be able to "train" the computer to use specific terminology when it does a rough draft translation of specific words or phrases.

C. Features Related to Input/Output

Please rate how useful the following input/output features would be to you or to the translators in your shop, on a scale of 1 to 10.

Feature	Not at all Useful									Extremely Useful
	1	2	3	4	5	6	7	8	9	10
1. Document Transfer Document transfer via phone or data networks from a customer or supplier to your computer systems. This would include use of standard protocols.										
2. Compatable Diskettes Document transfer via reading from and writing to diskettes which would be compatible with those used by suppliers and customers.										
3. Text File Transfer Transfer of text from one computer to another by either remote communications or by diskettes. This would be accomplished using standard protocols under development by international standards agencies.										
4. Annotated Drawings Transfer of annotated drawings to/from remote computer systems using either remote communications or diskettes and conforming to standard protocols.										

5. Machine Readable Text

Input of source text from hardcopy documents using an optical character recognition device.

6. Machine Readable Drawings

Input of drawings form hardcopy using an optical encoding device.

7. Typesetting

Output of target language documents, including both text and drawings, in a format suitable for use by computer typesetting systems.

8. Annotated Speech

Dictaphone speech storage, and input/output facilities. Speech files could be annotated at arbitrary points with text files and vice-versa. The computer would not have the ability to "understand" the speech files.

E. Other Word Processing And Administrative Functions

Please rate how useful the following word processing and administrative functions would be to you or the tranlators in your shop, on a scale of 1 to 10.

Feature	Not at all Useful									Extremely Useful
	1	2	3	4	5	6	7	8	9	10
1. Standard Word Processing										
Standard word processing features in both the source and target languages including search, replace, insert, delete, cut and paste, and formatting operations.										
2. Split Screen										
The use of adjustable, movable windows to view several files or applications on the screen at once. These would include source and target language test files as well as productivity tools which could be accessed without leaving the text processing application. Specific utilities are discussed as separate features.										
3. Administrative Information										
Database of associated administrative information residing on the user's work station. The database contents would include word counts, customer information, references to dictionaries, terminology database, text databases, and work scheduling information.										

4. Shared Working Environment 1 2 3 4 5 6 7 8 9 10

A shared working environment whereby translator's at two work stations could transfer information interactively. Both screens would have a shared visual space or "electronic blackboard" allowing information sharing during the course of translation work.

F. Translation Software Package For Existing Computers

Please read the following product description and answer the answers which follow.

A Software Package for Personal Computers providing access to terminological data while using a word processing package in the target language. The package would be sold as an add-on to, or a specialized upgraded version of, a successful high performance word processing product.

The source language words, target language words and associated terminological information would be linked by relationships based on common usage. The database area of previously translated text could be searched by key words directed at document titles and/or the body of the text.

The user could import single words or text portions from any of the databases into the target language text being worked on using special function keys. It would be possible to display both text processing and database information on the screen at once using movable adjustable windows.

The product would be sold with a basic amount of information in the database which the user could modify or add to as required.

The software package would be able to transfer documents via data communications using internationally accepted protocols for the source and target languages.

Please answer the following:

1. On a scale from 1 to 10 how useful would this product be to you or to the translators in your shop.

Not at all Useful									Extremely Useful
1	2	3	4	5	6	7	8	9	10

2. What is the maximum, per work station, you would be willing to pay for a product like this?

$ _____ one time purchase, or

$ _____ per month

G. Translator's Workstation

Please read the following product description and answer the questions which follow.

This personal work station product would be based on a combination of specialized hardware and software. It would posses all of the features of the previously described software package but could handle larger databases and more complex searching tools without a decrease in response time.

Hardware features would include an 8 1/2" x 11" screen; support for the full character set of both source and target languages on the screen, keyboard, and printer. It would also include interfaces for the following features:

1) Dictaphone: allowing storage and handling of text - annotated speech.

2) Optical Character Recognition: for input of hardcopy text.

3) Capabilities for input, output, and manipulation of text and drawings.

4) Capabilities to read/write to diskettes using standard PC and WP formats.

Please answer the following:

1. On a scale from 1 to 10 how useful would this product be to you or to the translators in you shop.

Not at all Useful									Extremely Useful
1	2	3	4	5	6	7	8	9	10

2. What is the maximum per work station you would be willing to pay for a product like this?

$ _____ one time purchase, or

$ _____ per month

H. Networking Software and Hardware

Please read the following product description and answer the questions which follow.

This would allow users of the previously described work stations to connect their machines together allowing them to:

1) Share specialized hardware, e.g., laser printer, OCR, etc.

2) Share terminological information.

3) Send and receive messages, files, and portions of text without disrupting translation work in progress.

Please answer the following:

1. On a scale from 1 to 10 how useful would this product be to you, or to the translators in your shop.

Not at all Useful									Extremely Useful
1	2	3	4	5	6	7	8	9	10

2. What is the maximum per work station you would be willing to pay for a product like this?

$ _____ one time purchase, or

$ _____ per month

I. AUTOMATED COMPUTER TRANSLATION

Please read the following product description and answer the questions which follow.

This would provide sophisticated features such as automated translation and large scale databases. These features would require minicomputer level hardware and a staff to maintain the hardware, software and terminological databases. The facility could be provided by a service bureau which would be accessed by data communications or an in-house facility if the user organization's workload could justify it. The user could access this central facility by terminals, PC's or translator's work stations.

Please answer the following:

1. On a scale from 1 to 10 how useful would this product be to you or to the translators in you shop.

Not at all Useful									Extremely Useful
1	2	3	4	5	6	7	8	9	10

2. What is the maximum per work station you would be willing to pay for a product like this?

$ _____ one time purchase, or

$ _____ per month

SUMMARY OF RESPONSES TO FOCUS GROUP QUESTIONNAIRE

	Respondent													
Feature	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**
A1	9	10	9	5	4	10	6	1	10	10	5	7	5	10
2	8	10	9	8	8	7	6	1	10	10	7	10	10	10
3	4	10	9	3	8	8	6	1	1	10	7	4	9	7
4	6	10	9	7	9	8	8	10	10	10	8	10	6	6
B1	5	4	6	2	4	8	5	7	1	8	6	2	5	4
2	3	2	7	4	8	5	3	2	8	5	4	2	1	2
3	5	7	9	1	2	1	3	8	8	5	4	2	1	3
4	8	4	9	5	9	8	3	NA	10	8	5	10	1	5
C1	9	10	9	5	10	3	8	10	10	8	9	10	10	10
2	8	7	9	6	10	4	6	10	6	8	9	10	10	8
3	9	10	9	6	10	8	8	10	10	8	9	10	10	9
4	4	10	9	9	9	3	1	7	10	3	9	4	10	9
5	9	4	9	9	10	6	6	3	5	5	8	2	10	9
6	9	10	9	9	9	5	1	8	1	2	8	2	6	9
7	9	10	8	8	6	4	6	9	10	5	9	10	8	9
8	7	4	8	8	9	5	1	3	1	?	9	5	5	9
E1	10	10	9	9	4	8	10	10	10	6	10	10	10	10
2	9	5	7	7	6	7	8	8	10	8	10	10	10	10
3	9	5	9	8	7	1	6	10	10	8	10	8	10	10
4	5	10	9	9	7	1	2	5	10	6	10	3	7	7
Scenario														
"F"	8	10	9	6	9	6	8	10	1	9	8	7	10	10
"G"	10	8	10	9	10	8	9	10	1	5	9	5	10	10
"H"	10	10	10	9	10	7	10	6	1	5	10	9	9	10
"I"	10	10	10	4	8	2	6	8	1	10	10	10	5	5

APPENDIX III

SURVEY DESIGN AND QUESTIONNAIRE

SURVEY DESIGN AND QUESTIONNAIRE

I. SURVEY DESIGN

The telephone survey was designed to obtain information about the following aspects of the translation industry in Canada relevant to the potential market for CAT systems:

a) The Market Size for CAT Systems:

- total annual volumes of words translated,
- fraction of work involving different language pairs and subject matter,
- potential for increased business by improving services, and
- relative amounts of business coming from different public and private sectors.

b) Economic Feasibility of CAT Systems:

- size of translation organizations,
- annual volumes of words translated per organization, and
- level of investment in automation.

c) Operational Feasibility of CAT Systems:

- nature of the process involved in translation,
- equipment used in translation, and
- level of computer experience among translation personnel.

d) Potential for CAT to Benefit Translation Service Suppliers:

- areas of translation work which are costly,
- causes of quality problems in translation work,

- aspects of translation work which are time-consuming,
- problems encountered exchanging translation work with clients, and
- importance to clients and potential to increase business through improvements in service.

To focus the design of the survey questionnaire, several research issues were formulated based on the focus group discussions, interviews with CAT suppliers, and a literature review. They are listed below.

The following four categories of respondents were surveyed:

a) Government Translation Departments

These respondents were selected from lists supplied by the Translation Bureau of SOS, and the membership directories of the translators' associations of Quebec, Ontario and New Brunswick. Federal, provincial and municipal government translators were included.

b) Translation Departments in Private Sector Organizations

These respondents were selected from the membership directories of the previously mentioned translators' associations and from a list prepared for a previous SOS survey (7). The organizations covered included private sector industrial and commercial firms, crown corporations, and non-profit organizations.

c) Translation Firms

These respondents were selected from the yellow pages the phone books of major Canadian cities.

d) Freelance Translators

These respondents were chosen from the previously mentioned membership directories and list used in a previous SOS survey.

In all cases, sampling within each respondent category was done randomly. Therefore, each sample should be considered representative of its category. Because the degree of sampling varied from one category to another, the total sample cannot be considered truly representative of the total translator population. The table on the following page provides background and sampling information on each respondent category.

II. COMPUTER ASSISTED TRANSLATION SURVEY QUESTIONNAIRE

ASK TO SPEAK TO PERSON NAMED ON LIST OR TO THE MANAGER OR PERSON IN CHARGE OF TRANSLATION SERVICES. ONCE YOU HAVE THE RIGHT PERSON: Hello, my name is _____ from The Coopers & Lybrand Consulting Group. We're conducting a survey for the federal Department of Communications and Secretary of State on the office automation needs of translators in Canada. Your organization has been randomly selected to participate in this study. All responses will be strictly confidential.

1. First, could you tell me which of the following best describes your translation services? Are you .... **READ LIST. CIRCLE ONLY ONE NUMBER.**

An independent translator	1
A translation department within a private company	2
A translation group within a government department	3
A private translation firm	4

INTERVIEWER: READ THE PARAGRAPH ASSOCIATED WITH THE NUMBER CIRCLED ABOVE.

1. From now on, when I say your organization, I mean you, as an independent translator. **SKIP TO QUESTION 4**

2. From now on, when I say your organization, I mean your translation department.

3. From now on, when I say your organization, I mean your immediate translation group within your government department.

4. From now on, when I refer to your organization, I mean your translation firm.

SECTION A: ORGANIZATIONAL PROFILE

2. How many people, including free-lancers, are currently working in your organization ... **READ LIST.**

	Number	Don't Know
In total	_____	88
Doing revision only	_____	88
Doing translation only	_____	88
Doing both revision and translation	_____	88
As word processors or typists	_____	88
Are actually working in your building rather than off-site	_____	88

3. About what percent of the translation jobs are ... **READ WHOLE LIST. REPEAT ONE AT A TIME. MAKE SURE TOTAL EQUALS 100%.**

	Percentage	Don't Know
Handled by one translator	_____	888
Handled by several translators	_____	888
TOTAL	100%	

4. Approximately, how many words per year does your organization translate?

_____	thousand words, or
_____	million words
Don't know	888888

5. What percentage of this translation work is ... **READ WHOLE LIST. REPEAT ONE AT A TIME. MAKE SURE TOTAL EQUALS 100%.**

	Percentage	Don't Know
English to French	_____	888
French to English	_____	888
Other	_____	888
TOTAL	100%	

6. What percent of your organization's translation work involves the following subject matter? **READ WHOLE LIST. REPEAT ON AT A TIME. MAKE SURE TOTAL EQUALS 100%.**

	Percentage	Don't Know
Legal	_____	888
Financial	_____	888
Advertising or Sales	_____	888
Scientific	_____	888
Engineering abd other technical material	_____	888
General administrative	_____	888
Others	_____	888
TOTAL	100%	

7. What percent of your translation work comes from clients in ... **READ WHOLE LIST. REPORT ONE AT A TIME. MAKE SURE TOTAL EQUALS 100%.**

	Percentage	Don't Know
Your building	_____	888
Elsewhere in your city	_____	888
Outside your city	_____	888
TOTAL	100%	

8. What percent of your translation work-load comes from the following types of organizations? **READ WHOLE LIST. REPEAT ONE AT A TIME. MAKE SURE TOTAL EQUALS 100%.**

	Percentage	Don't Know
The Federal Government	_____	888
Provincial Government	_____	888
Crown Corporations	_____	888
The Private Sector	_____	888
Other (specify)		
_____	_____	888
TOTAL	100%	

9. **ASK ONLY IF RESPONDENT DID NOT SAY "0%" TO THE PRIVATE SECTOR IN QUESTION 8 ABOVE.** What are the major industry sectors that your organization deals with? **PROBE.** Any others?

_____

_____

_____

10. How important are the following factors to the people you do your translations for? Is ... **READ ITEM, THEN SCALE. REPEAT FOR EACH ITEM.**

	Extremely Important	Somewhat Important	Not Very Important	Not At All Important	DO NOT READ Don't Know	DO NOT READ Not Applicable
Cost	4	3	2	1	8	9
Turnaround time	4	3	2	1	8	9
Quality	4	3	2	1	8	9
Compatibility of your word processors with those of the people you translate for	4	3	2	1	8	9
File transfer over phone lines	4	3	2	1	8	9

SECTION B: TRANSLATION METHOD

11. In your organization, approximately what percent of the initial translation work is ... **READ WHOLE LIST. REPEAT ONE AT A TIME. MAKE SURE TOTAL EQUALS 100%.**

	Percentage	Don't Know
Dictated	____	888
Handwritten	____	888
Simultaneously entered at the word processor or typewriter	____	888
TOTAL	100%	

12. Approximately what percent of your organization's productive time is spent ... **READ ITEMS ONE AT A TIME. TOTAL DOES NOT NEED TO EQUAL 100%.**

	Percentage	Don't Know
Consulting dictionaries, other terminological references, or examples of previous translations	____	888
Consulting dictionaries or previous translations located outside your building	____	888
Consulting other persons outside your organization	____	888
Maintaining dictionaries or other information sources used to help with translation terminology or translation subject matter	____	888

13. How frequently do people in your organization consult others about a particular translation job? Would you say ... **READ SCALE.**

More than 5 times a day	1
1 - 5 times a day	2
Less than once a day	3
Don't know	8

14. Which activities in your translation operation would you say are very costly in terms of the amount of time required to perform them? **PROBE.** Any others?

_____

_____

_____

15. What would you say are the major causes of quality problems in a typical translation project? **PROBE.** Any others?

_____

_____

_____

SECTION C: EQUIPMENT AND SYSTEMS USED

16. Do you use any of the following types of equipment or systems during translation work or document production? **FOR EACH ONE CIRCLED YES, ASK:** How many?

	Yes	No	Don't Know	Number	Don't Know
IBM compatible personal computers	1	2	8	___	8
Other types of personal computers	1	2	8	___	8
Dedicated word processors	1	2	8	___	8
Laser printers	1	2	8	___	8
Other types of printers	1	2	8	___	8
Dictaphones	1	2	8	___	8
Local area networks	1	2	8	___	8
Mini or mainframe computers	1	2	8	___	8
Desk publishing systems	1	2	8	___	8
Terminology databank	1	2	8	___	8
Modems	1	2	8	___	8
Facsimile	1	2	8	___	8
Word counting machine	1	2	8	___	8
Software packages to assist translation (if yes, specify which ones) _____	1	2	8	N.A.	N.A.
Software packages developed in-house (if yes, specify what for) _____	1	2	8	N.A.	N.A.

17. Do you find any of the products just discussed especially difficult or stressful to use?

Yes	1	
No	2	**SKIP TO QUESTION 19**
Don't know	8	

18. Please comment on the specific products and the type of problems you've experienced? **PROBE.** Any others?

_____

_____

_____

19. Approximately how much has your organization spent on automation of your translation services during the past year?

$_____

Don't know 888888

20. Could you briefly describe your plans for automation over the next year? **PROBE.** Anything else?

_____

_____

_____

SECTION D: I/O FORMAT

21. As you know, translators and their clients often exchange translation work as printed pages, that is hard copy. Alternatively, the text could be delivered on a diskette if the translator and client are using compatible Personal Computers or word processors. It is also possible to send text over the telephone lines, i.e., if both parties have the proper equipment. Does your organization ... **READ LIST. CIRCLE ONE NUMBER ON EACH LINE.**

	Yes	No	Don't Know
<u>Receive translation work from your customers as:</u>			
Hardcopy only	1	2	8
Diskette with or without hardcopy	1	2	8
Modem	1	2	8
Facsimile	1	2	8
<u>Deliver translated work to your customers as:</u>			
Hardcopy only	1	2	8
Diskette with or without hardcopy	1	2	8
Modem	1	2	8
Facsimile	1	2	8

22. How difficult is it to influence the means used by your clients to exchange translation work with you? Would you say it is ... **READ SCALE.**

Very difficult	4
Somewhat difficult	3
Not very difficult	2
Not difficult at all	1
Don't know	8

SKIP TO QUESTION 24

23. Could you comment on the major problems you experience with the different methods of exchanging translation work with your clients? **PROBE.** Any other problems?

_____

_____

_____

SECTION E: OPPORTUNITIES TO INCREASE BUSINESS

24. How likely is it that **(READ ITEM)** would result in an increase of 10% or more in the demand for your translation services? Would you say it is ... **READ SCALE. REPEAT FOR EACH ITEM.**

	Very Likely	Somewhat Likely	Not Very Likely	Not At All Likely	DO NOT READ Don't Know	DO NOT READ Not Applicable
* an increase in quality without any corresponding increase in time or costs	4	3	2	1	8	9
* having word processors or computers that could read all your customers' diskettes	4	3	2	1	8	9
* having word processors or computers which could communicate with all your customers' computers or word processors	4	3	2	1	8	9
* having the ability to produce a rough translation useful for a preliminary assessment of the subject matter only - at half the cost and turnaround time of normal translation work	4	3	2	1	8	9
* having the ability to handle a larger number of specialized areas such as automotive, medical and other translations	4	3	2	1	8	9
* having ability to handle more languages than you currently work in	4	3	2	1	8	9
* a 10% to 20% decrease in translation costs	4	3	2	1	8	9
* a 10% to 20% reduction in your turnaround time	4	3	2	1	8	9

25. Based on the requests you receive and your discussions with customers, please comment on any other capabilities that could increase demand or the efficiency of your operation? **PROBE.** Any other capabilities?

_____

_____

_____

26. There are some documents which have a simpler format or sentence structure than normal prose. Does your organization handle a lot of this type of work?

Yes	1	
No	2	**SKIP TO SECTION F**
Don't know	8	

27. What type of material is involved? **PROBE.** Any other type?

_____

_____

_____

SECTION F: BACKGROUND INFORMATION

Finally, the following questions are for classification purposes only.

28. How would you describe your level of experience with computers and office automation? Would you say you have ... **READ SCALE.**

Extensive experience	4
Some experience	3
Very little experience	2
No experience	1

29. **DO NOT ASK IF INDEPENDENT TRANSLATOR AT QUESTION 1 ON 1ST PAGE.** And, excluding yourself, how many of your translation and revision staff have ... **READ EACH ITEM.**

	Number	Don't Know
Extensive experience with computers and automation	____	88
Some experience with computers and automation	____	88
Very little experience with computers and automation	____	88
No experience with computers and automation	____	88

30. How many years of experience do you have in translation? **READ LIST.**

Less than 2 years	1
2 years or more but less than 5 years	2
5 years or more but less than 10 years	3
10 years or more but less than 20 years	4
20 years or more	5

31. **DO NOT ASK IF INDEPENDENT TRANSLATOR AT QUESTION 1 ON 1ST PAGE.** And, excluding yourself, how many of your translation and revision staff have ... **READ EACH ITEM.**

	Number	Don't Know
Less than 2 years of experience in translation	_____	88
2 years or more but less than 5 years of experience in translation	_____	88
5 years or more but less than 10 years of experience in translation	_____	88
10 years or more but less than 20 years of experience in translation	_____	88
20 years or more of experience in translation	_____	88

THANK AND TERMINATE

SUPPLEMENTARY TABLES

TABLE A:
SAMPLING DETAILS FOR TELEPHONE SURVEY

Respondent Type	Number of Respondents	Volume Translated (millions of words/year)	Estimated Population Size *	% of Total
Independent Translators	36	9.4	1020	4%
Translation Departments within a Private Company	45	35.3	260	17%
Translation Groups within government	24	70.5	110**	27%
Private Translation Firms	37	59.0	380	10%
Total	152	174.2		

* These are the populations of translation organizations represented by the respondents, not individual translators.

** Accounts for translators in the Federal Government (SOS) and the provincial governments of New Brunswick, Quebec, Ontario and Manitoba. Individual sections within translation bureaus are considered as separate translation units.

TABLE B

ANNUAL VOLUMES TRANSLATED FOR DIFFERENT TYPES OF CLIENT ORGANIZATION (WORDS PER YEAR)

Respondent Type / Type of Client	Independent Translator	Translation Department Within Private Company	Translation Group Within Government Departments	Private Translation Firm	All Respondents
Federal Government	1,698,300 18%	180,000 1%	58,158,150 82%	8,948,500 15%	68,976,950 40%
Provincial Government	371,250 4%	1,147,500 3%	10,725,600 15%	8,261,700 14%	20,506,050 12%
Crown Corporations	1,142,500 12%	7,245,000 21%	1,300,000 2%	4,616,500 8%	14,484,000 8%
Private Sector	5,593,950 60%	24,953,500 71%	351,200 0%	34,096,800 58%	64,995,450 37%
Other Types	575,000 6%	1,550,000 4%	1,050 0%	3,122,500 5%	5,248,550 3%
Total Volume per Respondent Category	9,381,000 100%	35,246,000 100%	70,528,000 100%	59,046,000 100%	174,211,000 100%
Number of Respondents	26	29	22	35	112

TABLE C

BREAKDOWN OF MATERIAL TRANSLATED BY SUBJECT TYPE

Subject Type	Percent of total volume		
	CEC Survey %	DEC Survey %	C & L Survey %
Commercial	35.4		
Industrial	21.0		
Financial			17
Marketing/Sales		23	7
Technical/Scientific	20.0	32	25
Legal		21	14
Press/Current Affairs	3.5		
Legislative	9.3		
Government		8	
General/Admin.			29
Educational	1.5	6	
Literature	0.3	4	
Other	9.0	6	8%
Total	100.0	100	100

CEC - Commission of the European Communities

DEC - Digital Equipment Corporation

TABLE D

BREAKDOWN OF MATERIAL TRANSLATED BY SUBJECT TYPE BASED ON C&L SURVEY

Subject Type	Volume and percent of total volume by respondent type				All Respondents
	Independent Translator	Translation Department Within Company	Translation Group Within Government Department	Private Translation Firm	
Legal	3,562,000 38%	4,769,100 14%	8,653,980 13%	7,430,100 13%	24,415,780 14%
Financial	859,000 9%	8,847,600 24%	12,007,480 18%	8,135,600 14%	29,489,680 17%
Advertising/Sales	341,000 4%	4,084,600 12%	970,480 1%	6,018,100 10%	11,414,180 7%
Scientific	1,013,770 11%	2,439,000 7%	9,085,480 14%	4,486,500 8%	17,024,750 10%
Engineering/Technical	597,750 6%	5,932,000 17%	5,225,100 8%	14,430,000 24%	26,184,850 15%
General/Administrative	2,115,050 23%	8,411,200 24%	243,060,480 36%	14,280,700 24%	48,857,430 29%
Other Types	891,830 10%	1,132,500 3%	7,025,000 10%	4,265,000 7%	13,314,330 8%
Total Volume per Respondent Category	9,381,000 100%	35,256,000 100%	67,028,000 100%	59,046,000 100%	170,701,000 100%
Number of Respondents	26	29	21	35	111

TABLE E

SHARE OF WORLD TRANSLATION MARKET HELD BY ENGLISH-FRENCH LANGUAGE PAIR

Source-target Language Pair	Percent of Total Volume of Material Translated		Weighted-Average
	CEC Survey	DEC Survey	
English to French	8.5	8.0	8
French to English	8.8	15.0	11
Number of Respondents	339	240	

TABLE F

AMOUNT SPENT AUTOMATING TRANSLATION SERVICES OVER THE PAST YEAR

	Type of Translation Service				All * Respondents
	Independent Translator	Translation Department Within Company	Translation Group Within Government Department	Private Translation Firm	
Amount spent on automating translation services in past year	$10,616	$20,400	N/A	$21,032	$17,647
Number respondents	31	30	N/A	40	101

* Government translation departments are not included in this average since SoS budgets are centralized.

TABLE G

PLANS FOR AUTOMATION OVER NEXT YEAR

Equipment Cited in Automation Plans	% Respondents Cited 1st	Cited 2nd	Cited 3rd
No Plans/No Change	38.5	0	0
Word Processing	8.4	15.0	0
System Printer	2.1	22.5	22.2
Modem	6.3	7.5	-
Personal Computer (s)	5.6	10.0	0
Facsimile Machine or Telex	4.9	2.5	11.1
Desktop Publishing System	4.2	10.0	0
Specialized Translation System	6.3	10.0	44.4
Other Computer Equipment	21.7	22.5	22.2
Other Plans	2.1	0	0
	100.0	100.0	100.0

TABLE H

MAJOR PROBLEMS EXPERIENCED EXCHANGING WORK WITH CLIENTS

	Type of Translation Service								Total	
	Independent Translator		Translation Department Within Company		Translation Group Within Government Department		Private Translation Firm			
	#	%	#	%	#	%	#	%	#	%
Major problems experienced with exchange with clients ...										
lack of system compatibility	3	23	6	67	5	42	8	44	22	42
difficult to influence client methods	3	23	0	0	1	8	3	17	7	13
other problems	3	23	2	22	4	33	4	22	13	25
no major problems	4	30	1	11	2	17	3	17	10	19
Total times cited	13	100	9	100	12	100	18	16	52	100
Total # respondents	12		8		10		16		46	

= Number of times cited

% = Percent of total times cited

TABLE I

RELATIVE IMPORTANCE OF FACTORS AFFECTING TRANSLATION SERVICE

Factor	% Respondents * Citing Factor as Somewhat/Extremely Important				All Respondents
	Independent Translators	Translation Department Within Companies	Translation Groups Within Government Departments	Private Translation Firms	
Cost	* 80	74	34	94	75
Turnaround Time	100	96	100	98	98
Quality	100	100	100	100	100
Word Processor Compatibility	44	63	76	65?	61

TABLE J

TYPE OF MATERIAL INVOLVING SIMPLER FORMAT

	Type of Translation Service								Total	
	Independent Translator		Translation Department Within Company		Translation Group Within Government Department		Private Translation Firm			
	#	%	#	%	#	%	#	%	#	%
Type of material involving simper format ...										
legal documents	6	43	3	17	2	40	4	18	15	25
promotional material/ pamphlets	3	21	0	0	0	0	5	23	8	14
notes, memos, correspondence	0	0	3	17	1	20	2	9	6	10
reports, articles, commentaires	1	7	5	28	0	0	2	9	8	14
other material	4	29	7	39	2	40	9	41	22	37
Total time cited	14	100	18	100	5	100	22	100	59	100
Total # respondents	9		13		4		14		40	

= Number of times cited

% = Percent of total times cited

APPENDIX V

ASSESSMENT OF CURRENTLY AVAILABLE CAT SYSTEMS

ASSESSMENT OF CURRENTLY AVAILABLE CAT SYSTEMS

GENERAL PURPOSE PRODUCTS

a. Foreign Language Kits

Example: L'Accent

Description

The language kit's function is to override the standard character set (7-bit ASCII, 8-bit ASCII, EBCDIC, etc.) normally associated with a specific computer system in order to include the complete alphabet of a non-English language.

A complete foreign language kit must include:

- the internal representation (code) of all characters needed;
- a character generator for CRT terminals;
- a keyboard enhancer or keyboard redefinition kit;
- a driver for the printers to be used with the system; and
- the definition of a collating sequence compatible with that of the language(s) used.

Normally, such a language kit should be memory-resident to facilitate access.

Price Range: between $50 and $300 on PC's, much more on large computers.

Present Situation

All language kits currently available contain many flaws and restrictions: unsatisfactory collating sequence, work on a limited number of printers,

tradeoffs in keyboard and pixel layouts due to hardware limitations, limited to one or very few foreign languages, etc. There is a real opportunity for one system that would comprehensively address the problem once for all.

b. Word Processing

Example: Word Perfect

Description

In addition to the basic editing functions, most significant word processing packages include sophisticated page format definitions, including formatting on two or more columns, auxiliary functions such as spell checking, word counting, etc.

Price Range: $100 to $800 on PC's

Present Situation

Word processing is now a mature application, except for some functions like spell checking in languages other that English. Several products have established market positions and will continue fighting for a market share. For the time being, Word Perfect is the market leader, but other products like WordStar, Microsoft Word, Palantir, XYWrite, Multimate, Display Writer, etc. have a visible market share.

This situation creates a challenge to the designers of other products to achieve compatibility with as many as possible of the major word processing packages.

The development of Natural Language Processing (NLP) techniques should open the door to significant enhancements like syntax sensitive Search and Replace, grammatically intelligent spell checking, etc.

c. Author Systems

Examples: More, MacProof

Description

Author systems usually include two kinds of functions: functions to help the writer of an original text to organize his/her thoughts and to structure the text being written accordingly; auxiliary functions such as grammar and style checking (looking for repetitions, suggesting synonyms, spotting ambiguous, racist or sexist terms, etc.).

Price Range: $150 to $800 on PC's

Present Situation

The second category of author systems functions could be useful to translators, but are very little used at the present time. There is certainly an opportunity for a system that would properly integrate author system functions into a word processing and terminology handling environment.

d. Diagramming Systems

Example: MacDraw

Description

Diagramming software packages combine line and surface drawing functions with simple word processing capabilities allowing for addition of a legend and comments. To be effective, they require a powerful CPU with enough main memory, a high-resolution screen and a pointing device such as a mouse.

Diagramming packages are helpful when translating diagrams of a source document, provided they are available in the proper format. The translator simply has to replace the source language indications, legend and comments with the corresponding ones in the target language, and possibly to adjust the size and position of some diagram component according to space requirements of the target language.

Price Range: $150 to $500 on PC's, much more on larger computers

Present Situation

There is currently no standard for diagramming packages and there are fewer conversion facilities than between word processing packages. This means that the translator must use the same package as the one used to produce the original diagrams. Moreover, full integration with other software including word processing is currently not available.

e. Desk-top Publishing

Example: Ventura

Description

Desk-top publishing is essentially an extension of word processing: use of a laser printer giving higher-quality output and a choice of several fonts, more powerful software and high-resolution screen for a true WYSIWYG (What You See Is What You Get) on some systems, page formatting or page making functions and better integration of graphics.

Those functionalities upgrade word processing to "document processing" but there is no fundamental change from the former.

Price Range: $500 to $1,500 on PC's, much more on larger computers

Present Situation

Desk-top publishing systems are relatively new products which are presented as being quite different and offered with little compatibility with less sophisticated word processing software. In the context of translation, they will have to be integrated with the other functions in much the same fashion as word processing has started to do in some systems like Termex and ABC word. (see section V.B.2).

f. Project Management

Example: Project Manager

Description

These software packages help manage several parallel projects by providing such functions as scheduling, resource allocation and time/cost monitoring. When properly used, they can help the manager of a translation service by providing early notice about projects that are taking too much time, by providing means for better cost accounting, etc.

Price Range: $200 to $1,000 on PC's, and much greater on larger computers

Present Situation

Project management packages now operate primarily as standalone applications. There is an opportunity to better integrate them in the translation environment by interfacing them with other functions such as word counting, billing etc. They could also become the keystone for the definition of a complete job environment that would include the background documentation, terminology and former translations relevant to each specific project.

g. Telecommunications

Example: GET, Red Ryder

Description

Communications software is intended to make a personal computer emulate a terminal to a remote system while taking advantage of its full computing power. For instance, it allows rapid downloading of information imported from the remote system onto local storage media, thus reducing considerably the connect time and associated costs.

Communications packages not only provide the link with remote systems, but they allow for defining and storing various communication and sign-on protocols including parameterized sessions to allow accelerated use. For instance, one can store the skeleton of a terminology lookup session with a given system and fill in the holes using a local program in order to reduce command keying errors and thus make more efficient use of the remote system and of the communications facilities.

Price Range: $50 to $200, often a nominal cost to be a registered user

Present Situation

Although most communications packages simplify access to a remote information retrieval system, their use is limited because they work work as a standalone application. The recent advent of core resident communications packages will further streamline the access to information by allowing connection to a remote system without having to exit the current application (such as word processing, diagramming, etc.).

h. General Office Automation Packages

Example: All-in-One

Description

Office automation packages integrate a number of office applications, typically word processing, spreadsheet, electronic mail, communications, personal agendas, etc. On larger systems (mini-computers), they can also provide a shared environment with sophisticated access restriction and security features.

Price Range: CPU dependent, up to $30,000 on large minicomputers

Present Situation

While they already integrate some functions of interest to translators, existing office automation packages are not flexible enough to truly integrate translation specific functions such as dictionary lookup. Moreover, they are more fully developed on large mini-computer systems such as VAX, HP3000, MV 2000, etc. and are therefore accessible only to large translation departments or to highly structured groups of individual translators.

i. Data Base Management Systems

Example: dBase III

Description

A database management system (DBMS) provides the framework for organizing data into a number of fields. Through the use of indexes, the DBMS allows retrieval, update, and logical operations on stored data.

Advanced DBMS packages also provide screen definition facilities for data entry and retrieval, reporting facilities, and volume processing functions for bulk data validation and/or update. The more advanced products provide complete development systems that allow the user to design and implement new applications.

Price Range: $200 to $1000 on PC's, much more on larger computers

Present Situation

Most general-purpose DBMS packages impose severe restrictions that make them inconvenient to use in terminology and information retrieval applications: no variable-length capability, capacity and performance problems, etc. These restrictions force language specialists to make several compromises when developing local term banks or information retrieval applications. Moreover, they are not integrated with other functions such as word processing and communications, and translators have to make further compromises to adapt them to their applications.

j. Information Retrieval Systems

Example: BASIS

Description

These products usually very large systems based on text management. They are typically used in bibliographic applications. For each document indexed, the basic author, title, and editor information is stored, together with a number of keywords and/or an abstract to identify the topics covered by the document.

Because of their size and cost, information retrieval systems are accessed remotely by terminal, usually through servers that can be reached via major packet switching networks.

Price Range: $10 to $150 per hour from a server

Present Situation

In spite of the rapid evolution of storage and retrieval devices, large information retrieval systems are here to stay because of the volumes involved. Automatic indexing has not yet reached a point where it will eliminate the need for sizeable organizations (often supported by public funds) to handle and index large numbers of publications. The opportunity here is to integrate information retrieval with other functions through incorporating communications software and a number of query protocols in the individual translator's working environment.

k. Full-Text Indexing and Retrieval Systems

Example: FUL/TEXT (Fulcrum Technologies Inc.)

Description

A full-text indexing and retrieval system is one that allows indexing of texts on all the words it contains. In most applications however, indexing will be limited to "significant words", with exclusion lists made up of grammatical and common words. Existing information retrieval systems are based on such systems.

Price Range: from several hundred to several thousand dollars

Present Situation

TDBMS (for text data base management system) is a new acronym, but the concept has been implemented for a long time. Recently these systems have become available on micro-computers and on local area networks, and could facilitate new applications for the translation environment.

One could, for instance, keep such a text data base of previously translated documents for reference or re-use portions of texts when translating a new document on the same topic or for the same organization. Through proper integration with other components, TDBMS could be used while translating a document or could lead to new applications such as computerized development of bilingual dictionaries.

Systems developed specifically for the translation market.

Not all systems in these categories are described here. For instance, we selected the two public term banks available in Canada, but there are similar systems in Europe. Also, we concentrated on those machine translation systems commercially available in Canada that will work on the English-French language pair, since they address the most significant market in Canada. In Europe, as well, English-French is the combination that represents the biggest market share, according to Common Market statistics.

TRANSLATION SPECIFIC PRODUCTS

a. TERMIUM

Description

TERMIUM is the public term bank system operated by the Secretary of State. In addition to the main terminology data base, it contains auxiliary data bases for official names, translation problems, and multilingual records. It has a total or nearly 1 million entries. The contents is mostly bilingual English-French, with a few Latin, Spanish and German entries.

TERMIUM can be accessed on-line through DATAPAC from everywhere in Canada. There are also a number of terminals installed abroad. There are many more users than the number of installations since there are numerous terminals shared by several translators (within SOS) and other accounts (e.g., in universities) used by several people. TERMIUM runs on VAX hardware. The underlying software is a version of the BASIS textual database management system that has been adapted especially for terminology applications.

Price: $15 per hour of connect time

General Assessment

While TERMIUM is recognized to be a very good system, capacity problems severely effect response time. To prevent further degradation of the response time, the Secretary of State has imposed limitations on issuing new subscriptions, admitting only those organizations that can make a significant contribution to the contents of the term bank.

For the time being, TERMIUM is not integrated with other productivity tools, and access means performing a separate procedure often using a separate piece of equipment.

b. Banque de terminologie du Quebec

Description

The Banque de terminologie du Quebec is another public term bank system, operated by the Office de la langue francaise du Quebec. It is also an English-French bilingual database, but it is aimed at standardization of French in Quebec rather than at translation per se.

The system runs on IBM mainframe equipment and IBM STAIRS is the underlying software. There are currently 75 to 100 active subscribers to the BTQ.

Price: $45 per hour of connect time

General Assessment

The Banque de terminologie du Quebec is less saturated than TERMIUM, but its contents are not at extensive of its Federal Government counterpart. It is also not integrated with other translation productivity tool.

c. Termex - Mercury

Description

Termex is a core-resident programme compatible with several major word processing packages on IBM PC, XT or AT hardware. It allows the computer user to dynamically call a glossary management programme in order to look for a term, insert it in the text being produced and/or update the glossary without having to exit from the word processing environment.

Termex also includes a keyboard enhancer that allows for two additional used-defined keyboards (which may include dead keys).

Approximately 200 to 300 copies of Termex have been sold Canada, Europe and the United States during the last 12 months.

Price: $275 retail in Canada

General Assessment

Termex has been the first of a new generation of products evolving towards an individual translator's workstation. It points to the feasibility of implementing a work station in software rather than hardware and demonstrates the utility of integrating general purpose and translation specific modules.

Termex is very fast and easy to use, but there is much room for increased functionality, e.g., integration of communications and definition of a shared environment. Linguatech has started developing such new additions to Termex and they should be available in future releases of Termex. Also, the investment in creating a glossary base will remain considerable unless ways are found to have several people share in the use and cost of individually developed dictionaries.

d. ABC WORD

Description

ABC WORD is a product very similar to Termex in its principle and functionality. The product has been released in Utah in early June 1987 and we did not have an opportunity to examine it in detail. People who have attended demonstrations confirmed its similarities with Termex (some said it might have been "reverse engineered" from Termex). It also runs on IBM PC, XT or AT compatible equipment.

Price: the price range would be between U.S. $150 and $200

General Assessment

It is not possible at this time to evaluate the product, except to say that it confirms the trend towards a software engineered integration of general-purpose and linguistic-specific tools.

e. INK Text Tools

Description

INK Text Tools is another individual productivity tool running on IBM PC/XT, PC/AT and compatibles. Unlike Termex and ABC WORD, it can be used in both single and multi-user environments. INK puts emphasis on text analysis for the purpose of creating dictionaries rather than on integration with word processing and other tools.

Price: $5,000 for 5 copies (minimum purchase)

General Assessment

INK Text Tools is also a very recent product (released one month ago in The Netherlands), so there is no reliable information on the user base or on the main strengths and weaknesses of the product.

f. ALPS AutoTerm

Description

ALPS Autoterm is a program that reads a text available in machine readable form and automatically looks up every word possible in a dictionary. The text to be translated is displayed one sentence at a time on the screen (window A). The translations found in the dictionary for the words in that sentence are displayed in window B, and the translator can paste them in the translated text he/she is writing in window C.

Autoterm runs on IBM PC/AT and Data General MV systems. A VAX version is in preparation and should be released soon.

Price: $2,000 to $6,000, depending on configuration

General Assessment

Although based on an interesting concept, AutoTerm requires a considerable investment in dictionary building, and the process of converting text produced with word processing software to the ALPS format is rather complicated. Therefore, only large volumes of text will justify the level of investment required, which may explain why there are less than 20 installations in the world.

Moreover, AutoTerm will appear more useful and productive if the translator remains close to the original text, which is often not compatible with quality translation.

g. ALPS TransActive

Description

ALPS TransActive is the machine translation system offered by ALPS. What makes TransActive unique is that it operates interactively, asking the user questions about the meaning and/or grammatical function of a word in case of ambiguity. The system contains several usual features including utilities to edit the raw output, create and update dictionaries, and reproduce in the translated text typesetting codes extracted from the original.

TransActive runs on IBM PC/AT and Data General MV systems. A VAX version is in preparation and should be released soon. There are between 10 and 12 installations in the world.

Price: $13,000

General Assessment

While the system is claimed to be a second generation system based on a linguistic theory called "junction grammars", it appears to be rather primitive in its syntactic and semantic processing. The information that one can place in the dictionaries is very poor the system asks many questions, and the quality of translation is poor unless much time and effort is spent developing situation-specific dictionaries.

The system is difficult to learn to use and does not have an automatic learning mechanism (it will ask the same questions all the time, even within the same text or sentence).

h. Logos

Description

Logos is a batch-oriented machine translation system that runs on Wang VS equipment. The post-editing and dictionary update functions are interactive, but the system does not interact with the user during translation. The system has been available for the German-English pair for several years, but the English-French version is rather new. There are only two installations of the English-French version at the present time.

Price: $6,500/month (one user), $10,500/month (multi-user)

General Assessment

Even in situations where the preliminary output is of good quality, there must be high volumes to justify the investment in this system. Assuming that machine translation can produce overall savings of $0.05 per word (25%) 210,000 words would have to be translated every month to pay for the software rental, exclusive of hardware and dictionary maintenance costs. Very few organizations will be in a situation where the system will be economically viable.

Price: $13,000/month per language pair

General Assessment

The absence of a general linguistic theory underlying the system, confirmed by the necessity to program individual macros as if every problem is a specific case, makes the system applicable only to situations where massive volumes of highly stereotyped texts have to be translated. Systran is also successfully used where the translation is used mainly for screening purposes. Moreover, it is difficult to acquire Systran, because of the intricate situation concerning the rights to sell, lease and/or distribute the various versions of the system.

k. Weidner

Description

Originally developed for PDP-11 and VAX-11 equipment for English-French, English-German and English-Spanish, the Weidner system is now implemented widely on IBM-PC compatible equipment. It combines batch oriented machine translation with an interactive environment for dictionary update and translation post-editing.

Price: $10,000 to $15,000 depending on configuration

General Assessment

While the system was not better or worse than others a few years ago, there is evidence that the firm is placing less emphasis on development work for language pairs of European origin. In addition, there was a massive layoff of research personnel last year. The company has been sold to a Japanese group and the Japanese-English counterpart of the system has had much more success in Japan (200 to 300 copies sold). It seems that this language pair is driving much of the total development effort on the Weidner system.

Price: $13,000/month per language pair

General Assessment

The absence of a general linguistic theory underlying the system, confirmed by the necessity to program individual macros as if every problem is a specific case, makes the system applicable only to situations where massive volumes of highly stereotyped texts have to be translated. Systran is also successfully used where the translation is used mainly for screening purposes. Moreover, it is difficult to acquire Systran, because of the intricate situation concerning the rights to sell, lease and/or distribute the various versions of the system.

k. Weidner

Description

Originally developed for PDP-11 and VAX-11 equipment for English-French, English-German and English-Spanish, the Weidner system is now implemented widely on IBM-PC compatible equipment. It combines batch oriented machine translation with an interactive environment for dictionary update and translation post-editing.

Price: $10,000 to $15,000 depending on configuration

General Assessment

While the system was not better or worse than others a few years ago, there is evidence that the firm is placing less emphasis on development work for language pairs of European origin. In addition, there was a massive layoff of research personnel last year. The company has been sold to a Japanese group and the Japanese-English counterpart of the system has had much more success in Japan (200 to 300 copies sold). It seems that this language pair is driving much of the total development effort on the Weidner system.