

ORIGINAL CONTAINS
COLOR ILLUSTRATIONS
7

NASW-4584
7N-82-CR
754
P-33

**MACHINE TRANSLATION PROJECT
ALTERNATIVES ANALYSIS**

Submitted to:

James Erwin

Scientific and Technical Information Program
National Aeronautics and Space Administration

Submitted by:

Catherine J. Bajis
Denise A. D. Bedford

July 15, 1993

(NASA-CR-200012) MACHINE
TRANSLATION PROJECT ALTERNATIVES
ANALYSIS (RMS Associates) 33 p

N96-17879

Unclas

G3/82 0098446

ALTERNATIVES ANALYSIS

- 1.0 Introduction
- 1.1 Alternatives Analysis
- 2.0 Definitions and Assumptions
 - 2.1 Definitions
 - 2.1.1 User Friendly Interface
 - 2.1.2 Vendor Experience and Support
 - 2.1.3 Multiple Standard Platforms
 - 2.1.4 Primary Language Pairs
 - 2.1.5 Reverse Language Pairs
 - 2.1.6 Dictionary Subjects
 - 2.1.7 Other Dictionary Features
 - 2.1.8 Input File Formats
 - 2.1.9 OCR Capabilities
 - 2.1.10 Source/ Object Language Handling
 - 2.1.11 Processing Features
 - 2.1.1.1 Related Terms Handling
 - 2.1.1.2 Dictionary Browsing
 - 2.1.1.3 Batch/ Interactive Mode
 - 2.1.1.4 Bilingual/ Split Screens
 - 2.1.1.5 Untranslatable Word Handling
 - 2.1.12 Target Language Handling
 - 2.1.13 Product and Delivery Options
 - 2.2 Assumptions Underlying the Machine Translation Procurement
- 3.0 Systems Descriptions
 - 3.1 Descriptions of Best Systems
 - 3.1.1 SYSTRAN
 - 3.1.2 Globalink
 - 3.1.3 STYLUS
 - 3.1.4 PC-Translator
 - 3.1.5 MicroCat/ MacroCat
 - 3.2 Overview of Other Systems
 - 3.2.1 Winger 92
 - 3.2.2 MicroTac's Language Assistants
 - 3.2.3 Socrata XLT
 - 3.2.4 TOLTRAN Professional Translation System 2.0
 - 3.2.5 ALPS
 - 3.2.6 LOGOS
 - 3.2.7 Tovna
- 4.0 Summary of Recommendations and Justifications
- 5.0 Table of Proposed Configurations and Costs

ALTERNATIVES ANALYSIS

1.0 Introduction

The Machine Translation Project consists of several components, two of which, the Project Plan and the Requirements Analysis, have already been delivered. The Project Plan details the overall rationale, objectives and timetable for the project as a whole. The Requirements Analysis is a description of the NASA STI Program's specific requirements for machine translation which must be satisfied by a given system or combination of systems. The Alternatives Analysis compares a number of available machine translation systems, their capabilities, possible configurations, and costs. The Alternatives Analysis has resulted in a number of conclusions and recommendations to the NASA STI Program concerning the acquisition of specific MT systems and related hardware and software.

1.1 Alternatives Analysis

Although fourteen (14) machine translation systems were evaluated, only four (4) are being recommended for immediate or future acquisition. The original intention of this project team was to select a single MT system, but upon careful consideration and evaluation, it was concluded that for a variety of reasons no single system meets all the requirements, although one system, SYSTRAN, meets all the critical requirements and many secondary ones as well.

NASA's most critical foreign-language-to-English machine translation requirements are: French, German, Japanese, and Russian. These four languages account for about 90% of the total words currently being translated by human translators for the NASA STI Program and other NASA users. Machine translation capability for additional foreign languages-into-English will supplement the existing human translation services and languages and satisfy some unmet user needs. The most critical English-to-foreign machine translation requirement is for English-into-Russian MT services to support NASA's ongoing cooperation with the Russian Space Agency.

In general, most MT requests will be processed by STI Program personnel through one of the MT systems; the translation will be returned to the user via FAX or e-mail. In some cases, though, the user/ requestor may wish to utilize one of these MT systems via dial-up access to process the translation without the direct assistance of STI Program personnel. In either case, MT will save considerable processing time and offer lower costs compared to human translation.

NASA also has critical requirements (rated with the highest factor weight of 3) concerning subject areas covered by the machine translation dictionaries. Translation requests which are produced primarily by human efforts and expertise under the current processing system cover a broad range of subject areas such as aeronautics, astronautics, astronomy, physics, chemistry and materials, life sciences, mathematics, computer science, engineering (mechanical, aeronautical, electrical), geosciences, and social sciences. In other words, NASA's translation requirements are driven by the agency's mission and cannot be any less broad than the agency's overall information requirements. For this reason, the number, subject areas and quality of the technical dictionaries supported by a given MT system are critical factors in that system's capability to satisfy NASA's overall translation requirements.

Furthermore, other dictionary characteristics are equally important. Quality translations, produced by machine or by humans, depend to a large extent on an understanding of phrases and idioms, abbreviations, and acronyms. No machine translation system can be complete without such linguistic information coded into its dictionaries any more than a human translator could produce quality translations without an understanding or knowledge of these linguistic elements.

2.0 Definitions and Assumptions

Definitions of the machine translation system factors which are referenced in the Alternatives Analysis are presented below. These definitions are listed in the order in which the factors are presented in Table 1 (Section 3) below. Assumptions are included to clarify the reasons why specific factors were selected and weighted as they were.

2.1 Definitions

2.1.1 User Friendly Interface

As the production of accurate translations is not a trivial effort regardless of the language skills of the personnel, i.e. whether the personnel are language-competent or simply trained to operate machine translation software, it is important that the interface of the selected system facilitate rather than encumber the translation process. User interface features may include clear commands and instructions, easy to read screens, and understandable error messages.

2.1.2 Vendor Experience and Support

Some machine translation products have been around since the 1970s while others are recent entries into the market. A vendor's longevity in the market does not necessarily imply a good system, since the fundamental MT approaches, technology, software and hardware have all changed considerably in the last ten years. Older systems may still be burdened with original problems that have been resolved in more recent systems.

The vendor of an MT system must be able to demonstrate commitment to the MT marketplace and a customer base that reflects technical or development interests comparable to NASA's. Future MT system developments from a vendor with established technical users will benefit NASA in the long-run. Whereas a company that has sold 250,000 small, PC-based general-dictionary systems to private citizens may have a very strong market share, the product will not be as useful, productive, or as good a long-term investment as the product from a company that has sold 100 technical-dictionary MT systems to scientific and technical agencies and corporations.

As machine translation is a complex task, vendor support is important for the resolution of software problems and the creation of a quality product. It is important that the vendor of any selected system has a commercial clientele of some size, demonstrates a consumer-orientation, and provides user services.

2.1.3 Multiple Standard Platforms

Due in part to the NASA STI Program's intention to shift from a mainframe environment to a client-server, distributed environment, a machine translation product that will function on a workstation or personal computer platform is preferred over a mainframe system. The hardware should be standard equipment which can be acquired through GSA schedules at competitive prices. To the extent possible, the selected product(s) should be able to function on equipment and software platforms already available to NASA. It should be compatible with existing operating systems, and should not require any major reconfiguration of hardware/ software in the NASA STI Program. This may not be the case for all technologies, since some such as the OCR, are state of the art and may require some additional components to implement. To the extent possible, though, the components should be standard hardware.

2.1.4 Primary Language Pairs

Primary language pairs for NASA means the availability of pairs where English is the target language, and the "foreign" language is the source language, e.g., Russian-to-English. A wide range of language pairs is offered by the systems commercially available.

2.1.5 Reverse Language Pairs

Reverse language pairs for NASA means the availability of pairs where English is the source language, and the target language is "foreign," e.g. English-to-Russian.

2.1.6 Dictionary Subjects

All systems which were considered have a general or core dictionary. These dictionaries generally include single-word entries and multiple-word entries. Specialized terminology dictionaries enhance translation quality for documents on technical subjects and are very valuable to NASA. Several developers offer subject-specific dictionaries for a variety of fields.

2.1.7 Other Dictionary Features

Other dictionary features include, (1) handling of phrases and idioms, (2) acronyms and abbreviations, (3) ability to customize or adjust definitions.

Some systems supply only general dictionaries which translate only single-words. These dictionaries may not accurately translate phrases and idioms. Likewise, translation of abbreviations and acronyms may not be supported. Both of these capabilities are important for scientific and technical translations.

Another important dictionary feature is the ability to customize a dictionary or to add terms and definitions. Some systems will generate lists of words not translated for a file, along with the capability to add these words and the user's definition to the dictionary.

The rate at which dictionaries are updated is also important to a scientific/ technical agency like NASA. In some fields such as computer science and engineering the terminology grows at a rapid rate. To achieve accuracy in translation, the dictionaries must be updated on a regular basis to reflect new terms or changing-definitions.

2.1.8 Input File Formats

The source document must be in a format that is compatible with the software. All of the software packages referenced in this report can read standard English ASCII format and some can accept foreign ASCII formats. Some of the systems can handle popular word processor files such as WordPerfect and Microsoft Word for English text. Some can accept foreign WordPerfect text as well. Some systems will accept the foreign file without further conversion, e.g. coding, whereas others require an additional conversion step.

2.1.9 OCR Capabilities

Since the majority of translations are performed on hard-copy rather than electronic files, some systems have been developed to accept scanned text. This increases the efficiency of processing tremendously, since the paper text does not have to be typed to create an electronic input file.

2.1.10 Source/ Object Language Handling

The source language, or object language, is the language being translated. Pre-editing is sometimes used to prepare a source document for translation by the computer. As a general rule, the better the quality of the source or object text, the better the quality of the translation product. Preparation can include running a spell-checker against the text, whether it is part of the machine translation software or part of the word processing software, searching for and correcting poor sentence construction or ambiguities using a grammar checker. Misspelled words cannot be translated correctly and may prevent the entire sentence from being properly translated at all. Poorly written text may generate translations with low conceptual accuracy.

2.1.11 Processing Features

Five processing features are discussed below.

2.1.1.1 Related Terms Handling

When a system encounters two or more acceptable translation terms for a source term, it may select one or it may present all and allow the user to choose. The need for user intervention and word selection could be reduced by utilization of more precise, technical dictionaries.

2.1.1.2 Dictionary Browsing

Some systems will allow the user to browse the dictionary components on-line to look up specific words or phrases. Generally, batch-processing systems do not provide this capability.

2.1.1.3 Batch/ Interactive Mode

Machine translations run in batch or interactive mode. Batch mode is unattended translation. Some systems offer an interactive mode, where the software will translate one sentence or paragraph at a time and pause for the user to edit the text on-screen. If a word has multiple meanings, some systems will pause during the translation and request the user to select a choice from a list of words. The user then selects from among the options and the translation continues. Some systems also permit the user to suspend translation, interrupt the process and recall the text later for further processing.

2.1.1.4 Bilingual/ Split Screens

Some systems offer postediting screens, which can be split vertically or horizontally with the source text beside the target text, above or below.

2.1.1.5 Untranslatable Word Handling

Another processing feature is the software's method of handling words which it cannot translate. In some systems the words will be marked or highlighted in the output. In other systems, the words are simply left as they were entered. If a system does not mark untranslated words, particularly where texts have to be converted from their original languages to a coding scheme, this can cause problems. Some systems also will generate lists of words-not-found, so that the dictionaries can be updated.

2.1.12 Target Language Handling

The target language is the language of the final translation. Most systems generate output in ASCII format which can be imported into word processing software. Grammar and spell checkers can be utilized on the resulting file to produce a more polished final product.

2.1.13 Product and Delivery Options

A time-saving feature offered in some systems is retention of formatting codes. If a source document was prepared for printing, it will contain certain embedded formatting codes such as bold face, underline, charts, and tables. Some software products can retain these codes in the target document for popular word processors such as WordPerfect or Microsoft Word.

2.2 Assumptions Underlying the Machine Translation Procurement

Nine assumptions are defined to provide a context for understanding the requirements specified for machine translation systems. These assumptions are discussed below.

Assumption 1: Machine translation services will be provided to the NASA Centers from a central operation and location. A distributed system configuration is not required at this time.

Assumption 2: The preferred machine translation systems are workstation or personal computer based systems which can be managed and maintained with a minimum of effort and systems expertise.

Assumption 3: The preferred machine translation systems will allow for dial-up access for submission of input files.

Assumption 4: The availability of OCR/ scanner technologies for machine translation will minimize the amount of labor required to prepare and process documents for translation.

Assumption 5: The need to pre-edit or code documents for translation should be minimal to maintain the potential efficiencies of machine translation.

Assumption 6: High quality, fine facsimile transmission will provide for faster turnaround times for translations.

Assumption 7: It should be possible to electronically transfer files from remote sites to the machine translation workstation, without interruption and/or interference.

Assumption 8: As a rule, if an MT product is "under development" it has not been evaluated as a product which is available and deliverable to NASA at this time. If a product is in a beta test stage in organizations, however, it was considered "deliverable."

Assumption 9: The Alternatives Analysis errs on the side of comprehensiveness, e.g. it evaluates all products referenced in published literature or referred to the STI Program even in cases where the system was not a good candidate for final selection.

It was important to take the time to evaluate as many products as possible because the market is not well-established. New products might offer capabilities and value that would have been overlooked otherwise. We were also motivated by a desire to provide as much information as possible to other NASA organizations which have expressed significant interest in MT and expect to acquire their own systems.

3.0 Systems Descriptions

Table 1 presents a summary graphic description of the characteristics and capabilities of all MT systems reviewed.

| Factor | Weight | PC-TRANSLATOR | | | | | M(A)ICROCAT | LEXITRANS | WINGER 92 | MICROTAC ASSISTANTS | SOCRATA/XLT | TOLTRAN | INTERGRAPH | ALPS | LOGOS | TOVNA/MTS |
|-----------------------------|--------|---------------|-----------|--------|---------------|-------------|-------------|-----------|-----------|---------------------|-------------|---------|------------|------|-------|-----------|
| | | SYSTRAN | GLOBALINK | STYLUS | PC-TRANSLATOR | M(A)ICROCAT | | | | | | | | | | |
| User Friendly Interface | 2 | | O | O | | | | | O | O | O | O | | | | |
| Vendor Experience & Support | 3 | O | O | d | O | | | | | | | | | | | |
| Multiple Standard Platforms | 3 | O | PC | PC | PC | O | PC | PC | PC | PC | PC | O | PC | PC | PC | |
| Primary Languages | | | | | | | | | | | | | | | | |
| Chinese | 2 | | d | | | | | | | | d | | | | | |
| French | 3 | O | O | | O | O | | | O | O | O | O | O | O | O | |
| German | 3 | O | O | | O | O | | | O | d | | O | | | d | |
| Italian | 2 | d | | | O | O | | | | O | d | | | | d | |
| Japanese | 3 | d | | | | O | | | | | | | | | | |
| Russian | 3 | O | O | O | | | O | d | | d | O | | | | O | |
| Spanish | 2 | O | O | | O | O | | O | O | O | | O | | | | |
| Danish | 1 | | | | O | O | | O | | | | | | | | |
| Norwegian | 1 | | | | O | O | | | | | O | | | | | |
| Swedish | 1 | | | | O | | | | | | | | | | | |
| Korean | 1 | O | | | | | | | | | | | | | | |
| Reverse Language Pairs | | | | | | | | | | | | | | | | |
| French | 2 | O | O | | O | O | | | O | O | | | O | O | | |
| German | 2 | O | O | | | O | | | O | | | | O | O | | |
| Japanese | 1 | O | | | | | | | | | | | | | | |
| Russian | 2 | O | O | O | | | | | | d | d | | | | | |
| Chinese | 1 | | d | | | | | | | | | | | | | |
| Portuguese | 1 | O | | | O | O | | | | | | | | | | |
| Norwegian | 1 | O | | | | | | | | | | | | | | |
| Korean | 1 | O | | | | | | | | | | | | | | |
| Greek | 1 | O | | | | | | | | | | | | | | |
| Dutch | 1 | O | | | | | | | | | | | | | | |
| Finnish | 1 | O | | | | | | | | | | | | | | |
| Swedish | 1 | O | | | O | | | | | | | | | | | |
| Danish | 1 | O | | | O | O | | | | | | | | | | |
| Italian | 1 | O | | | O | O | | | | O | d | | O | O | | |
| Spanish | 1 | O | O | | O | O | | | O | O | d | | O | O | | |
| Dictionary Subjects | | | | | | | | | | | | | | | | |
| General | 3 | O | O | O | O | O | O | O | O | O | O | O | O | O | O | |
| Aeronautics | 3 | O | O | | | | | | | | | | | | | |
| Astronautics | 3 | O | | O | | | | | | | | | | | | |
| Astronomy | 3 | O | | | | | | | | | | | | | | |
| Physics | 3 | O | | | | | | | | | | | | | | |
| Chemistry & Materials | 3 | O | | | | | | | | | | | | | | |
| Life Sciences | 3 | O | | | | | | | | | | | | | | |
| Mathematics | 3 | O | | | | | | | | | | | | | | |
| Computer Science | 3 | O | O | | | | | | | | | | | | | |
| Engineering | 3 | O | d | | | | | | | | | | | | | |
| Geosciences | 3 | O | | | | | | | | | | | | | | |
| Space Sciences | 3 | O | | | | | | | | | | | | | | |
| Social Sciences | 3 | O | | | | | | | | | | | | | | |

Table 1. Characteristics and Capabilities of All Machine Translation Products Evaluated

| Factor | Weight | SYSTRAN | GLOBALINK | STYLUS | PC-TRANSLATOR | M(A)ICROCAT | LEXITRANS | WINGER 92 | MICROTAC ASSISTANTS | SOCRATA/XLT | TOLTRAN | INTERGRAPH | ALPS | LOGOS | TOVNA/MTS |
|-------------------------------------|--------|---------|-----------|--------|---------------|-------------|-----------|-----------|---------------------|-------------|---------|------------|------|-------|-----------|
| Other Dictionary Features | | | | | | | | | | | | | | | |
| Phrases/Idioms | 3 | 0 | 0 | 0 | 0 | | 0 | | | | | | | | |
| Abbreviations | 3 | 0 | | 0 | 0 | | 0 | | | | | | | | |
| Acronyms | 3 | 0 | | 0 | | | 0 | | | | | | | | |
| Glossary Creation | 3 | 0 | 0 | 0 | 0 | | 0 | 0 | | | | | | | |
| Update Rates | 3 | 0 | 0 | 0 | | | | | | | | | | | |
| User Modifications | 3 | 0 | 0 | 0 | 0 | | 0 | 0 | | | | | | | |
| Input File Formats | | | | | | | | | | | | | | | |
| English ASCII | 3 | | 0 | | | | | 0 | | 0 | | | | | |
| English WordPerfect | 3 | | 0 | | | | | 0 | 0 | | 0 | | | | |
| Non-Cyrillic ASCII | 3 | | 0 | | 0 | 0 | | | | | | | | | |
| Foreign WordPerfect | 3 | 0 | 0 | 0 | | 0 | | | | | | | | | |
| Russian ASCII | 3 | | 0 | 0 | | | | | | | | | | | |
| Graphics Capabilities | 3 | | 0 | | | | | | | | | | | | |
| OCR Capabilities | 3 | 0 | 0 | 0 | | | 0 | | | | | | | | |
| Object Language Handling | | | | | | | | | | | | | | | |
| Spell Checkers | 1 | | 0 | 0 | | 0 | | | | 0 | | | | | |
| Grammar Checkers | 1 | | 0 | 0 | | 0 | | | | 0 | | | | | |
| Processing Features | | | | | | | | | | | | | | | |
| Related Terms | 1 | 0 | 0 | 0 | | | | 0 | | | | | | | |
| Dictionary Browsing | 1 | | | 0 | | | | | | | | | | | |
| Batch/Interactive | 1 | 0 | 0 | 0 | | 0 | | 0 | 1 | 0 | | | 0 | B | B |
| Bilingual Screens | 1 | 0 | 0 | 0 | | 0 | | 0 | 0 | | | | | | |
| Untranslatable Word Han | 1 | | 0 | 0 | | 0 | | 0 | | 0 | 0 | | | | |
| Target Language Handling | | | | | | | | | | | | | | | |
| Spell Checkers | 1 | | 0 | | | 0 | | | | | | | | | |
| Grammar Checkers | 1 | | 0 | | | 0 | | | | | | | | | |
| Product and Delivery Options | | | | | | | | | | | | | | | |
| English ASCII | 3 | 0 | 0 | 0 | | | | 0 | | | | | | | |
| English WP | 2 | | 0 | 0 | | | | 0 | | | | | | | |
| Graphics Format Retentic | 2 | | | | | | | | | | | | | | |
| TOTAL SCORE | 151 | 98 | 75 | 54 | 42 | 39 | 24 | 24 | 23 | 18 | 15 | 14 | 13 | 12 | 7 |

Table 1. Characteristics and Capabilities of All Machine Translation Products Evaluated (continued)

3.1 Description of Best Systems

3.1.1 SYSTRAN

Vendor: SYSTRAN Translation Systems, Inc., 7855 Fay Avenue, Suite 300, La Jolla, CA 92037

SYSTRAN is one of the best known MT systems. Originally developed as a Russian-to-English system under a contract for the U.S. Air Force, SYSTRAN now offers many language pairs, including a number of English-to-foreign pairs. Due to development under the Air Force contract, the language pairs with the largest technical dictionaries are Russian, German, and French.

Given the fact that NASA's primary machine translation requirements are for technical Russian, German, French, and Japanese into English, SYSTRAN is the best candidate to handle the majority of NASA's machine translation requests. The software versions for Russian, German, French and Spanish which were developed for the U.S. Air Force are currently available to U.S. government agencies at no cost; other SYSTRAN language pairs are only commercially available through the contractor. The Air Force is negotiating, however, with the contractor for a government-wide license agreement which will permit all U.S. government agencies to use other language pairs. Although the Japanese-to-English system is still under development, it is deployed to three agencies and is a working system.

SYSTRAN's capabilities and configuration options vary somewhat by language, but overall it offers a fairly broad language coverage, the best technical subject coverage, and the largest vocabularies of words, idioms, phrases and acronyms of any system which was reviewed by this project. In general, SYSTRAN is also more sophisticated and more complex to use than most of the other systems which were reviewed, particularly the PC-based ones. The technical dictionaries for Russian, German, and French cover aeronautics, astronautics, astronomy, physics, chemistry & materials, physics, life sciences, mathematics, computer science, engineering, geosciences, space sciences and social sciences. The language pairs which were not funded by the Air Force, however, do not contain the same type of technical dictionaries and specialized vocabularies as those found in the language pairs and versions developed for the Air Force.

The Japanese/English version of SYSTRAN is under development and currently in use by the Foreign Broadcast Information Service (FBIS), the Foreign Aerospace Science and Technology Center of the U.S. Air Force (FASTC), and the Department of Commerce. At the moment, utilization of this Japanese/English version of SYSTRAN is limited to these agencies but negotiations are underway to permit other agencies to acquire it. Although it has been deployed to these agencies, we did not list it as

deliverable for the above reason. As with the other language pairs developed for the U.S. government, the Japanese/English software will probably be cost free to other agencies.

Although the user generally must type/ rekey the foreign language text in order to prepare it for processing in SYSTRAN, some Optical Character Recognition (OCR) systems currently under development by FASTC and FBIS now permit the user to scan the foreign language text and convert it into a machine readable form. These OCR systems are sophisticated and accurate enough to read the printed foreign language and convert it ASCII. The ASCII file is then processed through SYSTRAN to produce a translation. The OCR recognition software has been offered to NASA at no cost from the developing agency.

FASTC, the primary government entity responsible for SYSTRAN's development under the Air Force contract, has funded the development of a new configuration for SYSTRAN that will run on a PS/2 with the assistance of an IBM Personal /370 Adapter card. The Personal /370 Adapter/A (P/370) is a co-processor for selected Micro Channel Architecture PS/2 computers. It emulates mainframe operations on a PS/2 by adding a standalone S/370 processor function to a PS/2 computer running OS/2. This configuration can be installed to permit users to access SYSTRAN via: 1) modem, 2) 3270 terminal emulation, 3) LAN. Any of the 3 access means will allow users to upload text files, select menu choices, process the translations, and download the translated text. The system can service 4-5 simultaneous users.

It is recommended that the NASA STI Program establish arrangements with the SYSTRAN contractor which are similar to the contractor's arrangements made with FASTC for the delivery and installation of the system. It is also recommended that the NASA STI Program acquire a workstation and configuration similar to the SYSTRAN workstation which was installed for FASTC at Wright-Patterson Air Force Base. The contractor was required to provide a fully bundled and tested system with 4 SYSTRAN MT systems pre-loaded and operational. Each system was assembled and tested by the contractor prior to delivery. The contractor delivered the systems to FASTC, installed them, and carried out the necessary testing and diagnosis to ensure that each system functioned properly. The contractor has the responsibility of debugging error situations, reloading software, or calling in hardware technicians if hardware problems occur.

The FASTC contractor is also required to instruct two FASTC systems analysts in the operation of the system, the procedure for loading new SYSTRAN software versions, and assist FASTC systems analysts in loading FASTC applications such as the Interactive SYSTRAN menu. The contractor also advises and consults FASTC in the connection of the system into a local area network. The system installation cost is included in the

purchase cost. The FASTC contractor was also required to separately price a one-time fixed trip charge that included travel time, air fare and per diem. It is recommended that NASA STI Program make similar arrangements with the SYSTRAN contractor in order to facilitate the installation and operation of SYSTRAN for the NASA STI Program in the most efficient manner possible.

Recommendation:
acquisition of the following:

- o IBM PS2-95
- o OS/2 Operating System
- o 370 processor functioning board
- o 1 IBM P/370 Adapter/A desktop workstation bundled with OS/2 operating system
- o SYSTRAN systems written in IBM assembler that run under VMS for the these languages: Russian/English, French/English, German/English, Spanish/English
- o Tiger OCR recognition system for Cyrillic from Cognitive Technologies, Office of Research and Development, and FASTC; no cost to U.S. Government agencies
- o HP Scanjet IIP
- o WordPerfect Russian module

3.1.2 Globalink

Vendor: Globalink, 9302 Lee Highway, Fairfax, VA 22031

Globalink is a PC-based machine translation software package for French/English, German/English, Spanish/English and Russian/English; these language pairs are all bi-directional and at no extra cost. The English/Russian version is due for release in July 1993; additional language pairs are under development. Globalink has a user-friendly interface, a split screen mode, and allows either batch or interactive processing of texts.

Globalink is an excellent PC-based MT system and has a dictionary of 60,000+ words. Although the dictionary is still relatively small and does not have sufficiently broad subject coverage to meet all of NASA's requirements for scientific and technical translations, it does have one of the largest dictionaries of the PC-based systems with the exception of SYSTRAN. Also, it is worth noting that Globalink is expanding the aerospace dictionary under a contract with a NASA organization and will probably increase the number of technical terms considerably. This aerospace dictionary should be available to the NASA STI Program.

Globalink has most of the capabilities and characteristics that are required by the NASA STI Program and ranked number two

in this analysis. Among the systems which were evaluated for this project, it has the unique capability to utilize an OCR system for German, French and Spanish; this characteristic is currently unavailable with the other systems which were reviewed or recommended. For this reason, it is recommended that the NASA STI Program acquire the French/English, German/English, Russian/English and Spanish/English language pairs. Although some of this language capability might seem to be redundant to the language coverage recommended for SYSTRAN, the acquisition of these language pairs for Globalink as well as SYSTRAN might provide the opportunity to utilize the Globalink OCR system in tandem with the SYSTRAN system. It might be possible to scan/convert the German, French, and Spanish texts to ASCII with the Globalink OCR system and subsequently use the SYSTRAN system for the more technical translations which Globalink is unable to handle. Furthermore, Globalink also has a business Russian dictionary which will probably contain additional words and phrases which are not likely to be found in SYSTRAN's scientific and technical dictionaries or in the other PC systems' small, general dictionaries.

Recommendation:
acquisition of these language pairs: French/English, German/English, Russian/English, Spanish/English; all language pairs are bi-directional

Proposed configuration:
PC; no additional hardware needed

3.1.3 STYLUS

Vendor: Sigma Technologies, Moscow, Russia

STYLUS is a PC-based MT system being marketed by Sigma Technologies, Moscow, Russia and Integration Communications International, Inc., Washington, DC. It has a user-friendly interface and a split screen mode. The English/Russian capability will probably be particularly useful in supporting NASA's joint cooperation with the Russian Space Agency. The French, German, Italian, and Spanish versions translate into Russian and vice-versa. Although this particular reverse capability for these languages will probably not be extremely useful, it might offer the opportunity to translate these languages, particularly Italian, into English, by processing the texts twice, once through the Italian/Russian module to produce the Russian version, and once through the Russian/English module to produce the final English.

Since STYLUS looks like an excellent MT product and the software has been offered to the STI Program at no cost, acquisition of all the language pairs available is a

recommendation. The English/Russian pair will certainly be useful to NASA; the installation and utilization of the foreign-to-Russian pairs is a question of storage and disk space rather than cost. Sigma Technologies has also indicated that additional development of the Russian/English aerospace dictionary could be done under contract with NASA.

Recommendation:
acquisition of Russian/English, Italian/Russian, German/Russian, Spanish/Russian, French/Russian; all language pairs are bi-directional

Proposed configuration:
PC; no additional hardware needed

3.1.4 PC-Translator

Vendor: Linguistic Products, P. O. Box 8263, The Woodlands TX 77387

PC-Translator offers the following primary language pairs: Spanish/English, French/English, Danish/English, Swedish/English, Italian/English, and German/English. It also offers the following reverse language pairs English/Spanish, English/French, English/Danish, English/Swedish and English/Italian. Language pairs are not bi-directional; capability for reverse translations must be purchased for a specific language pair and direction. Additional pairs are in development at this time, including English/Dutch, English/German, Portuguese/English, French/Spanish, Spanish/French, Dutch/German and German/Dutch. PC-Translator can read source documents in ASCII, WordPerfect, Microsoft Word, WordStar, WordStar 2000 files. Formatting codes from WordPerfect, Word, WordStar and WordStar 2000 files will be retained.

The general dictionary contains 40,000 to 70,000 terms, depending on the language. A separate user dictionary is included. Users may add, delete, or modify all dictionary entries. The dictionary is in ASCII format and permits the importation of word lists or glossaries of terms directly into the dictionary.

The software will allow stacking of up to ten single-word dictionaries and ten phrase dictionaries. The total dictionary sizes and entry lengths are unlimited. Wildcards can be used for dictionary updating and for conjugated verbs. A wildcard in a phrase can represent thousands of nouns, verbs, or adjectives. Multiple wildcards can also be placed in phrase dictionary entries.

Dictionaries can also be created from a words-not-found list. There is a utility to transfer the words-not-found directly into the dictionary. Dictionary coding indicates parts of speech, gender, and number. The software can automatically conjugate verbs. The Spanish, French, Italian, and Portuguese versions have gender and number agreement.

PC-Translator's primary users are major corporations, auto manufacturers, scientific organizations, and government agencies. Some specific users in the commercial sector are: British Telecom, Chemlink/Baker Hughes Company, Clorox International, Fluor Corporation, Ford Motor Company, General Electric, Hewlett Packard, Honeywell, IBM, John & Johnson, Kaiser Engineering, Massey-Ferguson, Marathon Oil, Michelin, Motorola, RCA Home video, Rockwell International, Spirotechnique, Volvo, University of Florida, Westinghouse, and Zenith Corporation.

The unique language pairs which are available with PC-Translator but not available from other system recommended for acquisition by this project, are: Danish/English, Swedish/English, Italian/English and many of the reverse pairs. It is recommended that NASA acquire the Italian/English language pair, and evaluate NASA's needs for additional language coverage. Each language pair sells for \$985. PC-Translator runs on the IBM-PC, with 640K RAM, DOS 3.1+, and requires 2.5 MB of hard disk space.

Recommendation:
acquisition of the Italian/English language pair

Proposed configuration:
PC; no additional hardware needed

3.1.5 MicroCat/MacroCat

Vendor: Weidner Communications, Northbrook, IL

Weidner offers two machine translation products: 1) MacroCat for a mainframe, and 2) MicroCat for a PC. The mainframe version, which runs on DEC VAX/VMS computers, was introduced to the market in the late 1970s. The personal computer version runs on IBM PC/XT platforms. Both products have many useful features and cover some important primary language pairs of interest to NASA. The language pairs available for MicroCat are: French/English, Spanish/English, German/English, and Japanese/English. Reverse language pairs available for both MicroCat and MacroCat include: English/French, English/Spanish, English/German, English/Italian and English/Portuguese. Subject-specific dictionaries are not available for any of these language pairs.

The company also sells a text editor that allows simultaneous split-screen viewing of the source and target text. The products operate in either batch or interactive mode and can accept documents in non-Cyrillic foreign ASCII and WordPerfect formats.

These products are ranked in the top five systems because of the language pairs and the interface features. There are two issues of concern, however, about MacroCat and MicroCat products. First, the price quoted in the literature for these systems is \$50,000 for a bi-directional system (i.e., English/French and French/English). For two pairs, the price is \$85,000. This price is far beyond the NASA budget for acquiring machine translation systems, since not even one language pair could be acquired with the funding available. Furthermore, there is no reason to conclude that the system would be superior in any way to the SYSTRAN system for the same language pairs which are already available to NASA at no charge.

The second concern is vendor viability. Although references in 1986 indicated that Weidner Communications was located in Northbrook, Illinois, the company could not be located in that area in 1993. Neither have there been subsequent references to the company in the literature since that time. There is concern that the product may not be actively supported by the vendor. Therefore, acquisition of this product is not recommended for these two reasons.

Recommendation:
acquisition not recommended

3.2 Overview of Other Systems

Several additional machine translation systems were reviewed in the Alternatives Analysis. Included were: Winger 92, Microtac's Language Assistant series, Socrata's XLT, Toltran's French Correspondent, Intergraph, ALPS, LOGOS, and Tovna's MTS.

These systems did not rate in the recommendations for acquisitions for several reasons. First, their language capabilities are limited in comparison to the top five systems. Second, their dictionaries were general and not scientific. Third, they were largely small-scale personal computer-based systems whose capabilities and features did not compare advantageously with Globalink's Linguistic Products', and Sigma's products. Fourth, the capabilities they offered were already covered by the top four systems. No capabilities are lost by excluding these systems from further consideration.

Finally, the capabilities of these systems, in terms of general system features, were not as competitive as those of the

first four. None offer specialized subject dictionaries. If a point score comparison is made (Table 1.b) these systems provided less than one-sixth of the machine translation capability required by NASA. In contrast, two systems, SYSTRAN and Globalink's GTS Professional, could provide half- to two-thirds of the capability required. PC-Translator provides additional capability in terms of language pairs not covered by the other systems. To illustrate these points a brief description of each of the systems is presented.

3.2.1 Winger 92

Vendor: Winger, Skodsborgvej 48 F1, DK-2830 Virum, Denmark

This is a personal computer based product which runs on an IBM-PC with 640K RAM and requires 40 MB of hard disk space. Currently only English/Danish, English/Spanish, Danish/English, and Spanish/English are available. Other language pairs are under development, including: English/French, French/English, English/Russian, Russian/English, Danish/Russian, and Russian/Danish. All of these available languages are covered by the top four systems.

The dictionary for these languages is general and not subject specific. The general dictionary contains between 15,000 and 40,000 entries. Single words and phrases may be added to the dictionary. Source text can be in ASCII, Wordperfect, DSI, or Ami Pro. Microsoft Word format will be supported in 1993. The product generates a list of not-found words in ASCII format, with guesses as to grammatical properties.

The price of the product is quoted at \$1,000, though it is not clear whether this price is for a single language pair or for all available languages. Currently there is no U.S. distributor for this product though it can be exported to the U.S. The product has only been on the market since 1992.

3.2.2 MicroTac's Language Assistants

MicroTac's Language Assistant products have been available since 1988. The user group consists of over 100,000 individuals; the product is utilized primarily for business correspondence and educational use. The price for each Assistant includes bi-directional translation capabilities. The language pairs available are: English/Spanish, English/French, English/Italian, and English/German. The Assistant series runs on IBM PC's with 640K RAM, DOS 2.1+, and 2.5 MB hard disk space. Each Assistant product is \$79.95.

The products began as simple verb conjugator programs. Each version has added new features and capabilities. The current versions offer sentence-by-sentence translations. The Assistants include Reference Tools, e.g. bilingual dictionaries, verb conjugators, grammar help topics and accent entry utilities. The general dictionary contains over 50,000 entries and the entries can be added, deleted, or modified. The verb conjugator contains conjugations for 2,000+ verbs. Verbs and bilingual dictionary entries can be pasted directly into a document.

The software will translate interactively or in batch mode. However, reviews of the software indicate that the Assistants cannot translate a full-document as submitted. The interactive mode allows users to select from as many as 15 translations for a given word. The translation options provide enough context to allow users with little or no knowledge of the target language to make a selection.

Source and target text can be viewed or printed in paragraph format, side-by-side, or line-by-line format. The software will automatically convert WordPerfect, Microsoft Word, or Word for Windows files to ASCII format. A word scan feature generates a list of words not found in the dictionary. Words that are not translated are placed in brackets within the text.

3.2.3 Socrata XLT

Vendor: Socrata, 5500 Royalmount Ave., #320, Town of Mount-Royal, Quebec, Canada H4P 1H7

The XLT product has only been on the market since 1992. The user group consists primarily of translation bureaus and departments. Four languages are offered, in any direction, including English, Spanish, French and Italian. The XLT product requires IBM-PC, UNIX or XENIX operating systems, a 386 processor with 200 MB of hard disk space and 4 MB RAM.

The initial subscription price for a single language pair is \$5,000. Additional pairs are \$2,500 each. Annual renewals of subscriptions, including upgrades, are available for \$1,000.

XLT translates interactively or in batch mode, running about 250,000+ words per hour on an AT-386. It is designed to support multiple users. The software requires a client name prior to translation of text. The system automatically creates a separate directory and glossary for each client. A quick scan is performed for text, and not-found-words are fed into the client dictionary. A spell-checker is provided.

Dictionary updating can be automatic or manual. The software will query the user on any terms not found. The general

dictionary contains over 70,000 entries and can be modified by the user. The total number of dictionary entries is unlimited. Dictionary entry lengths can be chained to allow for longer words or phrases. An unlimited number of translations for a single word are possible. Dictionaries can also be stacked.

3.2.4 TOLTRAN Professional Translation System 2.0

Vendor: Toltran, Ltd. 775 Oakwood Road, Suite S1A, Lake Zurich, IL 60047

Professional Translation System 2.0 is a personal computer based product that offers only one language pair at this time, English/Spanish, and Spanish/English. Under development are Russian, Chinese, Italian and Portuguese modules. The product requires an IBM-PC with a 286 processor or higher, 512K RAM, DOS 3.0 or higher, and 1 MB of hard disk space per module.

This system employs a modular approach to translation and has a great deal of potential since any source language module translates into any target language module. The graphical user interface (GUI) can be accessed via the keyboard or a mouse. Multiple windows may be opened on the screen. Words and phrases that are not translated are placed inside <> marks. TOLTRAN's PTS 2.0 cannot handle formatted word processed documents. All text must be in ASCII.

A portion of the text may be highlighted or marked for translation using a "smart paste" feature. The translation is inserted into the original document. Both source and target text can be edited on-line. The system also allows the user to queue up to sixteen documents for unattended batch processing.

The primary users of this product are international companies, government organizations, and educational institutions. The cost is only \$249 per module.

At this point in time, the product does not have much to offer NASA that isn't covered by another system. However, if the Chinese module is delivered to market, we recommend that NASA review that product for acquisition. NASA might consider acquiring the Chinese system, provided that some method of efficient text entry is available and the Chinese dictionaries cover some scientific and technical subjects.

3.2.5 ALPS

Vendor: Automated Language Processing Systems, Provo, Utah

The ALPS product was marketed as a machine translation tool to aid human translators. ALPS software runs on the IBM PC/AT under the Xenix operating system, on Data General MV series minicomputers running the AOS/VS operating system, and on IBM 4300 series computers running the VM/CMS operating systems.

ALPS currently supports translation from English/French, English/German, English/Italian, and English/Spanish. In addition, French/English is supported. Of these language pairs, NASA's primary interest would be French/English. This capability is already offered by other systems.

Two nice features of the ALPS products are multilingual word processor components, and interactive access to on-line dictionaries for each language pair.

Each one-way language pair costs \$13,000 per workstation. This cost would consume over half of the available funds for machine translation since acquisition of three pairs would cost \$39,000. Given the availability of other systems having similar language coverage for less cost, the acquisition of the ALPS product is not an efficient use of the NASA's funds.

In addition to the cost issue, when we contacted ALPS in Provo, Utah by telephone, they indicated the software was not being "actively marketed" at this time. The personnel appeared to discourage interest in the product. This does not speak well for vendor support or long-term product viability.

3.2.6 LOGOS

Vendor: Logos Computer Systems, Wellesley, Massachusetts

The LOGOS product was developed about the same time as SYSTRAN. In 1985 its clients included Nixdorf, IBM and Hewlett-Packard, clients who are now using the PC-Translator product described earlier in this report. In the late 1980s, LOGOS was used primarily by government translation bureaus in the U.S. and Canada and for the development of commercial product manuals. At this time only three language pairs are offered: English/French, German/English, and English/German. NASA's primary interest would be in the German/English pair. The dictionaries are general, not subject specific. The capabilities offered by this system are covered by other products. LOGOS runs on Wang VS and IBM VM/CMS systems. There is no personal computer version available at this time.

The LOGOS product is developed around a universal intermediate language, the Semantic Abstraction Language which serves as an intermediate code for all its translations. The source language is scanned once and changed to SAL. The SAL for

the text can then be read and changed to several languages without further reference to the source.

In addition, LOGOS does not sell its software, but licenses it. We could not find any pricing information for the product. Neither could prices be found in references to the product in the machine translation literature. In addition to being unable to find pricing information, we also were unsuccessful in locating the company.

3.2.7 Tovna MTS

Tovna's MTS product was introduced to the market in 1987. This is a foreign-developed product, which offered only one primary language pair, French/English. This pair is currently in development and is not available for purchase or license at this time. Two reverse pairs include English/French and English/Russian. The English/Russian pair is also in development. Theoretically, the English/French product is only a pilot system. It is in beta test at some organizations now.

Tovna claims to have the only machine translation system that learns from its users' editing and incorporates them into future translations. Tovna does not have a personal computer based product, so it does not meet NASA's configuration requirements. The MTS product is not available for purchase. When the products are delivered to market they will be available for permanent licensing.

4.0 Summary of Recommendations and Justifications

Although SYSTRAN clearly satisfies NASA's requirements more completely than any of the other systems reviewed by this project, it is advisable to acquire more than one MT system. Table 2a through 2d demonstrates the complementary and combined capabilities that are provided by the acquisition of the four top ranked MT systems. Although SYSTRAN provides the best scientific and technical subject coverage for specific languages, some of the other systems provide either additional language pairs or subject specific dictionaries related to law, finance, and business. These other language pairs and subject areas are occasionally required in the NASA translation environment. In addition, it is possible that some untranslated words which cannot be found in a given dictionary and system might be found in another.

Some of the PC-based packages are more user-friendly and less complex than the SYSTRAN software. Used in conjunction with telecommunications packages, it may be possible to allow individual requestors direct access to these user-friendly

translation systems. This would permit users to process their own translations directly without the intervention of STI Program personnel. This could be very popular in situations where turnaround time is extremely short and users are not willing or able to wait for STI Program personnel to process their texts.

Only SYSTRAN permits simultaneous users; all the other recommended systems are available to only a single user at any given moment. Therefore, multiple MT systems will allow users access to translation capabilities in the event that a particular system for a given language pair is already tied up.

Therefore, we recommend that the NASA STI Program install SYSTRAN, Globalink, PC-Translator and Stylus. Each system will confer unique capabilities and/ or specific advantages. SYSTRAN will be able to provide machine translation capability for extremely technical and scientific texts for Russian, German, French and Spanish. Globalink, PC-Translator, Stylus and every other system which was evaluated cannot match SYSTRAN's capabilities for technical and scientific texts. Globalink will provide, however, additional translation capability for general texts and reverse capability for each of these four languages. PC-Translator will provide general translation capability for Italian/ English, a language pair not available with SYSTRAN or Globalink. Stylus will provide general translation capability for Russian/ English and English/ Russian. Although some of Stylus' capabilities seem to duplicate SYSTRAN's and Globalink's, we feel that it may have dictionaries and capabilities, particularly for English/ Russian, which are not duplicated by any of the other systems.

The SYSTRAN software for Russian, French, German and Spanish is available to U.S. government agencies at no cost; only the hardware, installation and system support must be purchased. Stylus is available at no cost from Sigma Technologies. Globalink and PC-Translator must be purchased from the vendors.

ACQUISITION ACTIONS

Contact the following vendors and acquire the designated products. * items are available at no cost to NASA.

PRODUCT: SYSTRAN
SOURCE: SYSTRAN, Inc.
PURPOSE: to provide scientific and technical machine translation capability for Russian, French, German, Spanish
CONTACT: Chris Fitch, Steve Dakis
Tel. 619-459-6700
fax 619-459-8487
ADDRESS: 1055 Wall Street, Suite 213

P.O. Box 907
La Jolla, CA 92037

SYSTRAN does not have a GSA contract number.

ITEMS:

- 1) * SYSTRAN translation system software for Russian, French, German, and Spanish into English
- 2) IBM PS/2, model 95 acceptable, 486-33Mhz or better, 400 MB hard drive (model #9585-OXF or higher)
- 3) IBM P/370/A hard card, direct from SYSTRAN. (\$6,800.00 plus 10.58% G&A = \$6,966.54; includes VM operating system)
- 4) OS/2 operating system (2.0 or higher)
- 5) OS/2 extended services (approximately \$35.00)
- 6) IBM Communications Manager (includes 3270 emulation software)
- 7) DOS 5.0
- 8) WordPerfect, MS Word, and NotaBene software interface filters which SYSTRAN engineers have developed to allow seamless integration of text file formats with SYSTRAN translation software (\$200.00 each)
- 9) WordPerfect Russian module (approximately \$150.00)
- 10) 32 MB RAM (16 MB is the minimum, but additional RAM has been identified as a requirement)
- 11) additional 400 MB hard drive
- 12) VGA monitor (approximately \$250.00)
- 13) Microsoft PS/2 compatible mouse (approximately \$80.00)
- 14) 1 IBM "Magneto-Optical" disk (120 MB; approximately \$1400.00)
- 15) travel and per diem costs for one of SYSTRAN's systems engineers to install the system at a single location
- 16) HP ScanJet IIP
- 17) HP LaserJet Cyrillic font Cartridge (approximately \$255.00)
- 18) 3Comm Ethernet TCP/IP LAN Hard Card
- 19) Novell Netware LAN software
- 20) * Tiger OCR Russian software system (approximately \$750.00 if NASA obtains it from the vendor; NASA has been advised, however, that the software is available at no charge through the Office of Research and Development)

PRODUCT: Globalink Translation System (GTS)-Professional
SOURCE: Globalink, Inc.
PURPOSE: to provide bi-directional machine translation capability for Russian, French, German and Spanish for general texts; to provide additional translation capability for technical texts (see list of specific subject dictionaries)
CONTACT: Rich Perrotti
Tel. 703-273-5600

fax 703-273-3866
ADDRESS: 9302 Lee Highway, 12th Floor
Fairfax, VA 22031

GSA Contract Number: GS00K91AGS5233
product numbers for GSA schedule are in parentheses

ITEMS:

- 1) Globalink's GTS-Professional for:
Spanish (SE2P); French (FE2P); German (GE2P); Russian (RE2P)
- 2) Perceive OCR software (8001)
- 3) subject dictionaries:
French - Business, Legal and Finance (4009); Aviation
(4014); Chemical (4015)
German - Computer (4018); Telecommunications/Cable (4016)
Spanish - Petroleum and Mining (4008); Aviation and
Industrial (4011); Legal, Business, and Finance
(4012)
Russian - Business (4013)

PRODUCT: PC-Translator
SOURCE: Linguistic Products
CONTACT: Evelyn Smith
tel. 713-298-2565
fax 713-298-1911
ADDRESS: P.O. Box 8263
The Woodlands, TX 77387

Linguistic Products does not have a GSA contract number.

ITEMS:

- 1) PC-Translator version 3.4; language pair: Italian/ English
- 2) no additional hardware or software needed

PRODUCT: Stylus
SOURCE: Sigma Technologies, Inc.

PURPOSE: to provide English/ Russian, Russian/ English machine
translation capability for general texts

- 1) language pair(s): Russian/ English, Italian/ Russian,
German/ Russian, Spanish/ Russian, French/ Russian; all
language pairs are bi-directional
- 2) no additional hardware needed

SYSTRAN



| Subject Dictionaries | Primary Languages | Reverse Languages |
|----------------------|-------------------|-------------------|
| General | Russian | Russian |
| Aeronautics | French | French |
| Astronautics | German | German |
| Astronomy | Japanese | Japanese |
| Physics | Italian | Italian |
| Chemistry/Materials | Chinese | Chinese |
| Life Sciences | Spanish | Spanish |
| Mathematics | Danish | Danish |
| Computer Science | Norwegian | Norwegian |
| Engineering | Swedish | Swedish |
| Geosciences | Korean | Korean |
| Space Sciences | Portuguese | Portuguese |
| Social Sciences | Greek | Greek |
| Business | Dutch | Dutch |
| Legal | Finnish | Finnish |

| | | |
|-------------------------|---------------------|-----------------------|
| Pre-Processing | Dictionary Updates | Graphics Retention |
| User Friendly Interface | Abbreviations | Format Code Retention |
| PC PLATFORM | Acronyms | Bilingual Screens |
| Vendor Experience | Glossary Creation | Untranslatable |
| WP Compatibility | Phrases/Idioms | Word Handling |
| | Dictionary Browsing | |
| | Related Terms | |

Paper, Raw Output
Paper, Edited
Electronic, Raw Output
Electronic, Edited

 = SYSTRAN Capabilities

SYSTRAN

GLOBALINK

| | |
|----------------------|-----------------------|
| Paper Documents | FAX Document Delivery |
| Letters | Scanner |
| Reports | |
| Notes | |
| Articles | FTP |
| Drawings/Graphics | |
| Books | |
| Electronic Documents | |
| Letters | |
| Reports | |
| Notes | |
| Articles | |
| Drawings/Graphics | |
| Books | |

| Subject Dictionaries | Primary Languages | Reverse Languages |
|----------------------|-------------------|-------------------|
| General | Russian | Russian |
| Aeronautics | French | French |
| Astronautics | German | German |
| Astronomy | Japanese | Japanese |
| Physics | Italian | Italian |
| Chemistry/Materials | Chinese | Chinese |
| Life Sciences | Spanish | Spanish |
| Mathematics | Danish | Danish |
| Computer Science | Norwegian | Norwegian |
| Engineering | Swedish | Swedish |
| Geosciences | Korean | Korean |
| Space Sciences | Portuguese | Portuguese |
| Social Sciences | Greek | Greek |
| Business | Dutch | Dutch |
| Legal | Finnish | Finnish |

| Pre-Processing | Dictionary Updates | Graphics Retention |
|-------------------------|---------------------|------------------------------|
| User Friendly Interface | Abbreviations | Format Code Retention |
| PC PLATFORM | Acronyms | Bilingual Screens |
| Vendor Experience | Glossary Creation | Untranslatable Word Handling |
| WP Compatibility | Phrases/Idioms | |
| | Dictionary Browsing | |
| | Related Terms | |

Paper, Raw Output
Paper, Edited
Electronic, Raw Output
Electronic, Edited



SYSTRAN

GLOBALINK

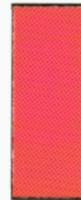
PC-TRANSLATOR

| | |
|----------------------|-----------------------|
| Paper Documents | FAX Document Delivery |
| Letters | |
| Reports | |
| Notes | |
| Articles | |
| Drawings/Graphics | Scanner |
| Books | |
| Electronic Documents | FTP |
| Letters | |
| Reports | |
| Notes | |
| Articles | |
| Drawings/Graphics | |
| Books | |

| Subject Dictionaries | Primary Languages | Reverse Languages |
|----------------------|-------------------|-------------------|
| General | Russian | Russian |
| Aeronautics | French | French |
| Astronautics | German | German |
| Astronomy | Japanese | Japanese |
| Physics | Italian | Italian |
| Chemistry/Materials | Chinese | Chinese |
| Life Sciences | Spanish | Spanish |
| Mathematics | Danish | Danish |
| Computer Science | Norwegian | Norwegian |
| Engineering | Swedish | Swedish |
| Geosciences | Korean | Korean |
| Space Sciences | Portuguese | Portuguese |
| Social Sciences | Greek | Greek |
| Business | Dutch | Dutch |
| Legal | Finnish | Finnish |

| Capabilities | GLOBALINK | PC-TRANSLATOR |
|-------------------------|---------------------|-----------------------|
| Pre-Processing | Dictionary Updates | Graphics Retention |
| User Friendly Interface | Abbreviations | Format Code Retention |
| PC PLATFORM | Acronyms | Bilingual Screens |
| Vendor Experience | Glossary Creation | Untranslatable |
| WP Compatibility | Phrases/Idioms | Word Handling |
| | Dictionary Browsing | |
| | Related Terms | |

| |
|------------------------|
| Paper, Raw Output |
| Paper, Edited |
| Electronic, Raw Output |
| Electronic, Edited |



= **SYSTRAN**
= Capabilities Only



= **GLOBALINK**
= Capabilities Only



= **PC-TRANSLATOR**
= Capabilities Only



= Capabilities
= Common to
= Two or More

SYSTRAN

GLOBALINK

PC-TRANSLATOR

STYLUS

| | |
|---|------------------------------|
| <p>Paper Documents</p> <ul style="list-style-type: none"> Letters Reports Notes Articles Drawings/Graphics Books | <p>FAX Document Delivery</p> |
| <p>Electronic Documents</p> <ul style="list-style-type: none"> Letters Reports Notes Articles Drawings/Graphics Books | <p>Scanner</p> |
| <p>FTP</p> | |

| Subject Dictionaries | Primary Languages | Reverse Languages |
|----------------------|-------------------|-------------------|
| General | Russian | Russian |
| Aeronautics | French | French |
| Astronomy | German | German |
| Physics | Japanese | Japanese |
| Chemistry/Materials | Italian | Italian |
| Life Sciences | Chinese | Chinese |
| Mathematics | Spanish | Spanish |
| Computer Science | Danish | Danish |
| Engineering | Norwegian | Norwegian |
| Geosciences | Swedish | Swedish |
| Space Sciences | Korean | Korean |
| Social Sciences | Portuguese | Portuguese |
| Business | Greek | Greek |
| Legal | Dutch | Dutch |
| | Finnish | Finnish |

| | | |
|-------------------------|---------------------|-----------------------|
| Pre-Processing | Dictionary Updates | Graphics Retention |
| User Friendly Interface | Abbreviations | Format Code Retention |
| PC PLATFORM | Acronyms | Bilingual Screens |
| Vendor Experience | Glossary Creation | Untranslatable |
| WP Compatibility | Phrases/Idioms | Word Handling |
| | Dictionary Browsing | |
| | Related Terms | |

| |
|------------------------|
| Paper, Raw Output |
| Paper, Edited |
| Electronic, Raw Output |
| Electronic, Edited |



= **SYSTRAN**
= Capabilities Only



= **GLOBALINK**
= Capabilities Only



= **PC-TRANSLATOR**
= Capabilities Only



= **STYLUS**
= Capabilities Only



= Capabilities
= Common to
= Two or More

5.0 Table of Proposed Configurations and Costs

Table 3 illustrates the proposed configurations and their associated costs for hardware, software and peripherals.

| SOFTWARE | | | | | | | | | | | | |
|---------------------------------------|--------------------|--------------------|--|--------------------------|-------|--------------------|---------------|--|--|--|---|----------------------------------|
| Systems | Processor | RAM | Hard Disk Space | | | Monitor | Scanner | Other | Operating System | Primary Languages | Other | Total Costs |
| | | | System | Dictionaries | OCR | | | | | | | |
| SYSTRAN | PS/2 or 4M OCR | 16MB | 2MB | 90 MB | 10 MB | VGA | HPScanjet IIP | Windows Mouse IBM Magneto Optical Disk IBM P/370 Card | OS/2 [Windows] | Russian German French Spanish | 3270 Emulation WP Filter, Tiger OCR Interface Filter, Cyrillic Fonts IBM Commun. Mgr. | \$13,380.51 |
| [COSTS] | \$3,267.64 | \$297.88 | (below) | | | (with PS/2) | (below) | \$89.86 \$1,514.39 \$6,966.54 | \$142.97 \$458.91 | No cost | | \$165.87 \$200.00 \$276.45 |
| Globalink | [same] | 2MB | 12 MB | 88 MB | 10 MB | VGA | HPScanjet IIP | Perceive OCR Software | DOS 5.0 [Windows] | (bi-direct.) Russian German French Spanish | Subject Dictionaries 3 French 2 German 3 Spanish 1 Russian | \$8,186.00 |
| [COSTS] | (included in PS/2) | (included in PS/2) | \$1,000.00 (Cost for additional 400 | | | (with PS/2) | \$1,342.00 | \$339.00 | (Available) | \$3,814.00 | | \$1,691.00 |
| PC-Translator | [same] | 4MB | 4MB | [integrated into system] | 10 MB | VGA | HPScanjet IIP | | DOS 5.0 [No Windows] | Italian | N/A | \$8,186.00 |
| [COSTS] | | | (below) | | | | | | (Available) | \$985.00 | | \$985.00 |
| STYLUS | [same] | | 2.3 MB | [Integrated with system] | | VGA | HPScanjet IIP | N/A | N/A | Russian | N/A | |
| [COSTS] | (included in PS/2) | (included in PS/2) | | | | (included in PS/2) | (above) | | No cost | | | \$0.00 |
| Requirements & Recommended Quantities | PS/2 | 16 MB | 527 | | | VGA | HPScanjet IIP | Windows Mouse IBM Magneto Optica Disk [120M] IBM P/370 Card | OS/2 [Windows] & DOS 5.0 partition | Russian German French Spanish Italian | 3270 Emulation Interface Filter WP Filters HP LaserJet Cyrillic Fonts | TOTAL COST \$22,551.51 |
| | | 32 MB | PS/2 Quote includes 400 MB Additional 400 recommended | | | | | | | | | |

Table 3. Resource Requirements and Costs for Proposed Machine Translation Configurations