

MACHINE TRANSLATION; PAST, PRESENT. AND FUTURE

S.-C. Loh (*Chinese University of Hong Kong*)

\* \* \* \*

Abstract

This article discusses the need for machine translation. It also describes briefly the history of machine translation from its beginnings to the present and its future development as seen by the author. As the article was written in Chinese and translated into English by means of a language translator called CULT, the quality of the translation may therefore provide some indication as to the capability of the machine translation system which has been developed.

\* \*

THE NEED FOR MACHINE TRANSLATION

THE RATE AT WHICH MAN HAS CONTINUOUSLY BEEN ACCUMULATING USEFUL KNOWLEDGE ABOUT HIMSELF AND THE ENVIRONMENT HAS ACCELERATED GREATLY, PARTICULARLY DURING THE LAST 100 YEARS. THERE ARE MANY REASONS FOR THIS INCREASE : THERE ARE MORE SCIENTISTS TO-DAY THAN EVER BEFORE, SCIENTIFIC RESEARCH IS RECEIVING SUBSTANTIAL GRANTS AND FINANCIAL SUPPORT FROM GOVERNMENT AND INDUSTRY, THE COLD WAR HAS RESULTED IN A COMPETITIVE ATMOSPHERE IN WHICH NATIONS ARE ENDEAVOURING FOR MILITARY AND TECHNOLOGICAL SUPREMACY, ETC.

BY THE MOST CONSERVATIVE ESTIMATES, EUROPE WAS PUBLISHING 1000 NEW BOOKS PER YEAR BEFORE 1500. BY 1950, FOUR AND A HALF CENTURIES LATER, THE RATE HAS ACCELERATED SO SHARPLY THAT EUROPE WAS PUBLISHING 120,000 NEW BOOKS PER YEAR. BY THE MID-SEVENTIES, THE PUBLICATION OF BOOKS ON A WORLD SCALE, EUROPE INCLUDED, REACHED THE FIGURE OF APPROXIMATELY 1.500 NEW BOOKS PER DAY.

THERE ARE MORE SCIENTIFIC SOCIETIES IN EXISTENCE TODAY THAN EVER BEFORE. AND THESE SOCIETIES ARE HOLDING MORE MEETINGS. CONVENTIONS AND SEMINARS, AND PUBLISH MORE JOURNALS AND PROCEEDINGS THAN EVER BEFORE. THE FIRST TWO SCIENTIFIC JOURNALS APPEARED 300 YEARS AGO. IN 1958, THE NUMBER OF RESPONSIBLE TECHNICAL JOURNALS APPROACHED 100,000; THERE ARE NOW FOUR MILLION JOURNAL ARTICLES, 120,000 TECHNICAL BOOKS AND 100,000 TECHNICAL REPORTS BEING PUBLISHED ANNUALLY ( ON A WORLDWIDE BASIS APPROXIMATELY 60.000,000 PAGES EVERY YEAR ). AND THIS RATE OF PUBLICATION IS INCREASING EVERY YEAR. TO-DAY, FOR EXAMPLE, THE NUMBER OF SCIENTIFIC AND TECHNOLOGICAL JOURNALS AND ARTICLES IS DOUBLING, LIKE INDUSTRIAL PRODUCTION IN THE INDUSTRIALISED COUNTRIES, ABOUT EVERY FIFTEEN YEARS.

HOWEVER, WE CAN ARGUE THAT EVERY BOOK IS A GAIN FOR THE ADVANCEMENT OF KNOWLEDGE. NEVERTHELESS, IT IS TRUE THAT THE ACCELERATION RATE OF BOOK PUBLICATION IS, IN FACT, PARALLEL TO THE RATE AT WHICH MAN DISCOVERS NEW KNOWLEDGE.

THE RESULT OF ALL THESE PUBLICATION ACTIVITIES IS THAT THE INDIVIDUAL SCIENTIST IS STRUGGLING IN AN OCEAN OF INFORMATION AND IS ENDEAVOURING TO SEEK THE INFORMATION HE NEEDS. FURTHERMORE. ALL JOURNALS AND BOOKS ARE PUBLISHED IN DIFFERENT NATIONAL LANGUAGES. BECAUSE NO ONE COUNTRY CAN MONOPOLIZE THE DEVELOPMENT OF SCIENCE AND TECHNOLOGY. THEN, HOW CAN THE SCIENTIST OF ONE

COUNTRY BE FAMILIAR WITH THE SCIENTIFIC AND TECHNOLOGICAL DEVELOPMENTS AND RESEARCH FINDINGS TAKING PLACE IN OTHER COUNTRIES ? SHOULD WE INSIST THAT THE SCIENTIST BE ABLE TO READ ONE, TWO, OR MORE FOREIGN LANGUAGES, SO HE CAN READ THE CURRENT LITERATURE RELATING TO HIS OWN FIELD, TO SEARCH THE LITERATURE ACCUMULATED IN THE PAST, AND TO BE MADE AWARE OF WORK RELATED TO HIS FIELD SO THAT HE CAN PROCEED CONFIDENTLY IN HIS WORK TO AVOID UNINTENTIONAL DUPLICATION.

THE SO-CALLED " INFORMATION EXPLOSION " HAS ARRIVED. SUITABLE WAYS MUST BE ESTABLISHED TO ASSIST THE SCIENTIST AND RESEARCH WORKER TO READ CURRENT LITERATURE IN THE FIELD RELATING TO HIS WORK, TO SHARE KNOWLEDGE AMONG PEOPLE AND AMONG NATIONS. I BELIEVE THAT MACHINE TRANSLATION OFFERS THE ONLY REALISTIC SOLUTION TO THE PROBLEM WE ARE FACING. A GREAT DEAL OF TIME AND RESOURCES COULD BE SAVED IF DUPLICATION OF RESEARCH EFFORT WERE TO BE ELIMINATED, OR AT LEAST REDUCED TO A MINIMUM BY MUTUAL REPORTING OF PROGRESS AND EXCHANGING IDEAS. THE WORK OF SCIENTISTS SPEAKING DIFFERENT LANGUAGES CAN BE ACCELERATED BY LEARNING WHAT SCIENTISTS IN OTHER COUNTRIES ARE DOING, THEIR ACHIEVEMENTS AND FAILURES, AND THEY CAN PLAN A BETTER COURSE OF ACTION TO SOLVE THEIR OWN PROBLEMS. THE ONLY MEANS BY WHICH THESE SCIENTISTS CAN COMMUNICATE WITH EACH OTHER IS BY PUBLISHING PAPERS OUTLINING THE METHOD USED AND THE RESULTS OBTAINED, AND THESE PAPERS HAVE TO BE TRANSLATED BEFORE THEY BECOME UNDERSTANDABLE TO THE OTHER SCIENTISTS. AUTOMATIC TRANSLATION IS DEFINITELY NEEDED SINCE HUMAN TRANSLATORS CANNOT POSSIBLY HANDLE THE VOLUME OF WORK OR EVEN HOPE TO ACQUIRE THE SPECIAL VOCABULARY TO MAKE GOOD TRANSLATIONS IN A WIDE RANGE OF TECHNICAL SUBJECTS.

AUTOMATIC TRANSLATION CAN NOT BE PERFECT. WHETHER IT COULD EVEN BE HIGH QUALITY OR NOT IS DEPENDENT ON HOW HIGH THE STANDARDS ARE SET. THE IMMEDIATE GOAL IS NOT TO DESIGN A PERFECT AUTOMATIC TRANSLATION SYSTEM OR TO ACHIEVE HIGH QUALITY MACHINE TRANSLATION, BUT TO DESIGN A MACHINE TRANSLATION SYSTEM THAT IS BETTER AND MORE EFFICIENT THAN THE ONES WE HAVE TO-DAY.

#### PAST

THE SUGGESTION THAT MODERN COMPUTING MACHINES COULD BE USED FOR TRANSLATION ORIGINATED FROM A.D. BOOTH. IN 1946, VARIOUS NEW USES FOR AUTOMATIC DIGITAL COMPUTERS WERE BEING CONSIDERED. THESE APPLICATIONS RANGED FROM THE MORE OBVIOUS APPLICATIONS TO PROBLEMS IN MATHEMATICS AND PHYSICS, TO PHILOSOPHICAL PROBLEMS SUCH AS MECHANIZATION OF HUMAN THOUGHT PROCESSES, THE PLAYING OF GAMES AND THE TRANSLATION OF LANGUAGE. AT THIS EARLY STAGE OF AUTOMATIC TRANSLATION, MOST OF THE RESEARCH WORK DONE IN AMERICA WAS FOCUSED ON GERMAN INTO ENGLISH, WHILE IN RUSSIA THEY FOCUSED ON GERMAN INTO RUSSIAN. THIS WAS PROBABLY BECAUSE GERMANY HAD DEVELOPED VERY ADVANCED TECHNOLOGY DURING WORLD WAR II. HOWEVER, ATTENTION WAS SOON DIRECTED TOWARDS RUSSIAN-ENGLISH TRANSLATION IN AMERICA. WITH REGARDS TO THE AUTOMATIC TRANSLATION OF CHINESE, RUSSIA HAD SPENT A LOT OF TIME AND RESOURCES IN THIS RESPECT BEFORE THE AMERICAN REALISED ITS IMPORTANCE.

GENERALLY SPEAKING, MACHINE TRANSLATION HAS PASSED THROUGH THREE PHASES OF DEVELOPMENT. THE FIRST, OR HEURISTIC PERIOD, BEGAN WITH BOOTH'S SUGGESTION IN 1946. DURING THIS PERIOD, BOOTH AND RICHENS PERFORMED EXPERIMENTS ON MICRO-GLOSSARY DESIGN WHILST OTHER RESEARCH GROUPS STUDIED MACHINE PRODUCED GRAMMATICAL RULES TO ASSIST IN TRANSLATION.

THE SECOND PERIOD STARTED AROUND 1950 WHEN ERWING REIFLER ( WASHINGTON UNIVERSITY ) SUGGESTED A LINGUISTIC ANALYSIS AND PROPOSED SOME TRANSLATION

PROCEDURES. SUGGESTIONS ON THE USE OF PRE-EDITORS AND POST-EDITORS AND A LOT OF LINGUISTIC TRICKS TO HELP IN SOLVING MACHINE TRANSLATION PROBLEMS WERE PROPOSED. WITH SUBSTANTIAL GOVERNMENT SUPPORT. THE AMERICAN SCIENTISTS MADE PROGRESS ON A NUMBER OF FRONTS SIMULTANEOUSLY. HOWEVER, THE END OF THE SECOND PERIOD WAS A TIME OF GREAT DISAPPOINTMENT. THERE WERE A LOT OF UNSOLVED PROBLEMS WHICH SEEMED UNLIKELY TO BE SOLVED. MANY GAVE UP HOPE OF ACHIEVING ANYTHING CLOSE TO A PERFECT TRANSLATION AND DIRECTED THEIR EFFORTS TO MACHINE-AIDED TRANSLATION.

IN THE THIRD PERIOD, FROM EARLY 1960'S TO 1966, SOME LIMITED PROGRESS WAS MADE AND THE IMPORTANCE OF LINGUISTIC RESEARCH WAS REALISED. MORE EFFORT WAS DIRECTED TOWARDS RESEARCH INTO SYNTAX AND SEMANTICS IN ORDER TO ESTABLISH MORE PRECISE GRAMMATICAL RULES.

IN 1967, THE AUTOMATIC LANGUAGE PROCESSING ADVISORY COMMITTEE AT THE REQUEST OF DR. HARWORTH, DIRECTOR OF THE NATIONAL SCIENCE FOUNDATION, ANNOUNCED ITS FAMOUS REPORT RECOMMENDING THAT RESEARCH SUPPORT SHOULD BE GIVEN FOR COMPUTATIONAL LINGUISTICS, BUT NOT FOR MACHINE TRANSLATION WHICH, IN THEIR VIEW, WOULD HAVE NO POSSIBILITY OF TANGIBLE SUCCESS IN THE NEAR FUTURE. ALL RESEARCH IN MACHINE TRANSLATION WAS THUS TERMINATED.

#### PRESENT

THE MACHINE TRANSLATION RESEARCH AT THE CHINESE UNIVERSITY OF HONG KONG BEGAN IN 1969. WE CAREFULLY STUDIED AND REVIEWED RESEARCH REPORTS CONCERNING MACHINE TRANSLATION. THE ACCOMPLISHMENTS AND FAILURES OF PREVIOUS WORKS WERE EXAMINED. A NEW APPROACH TO AUTOMATIC TRANSLATION OF LANGUAGE BY COMPUTER WAS PLANNED.

WE RECOGNISED AT THE OUTSET THAT FULLY AUTOMATIC HIGH QUALITY TRANSLATION (FAHQT) IS NOT FEASIBLE. A GREAT DEAL OF RESEARCH WORK, BOTH IN THEORETICAL LINGUISTICS AND COMPUTER PROGRAMMING HAS TO BE DEVELOPED BEFORE THE PROBLEMS OF SPEED, ACCURACY, READABILITY, QUALITY AND ECONOMIC FACTORS OF AUTOMATIC TRANSLATION CAN BE SOLVED. ACCORDING TO BAR-HILLEL, "HIGH QUALITY HAS TO BE RELATIONIZED WITH REGARD TO USERS AND SITUATIONS. THOSE WHO ARE INTERESTED IN MACHINE TRANSLATION PRIMARILY AS PRACTICAL DEVICE MUST REALIZE THAT FULL AUTOMATION OF TRANSLATION PROCESSES IS INCOMPATIBLE WITH HIGH QUALITY. THERE ARE TWO POSSIBLE DIRECTIONS IN WHICH A COMPROMISE COULD BE REACHED; ONE COULD SACRIFICE QUALITY OR ONE COULD REDUCE THE SELF-SUFFICIENCY OF THE MACHINE OUTPUT. " FAHQT MUST THEREFORE BE SACRIFICED FOR THE TIME BEING, AND EFFORT SHOULD BE FOCUSED MORE ON A MACHINE-MAN PARTNERSHIP FOR TRANSLATION.

RESEARCH INTO MACHINE TRANSLATION AT THE CHINESE UNIVERSITY TAKES A DIFFERENT APPROACH THAN THE OTHERS IN THAT THE CHINESE UNIVERSITY OF HONG KONG PLACES A HEAVY EMPHASIS ON PRE-EDITING THE SOURCE TEXT INSTEAD OF POST-EDITING THE TARGET TEXT. IT IS THE ONLY GROUP TAKING THIS APPROACH OF COMPUTER-PRE-EDITOR PARTNERSHIP. ALL THE OTHER GROUPS, WHO REALISED THE FAHQT IS NOT REALLY ATTAINABLE IN THE NEAR FUTURE AND SO A LESS AMBITIOUS AIM IS DEFINITELY INDICATED, HAD ADOPTED A TENDENCY TO COMPROMISE IN FINDING SOME COMPUTER-POST-EDITOR PARTNERSHIP.

ACCORDING TO BAR-HILLEL AGAIN. " THE ONLY REASONABLE AIM, THEN, FOR SHORT-TERM RESEARCH INTO MACHINE TRANSLATION SEEMS TO BE THAT OF FINDING SOME MACHINE-POST-EDITOR PARTNERSHIP THAT WOULD BE COMMERCIALY VIABLE WITH EXISITING HUMAN TRANSLATION AND THEN TO TRY TO IMPROVE THE COMMERCIAL EFFECTIVENESS OF THIS PARTNERSHIP BY IMPROVING THE PROGRAMMING IN ORDER TO DELEGATE TO THE

COMPUTER MORE AND MORE OPERATIONS IN THE TOTAL TRANSLATION PROCESS WHICH IT CAN PERFORM MORE EFFECTIVELY WITH A POST-EDITOR. THESE IMPROVEMENTS WILL, OF COURSE, UTILIZE NOT ONLY DEVELOPMENTS IN HARDWARE, PROGRAMMING AND LINGUISTIC ANALYSIS, BUT ALSO THE EXPERIENCE GAINED BY ANALYSING THE MACHINE OUTPUT ITSELF. THE RESEARCHERS AT CUHK FEEL STRONGLY THAT IF " POST-EDITORS " MENTIONED ABOVE ARE CHANGED TO " PRE-EDITORS ", THEN THE SEMANTIC AND LEXICAL ANALYSIS WILL BE MUCH SIMPLER AND EFFECTIVE. IN THEORY, IF WE CAN PRE-EDIT THE SOURCE LANGUAGE TO MAKE IT AS COMPATIBLE TO THE STRUCTURE OF THE TARGET LANGUAGE AS POSSIBLE AND THEN PROGRAM A SET OF RULES FOR ANALYSIS AND SYNTHESIS, THEN THE TRANSLATED TEXTS OBTAINED WILL BE READABLE AND ACCURATE. TO QUOTE A.G. OETTINGER : " BEFORE MACHINE INSTRUCTIONS FOR TRANSLATION CAN BE FORMULATED, THE SIGNIFICANT PATTERNS OF THE SOURCE LANGUAGE AND TARGET LANGUAGE MUST BE PRECISELY IDENTIFIED AND PUT INTO CORRESPONDENCE . . . CORRESPONDING PATTERNS, THEREFORE, MUST BE DEFINED AS CONVEYORS OF EQUIVALENT MEANINGS SINCE, WHATEVER MEANING IS, IT IS GENERALLY AGREED THAT IT MUST BE PRESERVED IN TRANSLATION. "

A FIXED SET OF PRE-EDITING RULES MUST THEREFORE BE FORMULATED TO ENABLE INEXPERIENCED AND EVEN MONO-LINGUAL PEOPLE TO TRANSFORM QUICKLY THE INPUT INTO MACHINE-TRANSLATABLE FORM. WITH THIS ARRANGEMENT, POST-EDITING CAN BE KEPT TO A MINIMUM, IF NOT ALL TOGETHER ELIMINATED. GIVEN TIME AND BETTER PROGRAMMING TECHNIQUES, THESE PRE-EDITING RULES WILL GRADUALLY BE REDUCED SO THAT THE COMPUTER WILL EVENTUALLY TAKE UP THIS ROUTINE WORK. PRE-EDITING CAN THEREFORE SOLVE MANY OF THE PRESENT LINGUISTIC PROBLEMS THAT ARE OTHERWISE DEPENDENT ON FURTHER RESEARCH IN NATURAL LANGUAGE, COMPUTATIONAL LINGUISTICS AND TRANSFORMATION MATHEMATICS. IN OTHER WORDS, MODELS THAT ARE MUCH MORE COMPREHENSIVE AND SOPHISTICATED THAN THE PRESENT ONES HAVE TO BE DESIGNED. THESE MODELS MAY TAKE YEARS TO PERFECT AND, AT PRESENT, PRE-EDITING IS ABSOLUTELY ESSENTIAL IN ORDER TO ACHIEVE THE GOAL. IN PRESENT STAGE OF OUR DEVELOPMENT, VERY COMPLEX SENTENCES CAN BE TRANSLATED WITH THE AID OF PRE-EDITING. A SENTENCE WHICH HAS A COMPLICATED STRUCTURE CAN BE ANALYSED BY THE EXISTING PROGRAM IF IT IS BROKEN UP INTO SIMPLER SENTENCES WHICH ARE THEN READILY TRANSLATED BY THE COMPUTER. SENTENCES IN CHINESE ARE OFTEN WITHOUT VERBS OR SUBJECTS AND PRE-EDITING CAN ADD THE VERBS OR SUBJECTS SO THAT THESE SENTENCES CAN THUS BE ANALYSED AND TRANSLATED.

CULT ( CHINESE UNIVERSITY LANGUAGE TRANSLATOR ) WAS DEVELOPED BASED ON THE PRINCIPLE MENTIONED ABOVE AND HAS BEEN RIGOROUSLY EXAMINED AND TESTED. THE TRANSLATION ALGORITHM IS GIVEN IN FIG. 1. SINCE THE BEGINNING OF THIS YEAR, THE CULT SYSTEM HAS BEEN USED ON A REGULAR BASIS TO TRANSLATE TWO CHINESE SCIENTIFIC JOURNALS, ACTA MATHEMATICA SINICA AND. ACTA PHYSICA SINICA, WHICH ARE PUBLISHED BY THE PEKING ACADEMY OF SCIENCE. THIS ACCOMPLISHMENT INDICATE THE CORRECTNESS OF OUR APPROACH AND THE POTENTIAL CAPABILITY OF CULT.

#### FUTURE

IN THE NEXT FIVE TO TEN YEARS, DEVICES MIGHT BE AVAILABLE AT A REASONABLE COST FOR INPUTTING AND OUTPUTTING CHINESE OR OTHER NON-ALPHABETIC CHARACTERS. THIS IS, AT PRESENT, THE PRINCIPAL PROBLEM FACING THE PROCESSING OF LANGUAGES OTHER THAN THOSE USING ALPHABETS.

MORE LINGUISTIC RESEARCH APPLICABLE TO MACHINE TRANSLATION MAY YIELD USEFUL RESULTS SO THAT MORE COMPREHENSIVE GRAMMATICAL RULES AND SENTENCE STRUCTURES BETWEEN LANGUAGES CAN BE FORMULATED

AT THE CHINESE UNIVERSITY, WE ARE AT THE PRESENT WORKING ON A LANGUAGE TRANSLATOR WHICH WILL ENABLE US TO TRANSLATE CHINESE INTO ENGLISH AS WELL AS

TO TRANSLATE ENGLISH INTO CHINESE. A SIMULATION STUDY HAS BEEN DONE WHICH INDICATES THE POSSIBILITY OF SUCCESS. WE ARE HOPING THAT, IN A FEW MONTHS, A MODEL MAY BE DESIGNED TO DEMONSTRATE THE SYSTEM. IF IT IS SUCCESSFUL, WE ARE OF THE OPINION THAT A UNIVERSAL TRANSLATOR MAY BE DESIGNED SO THAT ONLY ONE TRANSLATOR WILL BE REQUIRED TO DO TRANSLATION BETWEEN LANGUAGES.

" ASK NOT WHAT THE COMPUTER CAN DO FOR YOU. ASK WHAT YOU, WITH THE AID OF THE COMPUTER, CAN DO FOR THE BETTERMENT OF THE MANKIND. "

### TRANSLATION ALGORITHM

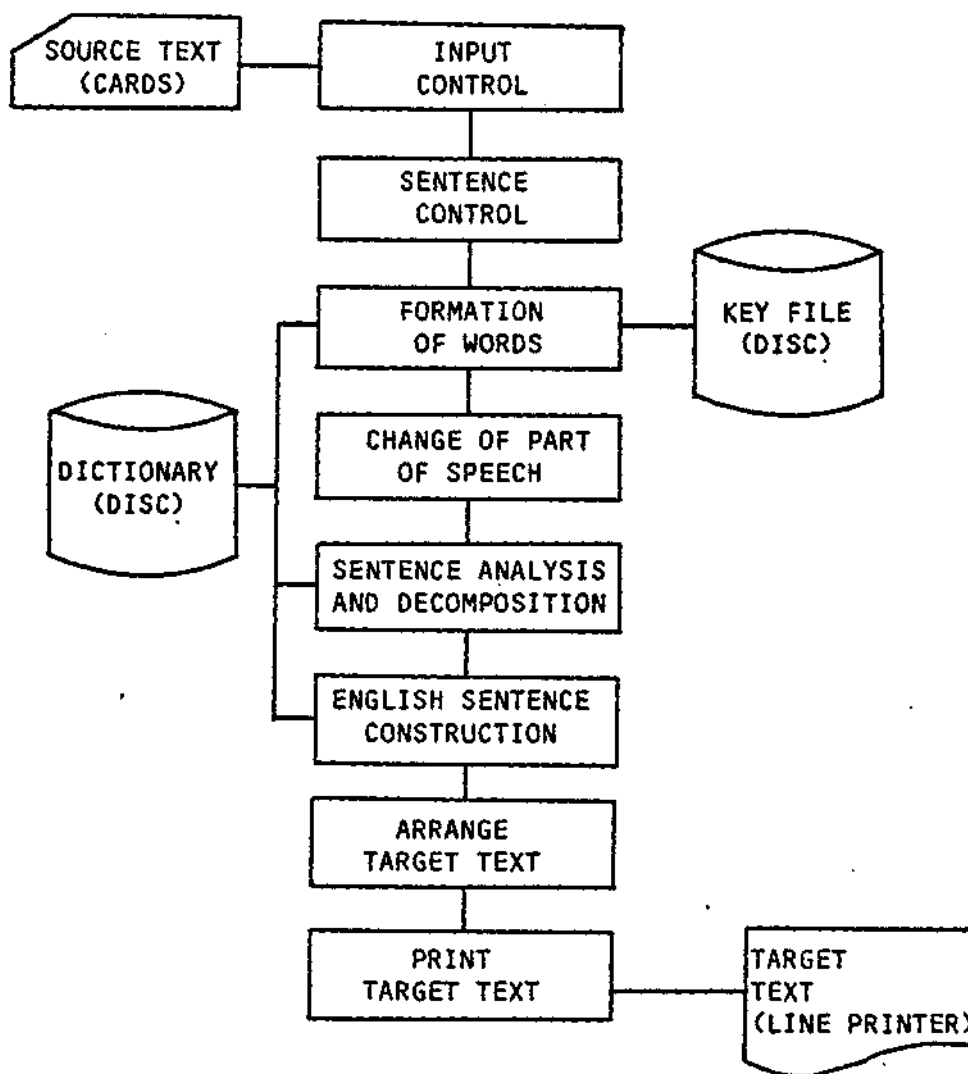


FIG. 1