



DEUTSCHE DEMOKRATISCHE REPUBLIK
AMT FÜR ERFINDUNGS- UND PATENTWESEN

PATENTSCHRIFT 143 667

Ausschließungspatent

Erteilt gemäß § 5 Absatz 1 des Änderungsgesetzes zum Patentgesetz

In der vom Anmelder eingereichten Fassung veröffentlicht

Int. Cl.³

(11) 143 667 (44) 03.09.80 3(51) G 06 F 15/38
(21) AP G 06 F / 211 497 (22) 09.03.79

-
- (71) siehe (73)
(72) Toma, Peter, Dr., US
(73) Systran Institut, Gesellschaft für Forschung und Entwicklung
maschineller Sprachübersetzungssysteme mbH, München, DE
(74) Internationales Patentbüro Berlin, 1020 Berlin, Wallstraße 23/24
-

- (54) Verfahren zur Verwendung eines Digitalcomputers zur
Übersetzung natürlicher Sprachen
-

(57) Ziel und Aufgabe der Erfindung bestehen darin, ein Verfahren der oben genannten Art anzugeben, das es ermöglicht, eine genaue und schnelle, fast augenblickliche Übersetzung eines großen Textvolumens von einer Ursprungssprache, d.h. einer natürlichen Sprache, in eine andere natürliche Sprache, die sogenannte Zielsprache, durchzuführen. Zur Lösung dieser Aufgabe umfaßt die vorliegende Erfindung folgende Schritte: Übertragen des Ursprungstextes von visuellen Zeichen in computerverständliche Binärsignale; Speichern des übertragenen Ursprungstextes in einen Hauptspeicher; Abtasten und Vergleichen der übertragenen Ursprungstextworte mit einem in einem Speicher gespeicherten Wörterbuch der Ursprungssprache und für jedes Ursprungswort; Speichern jedes Wortes in einer Datei im Hauptspeicher und in Verbindung mit jedem dieser Worte einer kodierten Information; Analysieren des übertragenen Ursprungstextwortes in die Datei der Worte, einen kompletten Satz auf einmal und dessen Übertragung in einen Satz der Zielsprache unter Verwendung der kodierten Information.

Verfahren zur Verwendung eines Digitalcomputers zur Übersetzung natürlicher Sprachen

Anwendungsgebiet der Erfindung

Die Erfindung betrifft ein Verfahren zur Verwendung eines Digitalcomputers für die Übersetzung zwischen natürlichen Sprachen.

Charakteristik der bekannten technischen Lösungen

Versuche wurden unternommen, um einen Digitalcomputer für die Übersetzung von einer Sprache in eine andere zu verwenden, d.h. von einer Ursprungssprache in eine Zielsprache. Das Übersetzungssystem verwendet ein programmierbares Digitalcomputersystem mit einem Programm für die Ausführung der Übersetzung. Die verwendeten Methoden waren theoretisch. Die theoretische Sprachmethode zur Syntax-Analyse war nicht akzeptabel, weil sie von einer sprachwissenschaftlichen Annahme ausging, anstatt die Kapazität eines programmierbaren Digitalcomputers zu bedenken und die Übersetzung vom Standpunkt des Computers aus anzugehen. Die Idee einer Maschinenübersetzung wurde erstmals 1946 durch Warren Weaver und A.D. Booth angegeben. Viele Versuche, ein Maschinensystem zu erreichen und es in Anwendung zu bringen, wurden unternommen. Die Projekte

waren auf die Entwicklung einer sprachwissenschaftlichen Theorie gerichtet, welche die gesamte natürliche Sprache umfaßt, die dann auf den Computer übertragen wurde. Diese Methode schlug unvermeidlich fehl, weil der menschliche Verstand nicht die Gesamtheit der Sprache umfassen kann.

Nachstehend folgt eine kurze Zusammenfassung der Methoden und der zugehörigen Theorien.

Die "General Analysis Technique"-GAT - (Instituts of Languages and Linguistics Machine Translation Research Center, Washington DC Juni 1959, von Peter Toma, herausgegeben von J.A. Mayne) verwendet ein sprachwissenschaftlich orientiertes Computer-Codierschema. Ad hoc-Lösungen wurden eingeführt, und das ganze System war innerhalb eines stark beschränkten sogenannten Simulated-Linguistic-Computer-(SLC)Systems, geschrieben in oktaler Schreibweise, begrenzt und war nicht offen für irgendwelche weiteren Verbesserungen.

Die Stützpunkttheoriemethode, entwickelt von 1959 bis 1967 von der "Bunker-Ramo Corporation", USA, war gerichtet auf die Lösung der Probleme, die in einem begrenzten russischen Text auftraten, mittels eines relativ kleinen Wörterbuches. Es wurde kein Versuch dabei unternommen, eine Lösung für Mehrfachbedeutungen einzuführen. Vielmehr wurden die verschiedenen Bedeutungen auf der Ausgabe getrennt gedruckt. Ein prädikatives Syntaxsystem wurde von National Bureau of Standards und dem Massachusetts Institute of Technology in den Jahren 1960 bis 1964 entwickelt. Diese Methode schlug fehl, weil sie nur einen begrenzten Weg zu einem Satz vorsah. Dieses System wurde niemals in größerem Maßstab angewendet, wurde aber nur innerhalb eines begrenzten Experimentierumfeldes verwendet.

Eine Transformationsgrammatik war ein weiterer Versuch. Doch diese Methode stellte sich als absolut inkompatibel mit den Computerübersetzungsanforderungen heraus. Auf der Basis dieser Theorie wurden nur kleine Experimentiersysteme entwickelt, und diese mußten aufgegeben werden, bevor irgendeine signifikante Übersetzung entstanden war.

Ziel der Erfindung

Ziel der Erfindung ist es, die vorgenannten Nachteile des Standes der Technik zu vermeiden.

Darlegung des Wesens der Erfindung

Der Erfindung liegt die Aufgabe zugrunde, ein Verfahren zur Verwendung eines Digitalcomputers für die Sprachübersetzung anzugeben, das es ermöglicht, eine genaue und schnelle, fast augenblickliche Übersetzung eines großen Textvolumens von einer Ursprungssprache, d.h. einer natürlichen Sprache, in eine andere natürliche Sprache, die sogenannte Zielsprache, durchzuführen. Anstatt einer natürlichen Sprache kann eine künstliche Sprache, wie z.B. Esperanto, als Ursprungs- oder Zielsprache fungieren. Um diese Aufgabe zu lösen, besteht das Wesen der vorliegenden Erfindung in einem Verfahren, welches ein programmierbares Digitalcomputersystem verwendet, welches folgende Schritte umfaßt: Übertragen des zu übersetzenden Ursprungstextes visuellen Zeichen in computerverständliche Binärsignale: Speichern des übertragenen Ursprungstextes in einen Hauptspeicher des Computersystems: Abtasten und Vergleichen der übertragenen Ursprungstextworte mit einem in einem Speicher gespeicherten Wörterbuch für die Wörter der Ursprungssprache und für jedes Ursprungswort, für welches ein Gegenstück gefunden wurde. Speichern jedes Wortes in einer Datei im Hauptspeicher und in Verbindung mit jedem dieser

Worte einer kodierten Information, abgeleitet von dem Wörterbuch für die Verwendung des Wortes in der Übersetzung, wobei die kodierte Information eine Speicheradresseverbindung zu einem Speicher in dem Computersystem aufweist, in welchem Grammatik und Zielsprachübersetzungen für dieses Wort gespeichert sind; Analysieren des übertragenen Ursprungstextwortes in die Datei der Worte, einen kompletten Satz auf einmal, und Übertragen desselben in einen Satz in der Zielsprache unter Verwendung der kodierten Information und Umfassen folgender Schritte: 1) Verwendung der Speicheradresseverbindung zur Erzielung der Zielsprachübersetzung des Wortes von einem Speicher und 2) Reorganisieren der Zielsprachübersetzung in einen geeigneten Zielsprachsatz; Übertragen des Zielsprachsatzes von computerverständlichen Binärsignalen zurück in lesbare Zeichen und Drucken der Zielsprache.

Das Verfahren ist weiterhin durch folgende Schritte gekennzeichnet:

- Speichern eines Wörterbuches von oft verwendeten Ursprungsworten und angeschlossener Zweigadresseverbindungen (offset address linkages); die Zweigadresseverbindungen identifizieren den Speicherplatz für Grammatik und Bedeutungsinformationen für die Ursprungsworte;
- Speichern eines Wörterbuches für wenig verwendete Ursprungsworte in Verbindung mit Grammatikcodebedeutungen für jedes Wort;
- Identifikation des Platzes von Zielsprachgrammatikcodes und Bedeutungen mittels eines Stammwortes für die Ursprungssprache zusammen mit Abzweigverbindungen.
- Vergleichen der Ursprungstextworte mit dem Wörterbuch für oft verwendete Worte und nach Entdecken einer Gleichheit erfolgt Speichern des Wortes und der angeschlossenen Zweig-

adressverbindung zusammen in einem oft verwendeten Speicher;

- Vergleichen der Ursprungstextworte mit dem Wörterbuch für selten verwendete Worte und nach Entdecken einer Gleichheit Speichern des Wortes und der angeschlossenen Grammatikcodebedeutung in einem selten verwendeten Speicher.

Zweite und nachfolgende Worte von Phrasen, falls vorhanden, sind in einer Ursprungsphase zumindest in einem Wörterbuch im Speicher gespeichert. Ein oft verwendetes Wörterbuch umfaßt die ersten Worte einer Phrase der Ursprungssprache, ebenso daran angeschlossene Adressverbindungen zu zweiten und nachfolgenden Worten derselben Phrase, die im Phrasenwörterbuch lokalisiert sind. Das erste Wort und nachfolgende Worte falls vorhanden, einer Phrase, die eine daran anschließende Bedeutung in der Zielsprache hat, ist in Verbindung damit gespeichert und durch folgende, während des Vergleichens mit dem oft verwendeten Wörterbuch, vorgesehene Schritte gekennzeichnet:

- Feststellung jener Worte, die gleich sind, und erste Worte von Phrasen darstellen;
- Benutzung der Abzweigadressverbindung zur Lokalisierung von einem zusätzlichen Wort oder Wörtern derselben Phrase, die im Phrasenwörterbuch lokalisiert sind;
- Vergleichen der festgestellten weiteren Worte in der Phrase mit den Wörtern des Ursprungstextes, der dem ersten Wort der Phrase folgt, auf Gleichheit;
- nach Feststellung einer fertigen Gleichheit Speichern des ersten Phrasenwortes zusammen mit der Zielsprachbedeutung in einen oft benutzten Speicher;
- Speichern eines jeden der zweiten und nachfolgenden Worte der Phrase zusammen mit einer Phrasenmarkierung in den oft benutzten Speicher, wobei die Phrasenmarkierung bestimmt,

- daß dieses Wort nicht ein erstes Wort einer Phrase ist;
- Speichern des ersten Wortes einer Phrase zusammen mit einer Markierung in den oft verwendeten Speicher, wobei die Markierung bestimmt, daß das Wort ein erste Wort einer Phrase ist;
 - Aussuchen der Zielbedeutung während des Analysierens, verbunden mit einem ersten Phrasenwort als Übersetzung des ersten und der nachfolgenden Worte der Phrase.

Das Verfahren ist ferner durch den folgenden Schritt gekennzeichnet: Speichern einer Indizierung in Verbindung mit den Worten im gleichen benutzten Speicher gleich, ob die Worte Hauptworte und/oder Partizipien sind, in begrenzt semantischen Verbindungen, wobei die Verbindungen in eigenen Speichern gespeichert sind und begrenzt semantische Nummern für jedes Hauptwort enthalten. Die begrenzt semantischen Nummern indizieren alle Partizipien, wobei die begrenzt semantischen Nummern der Partizipien mit den Hauptwörtern in derselben Folge verknüpft sind, in der die Partizipien Ausdrucksverbindungen mit dem Hauptwort bilden. Während des Ab tastens und Vergleichens mit den Wörterbüchern wird eine Information mit dem Hauptwort verbunden, welche alle die Ausdrucksverbindungen, innerhalb der das Hauptwort als Schlüsselwort in der Verbindung fungiert, enthält. Die Information enthält entweder Folgen von begrenzt semantischen Nummern oder eine Adresse für einen Hilfsspeicher, in dem die Information lokalisiert ist. Durch weitere Schritte werden die dem Ursprungstext einer "subject field code" zuweisenden begrenzt-semantischen Nummern während des Analysierens der eintreffenden Worte vom Hilfsspeicher in einen Hauptarbeitsspeicher übertragen, wenn die gleiche Ausdrucksverbindung der Ursprungssprache durch die begrenzten semantischen Nummern, die unterschiedliche Zielbedeutungen

haben, angezeigt wird infolge der Auswahl der richtigen Bedeutung für die Verbindung, abhängig von dem "subject field code", in dem die Übersetzung stattfindet.

Der Verfahrensschritt Analysieren umfaßt den Schritt Analysieren des Ursprungstextes und den Schritt Übertragen des Ursprungssprachsatzes in einem speziellen Bereich, welcher dafür im Computerspeicher reserviert ist, wobei sich die Analyse von Anfang bis Ende und vom Anfang bis zum Anfang des Ursprungssprachplatzes bewegt. Dabei werden Markierungen von speziellen Beziehungen der Satzumgebung aufgestellt; die Markierer werden durch Routiner im Computerprogramm aufgestellt, welche unabhängig von unterschiedlichen Teilen der Sprache unter den Wörtern hervorgerufen werden.

Während der Analyse ist die Information gespeichert, um sie erinnernd von einem Ursprungstextsatz zum anderen zu verwenden, um die Übersetzung jener Worte, für die eine Rückbeziehung auf den vorhergehenden Satz notwendig ist, zu ermöglichen. Während der Analyse erfolgen semantische und kontaktische Analysedurchgänge durch jeden Ursprungssatz, wobei eine kodierte Information verknüpft wird. Die Durchgänge erfolgen entweder vom Anfang zum Ende oder vom Ende zum Anfang, abhängig von den Codes, die mit den einzelnen Worten verbunden sind und abhängig von der Charakteristik der Ausgangssprache. Während der Schritte Abtasten, Vergleichen und Speichern werden Teile des Sprachcodes den Ursprungstextworten zugeordnet. Während der Analyse werden die zugewiesenen Teile zum Verzweigen in gewünschten Routinen verwendet, indem der Wert jedes Teiles der Sprache zu einer fixen Adresse addiert wird, um die Analyse auszuführen. Während der Analyse der Ursprungssprachworte hält der Computer eine Spur von vorher

angetroffenen Situationen aufrecht, indem er eine Reihe von Speicherschaltern (memory switches) setzt, wobei jedes Mal, wenn ein neuer Teil der Sprache angetroffen wird, eine Reihe von vorher gesetzten Schaltern selektiv ausgeschaltet werden, während einer oder mehrere zusätzliche Schalter selektiv eingeschaltet werden, mit einer Routine zum Abfragen der Schalter in einer vorher bestimmten Reihenfolge. Auch während der Analyse ist jedes Wort des Satzes mit einer bestimmt kodierten Information versehen, ob es zu einem Hauptsatz oder zu einem Nebensatz gehört. Es ist eine Routine bestimmt für das Aufsuchen von Ursprungswörtern innerhalb derartiger Sätze für Prädikate und Subjekte, wobei nachfolgend während der Analyse die Prädikate und Subjekte der Ursprungssprache von Routinen gesucht und entsprechend mit Codes versehen werden; die Prädikate und Subjekte werden von Routinen innerhalb von in Sätzen enthaltenen Sätzen gesucht. Dabei wird auf Routinen geachtet, die bestimmen, ob die Frage besteht, ob die Funktion eines Kommas oder einer Konjunktion einen neuen Satz beginnt und eigenes Suchen nach Prädikaten und Subjekten durchführen und die Antwort als Ergebnis der Analyse jedes Ursprungswortes in einem Satz mit Marken versehen ist, welche die Abhängigkeitsbeziehungen unter allen Worten des Satzes ausdrücken. Gleichzeitig wird jedem Wort im Satz eine syntaktische Funktion zugewiesen, die in einer verdichteten Form unter Verwendung von "coded bis" ausgedrückt ist. Die Syntaxanalyse wird fortgesetzt, bis jedem Wort im Satz zumindest eine Funktion zugewiesen ist.

Ein lokales Spezialwörterbuch ist gespeichert, das unterschiedliche Zielsprachbedeutungen für ein Ursprungswort in Übereinstimmung mit verschiedenen technischen Themen enthält. In diesem Schritt ist eine Programm-Routine zur Überprüfung jedes Ursprungssatzes in der Analyse und zur Auswahl

von Bedeutungen eines aktuellen Spezialwörterbuches in Übereinstimmung mit dem technischen Sachinhalt des Satzes enthalten. Die Ursprungssprachsätze werden in einer hierarchischen Ordnung analysiert. Im ersten Durchgang werden grundlegende Zweideutigkeiten aufgelöst, in einem zweiten folgenden Durchgang werden primäre syntaktische Zwischenverbindungen aufgestellt und in einem dritten folgenden Durchgang werden diese Verbindungen auf abgegrenzte Phrasen und Sätze ausgedehnt. In einem vierten Prädikat-Subjekt-Durchgang, in dem in hierarchischer Ordnung nach Prädikaten und Subjekten gesucht wird, werden verbleibende Satztrennungs-Zeideutigkeiten aufgelöst. In einem fünften Durchgang wird sichergestellt, daß die Funktion jedes Wortes in dem Satz richtig erkannt wurde. Im Hauptsyntax-Durchgang werden generalisierte Unterprogramme abgerufen, um syntaktische Beziehungen zu bestimmen, und zu kodieren. Daran anschließend werden Unterprogramme zur Ausführung der Basisaufgaben abgerufen.

Ein eigenes Programm bestimmt und analysiert die Präpositionen der Ausgangssprache und bestimmt dabei die den Präpositionen eigene Bedeutung, wobei die Bedeutung der Präpositionen entweder von Worten, welche die Präpositionen regeln, bestimmt wird oder durch Worte, die als Präpositionen fungieren. Dabei wird eine kodierte Information für die Übersetzung von einer oder mehreren Präpositionen der Wörter vom Wörterbuch zugeordnet.

Das Verfahren ist weiterhin durch folgende Schritte gekennzeichnet

- Analysieren jedes Wortes durch mehrfache Durchgänge durch jeden Satz, wobei Codes zugewiesen werden;
- Bedenken all der Codes, welche in vorhergehenden Durchgängen mit einem Wort verknüpft wurden und Zuweisen von Ziel-

- sprach-Synthesecodes, welche mit der Bedeutung verknüpft sind, mit der die Code-Funktion im Satz existiert;
- Setzen des Wortes in eine Form entsprechend der Zielsprache abhängig von der Analyse und dem Inbetrachtziehen aller relevanten und dem Wort zugewiesenen Codes, wobei die übersetzten Worte gebildet nach der Analyse, in einem Druckbereich in einer Folge spezifiziert durch die Beziehung zwischen der Ausgangs- und der Zielsprache und vorher aufgestellt nach den nach jedem Wort gegebenen Nummern und Übertragen der Übersetzung, wenn der Druckbereich aufgefüllt ist, entweder direkt auf einem Drucker oder auf einem temporären Hilfsspeicher.

Vorteilhafterweise umfaßt das Verfahren einen Schritt, bei welchem ein spezielles Unterprogramm des Wörterbuchnachschlagverfahrens eine Liste von Worten (lexikalische Liste) sucht, um zu bestimmen, welche Worte die Verwendung eines speziellen lexikalischen Unterprogramms in der Wörterbuchdatei erfordert, wobei ein getrenntes lexikalisches Steuerprogramm die Worte des Ursprungssprachsatzes analysiert, für jene Fälle, bei denen nur das Ergebnis einer syntaktischen-semanticen Analyse des Satzes oder der Mitgliedschaft eines Wortes in einer grammatikalischen oder semanticen Klasse bestimmen kann, ob oder ob nicht ein lexikalisches Unterprogramm hereingeholt werden muß, zum Zeitpunkt der Übersetzung eines speziellen Wortes oder Ausdruckes, um die Bedeutung des Wortes oder Ausdruckes durch Überprüfung der syntaktischen Beziehungen, welche für dieses Wort oder für diesen Ausdruck etabliert wurden, zu bestimmen unter Verwendung der syntaktischen oder semanticen Regeln, welche nur diesem Wort oder dieser Klasse von Wörtern, zu welchen es gehört, zukommt.

Weiterhin ist vorteilhafterweise ein Verfahren zur Lösung semantischer Zweideutigkeiten vorgesehen, wobei jedem Wort im Stammwörterbuch (Einzelwort) eine eigene begrenzte Semantikzahl (LS) zugewiesen ist und die speziellen Mehrfachausdrücke des LS-Verbundwörterbuches (LS-Ausdrücke), welches von den LS-Nummern-Darstellungen ihrer individuellen Wörter im Stammwörterbuch gebildet wird, werden dann automatisch in einen Wörterbuchblock gruppiert, gemäß dem Hauptwort des Ausdruckes, wobei einer für jedes eigene Hauptwort existiert. Der Block wird dann gesucht, wann immer das Hauptwort in einem zu übersetzenden Satz auftritt, um festzustellen, ob ein Gegenstück zwischen einer Gruppe von benachbarten Worten im Text und einer der Ausdrücke in dem Block existiert. Das längste Gegenstück oder der Ausdruck mit der höchsten Priorität wird verwendet, um die spezielle Bedeutung dieser Wortgruppe zu bestimmen, welche von der Summe der Bedeutungen der Einzelworte sich unterscheidet. Eine Teilmenge eines LS-Ausdrucks ist der bedingt begrenzte semantische Ausdruck (CLS), welcher den Einschluß aller definierbaren syntaktischen und semantischen Regeln erlaubt, als auch einfache Programmierinstruktionen, welche dazu verwendet werden können, die in den Bits und Bytes des Satzanalysenbereiches gespeicherten Informationen auszuwechseln, u.zw. im LS-Ausdruck selbst (d.h. im Wörterbuch), so daß der Wörterbuchausdruck nur dann verglichen wird, wenn die syntaktischen und/oder semantischen Bedingungen, ausgedrückt in den Regeln, von den Worten im Text getroffen werden. Dieser Umstand kann nur dann bestimmt werden, wenn eine vollständige syntaktische und semantische Analyse des Satzes durch die sprachwissenschaftlichen Analyseprogramme ausgeführt worden ist.

Bei dem Verfahren können vorteilhafterweise alle Worte der Ursprungssprache mit semantischen Kategoriecodes variabler

Länge ausgestattet werden, welche dann durch die Analyseprogramme für die Ursprungssprache abgefragt werden, ferner durch die lexikalischen Routinen und die CLS-Wörterbuchnachschragsroutinen als Hilfe bei der Lösung semantischer Zweideutigkeiten. Diese Codes sind in einer Hierarchie (Satz von Baumstrukturen) in der Weise ausgedrückt, daß das Kodieren einer Kategorie, welche auf einem niedrigeren Niveau eines semantischen Baumes existiert, automatisch die Kodierung mit geeigneten Codes eines höheren Niveaus durch das System selbst erzeugt.

Eine Folge von Wörtern wird in der Ausgangssprache innerhalb von Phrasen und Sätzen in Beziehung zur Wortfolge der Zielsprache analysiert, wobei die Wortfolge der Zielsprache symbolisch durch zugewiesene Nummern ausgedrückt wird. Beim Drucken der Übersetzung wird auf jedes Ursprungswort Bedacht genommen.

Zusammenfassend beschreibt die Erfindung ein Verfahren für den Betrieb eines Digitalcomputers zur Übersetzung von einer natürlichen Ursprungssprache, z.B. Russisch, in eine natürliche Zielsprache, z.B. Englisch. Das Verfahren umfaßt drei Phasen. Die Wörterbuchnachschragsphase etabliert die Zielsprachbedeutung jedes Wortes oder Ausdruckes im Ursprungstext. Die Wörterbuchnachschragsphase verknüpft grammatikalische Codes und Zielsprachäquivalente mit jedem Wort und Ausdruck in der Ursprungssprache. Die syntaktische Analysephase identifiziert syntaktische Informationen auf der Basis der mit den Worten und Ausdrücken verknüpften grammatikalischen Information und macht sich ebenfalls die Beugung und Stellung des Wortes im Ursprungstext zu Nutze. Die Synthesephase nimmt die Bedeutung und die syntaktische Information aller Worte

des Satzes im Ursprungstext und formt einen Satz in der Zielsprache.

Mehr im einzelnen beginnt das Verfahren mit dem Eingeben des Ursprungstextes in den Speicher eines Computers. Jedes Ursprungswort wird dann transformiert in ein umgewandeltes Ursprungstextwort. Das umgewandelte Ursprungstextwort besteht aus dem Ursprungstextwort und kodierter Information. Die kodierte Information kann eine Speicheradreßverbindung umfassen, welche den Zugriff zu einem Speicherplatz vorsieht, der eine syntaktische Information und die Übersetzung des Ursprungstextwortes enthält. Die umgewandelten Ursprungstextwörter, welche sich vom Ursprungstextsatz ableiten, werden dann in eine Zielsprachübersetzung des Satzes synthetisiert. Die Synthese etabliert beides korrekt: die Wortbedeutung und die Wortstellung im Zielsprachsatz.

Ein wesentliches Merkmal der Erfindung ist die getrennte Behandlung von häufig verwendeten Wörtern gegenüber selten verwendeten Wörtern. Um die effektive Kapazität des Kernspeichers des Computers zu maximieren, tragen die selten verwendeten Wörter ihre Übersetzungsinformation mit sich, während jedes der viel verwendeten Wörter eine Speicheradreßverbindung trägt, welche einen leichten Zugriff zur Übersetzungsinformation erlaubt, welche im Kernspeicher gespeichert ist. Auf diese Weise wird die Übersetzungsinformation für häufig verwendete Wörter auf einem leichter zugreifbaren Platz im Computer gehalten, mehr als dies während jedes Auftretens des Wortes für die selten verwendeten Wörter geschieht.

Während die obige Beschreibung eine menschliche Analogie der beanspruchten Funktionen darstellt, versteht es sich von

selbst, daß die tatsächliche Durchführung des Verfahrens mittels des Computers deutlich unterschiedlich ist. Von dem Zeitpunkt an, von dem der Ursprungstext in maschinenlesbare Eingabedaten übersetzt ist, bis zu dem Zeitpunkt, bei dem die maschinenlesbaren Ausgabedaten in einen menschenlesbaren Übersetzungstext umgewandelt sind, schreitet das beanspruchte Verfahren unter der Kontrolle eines Computerprogrammes fort. Obwohl es zweckmäßig ist, die Schritte des Programmes so zu beschreiben, als wenn sie von einem menschlichen Übersetzer ausgeführt würden, geschieht tatsächlich nichts dergleichen. Eher führt der Computer, eine Reihe gedankenloser abstrakter mathematischer Operationen auf den abstrakten Werten, die im Speicher des Computers gespeichert sind, aus. Das Programm funktioniert unabhängig von der Bedeutung oder Wichtigkeit der Daten, an denen es ausgeführt wird. Die Tatsache, daß das Programm in einer hochorganisierten Programmiersprache geschrieben ist, welche das Programm scheinbar bezüglich der Maschinenoperation verdeutlicht, ändert nichts an der Tatsache, daß die Maschine tatsächlich eine Reihe von abstrakten Schritten ausführt, welche nichts mit der Übersetzung zwischen zwei natürlichen Sprachen zu tun hat. Wenn eine unterschiedliche Art von Information in den Computer eingegeben wird, so kann das in der vorliegenden Erfindung verwendete Programm begreiflicherweise eine von der Übersetzung völlig unterschiedliche Funktion durchführen. Die vorliegende Erfindung umfaßt auch jenes Verfahren, mittels welchem Informationen vom Computer ausgegeben werden, inklusive des Ausdrucks der Übersetzung, d.h. der Schritt des Übersetzens des Zielsprachsatzes von computerlesbaren Binärsignalen zurück in lesbare Zeichen. Es ist ein großer Vorteil, daß mittels der vorliegenden Erfindung irgendeine Ausgangssprache in eine Zielsprache übersetzt werden kann,

indem das vorgeschlagene Verfahren und dieselbe Ausrüstung verwendet wird, wobei nur die sprachabhängigen Teile der Ausrüstung, wie z.B. Wörterbücher und sprachwissenschaftliche Programme, welche die gesamten syntaktischen und semantischen Regeln beider Sprachen enthält, ausgewechselt werden.

Ein weiterer Vorteil des erfindungsgemäßen Verfahrens besteht in der Folgerichtigkeit der Übersetzung, d.h., ein besonderer Ausdruck wird von der Ausgangssprache in die Zielsprache mit derselben Umgebung übersetzt. Wenn die Umgebung sich ändert, verändert die vorliegende Erfindung übereinstimmend die Bedeutung des Ausdruckes mit der oben genannten Folgerichtigkeit. Ebenso ist das Wiedererkennen aller syntaktischen und semantischen Eigenschaften der Ursprungssprache, welche in der Zielsprache wiedergegeben werden, von Vorteil.

Ausführungsbeispiel

Nachstehend werden spezielle Ausführungsformen der Erfindung beispielsweise anhand der Zeichnungen näher erläutert. Es zeigen:

Fig. 1: die Illustration des Übersetzungsvorganges,

Fig. 2: schematisch das Wörterbuchnachschiagen und

Fig. 3: schematisch die Zusammenfassung der Programme.

Wie in Fig. 1 gezeigt, werden Texte dem Übersetzungssystem entweder on-line über Terminals oder off-line über Platten oder Bänder eingegeben. (Eingabe über Lochkarten ist ebenfalls möglich) Der eingegebene Text und die Übersetzungswörterbücher und die Programme befinden sich auf Plattenspeichern, auf welche der Computer, der die Übersetzung

durchführt, zurückgreifen kann. Die Ergebnisse werden dann im gewünschten Format am Drucker ausgedruckt, wobei irgendwelche Formatcodes im Eingabetext berücksichtigt werden.

Wenn gewünscht, kann das nachträgliche Ausgeben des Ausgabetextes und irgendein sich ergebender Wörterbuchänderungsdienst über ein on-line-Terminal erfolgen (oder es kann auch off-line durch nachträgliche Ausgeber und Wörterbuchcodes erfolgen).

Diagramme und Daten, welche in derselben Form in der übersetzten Ausgabe enthalten sein sollen, können in den Text über eine Lichtsetzvorrichtung eingefügt werden. Die Fig. 2 zeigt ein Programm, welches den Eingabetext in den Computerspeicher (LOADTXT) eingibt, die häufig gebrauchten Wörter identifiziert und ebenso die Phrasen im Text, und welches jedem Wort im Text eine eigene Seriennummer zuteilt. Es verbindet unmittelbar eine Übersetzung mit den Phrasen und stellt sicher, daß die Information für alle oft gebrauchten Wörter sofort im Kernspeicher des Computers während des Übersetzungsprozesses verfügbar ist. Die verbleibenden (wenig verwendeten) Worte werden dann alphabetisch sortiert, um den Prozeß des Nachschlagens im Hauptwörterbuch (MDL) zu erleichtern. Das Hauptwörterbuchnachschiagen verknüpft mit jedem wenig verwendeten Wort die notwendige grammatikalische Information ebenso wie eine Querbeziehung zum Block der Vielwort-begrenzt-semanticen Verbindungen (LS), die dieses Wort enthalten. Nach dem Hauptwörterbuchnachschiagen werden alle Worte im Text in ihre ursprüngliche Reihenfolge zurücksortiert. Gemäß Fig. 3 kontrolliert das Programm IMITCALL den momentanen Übersetzungsvorgang. Es ruft zuerst das Programm GETSEWTW ab, welches einen Satzanalysebereich im Computerspeicher etabliert, bestehend aus Informationen von 160

Byts fester Länge für jedes Wort, mit Querverbindungen zu zusätzlichen Bereichen für Informationen variabler Länge. Zuerst enthält der Satzanalysebereich nur die Wörterbuchinformation für jedes Wort. Während des Übersetzungsvorgangs werden zusätzliche Informationen durch das Ausgangsspracheanalyseprogramm (bezeichnet als "passes") hinzugefügt, welches die syntaktischen und semantischen Zweideutigkeiten auflöst, Satzgrenzen etabliert, syntaktische Grundbeziehungen zwischen Wörtern beschreibt, Subjektprädikatbeziehungen identifiziert und die Funktion von Präpositionen analysiert.

Begrenzt semantisches (LS) und bedingt begrenzt semantisches (CLS) Wörterbuchnachschiagen findet an geeigneten Stellen in der Analyse statt, wobei bedingt begrenzt semantisches Nachschlagen erst dann möglich ist, wenn der gesamte Satz analysiert worden ist. Die verbleibenden vier Schritte, welche die Zielsprache auf der Basis der in dem Satzanalysebereich enthaltenen gesamten Information synthetisieren, sind: Die Übersetzung der Präpositionen, Lösen der wortspezifischen Probleme durch lexikalische Routinen, tatsächliche Synthese der Zielsprachworte und Reorganisieren des Satzes in die für die Zielsprache geeignete Wortordnung. Der abschließende Schritt im Übersetzungsprozeß ist das Drucken der übersetzten Aufgabe durch das Programm TRPRINT.

Erfindungsanspruch

1. Verfahren zur Verwendung eines Digitalcomputers zur Übersetzung natürlicher Sprachen, gekennzeichnet durch die folgenden Schritte:
 - a) Übertragen des zu übersetzenden Ursprungstextes von lesbaren Zeichen in computerverständliche binäre Signale;
 - b) Speichern des übertragenen zu übersetzenden Ursprungstextes in einen Hauptspeicher des Computersystems;
 - c) Abtasten und Vergleichen der so übertragenen Ursprungstextworte mit Wörterbüchern der Ursprungssprachworte, gespeichert in einem Speicher und für jedes Ursprungstextwort, für welches ein Gegenstück gefunden wurde, Speichern in einer Datei im Hauptspeicher jedes Wort und (so in) Verbindungen mit jedem derartigen Wort, eine kodierte Information abgeleitet vom Wörterbuch, für die Verwendung in der Übersetzung eines solchen Wortes, wobei die kodierte Information einer Speicheradressbeziehung zu einem Speicher im Computer-System umfaßt, wo Grammatik und die Zielsprachübersetzungen für dieses Wort gespeichert sind;
 - d) Analysieren der übertragenen Ursprungstexte in der Wortdatei, ein vollständiger Satz auf einmal, und Übertragen desselben in einen Satz in der Zielsprache unter Verwendung der kodierten Information und umfassend folgende Schritte:
 - 1) Benützen der Speicheradressbeziehung, um die Zielsprachübersetzungen der Worte von einem Speicher zu erzielen, und
 - 2) Wiederordnen der Zielsprachübersetzung in der geeigneten Zielsprachfolge;
 - f) Drucken der Zielsprache.
2. Verfahren nach Punkt 1, gekennzeichnet durch die Schritte:

Speichern eines Wörterbuches von oft verwendeten Ursprungsworten und angeschlossener Zweigadreßverbindungen (offset address linkages), die Zweigadreßverbindungen identifizieren den Speicherplatz für Grammatik und Bedeutungsinformationen für die Ursprungsworte: Speichern eines Wörterbuches für wenig verwendete Ursprungsworte in Verbindung mit Grammatikcodebedeutungen für jedes Wort; ein Stammwort für die Ursprungssprache zusammen mit Abzweigverbindungen identifizieren den Platz von Zielsprachgrammatikcodes und Bedeutungen; Vergleichen der Ursprungstextworte mit dem Wörterbuch für oft verwendete Worte und nach Entdecken einer Gleichheit mit einem Wort Speichern des Wortes und der angeschlossenen Zweigadreßverbindung zusammen in einem oft verwendeten Speicher und Vergleichen der Ursprungstextworte mit dem Wörterbuch für selten verwendete Worte und nach Entdecken einer Gleichheit Speichern des Wortes und der angeschlossenen Grammatikcodebedeutung in einem selten verwendeten Speicher.

3. Verfahren nach Punkt 2, gekennzeichnet dadurch, daß zweite und nachfolgende Worte von Phrasen, falls vorhanden, in einer Ursprungsphrase zumindest in einem Wörterbuch im Speicher gespeichert sind, daß ein oft verwendetes Wörterbuch die ersten Worte einer Phrase der Ursprungssprache umfaßt, ebenso daran angeschlossene Adreßverbindungen zu zweiten und nachfolgenden Worten derselben Phrase, welche im Phrasenwörterbuch lokalisiert sind, das erste Wort und nachfolgende Worte, falls vorhanden, einer Phrase, welche eine daran anschließende Bedeutung in der Zielsprache hat, ist in Verbindung damit gespeichert und durch folgende, während des Vergleichens mit dem oft verwendeten Wörterbuch, vorgesehene Schritte gekennzeichnet:

- a) Feststellen jener Worte, welche gleich sind, und erste Worte von Phrasen darstellen;
 - b) Benutzen der Abzweigadreßverbindung zur Lokalisierung von einem zusätzlichen Wort oder zusätzlichen Worten derselben Phrase, welche im Phrasenwörterbuch lokalisiert sind;
 - c) Vergleichen der festgestellten weiteren Worte in der Phrase mit den Worten des Ursprungstextes, welcher dem ersten Wort der Phrase folgt, auf Gleichheit und
 - d) nach Feststellen einer fertigen Gleichheit Speichern des ersten Phrasenwortes zusammen mit der Zielsprachbedeutung in einen oft benutzten Speicher;
 - e) Speichern eines jeden der zweiten und nachfolgenden Worte der Phrase zusammen mit einer Phrasenmarkierung in den oft benutzten Speicher, wobei die Phrasenmarkierung bestimmt, daß dieses Wort nicht ein erstes Wort einer Phrase ist;
 - f) Speichern des ersten Wortes einer Phrase zusammen mit einer Markierung in den oft verwendeten Speicher, wobei die Markierung bestimmt, daß das Wort ein erstes Wort einer Phrase ist;
 - g) während des Analysierens Aussuchen der Zielbedeutung, verbunden mit einem ersten Phrasenwort als die Übersetzung des ersten und der nachfolgenden Worte der Phrase.
4. Verfahren nach Punkt 1, gekennzeichnet durch folgenden Schritt: Speichern einer Indizierung in Verbindung mit den Worten im selben gebrauchten Speicher, gleich, ob die Worte Hauptworte und/oder Partizipien sind, in begrenzt-semanticischen Verbindungen, wobei die Verbindungen in eigenen Speichern gespeichert sind und begrenzt semantische Nummern für jedes Hauptwort enthalten, wobei die begrenzt-

semantischen Nummern alle Partizipien indizieren, wobei die begrenzt semantischen Nummern der Partizipien mit den Hauptwörtern in derselben Folge verknüpft sind, in welcher die Partizipien Ausdrucksverbindungen mit dem Hauptwort bilden, wobei während des Abtastens und Vergleichens mit Wörterbüchern eine Information mit dem Hauptwort verbunden wird, welche alle die Ausdrucksverbindungen, innerhalb welcher das Hauptwort als Schlüsselwort in der Verbindung fungiert, enthält, wobei die Information entweder Folgen von begrenzt semantischen Nummern oder eine Adresse für einen Hilfsspeicher, in welchem die Information lokalisiert ist, enthält, und daß durch weitere Schritte, die dem Ursprungstext ein "Subject field code" zuweisenden begrenzt-semantischen Nummern während des Analysierens des Wortspeichers vom Hilfsspeicher in einen Hauptarbeitsspeicher übertragen werden, wobei, wenn dieselbe Ausdrucksverbindung der Ursprungssprache, indiziert wird, durch die begrenzt-semantischen Nummern, die verschiedene Zielbedeutungen haben, durch die Auswahl der richtigen Bedeutung für die Verbindung abhängig von dem Subject field code, in welchem die Übersetzung stattfindet.

5. Verfahren nach Punkt 1, gekennzeichnet dadurch, daß der Verfahrensschritt Analysieren, den Schritt Analysieren des Ursprungstextes und dem Schritt Übertragen des Ursprungssprachsatzes in einem speziellen Bereich umfaßt, welcher dafür im Computerspeicher reserviert ist, wobei sich die Analyse vom Anfang bis zum Ende und vom Anfang bis zum Anfang des Ursprungssprachsatzes bewegt, wobei Markierungen von speziellen Beziehungen der Satzumgebung aufgestellt werden, wobei die Markierungen durch Routinen im Computerprogramm aufgestellt werden, welche abhängig von

unterschiedlichen Teilen der Sprache unter den Wörtern hervorgerufen werden, wobei während der Analyse die Information gespeichert ist, um sie erinnernd von einem Ursprungstextsatz zum anderen zu verwenden, um die Übersetzung jener Worte, für welche eine Rückbeziehung auf den vorhergehenden Satz notwendig ist, zu ermöglichen, wobei während der Analyse semantische und kontaktische Analyse-durchgänge durch jeden Ursprungssatz erfolgen, wobei eine kodierte Information verknüpft wird, wobei die Durchgänge entweder vom Anfang zum Ende oder vom Ende zum Anfang erfolgen, abhängig von den Codes, die mit den einzelnen Worten verbunden sind, und abhängig von der Charakteristik der Ausgangssprache, und daß während der Schritte Abtasten, Vergleichen und Speichern Teile des Sprachcodes den Ursprungstextworten zugeordnet werden, und wobei während der Analyse die zugewiesenen Teile zum Verzweigen in gewünschte Routinen verwendet werden, indem der Wert jedes Teiles der Sprache zu einer fixen Adresse addiert wird, um die Analyse auszuführen, daß während der Analyse der Ursprungssprachworte der Computer eine Spur von vorher angetroffenen Situationen aufrechterhält, indem er eine Reihe von Speicherschaltern (memory switches) setzt, wobei jedes Mal, wenn ein neuer Teil der Sprache angetroffen wird, eine Reihe von vorher gesetzten Schaltern selektiv ausgeschaltet wird, während einer oder mehrere zusätzliche Schalter selektiv eingeschaltet werden, mit einer Routine zum Abfragen der Schalter in einer vorherbestimmten Reihenfolge, wobei auch während der Analyse jedes Wort des Satzes mit einer bestimmt kodierten Information versehen ist, ob es zu einem Hauptsatz oder einem Nebensatz gehört, daß eine Routine bestimmt ist für das Aufsuchen von Ursprungswörtern innerhalb derartiger Sätze für Prädikate und Subjektē, wobei nachfolgend während der

Analyse die Prädikate und Subjekte der Ursprungssprache von Routinen gesucht werden und entsprechend mit Codes versehen werden, die Prädikate und Subjekte gesucht von Routinen innerhalb von in Sätzen enthaltenen Sätzen, wobei Bedacht genommen wird auf die Routinen, welche bestimmen, ob die Frage besteht, ob die Funktion eines Kommas oder einer Konjunktion einen neuen Satz beginnt und eigenes Suchen nach Prädikaten und Subjekten durchführen, und die Antwort als Ergebnis der Analyse jedes Ursprungsworts in einem Satz ist mit Marken versehen, welche die Abhängigkeitsbeziehungen unter allen Worten des Satzes ausdrücken, wobei gleichzeitig jedem Wort im Satz eine syntaktische Funktion zugewiesen wird, welche in einer verdichteten Form unter Verwendung von "coded bis" ausgedrückt ist, wobei die Syntaxanalyse fortgesetzt wird, bis jedem Wort im Satz zumindest eine Funktion zugewiesen ist.

6. Verfahren nach Punkt 1, gekennzeichnet dadurch, daß ein lokales Spezialwörterbuch gespeichert ist, welches unterschiedliche Zielsprachbedeutungen für ein Ursprungswort in Übereinstimmung mit verschiedenen technischen Themen enthält, wobei in diesem Schritt eine Programm-Routine zur Überprüfung jedes Ursprungssatzes in der Analyse und zur Auswahl von Bedeutungen eines aktuellen Spezialwörterbuches in Übereinstimmung mit dem technischen Sachinhalt des Satzes enthalten ist.
7. Verfahren nach Punkt 1, gekennzeichnet dadurch, daß die Ursprungssprachsätze in einer hierarchischen Ordnung analysiert werden, wobei im ersten Durchgang grundlegende Zweideutigkeiten aufgelöst werden, in einem zweiten folgenden Durchgang primäre syntaktische Zwischenverbindungen

etabliert werden und in einem dritten folgenden Durchgang diese Verbindungen auf abgegrenzte Phrasen und Sätze ausgedehnt werden, daß in einem vierten folgenden Prädikat-Subjekt-Durchgang, in welchem in hierarchischer Ordnung nach Prädikaten und Subjekten gesucht wird, verbleibende Satztrennungs-Zweideutigkeiten aufgelöst werden, und daß in einem fünften folgenden Durchgang sichergestellt wird, daß die Funktion jedes Wortes in dem Satz richtig erkannt wurde.

8. Verfahren nach Punkt 1, gekennzeichnet dadurch, daß im Haupt-Syntax-Durchgang generalisierte Unterprogramme abgerufen werden, um syntaktische Beziehungen zu bestimmen und zu kodieren und daß daran anschließend die Unterprogramme selektiv kleinere oder eingenistete Unterprogramme, zur Ausführung der Basisaufgaben abgerufen werden.
9. Verfahren nach Punkt 1, gekennzeichnet dadurch, daß ein eigenes Programm die Präpositionen der Ausgangssprache bestimmt und analysiert und dabei die den Präpositionen eigene Bedeutung bestimmt, wobei die Bedeutung der Präpositionen entweder von Worten, welche die Präpositionen regeln bestimmt wird oder durch Worte, welche als Präpositionen fungieren, wobei eine kodierte Information für die Übersetzung von einer oder mehreren Präpositionen den Wörtern vom Wörterbuch zugeordnet wird.
10. Verfahren nach Punkt 1, gekennzeichnet durch die Schritte: Analysieren jedes Wortes durch mehrfache Durchgänge durch jeden Satz, wobei Codes zugewiesen werden, Bedenken aller Codes, welche in vorhergehenden Durchgängen mit einem Wort verknüpft wurden, und Zuweisen von Zielsprach-Synthesecodes, welche mit der Bedeutung verknüpft sind, mit

der die Code-Funktion im Satz existiert, Setzen des Wortes in eine Form entsprechend der Zielsprache abhängig von der Analyse und dem Inbetrachtziehen aller relevanten und dem Wort zugewiesenen Codes, wobei die übersetzten Worte gebildet nach der Analyse, in einem Druckbereich in einer Folge spezifiziert durch die Beziehung zwischen der Ausgangs- und der Zielsprache und vorher aufgestellt nach den nach jedem Wort gegebenen Nummern und Übertragen der Übersetzung, wenn der Druckbereich aufgefüllt ist, entweder direkt auf einem Drucker oder auf einem temporären Hilfsspeicher.

11. Verfahren nach Punkt 1, gekennzeichnet dadurch, daß ein spezielles Unterprogramm des Wörterbuchnachsschlagverfahrens eine Liste von Worten (lexikalische Liste) durchsucht, um zu bestimmen, für welche Worte die Aufnahme eines speziellen lexikalischen Unterprogrammes in den Wörterbuchspeicher erforderlich ist, wobei das eigene lexikalische Kontrollprogramm die Wörter des Ausgangssprachsatzes für jene Fälle analysiert, für welche nur die Ergebnisse der syntaktischen und semantischen Analyse des Satzes oder der Mitgliedschaft eines Wortes in einer grammatikalischen oder semantischen Klasse bestimmen kann, ob oder ob nicht ein lexikalisches Unterprogramm zum Zeitpunkt der Übersetzung eines speziellen Wortes oder eines speziellen Ausdruckes abgerufen werden muß, um die Bedeutung dieses Wortes oder dieses Ausdruckes zu bestimmen, indem die syntaktischen Beziehungen, welche für dieses Wort oder für diesen Ausdruck etabliert wurden, zu überprüfen, wobei jene syntaktischen und/oder semantischen Regeln benutzt werden, welche nur diesem Wort oder dieser Klasse von Worten, zu denen es gehört, zugewiesen ist.

etabliert werden und in einem dritten folgenden Durchgang diese Verbindungen auf abgegrenzte Phrasen und Sätze ausgedehnt werden, daß in einem vierten folgenden Prädikat-Subjekt-Durchgang, in welchem in hierarchischer Ordnung nach Prädikaten und Subjekten gesucht wird, verbleibende Satztrennungs-Zweideutigkeiten aufgelöst werden, und daß in einem fünften folgenden Durchgang sichergestellt wird, daß die Funktion jedes Wortes in dem Satz richtig erkannt wurde.

8. Verfahren nach Punkt 1, gekennzeichnet dadurch, daß im Haupt-Syntax-Durchgang generalisierte Unterprogramme abgerufen werden, um syntaktische Beziehungen zu bestimmen und zu kodieren und daß daran anschließend die Unterprogramme selektiv kleinere oder eingestufte Unterprogramme, zur Ausführung der Basisaufgaben abgerufen werden.
9. Verfahren nach Punkt 1, gekennzeichnet dadurch, daß ein eigenes Programm die Präpositionen der Ausgangssprache bestimmt und analysiert und dabei die den Präpositionen eigene Bedeutung bestimmt, wobei die Bedeutung der Präpositionen entweder von Worten, welche die Präpositionen regeln bestimmt wird oder durch Worte, welche als Präpositionen fungieren, wobei eine kodierte Information für die Übersetzung von einer oder mehreren Präpositionen den Wörtern vom Wörterbuch zugeordnet wird.
10. Verfahren nach Punkt 1, gekennzeichnet durch die Schritte: Analysieren jedes Wortes durch mehrfache Durchgänge durch jeden Satz, wobei Codes zugewiesen werden, Bedenken aller Codes, welche in vorhergehenden Durchgängen mit einem Wort verknüpft wurden, und Zuweisen von Zielsprach-Synthesecodes, welche mit der Bedeutung verknüpft sind, mit

der die Code-Funktion im Satz existiert, Setzen des Wortes in eine Form entsprechend der Zielsprache abhängig von der Analyse und dem Inbetrachtziehen aller relevanten und dem Wort zugewiesenen Codes, wobei die übersetzten Worte gebildet nach der Analyse, in einem Druckbereich in einer Folge spezifiziert durch die Beziehung zwischen der Ausgangs- und der Zielsprache und vorher aufgestellt nach den nach jedem Wort gegebenen Nummern und Übertragen der Übersetzung, wenn der Druckbereich aufgefüllt ist, entweder direkt auf einem Drucker oder auf einem temporären Hilfsspeicher.

11. Verfahren nach Punkt 1, gekennzeichnet dadurch, daß ein spezielles Unterprogramm des Wörterbuchnachsschlagverfahrens eine Liste von Worten (lexikalische Liste) durchsucht, um zu bestimmen, für welche Worte die Aufnahme eines speziellen lexikalischen Unterprogrammes in den Wörterbuchspeicher erforderlich ist, wobei das eigene lexikalische Kontrollprogramm die Wörter des Ausgangssprachsatzes für jene Fälle analysiert, für welche nur die Ergebnisse der syntaktischen und semantischen Analyse des Satzes oder der Mitgliedschaft eines Wortes in einer grammatikalischen oder semantischen Klasse bestimmen kann, ob oder ob nicht ein lexikalisches Unterprogramm zum Zeitpunkt der Übersetzung eines speziellen Wortes oder eines speziellen Ausdruckes abgerufen werden muß, um die Bedeutung dieses Wortes oder dieses Ausdruckes zu bestimmen, indem die syntaktischen Beziehungen, welche für dieses Wort oder für diesen Ausdruck etabliert wurden, zu überprüfen, wobei jene syntaktischen und/oder semantischen Regeln benutzt werden, welche nur diesem Wort oder dieser Klasse von Worten, zu denen es gehört, zugewiesen ist.

12. Verfahren nach Punkt 1, gekennzeichnet dadurch, daß zur Lösung semantischer Zweideutigkeiten jedes Wort im Stammwörterbuch (Einzelwort) mit einer eigenen begrenzt-semantischen Nummer (LS) versehen ist, und daß die speziellen Mehrfachbedeutungen des LS-Verbundwörterbuches (LS-Ausdrücke), welches zusammengesetzt ist aus den LS-Nummern-Repräsentanten der Einzelwörter im Stammwörterbuch, dann automatisch in den Wörterbuchblock gruppiert werden, gemäß dem Hauptwort des Ausdruckes, wobei einer für jedes eigene Hauptwort existiert, und dieser Block abgesucht wird, wann immer das Hauptwort in einem zu übersetzenden Satz auftritt, um zu bestimmen, ob eine Übereinstimmung existiert zwischen einer Gruppe aufeinanderfolgender Worte im Text und irgendeinem Ausdruck in dem Block, wobei die längste Übereinstimmung oder der Ausdruck mit der höchsten Priorität verwendet wird, um die spezielle Bedeutung dieser Wortgruppe zu bestimmen, welche von der Summe der Bedeutungen der Einzelworte abweicht, daß ein untergeordneter Satz für LS-Ausdrücke von bedingt begrenzten semantischen Ausdrücken gebildet wird (CLS), welcher erlaubt, alle definierbaren syntaktischen und semantischen Regeln als auch einfache Programmierinstruktionen einzuschließen, welche dazu verwendet werden können, die Informationen in den Bits und Bytes des Satzanalyse-Bereiches im LS-Ausdruck selbst (d.h. im Wörterbuch) auszuwechseln, so daß der Wörterbuchausdruck nur dann übereinstimmt, wenn die syntaktischen und/oder semantischen Bedingungen, ausgedrückt in den Regeln, durch die Worte im Text erfüllt sind, wobei diese Tatsache nur nach einer vollständigen syntaktischen und semantischen Analyse des Satzes, durchgeführt durch das sprachwissenschaftliche Analyseprogramm bestimmbar ist.

13. Verfahren nach Punkt 1, gekennzeichnet dadurch, daß alle Ursprungssprachworte mit semantischen Kategorie-Codes mit variablem Längenformat versehen werden können, welche dann durch die Ursprungssprachanalyse-Programme, die lexikalischen Routinen und die CLS-Wörterbuchnachschragsroutinen abgefragt werden können, als Hilfe bei der Lösung semantischer Zweideutigkeiten, wobei diese Codes in einer hierarchischen Klassifizierung (einem Satz von Baumstrukturen) in der Weise ausgedrückt sind, daß die Kodierung einer Kategorie, welche auf einem niedrigeren Niveau eines semantischen Baumes existiert, automatisch die Kodierung eines geeigneten Codes eines höheren Niveaus durch das System selbst erzeugt.
14. Verfahren nach Punkt 1, gekennzeichnet dadurch, daß eine Folge von Worten in der Ausgangssprache innerhalb von Phrasen und Sätzen in Beziehung zur Wortfolge der Zielsprache analysiert wird, wobei die Wortfolge der Zielsprache symbolisch durch zugewiesene Nummern ausgedrückt wird und daß beim Drucken der Übersetzung auf jedes Ursprungswort Bedacht genommen wird.

Hierzu 3 Seiten Zeichnungen





