

Universität Bremen  
Studiengang Linguistik  
Wintersemester 2002/2003

# **DLT – Distributed Language Translation**

**Grundlagen der maschinellen Übersetzung**



Carolin Hinz  
Wilhelm-Leuschner-Straße 10  
28329 Bremen  
chinz@uni-bremen.de

# Inhaltsverzeichnis

<b>0. Einleitung</b> .....	3
<b>1. Architekturen von Übersetzungssystemen</b> .....	3
1.1 Direkte Übersetzung .....	4
1.2 Transfer .....	4
1.3 Interlingua .....	4
1.4 Vergleich der Architekturen.....	5
<b>2. Allgemeines und Entwicklungsgeschichte</b> .....	7
<b>3. Linguistische Grundlagen</b> .....	7
3.1 Dependenz mit Metataxis.....	7
3.2 Esperanto.....	10
<b>4. Übersetzungsprozess</b> .....	17
<b>5. Übersetzungsschema</b> .....	23
<b>6. Zusammenfassung</b> .....	24
Bibliographie als Anhang	

## **0. Einleitung**

Wir haben in diesem Semester eine ganze Reihe von ambitionierten Projekten zum Thema Maschinelle Übersetzung (MÜ) kennen gelernt. Viele von ihnen sind gescheitert bzw. wurden eingestellt.

Auch das DLT-Projekt zählt zu diesen. Die Urheber dieses Programms erkannten schon früh, das FAHQT (Fully Automatic High Quality Translation) ein unerreichbares Vorhaben ist, und so versuchten sie sich an SAHQT (Semi Automatic High Quality Translation). Dieses Anfang der 80er Jahre initiierte Projekt schien darin auch sehr erfolgreich zu sein, wurde jedoch Anfang der 90er Jahre plötzlich eingestellt. DLT war immer nur ein Prototypprojekt, es gab nie eine marktfähige Version. Die linguistische, wie auch technische, Begleitliteratur stellte immer den gerade aktuellen Stand und zukünftige Planungen dar. Daher wurde dieses Programm in seiner Architektur und Funktionsweise auch jedes Mal etwas anders beschrieben. Ich habe nun versucht, die Gemeinsamkeiten aller Texte herauszufiltern und dadurch eine angemessene Beschreibung des Programms zu liefern. Allerdings kann ich mit der mir vorhandenen Literatur nicht dafür garantieren, dass meine Beschreibung auch den realen Stand am Abschluss des Projektes wiedergibt.

Der Schwerpunkt dieser Arbeit soll auf der Erläuterung der linguistischen Grundlagen des Programms liegen. Dabei sollen auch die Schwierigkeiten, mit denen automatische Übersetzungsprogramme zu kämpfen haben, betrachtet werden, da sie letztendlich auch zu den Gründen zählen, warum das Projekt eingestellt wurde.

Vorab jedoch werde ich noch einige allgemeine Zusammenfassungen zur Architektur von MÜ-Systemen und auch zu DLT geben.

## **1. Architekturen von Übersetzungssystemen**

Wenn man privat einen fremdsprachlichen Text übersetzen will, sitzt man in der Regel mit mehreren Wörterbüchern und gegebenenfalls auch mit einer Grammatik an seinem Schreibtisch und übersetzt Satz für Satz einzeln. Ähnlich arbeiten auch maschinelle Übersetzungssysteme. Doch während wir als Mensch den gesamten Text vor Augen haben und somit auf Kontextwissen zugreifen können, gibt es bis jetzt noch keine Möglichkeit, dass auch MÜ-Systeme den Text als Ganzes in die Übersetzung einbeziehen können. Dieses Problem wird meistens dann deutlich, wenn es um Anapherauflösung geht. Aber auch unter

den verschiedenen Übersetzungsprogrammen gibt es mehrere Vorgehensweisen und Unterschiede.

### **1.1 Direkte Übersetzung**

Bei direkten Übersetzungssystemen wird der Ausgangssprachliche Text mit Hilfe eines Wörterbuchs und einer Minigrammatik analysiert und unter Einbeziehung eines zweisprachigen Wörterbuchs Wort-für-Wort übersetzt.

### **1.2 Transfer**

Transfersysteme sind wesentlich umfangreicher als direkte Systeme und arbeiten auch genauer als diese. Der Ausgangssprachliche Satz wird zuerst syntaktisch analysiert und es wird eine Struktur von ihm erzeugt. Diese Ausgangssprachliche Struktur wird anschließend über bilinguale Regeln in eine Zielsprachliche Struktur transferiert. Zum Schluss wird aus dieser Struktur unter Einbeziehung der Zielsprachlichen Grammatik der linearisierte Satz erzeugt.

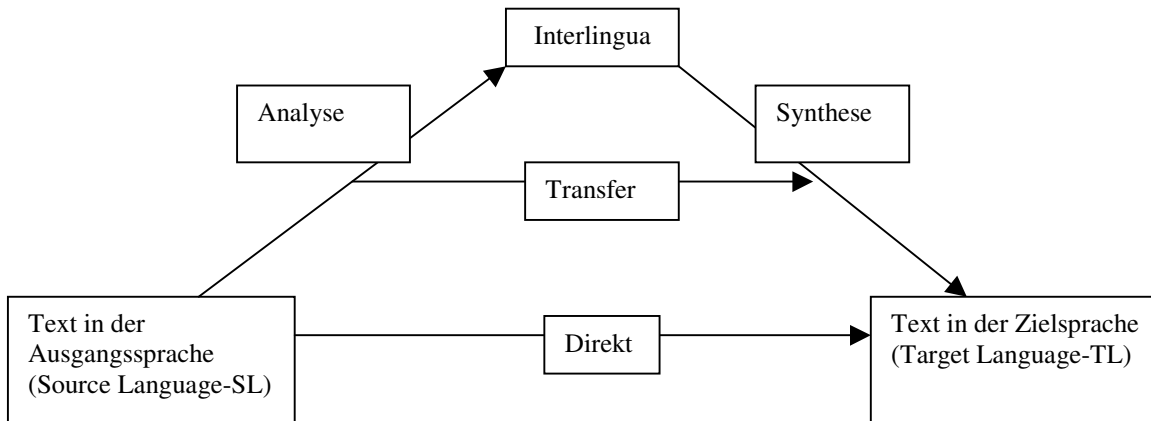
### **1.3 Interlingua**

Interlinguasysteme arbeiten ähnlich wie Transfersysteme. Nach der syntaktischen Analyse des Ausgangssprachlichen Textes wird eine Interlinguarepräsentation erzeugt. Aus dieser Zwischenrepräsentation wird in der Synthese mit Hilfe der Grammatik der Zielsprache der Zielsprachliche Satz erzeugt.

DLT wird allgemein zu den Interlinguasystemen gezählt, weil der Ausgangssprachliche Text erst in Esperanto und anschließend in die Zielsprache übersetzt wird. An Hand der Beschreibung des Übersetzungsprozesses wird auf die spezielle Architektur von DLT noch einmal eingegangen.

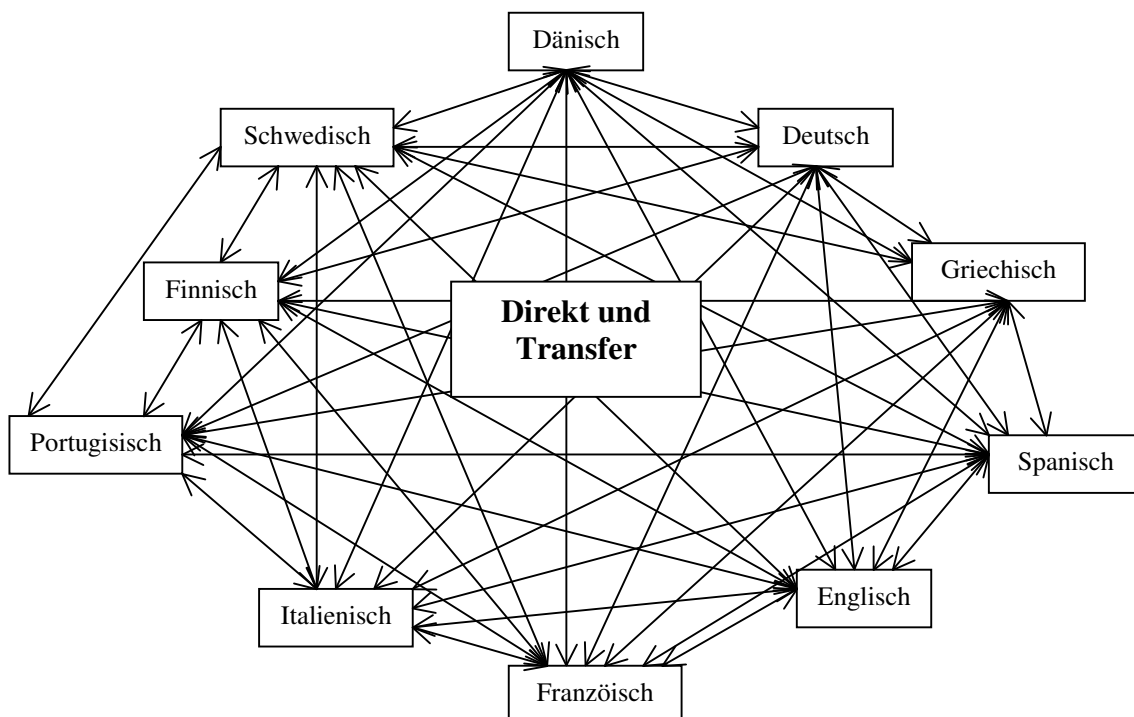
## 1.4 Vergleich der Architekturen

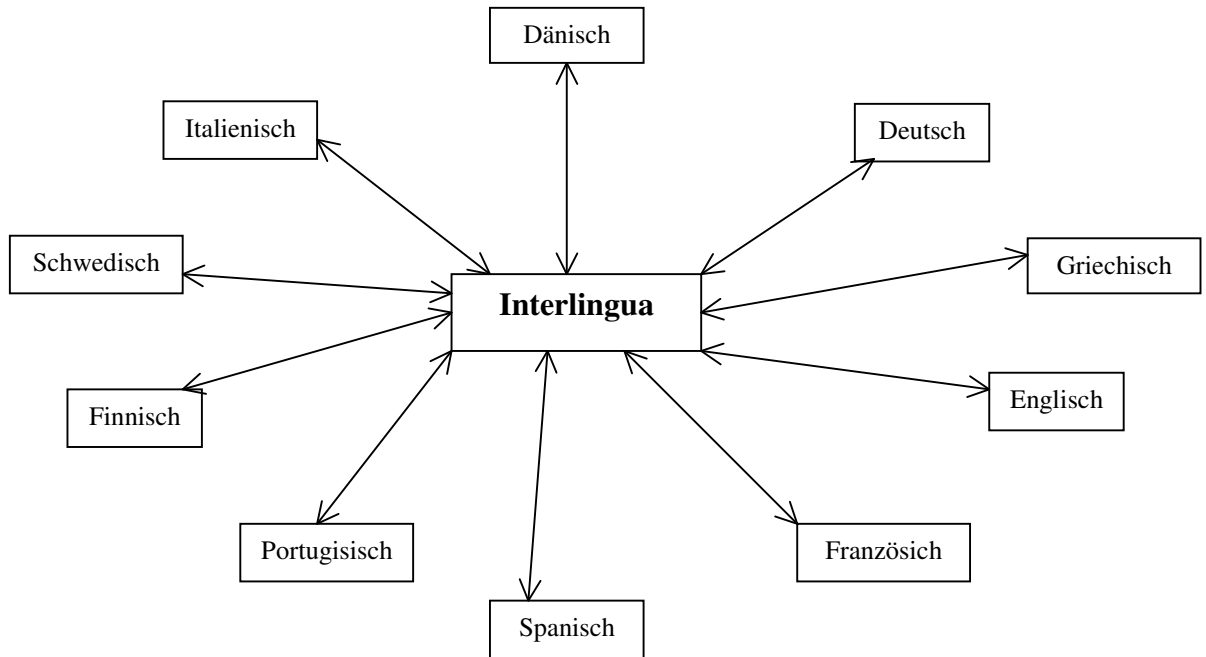
Graphisch kann man diese drei Varianten folgendermaßen darstellen:



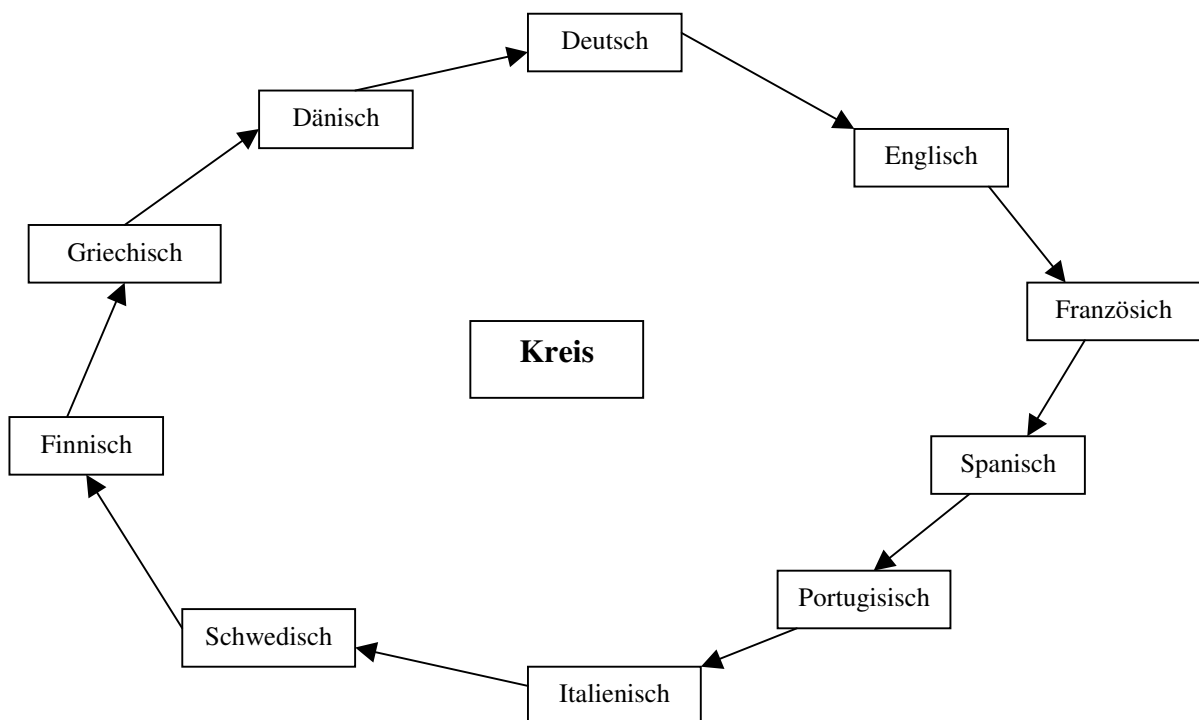
Interlinguasysteme eignen sich besonders gut zur Übersetzung in multilingualen Systemen, wie zum Beispiel in der Europäischen Union. Es gibt in der EU zehn offizielle Sprachen. Jedes Schriftstück, das die EU verfasst, muss in allen zehn Sprachen wiedergegeben werden. Bei direkten oder Transfersystemen ergibt das eine Summe von 90 Sprachpaaren. Bei einem Interlinguasystem reduziert sich diese Zahl auf 19.

Folgende Graphiken sollen diesen Vorteil noch verdeutlichen:





Eine etwas extreme Möglichkeit, die Anzahl der Sprachpaare zu vermindern, ist die kreisförmige Übersetzung:



Bei dieser Variante stellt sich die Frage, ob nicht auf dem Weg von einer SL zur TL Informationen verloren gehen. Bei Interlinguaübersetzungen gibt es maximal zwei Schritte von der SL zur TL, bei der kreisförmigen Übersetzung mit zehn Sprachen gibt es bis zu neun Zwischenschritte.

## **2. Allgemeines und Entwicklungsgeschichte**

Während das Eurotraprojekt in akademischen Kreisen entwickelt wurde, hat DLT seinen Ursprung in der Softwareindustrie, genauer bei BSO (Buro voor Systemontwikkeling) in Utrecht/Holland. Es war geplant als multilinguales Übersetzungssystem in Videotext-Netzwerken. Dabei war ausschließlich die Übersetzung von informativen Texten wie Bedienungsanleitungen, wissenschaftliche Arbeiten und Wirtschaftsberichte, die von Natur aus frei von rhetorisch gewollten Ambiguitäten sind, vorgesehen.

DLT wurde als halb-automatisches Interlinguasystem konzipiert mit Esperanto als Interlingua (IL) und auf der Basis von Dependenzgrammatik mit Metataxis. Programmiert wurde es in Prolog und es lief unter dem Betriebssystem Unix auf Sun-Rechnern.

Der Name Distributed Language Translation bezieht sich auf die Art des Übersetzungsprozesses: "distributed" meint hier die Verteilung des Gesamtprozesses auf mehrere Terminal (Sender-Empfänger) und die damit verbundene Trennung von Raum und Zeit.

Erste Untersuchungen zu DLT wurden 1979 vom Chefsingenieur Toon Witkam unternommen. Dieser veröffentlichte 1983 auch eine Durchführbarkeitsstudie, in der DLT große Chancen auf Erfolg beschienen wurde. 1985 wurde mit dem niederländischen Finanzministerium ein 6-Jahresvertrag abgeschlossen. Langfristige Ziele waren die Erstellung einer Prototypversion bis 1987 und einer marktfähigen Version bis 1993. Zwei Jahre später wurde der erste Prototyp auch vorgestellt. Er übersetzte Texte, in denen es um Flugzeugwartung ging, aus einer vereinfachten Form des Englischen (Simplified English) ins Französische. 1991 wurde das Projekt dann trotz weitreichender Planungen plötzlich eingestellt. Auf die Gründe werde ich am Ende dieser Arbeit noch eingehen.

## **3. Linguistische Grundlagen**

### **3.1 Dependenz mit Metataxis**

Bevor ich die Prinzipien der Dependenzgrammatik erläutere, möchte ich noch einmal kurz auf die wesentlichen Aufgaben von Syntax eingehen. Syntax dient der Identifikation der Einheiten, die den Satz gliedern, und ihrer jeweiligen weiteren Bestandteile. Außerdem werden die Relationen, die zwischen den verschiedenen Bestandteilen eines Satzes bestehen

bestimmt. Mit Hilfe von Syntax werden die Muster und Regeln ermittelt, nach denen die Bestandteile miteinander kombiniert werden und in welcher Reihenfolge sie erscheinen. Für diese Aufgaben gibt es verschiedene Grammatikmodelle und Theorien, von denen eine die Dependenzgrammatik ist.

Als Vater der Dependenzgrammatik gilt Lucien Tesnière, der eine didaktische Zielsetzung verfolgte. In bezug auf den Fremdsprachenunterricht war er der Meinung, dass eine Satzanalyse zu Kenntnissen zur Satzbildung führt.

In einem Satz gibt es zwei Ordnungsprinzipien: linear und strukturell. Linear verschiedene Sätze können gleich strukturiert sein. Zum Beispiel: „Pascal gibt dem Kind den Apfel.“ – „Den Apfel gibt Pascal dem Kind.“ Deshalb richtet die Dependenzgrammatik ihren Fokus auf die Analyse der strukturellen Ordnung.

In einem Satz gibt es fakultative und obligatorische Konstituenten. Die fakultativen Elemente können nur gesetzt werden, wenn die obligatorischen vorhanden sind und sind somit von diesen abhängig. Dieses Abhängigkeitsverhältnis bezeichnet man als Dependenz. Der obligatorische Bestandteil einer Konstituente wird Kopf genannt, der fakultative oder abhängige Bestandteil Dependens.

Wenn ein Kopf festlegt, welche Wortform sein Dependens haben darf, dann ist dies ein Indiz dafür, dass der Kopf das Dependens regiert. Diese Wirkung nennt man Rektion. Das regierende Element wird oft als Regens, das regierte als Rectum bezeichnet.

In der Dependenzgrammatik nimmt man an, dass alle Satzglieder vom finiten Verb desselben Satzes unmittelbar oder mittelbar abhängen. Dies lässt sich zwar nicht empirisch beweisen, es gibt jedoch verschiedene Argumente für die Verbzentrierung. Semantisch betrachtet gibt das Verb einem Satz erst einen Sinn. Außerdem ist das Verb von allen sprachlichen Kategorien am stärksten an die Sprechsituation (z.B. Tempus) gebunden. Syntaktisch gesehen ist das Verb obligatorisch für den Satz und dessen Valenz bestimmt die Zahl der Aktanten.

Metataxis ist ein Teilbereich der Konnexion. Als Metataxis wird die Übersetzung der Beziehungen innerhalb eines Satzes bzw. die Struktur eines Textes bezeichnet. Die Metataxisbeschreibung eines bestimmten Sprachpaares ist eine Reihe von Regeln, die Texte einer Sprache in eine andere Sprache transformieren. Diese ist an eine bestimmte Übersetzungsrichtung mit einer Ausgangssprache (SL) und einer Zielsprache (TL) gebunden.



Sowohl für die SL als auch für die TL gibt es je eine autonome Dependenzgrammatik, die durch Metataxis verbunden werden.

Durch Metataxis werden bereits analysierte Dependenzbäume übersetzt. Semantische und pragmatische Schwierigkeiten des lexikalischen Transfers stehen außerhalb des Metataxissystems. Das heißt, es werden keine Ambiguitäten aufgelöst: Bei mehrdeutiger Übersetzung werden auch mehrere Alternativbäume generiert.

Metataxis bedeutet damit in erster Linie strukturelle Veränderung auf allen Ebenen von Struktur: Wort, Morphem, Wortart. Wo es nötig wird, werden Morpheme hinzugefügt oder entfernt, Dependenzlabel, die die syntaktische Funktion angeben, verändert oder es kommt zu neuen Dependenzbeziehungen, also neuen Regens-Dependens-Relationen. Diese Veränderung in den Dependenzbäumen ist allerdings durch die Wohlgeformtheitsbedingungen der Dependenzgrammatiken von SL und TL beschränkt. Um die Korrektheit der Veränderung der Dependenzbäume zu gewährleisten, müssen die Lexikoneinträge bestimmte Voraussetzungen erfüllen. Bei einer Übersetzung muss immer auch der Kontext berücksichtigt werden, in dem bestimmte Wörter vorkommen. Somit sollten nicht nur einzelne Wörter, sondern auch Wortgruppen in das Lexikon übernommen werden. Das könnte jedoch bei der Übersetzung zu Problemen führen. Deshalb ist es wichtig, auch die syntaktischen Beziehungen zwischen den Wörtern anzugeben. In unserem Fall heißt das, dass die Lexikoneinträge Dependenzbäume darstellen. Wenn jeder Eintrag ein Dependenzbaum ist, kann daraus der gesamte Baum (eines Satzes) generiert werden. Damit wäre die komplette Übersetzungssyntax im zweisprachigen Wörterbuch verankert. Dies wäre jedoch viel zu aufwendig und vieles wäre redundant. Die Lösung für dieses Problem sind kontrastive lexikalische Redundanzregeln.

Im Allgemeinen sind grammatische Regeln nicht nur kompliziert, sondern haben auch Ausnahmen. Bei der Vernetzung zweier Sprachen werden die Unregelmäßigkeiten einer Sprache zu denen der anderen addiert. Man muss also eine Möglichkeit finden, damit Regeln und Ausnahmen flexibel nebeneinander existieren. In der Metataxis unterliegen die Regeln einer festen Hierarchie. Die niedrigste Stufe in der Hierarchie sind die unmarkierten Metataxisregeln, die am häufigsten auftreten: Ein Dependenzlabel der SL wird zu dem äquivalenten Dependenzlabel der TL. Von Stufe zu Stufe werden die Regeln immer spezieller. Am Ende gibt es Regeln, die sich nur noch auf einen spezifischen Eintrag beziehen. Konkret enthalten die Metataxisregeln meist Variablen, die mit verschiedenen Wörtern, die bestimmte Bedingungen erfüllen, besetzt werden können. Eine Variable, die für

alle möglichen Wörter steht, ist weniger spezifisch als eine Variable für eine bestimmte Wortklasse. Eine wichtige Zusatzregel besagt, dass bei der Wahl der entsprechenden Metataxisregel eine spezifische einer generellen vorzuziehen ist. Wenn es mehrere gleichrangige Regeln gibt (parallele Regeln), wird metataktische Ambiguität erzeugt, die erst später aufgelöst werden kann.

Bei der Wahl der Regel wird im Metataxisprozess jeweils von der SL ausgegangen. Im Metataxisprozess werden die Dependenzbäume schrittweise transformiert: top-down und left-to-right. Somit gibt es als Zwischenergebnisse mehrere hybride Bäume.

Die Metataxis eines bestimmten Sprachpaares enthält somit eine relativ große Anzahl komplexer Regeln. Zur ihrer Vereinfachung und Verringerung der Anzahl werden in DLT SL- und TL-Filter eingesetzt. Für die SL bedeutet dies, dass syntaktische Variationen, die für die Übersetzung irrelevant sind, auf eine Standardstruktur reduziert werden. So wird beispielsweise aus can't / cannot → can-Circ-not. Bei der TL werden wort-spezifische Beschränkungen erst ganz am Ende des Prozesses verändert. Somit gibt es drei Blöcke von Metataxisregel, die im Übersetzungsprozess zu inkorrekten Bäumen als Zwischenergebnissen führen: SL-Filter (SL→SL), Transformationen (SL→TL) und TL-Filter(TL→TL).

Das Ergebnis des Metataxisprozesses sind mehrere verschiedene Dependenzbäume, die alle möglichen (semantischen/syntaktischen) Übersetzungen repräsentieren.

## **3.2 Esperanto**

Esperanto ist eine Plansprache, die zum Einsatz als internationale Verkehrssprache von Ludovic Lazar Zamenhof 1887 konstruiert wurde. Das Vokabular stammt aus europäischen Sprachen, weswegen ihr Eurozentrismus vorgeworfen wird. Dennoch gibt es mittlerweile eine große Sprachgemeinschaft. Durch den verbreiteten Gebrauch über längere Zeit kam es wie bei anderen natürlichen Sprachen zu sprachlichen Entwicklungen und Innovation.

### 3.2.1 Morphologie

Esperanto ist eine agglutinierende Sprache, d. h. jedes grammatische Merkmal wird durch ein Morphem dargestellt, das an einen Stamm angehängt wird. Jedes Morphem ist unveränderlich. Komposita sind hinsichtlich ihrer Bedeutung transparent, da sich diese aus der Bedeutung der einzelnen Bestandteile ergibt.

Die Wörter des Esperanto werden aus Wurzeln und Affixen zusammengesetzt. Es gibt auch Komposita, die aus zusammengeführten Wörtern oder zusammengeführten Wurzeln bestehen.

### 3.2.1.1 Affigierung:

Die Affixe sind sowohl Präfixe als auch Suffixe und dienen einerseits zur Bildung der verschiedenen grammatischen Formen der Wörter, andererseits zur Bildung von Wortstämmen mit lexikalischer Bedeutung. Man kann daher zwischen grammatischen und lexikalischen Affixen unterscheiden.

### 3.2.1.2 Wurzeln:

Es gibt freie und gebundene Wurzeln. Freie Wurzeln, die sowohl als selbständige Wörter vorkommen als auch in Wortbildungsprozesse eingeschlossen werden können, sind in der Regel unveränderliche Funktionswörter wie Determinatoren, Präpositionen, Konjunktionen u.ä. Gebundene Wurzeln können dagegen nur mit Affixen oder anderen Wurzeln vorkommen, mit denen sie durch anfügen eines grammatischen Suffixes zu selbständig gebrauchten Wörtern werden.

## 3.2.2 Wortarten

Man unterscheidet vier lexikalische Hauptkategorien: Nomina, Adjektive, Adverbien und Verben. Jede dieser Wortarten wird durch anhängen des entsprechenden Suffixes an eine Wurzel gebildet.

Bsp.: Wurzel „telefon-„

```
\t telefono          sonoras.  
\m telefon -o      sonor  -as  
\g "Telefon-" -Subst. "Klang-" -Präs.  
\p Wurzel      -Suff  Wurzel  -Suff  
\f ein Telefon klingelt
```

```
\t mi                telefonas.  
\m mi                telefon -as  
\g 1.Sg.            "Telefon-" -Präs.  
\p Pers.Pron.      Wurzel  -Suff  
\f ich telefoniere
```

```
\t telefonajn  
\m telefon -a      -j      -n  
\g "Telefon-" -Adj. -Pl.  -dir.Obj.  
\p Wurzel        -Suff -Suff -Suff
```

```

\t mesagxojn
\m mesagx      -o      -j      -n
\g "Nachricht" -Subst. -Pl.  -dir.Obj.
\p Wurzel      -Suff   -Suff  -Suff
\f telefonische Nachrichten

\t la instruisto          parolos
\m la instru -ist        -o      parol      -os
\g Def "Lehr-" -Tätigkeit -Subst. "Sprach-" -Fut.
\p Det Wurzel -Suff      -Suff  Wurzel     -Suff

\t telefone          kun sxi.
\m telefon -e      kun sxi
\g "Telefon-" -Adv. mit 3.Sg.Fem.
\p Wurzel      -Suff P   Pers.Pron.
\f der Lehrer wird telefonisch mit ihr sprechen

```

### 3.2.2.1 Nomina:

Das Nomen bekommt immer die Endung –o. Die Deklination erfolgt nach Numerus (Pluralendung –j) und Kasus, wobei nur zwei Fälle unterschieden werden. Dabei ist der Nominativ unmarkiert und der Kasus zur Kennzeichnung des direkten Objekts (Akkusativ) wird durch die Endung –n repräsentiert, die immer am Ende des Wortes steht. Andere Kasusrelationen wie Genitiv, Dativ und Ablativ werden durch Präpositionen ausgedrückt.

Bsp.:

```

\t telefono          sonoras.
\m telefon -o      sonor      -as
\g "Telefon-" -Subst. "Klang-" -Präs.
\p Wurzel      -Suff   Wurzel    -Suff
\f ein Telefon klingelt

\t la telefonoj          sonoras.
\m la telefon -o      -j      sonor      -as
\g Def "Telefon-" -Subst. -Pl.  "Klang-" -Präs.
\p Det Wurzel      -Suff   -Suff  Wurzel    -Suff
\f die Telefone klingeln

\t li          purigis          la
\m li          purig            -is   la
\g 3.Sg.Mask. "Reinigung-" -Prät. Def
\p Pers.Pron. Wurzel          -Suff Det

\t telefonojn.
\m telefon -o      -j      -n
\g "Telefon-" -Subst. -Pl.  -dir.Obj.
\p Wurzel      -Suff   -Suff  -Suff
\f er reinigte die Telefone

```

### 3.2.2.2 Adjektive:

Das Adjektiv endet immer auf –a. Ansonsten ist die Deklination wie beim Nomen. Die Steigerung und der Vergleich erfolgen durch freie Wurzeln.

Bsp.:

```
\t la patro estas en la
\m la patr -o est -as en la
\g Def "Elter-" -Subst. "sein" -Präs. in Def
\p Det Wurzel -Suff Wurzel -Suff P Det
```

```
\t nova domo.
\m nov -a dom -o
\g "neu" -Adj. "Haus" -Subst.
\p Wurzel -Suff Wurzel -Suff
\f der Vater ist im neuen Haus
```

### 3.2.2.3 Adverbien:

Abgeleitete Adverbien enden immer auf –e.

Bsp.:

```
\t la auxtomobiloj rapide
\m la auxtomobil -o -j rapid -e
\g Def "Auto" -Subst. -Pl. "schnell" -Adv.
\p Det Wurzel -Suff -Suff Wurzel -Suff
```

```
\t veturas.
\m vetur -as
\g "Fahr-" -Präs.
\p Wurzel -Suff
```

```
\f die Autos fahren schnell
```

### 3.2.2.4 Verben:

Das Verb wird durch verschiedene Verbalaffixe repräsentiert. Dabei wird das Verb nur nach Tempus und Modus konjugiert. Die Unterscheidung nach Numerus und Person erfolgt durch entsprechende Personalpronomina.

Bsp.:

```
\t mi telefonas.
\m mi telefon -as
\g 1.Sg. "Telefon-" -Präs.
\p Pers.Pron. Wurzel -Suff
\f ich telefoniere
```

```

\t mi          telefonis.
\m mi          telefon -is
\g 1.Sg.       "Telefon-" -Prät.
\p Pers.Pron. Wurzel -Suff
\f ich telefonierte

```

```

\t mi          telefonos.
\m mi          telefon -os
\g 1.Sg.       "Telefon-" -Fut.
\p Pers.Pron. Wurzel -Suff
\f ich werde telefonieren

```

### 3.2.2.5 Partikel:

Partikel sind unflektierte freie Wurzeln.

#### 3.2.2.5.1 Artikel:

Es gibt im Esperanto nur einen unveränderlichen Determinatoren –la. Dieser wird also nicht nach Genus, Numerus oder Kasus unterschieden. Indefinite Nominalphrasen haben keinen Artikel.

Bsp.:

```

\t telefono          sonoras.
\m telefon -o        sonor -as
\g "Telefon-" -Subst. "Klang-" -Präs.
\p Wurzel -Suff Wurzel -Suff
\f ein Telefon klingelt

```

```

\t la telefono          sonoras.
\m la telefon -o      sonor -as
\g Def "Telefon-" -Subst. "Klang-" -Präs.
\p Det Wurzel -Suff Wurzel -Suff
\f das Telefon klingelt

```

```

\t la telefonoj          sonoras.
\m la telefon -o      -j sonor -as
\g Def "Telefon-" -Subst. -Pl. "Klang-" -Präs.
\p Det Wurzel -Suff -Suff Wurzel -Suff
\f die Telefone klingeln

```

#### 3.2.2.5.2 Personalpronomina:

Die Personalpronomina werden nach den grammatischen Kategorien Numerus, Person und Genus unterschieden.

Bsp.:

```
\t la instruisto parolos
\m la instru -ist -o parol -os
\g Def "Lehr-" -Tätigkeit -Subst. "Sprach-" -Fut.
\p Det Wurzel -Suff -Suff Wurzel -Suff
```

```
\t telefone kun sxi.
\m telefon -e kun sxi
\g "Telefon-" -Adv. mit 3.Sg.Fem.
\p Wurzel -Suff P Pers.Pron.
\f der Lehrer wird telefonisch mit ihr sprechen
```

```
\t la instruistino parolos
\m la instru -ist -in -o parol -os
\g Def "Lehr-" -Tätigkeit -Fem. -Subst. "Sprach-" -Fut.
\p Det Wurzel -Suff -Suff -Suff Wurzel -Suff
```

```
\t telefone kun li.
\m telefon -e kun li
\g "Telefon-" -Adv. mit 3.Sg.Mask.
\p Wurzel -Suff P Pers.Pron.
\f die Lehrerin wird telefonisch mit ihm sprechen
```

Als weitere Partikel gibt es Adverbialpartikel, Präpositionen, Korrelativa und Konjunktionen.

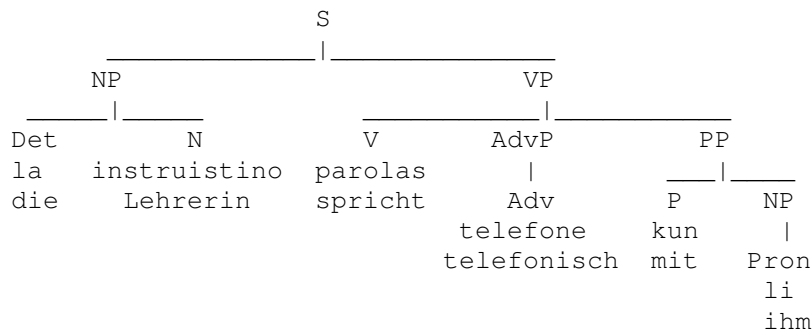
### 3.2.3 Wortstellung

Die normale Wortstellung des Esperanto ist SVO. Determinatoren und Adjektive stehen gewöhnlich vor dem Nomen. Des Weiteren ist die Sprache präpositional. D.h. Adpositionen stehen vor dem Kopf.

Bsp.:

la instruistino parolas telefone kun li

1:



### 3.2.4 Zusammenfassung

Esperanto gilt als eine leicht zu erlernende Sprache. Dies folgt vor allem aus der vollständigen Regelmäßigkeit der Sprache. Grammatische Regeln haben keine Ausnahmen und die agglutinierende morphologische Struktur erleichtert den Wortschatzerwerb gegenüber anderen Sprachen.

In bezug auf die Rolle Esperantos als Interlingua in einem Übersetzungsprozess muss aber doch auf einige Schwierigkeiten hingewiesen werden, die auftreten könnten. So gibt es im Esperanto nur einen definiten Artikel und keinen indefiniten. Auch gibt es im Esperanto nur drei einfache Zeitformen. Dies kann bei der Übersetzung ins Deutsche, Französische und Englische zu Problemen führen.

Der Versuch eine IL zu kreieren, die alle TL-Aspekte berücksichtigt würde in einer „explosiven“ IL resultieren. Probleme, wie ich sie oben kurz angesprochen habe, sollen durch die TL-Grammatik und ein weites IL-TL-Lexikon gelöst werden.

Zur Rechtfertigung von Esperanto als IL zitiert Witkam Andreyev: „Human Languages are much nearer to each other, than to symbols of any variation of a logical system, and consequently an effective Interlingua must be sufficiently similar to spoken human language“. [Andreyev 1967:5 zitiert nach Witkam 1983: III-39]



#### 4. Übersetzungsprozess

Ich werde den Übersetzungsprozess exemplarisch an Hand des folgenden Ausgangssatzes demonstrieren: *He watched the girl with the telescope.* Dieser Satz ist syntaktisch ambig, wie es in der Zeichnung [Witkam 1983: III-17] dargestellt ist.



Die Anzahl der einzelnen Schritte des Übersetzungsprozesses wird in der Literatur verschieden angegeben. Witkam und Schubert haben mehrere kleinere Schritte zu einem großen zusammengefasst und haben somit eine Gesamtzahl von sechs Schritten erreicht. Dies liegt vor Allem daran, dass mitunter mehrere Dinge gleichzeitig ablaufen. Ich habe mich jedoch entschlossen, ähnlich wie Hutchins, zur besseren Erläuterung auch die Zwischenschritte einzeln anzugeben.

Schritt 1:

Im ersten Schritt wird der Satz durch einen ATN-Parser syntaktisch analysiert. ATN (Augmented Transition Network) wurde 1970 von Woods auf der Basis von Konstituenz entwickelt. Im Rahmen des DLT-Projektes wurde das Design jedoch so verändert, dass es auch mit Abhängigkeitsgrammatiken umgehen kann. Bei der syntaktischen Analyse wird auf die SL-Syntaxregeln und das SL-Wörterbuch zugegriffen.

Das Ergebnis sind eine oder mehrere Baumstrukturen, je nachdem ob der Ausgangssatz ambig ist oder nicht. In unserem Fall erhalten wir zwei Bäume.

Legende der Abhängigkeitslabel:

Subj = Subjekt

Obj = Objekt

Circ = Circumstantial

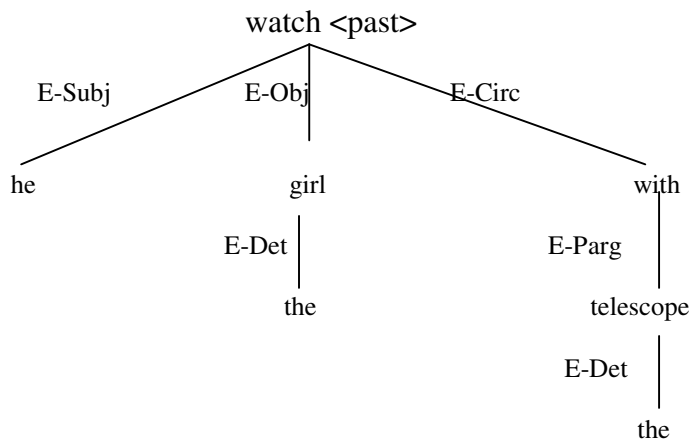
Atr2 = Postnominal Attribute

Det = Determiner

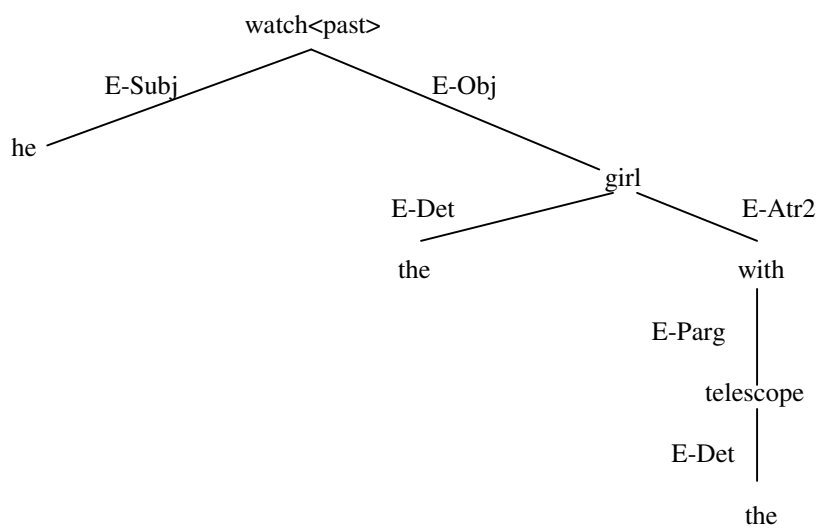
Parg = Prepositional Argument

E = Englisch

erster Baum:



zweiter Baum:

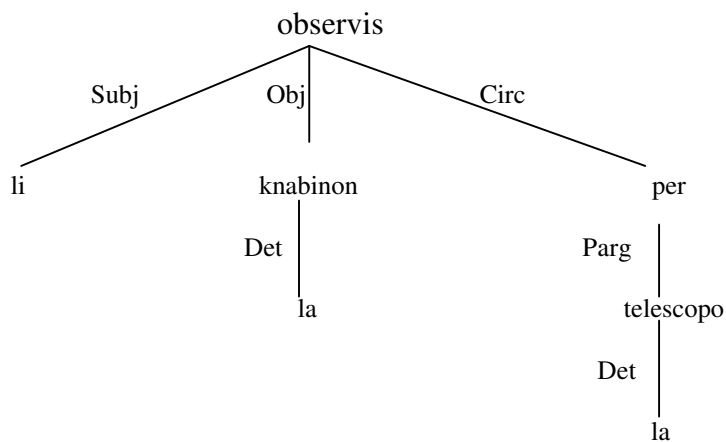


## Schritt 2:

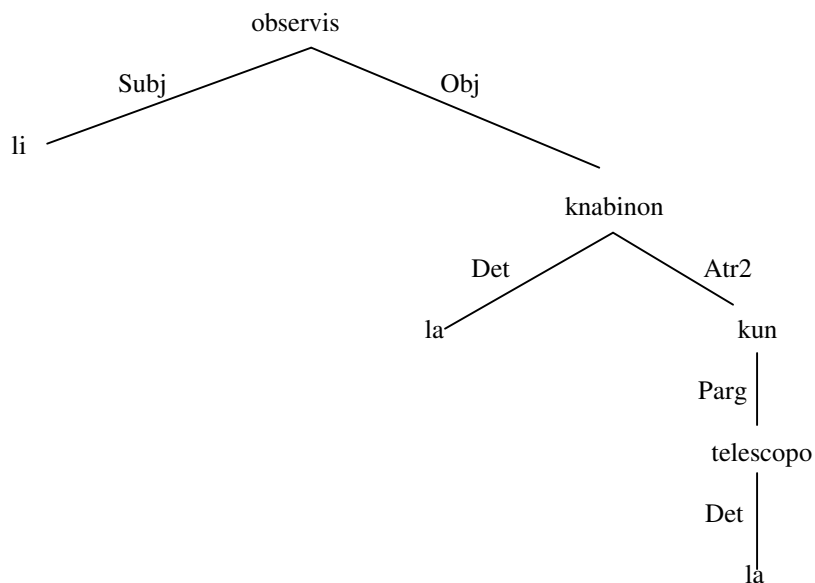
Im nächsten Schritt werden die einzelnen Wörter übersetzt. Dabei werden die englischen Dependenzlabel durch äquivalente in Esperanto ersetzt. Das Programm greift dabei auf die Metataxisregeln und das zweisprachiges Wörterbuch (SL-IL) zu.

Das Ergebnis sind in unserem Fall mehrere IL-Bäume.

erster Baum: Li observis la knabinon **per** la telescopo.



zweiter Baum: Li observis la knabinon **kun** la telescopo.



In unserem Beispielsatz haben wir strukturelle Ambiguitäten. Lexikalische würden an den jeweiligen Knoten eines Baumes generiert werden.

### Schritt 3: Semantische Wortwahl durch SWESIL

SWESIL ist ein separates Modul (Semantic Word Expert System in the Intermediate Language), das gewissermaßen Gebrauch von Weltwissen macht. Dieses bezieht es aus einer "Lexical Knowledge Bank" (LKB).

Eine LKB ist eine Sammlung von Wortpaaren mit jeweils einem syntaktischen Bezugspunkt in der IL.

Ein Wortpaar wird durch ein "content word" und einen "relator" gebildet. Als "relator" treten zum Beispiel Präpositionen, Pronomen und Artikel mit ihren jeweiligen Inhalten (content words) auf.

Aus dem Dependenzbäumen heraus werden alle möglichen Kombinationen von Wortpaaren gebildet. Anschließend sucht SWESIL in der LKB nach identischen oder ähnlichen Paaren. Die gefundenen Ergebnisse werden nach einem bestimmten Algorithmus mit Zahlen bewertet. Das Wortpaar mit dem höchsten Wert ist am wahrscheinlichsten und wird in die Übersetzung übernommen.

### Schritt 4: Interaktiver Dialog

Wenn Ambiguitäten durch den vorherigen Schritt nicht aufgelöst werden konnten, wird ein interaktiver Dialog initiiert.

Dabei gibt der Computer mehrere Auswahlmöglichkeiten in der SL, ohne linguistische Fachtermini zu verwenden. Der jeweilige Nutzer wählt anschließend die richtige Variante aus.<sup>1</sup>

In unserem Beispiel würde folgendes zur Auswahl stehen:

*with the telescope goes with:*

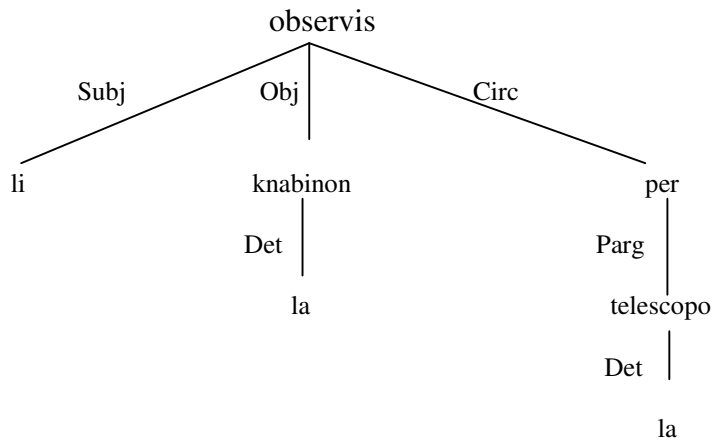
1. the girl
2. watched

---

<sup>1</sup> In der Studie von Witkam wahr geplant, dass alle bis jetzt erläuterten Schritte als ein gemeinsamer Schritt bereits während der Eingabe eines Textes durchgeführt werden. Das heißt, dass im Rhythmus der Texteingabe die Sätze intervallweise syntaktisch analysiert und auch bereits in die IL übersetzt werden. Je nach Voreinstellung unterbricht der Dialog die Texteingabe oder wird am Ende eines kompletten Textes initiiert. Der Nutzer von DLT ist somit auch der Autor des Textes und kann mit seinem Wissen Ambiguitäten im Dialog auflösen.

Der Nutzer entschließt sich in diesem Fall für die zweite Variante, die mit Weltwissen wahrscheinlich ist.

Als Endergebnis dieses Schrittes liegt nur noch ein Baum in der IL vor.



#### Schritt 5 : IL-Linearisierung

Durch Zugriff auf Linearisierungsregeln, die wieder auf strukturellen Beziehungen (Regens-Dependens) beruhen, wird der Ausgangssatz in der korrekten Wortstellung im Esperanto repräsentiert: *Li observis la knabinon per la telescopo*. Dabei gibt es als default-Regel folgende Festlegung: Ein Dependens folgt seinem Regens, wenn es keine spezifischere Regel gibt, die etwas anderes fordert.

#### Schritt 6:

Alle Sätze, die in den vorhergehenden Schritten produziert wurden, werden noch einmal durch einen Parser auf ihre Wohlgeformtheit überprüft.

Der akzeptierte Esperantotext wird an den Zielterminal geschickt.<sup>2</sup>

---

<sup>2</sup> Die Linearisierung des Esperantotextes diene in erster Linie der Kompaktheit. Für den Endnutzer sollte die Linearisierung nicht sichtbar gemacht werden. In der Entwicklungsphase diene sie jedoch der Überprüfbarkeit der Genauigkeit der anderen Schritte. Es war vorgesehen, dass die IL-Texte in linearisierter Form wie jede andere Datei auch gespeichert werden konnten, bis sie zur Übersetzung weiterverwendet werden sollten.

Schritt 7 und folgende:

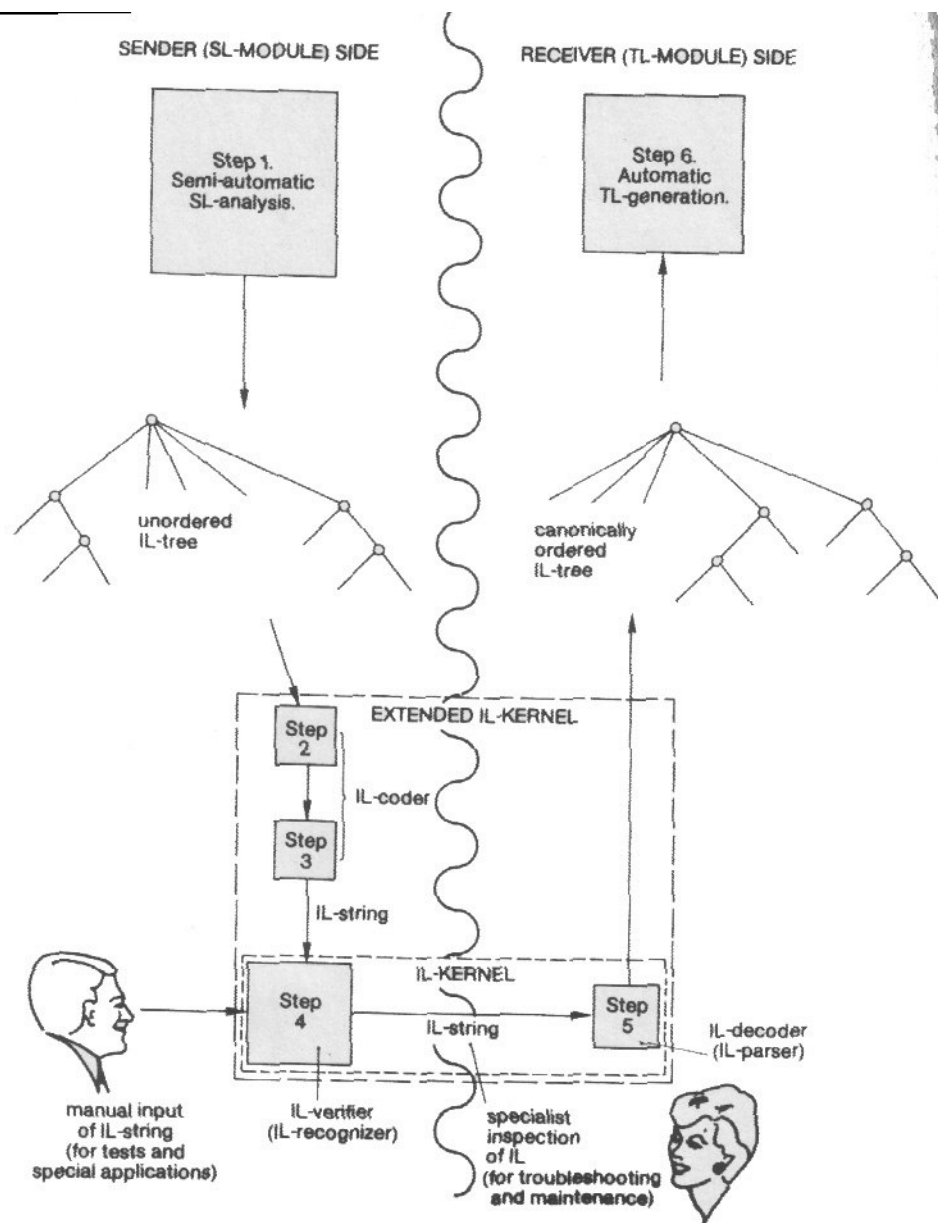
Bei der Übersetzung IL-TL werden äquivalent alle Schritte wie bei SL-IL durchlaufen. Da jedoch davon ausgegangen wird, dass der IL-Text keinerlei Ambiguitäten enthält, entfällt der Schritt des Dialoges.

Das Ergebnis ist ein linearisierter zielsprachiger Text bzw. Satz.

Beispiel: Il regarde la fille avec le telescope.

## 5. Übersetzungsschema

Ich habe am Anfang mehrere Architekturen von Übersetzungssystemen vorgestellt. Dabei haben wir gesehen, dass DLT offiziell zu den Interlinguasystemen zählt. Linguistisch betrachtet kann man aber auch sagen, dass der gesamte Übersetzungsprozess eine doppelte Transfer-Übersetzung beinhaltet. Witkam weist aber ausdrücklich darauf hin, dass die Übersetzung in die IL lediglich eine Hälfte des gesamten Prozesses darstellt. Damit untermauert er die Einordnung von DLT als Interlinguasystem. Zur Verdeutlichung findet sich in seiner Studie folgendes Schema: [Witkam 1983: III-48]



## 6. Zusammenfassung

Ich habe am Anfang dieser Arbeit daran erinnert, dass wir eine Vielzahl von Übersetzungsprogrammen kennengelernt haben, einige wenige auch in ihrer Anwendung. Das kommerziell erfolgreichste dieser Programme ist zweifelsohne Systran. Dieses Direktübersetzungsprogramm arbeitet mit dem System, für jedes neu auftretende Problem, einen neuen Lexikoneintrag zu erzeugen. In seiner Anwendung hat sich gezeigt, dass dieses System nicht besonders effektiv ist. Systran erledigt Übersetzungsaufträge zwar sehr schnell, aber dafür mit teilweise gravierenden Fehlern.

Bereits in der Durchführbarkeitsstudie von 1983 stellte Witkam fest, dass der Markt für MÜ-Programme gerade für exportierende Industriezweige wachsend ist. In einer Zeit, in der das Internet in seiner heutigen Form noch eine Vision war, zitiert er Rolling: "the most valuable assistance to machine translation will be provided by the new techniques of office automation and telecommunications..." [Rolling 1983: 173 zitiert nach Witkam 1983: I-3]. Witkam stellt aber auch fest, dass diesen Branchen eine "quick and dirty"-Übersetzung zu genügen scheint, also eine Übersetzung, die nur den ungefähren Inhalt wiedergibt.

Das Ziel des DLT-Projektes war jedoch die Entwicklung eines multilingualen qualitativ hochwertigem Übersetzungsprozess mit Esperanto als Interlingua.

Während der Arbeit an dem Projekt wurde jedoch die Erfahrung gemacht, dass Esperanto (mittlerweile) genauso lexikalisch und strukturell ambig ist wie andere natürliche Sprachen.

Eine qualitativ hochwertige Übersetzung hätte dadurch viel Zeit beansprucht ebenso die Weiterentwicklung dieses Projektes. Doch fanden sich dafür keine Interessenten und vor allem keine Geldgeber.

Das Fazit dieser Gründe ist eher pessimistisch: „In our ‚information society‘ these days, speed is often considered more important than quality.“ [Witkam 2003: 1]

Witkam hat aber auch die Hoffnung, dass irgendwann einmal die Computer selbst in der Lage sein werden, aus allen Publikationen der Welt, ihre "knowledge banks" zu aktualisieren und ein Gefühl für Kontext und Wortwahl zu entwickeln. Dies würde die automatische Übersetzung erheblich verbessern. Allerdings werden wir wohl auf die Erfüllung dieser Hoffnung noch einige Jahrzehnte warten müssen.



## **Bibliographie**

Andreyev, N. D.: The Intermediate Language as the focal point of machine translation. in: Booth, A.D. (Hg.): Machine Translation. Amsterdam: North-Holland Publ. Co. 1967. S. 1-27. zitiert nach: Witkam 1983.

Baum, Richard: Dependenzgrammatik. Tesnières Modell der Sprachbeschreibung in wissenschaftsgeschichtlicher und kritischer Sicht. Tübingen: Niemeyer. 1976.

Hutchins, W. John & Somers, Harold L.: An introduction to Machine Translation. London: Academic Press. 1992. S. 297-311.

Maxwell, Dan: Distributed Language Translation: A multilingual translation project. Utrecht: BSO Research. Reproduced by: The Indiana University Linguistics Club. Bloomington. 1990.

Rolling, L.: Machine Translations big ten years: 1980-1990. in: Taylor, P.J. & Cronin, B. (Hgs.): Proceedings of the EURIM 5 Conference. London: Aslib. 1983. S. 169-176. zitiert nach: Witkam 1983.

Schubert, Klaus: Metataxis. Contrastive dependency syntax for machine translation. Dordrecht: Foris. 1987.

ders.: Syntactic Tree Structures in DLT. Utrecht: BSO Research. 1986.

Witkam, A.P.M. (Toon): Distributed Language Translation: Feasibility study of a multilingual facility for videotex information networks. Utrecht: BSO. 1983.

ders.: Why the DLT project had to be abandoned. (unveröffentlicht) Januar 2003.